Aus der Abteilung für
Psychosomatische Medizin
Prof. Dr. Thomas Loew
der Fakultät für Medizin
der Universität Regensburg

Konzeption einer elektronischen Studienplattform zur Durchführung umfangreicher internetbasierter Studien in der Psychosomatik am Beispiel SURE

Inaugural – Dissertation
zur Erlangung des Doktorgrades
der Medizin

der
Fakultät für Medizin
der Universität Regensburg

vorgelegt von
Christian Hanshans

2015

Dekan: Prof. Dr. Dr. Torsten E. Reichert

1. Berichterstatter: Prof. Dr. med. Thomas Loew
2. Berichterstatter: Prof. Dr. med. Michael Nerlich

Tag der mündlichen Prüfung: 11.07.2016

1 **Vorbemerkung** **5**
2 **Einleitung** **6**
3 **Allgemeine Anforderungen an Onlineplattformen** **10**
3.1 Anforderungen an den Datenschutz 11
3.2 Anforderungen an Datensicherheit 12
3.3 Gebrauchstauglichkeit und Ergonomie 13
4 **Marktüberblick** **14**
4.1 Grundlegendes zu Fragebogensoftware 14
4.2 Fragebogensoftware Unipark 15
4.3 Surveymonkey 17
4.4 SoSci Survey 18
4.5 EvaSys 20
4.6 LimeSurvey 22
4.7 Google Forms 23
4.8 Sphinx Survey 24
4.9 NetQ 25
4.10 Marktforschungs-Software und Dienstleister 25
4.11 Eigenentwicklungen von Studienportalen 26
4.12 Bewertung des Marktüberblicks 28
5 **Modell eines Studienportals** **32**
5.1 Personas 32
5.2 Anwendungsfälle (Use Cases) 34
5.3 Prozessmodelle 37
5.3.1 Registrierung der Teilnehmer 37
5.3.2 Studienteilnahme 37
5.3.3 Zugriff auf Studiendaten 38
6 **Problemanalyse und Problemlösung** **39**
6.1 Rekrutierung großer Studienpopulationen 39
6.1.1 Email 39
6.1.2 Medien und Mundpropaganda 39
6.1.3 Soziale Netze 39
6.1.4 Suchmaschine und eigene Webseite 40
6.1.5 Verlinkung auf externen Seiten und Forschungsdatenbanken 41
6.2 Adhärenz der Studienteilnehmer 41
6.2.1 Einsatz von individuellen Startseiten (Landing-Pages) 41
6.2.2 multimediale Unterstützung 43
6.2.3 Erinnerungsmöglichkeit bei nicht eingereichten Fragebögen 43
6.2.4 Fallstricke elektronischer Formulare 43
6.2.5 Druck in Papierform 44
6.3 Management großer Studienpopulationen und Datenmengen 44
6.3.1 Kommunikation mit Studienteilnehmern 44
6.3.2 Ökonomische Aspekte 46
6.3.3 Auswertung 47
6.3.4 Archivierung 48
6.4 Auswertung 48
6.4.1 Grafische Auswertung (Histogramm) 48
6.4.2 Datenexport für externe Programme (Excel, SPSS) 49
6.5 Datenschutz 49
6.5.1 Verschlüsselte Datenübertragung via SSL/TLS 49
6.5.2 Perfect Forward Secrecy (PFS) 53
6.5.3 Webseiten-Statistik 58

6.5.4 Daten innerhalb des Studienportals ... *59*
6.5.5 Anonymisierung der Fragebögen ... *60*
6.5.6 Vollständige Anonymität – eine Gefahr? ... *62*
6.5.7 Cookies ... *63*
6.6 Datensicherheit ... 64
6.6.1 Server und Hardware ... *65*
6.6.2 Backup ... *73*
6.6.3 Monitoring ... *75*
7 Implementierung des Prototypen ... 78
7.1 Weboberfläche und Benutzerverwaltung ... 78
7.1.1 Moodle ... *78*
7.1.2 Debian Linux ... *79*
7.1.3 Webserver (LAMP) ... *80*
7.1.4 Benutzeroberfläche ... *82*
7.1.5 Fragebogen erstellen ... *85*
7.1.6 Datenauswertung und Datenexport ... *90*
7.1.7 Videounterstützung (Browserweiche) ... *92*
7.1.8 Registrierung der Studienteilnehmer ... *98*
7.1.9 Studienverwaltung ... *100*
7.1.10 Rollen und Benutzerrechte ... *101*
7.2 Sicherheitsimplementierungen ... 103
7.2.1 Moodle Sicherheitseinstellungen ... *105*
7.2.2 Sicherheitsanalyse von Moodle ... *109*
7.2.3 Sicherheitsmängel von Moodle ... *110*
7.2.4 Verschlüsselung des Transportwegs ... *113*
7.2.5 DNSSec ... *117*
7.2.6 DANE/TLSA ... *119*
7.2.7 Intrusion detection & Web-Application Firewall ... *123*
7.2.8 Schutz durch Faktor-2-Authentifizierung ... *124*
7.2.9 Sicherheit des Mailsystems ... *126*
7.3 Namensserver ... 130
7.4 Hochverfügbarkeitscluster ... 132
7.4.1 Hardware ... *132*
7.4.2 Hochverfügbarkeitscluster (HA-Cluster) ... *132*
7.4.3 Rechenzentrum ... *137*
7.5 Monitoring ... 138
7.6 Backup ... 139
7.7 Benchmarks ... 141
7.7.1 DNS Server ... *141*
7.7.2 Webserver ... *142*
8 Diskussion ... 147
9 Ausblick ... 151
10 Zusammenfassung ... 154
11 Anhang ... 156
11.1 Abbildungsverzeichnis ... 156
11.2 Tabellenverzeichnis ... 158
12 Quellenangabe ... 159
13 Lebenslauf

1 Vorbemerkung

Das Thema zu dieser für eine medizinische Dissertation unüblichen Arbeit entstand durch Prof. Dr. med. Thomas Loew, zu dessen Qualifikationen nicht nur die Medizin, sondern auch die der Medizininformatik zählt. In der Psychosomatik sieht er sich häufig mit dem Problem konfrontiert, Studien in Form von Befragungen bzw. Fragebogenerhebungen durchzuführen. So drängt sich zwangsweise der Wunsch auf, Fragebögen automatisiert über das Internet zu erfassen. Als Ingenieur und Medizininformatiker mit mehrjähriger Erfahrung im Umgang mit eLearning, datenbankgestützten Informationssystemen, Internet- und Servertechnologien, sowie einem abgeschlossenem Medizinstudium bietet das Thema für den Verfasser dieser Arbeit einen ausgezeichneten Anwendungsfall zur Verknüpfung der technischen und medizinischen Ausbildung.
Kernpunkte dieser Arbeit stellen insbesondere die Problemanalyse und Anforderungsanalyse dar. Der Fokus liegt hierbei auf organisatorischen und technischen Details bei der Durchführung großer internetbasierter Studien. Es werden explizit komplexe Bereiche der verschlüsselten Datenübertragung und Anonymisierung/Pseudonymisierung als wesentliche technische Grundlagen des Datenschutzes und Einsatz von Hochverfügbarkeits-Techniken zur Sicherstellung von Datensicherheit im Konzept verankert. Da sich diese Teile der Arbeit sehr stark mit Aspekten der Informatik beschäftigt, sind sie auf einen technisch interessierten und versierten Leserkreis abgestimmt. Aufgrund des techniklastigen Anteils der Themenstellung bilden einen großen Teil der Quellen technische Beschreibungen von Herstellern oder OpenSource Projekten. Wissenschaftliche Publikationen, wie aus medizinischen Datenbanken (z.B. PubMed) gewohnt, sind in diesem Bereich eher rar. Sie werden durch Whitepapers, Howtos oder Referenzhandbücher, also praxisnahe Veröffentlichungen oder technische Beschreibungen ersetzt. Als Konvention werden spezielle technische Fachbegriffe oder Algorithmen sowie Schaltflächen von Benutzeroberflächen mit *kursiver Schrift* gekennzeichnet.

2 Einleitung

Nach traumatisierenden Situationen, z.B. nach Unfällen, Kriegs- oder Rettungseinsätzen, besteht ein großer Bedarf an psychologischen Interventionen, die unmittelbar beruhigen können. Die Holding- oder Containing- Funktion des Psychotherapeuten, also die Präsenz und die Bereitschaft zuzuhören, sowie Imaginations- und Entspannungstechniken galten bisher als Mittel der ersten Wahl. Oft sind Opfer wie Helfer in Methoden wie Progressiver Muskelrelaxation oder autogenem Training nicht ausgebildet.
Eine einfach zu erlernende und anzuwendende Alternative zu den klassischen Entspannungstechniken stellt SURE (sprich: ʃʊr) dar. Diese Methode basiert auf einer schwingenden Bewegung des Oberkörpers von den Hüften aufwärts, die sich aus den Meditationstechniken der Sufi-Mönche ableitet und erstmalig von Alliev in Russland zur Stressbewältigung von Astronauten genutzt wurde. Das Schwingen in der Sagittal- oder Frontal-Ebene für ungefähr 5 Minuten versetzt den Anwender in eine leichte Trance. In einer kontrollierten Studie wurden 28 Probanden, die in mehreren Gruppen das autogene Training gelernt hatten, mit 20 verglichen, die zunächst in SURE eingeführt wurden. Es wurde hierbei die visuelle Analogskala genutzt, um die jeweilige psychische Anspannung vor und nach der Intervention zu messen. Die letzte Sitzung mit autogenem Training wurde mit der SURE-Intervention verglichen; beide dauerten etwa 5 Minuten. Unter beiden Bedingungen konnte die psychische Anspannung signifikant gesenkt werden. Im Wilkoxon Test für verbundene Stichproben unterschieden sie sich nicht voneinander. Es konnte bereits 2010 in einer randomisierten kontrollierten Studie gezeigt werden, dass Rettungskräfte davon profitieren können, SURE zu erlernen, um sich selbst besser emotional zu regulieren. [71] Es ist vergleichbar wirksam zur progressiven Muskelrelaxation jedoch einfacher und deutlich schneller zu erlernen. Weiterhin konnte gezeigt werden, dass SURE zur Selbst-Entspannung gesunder Probanden dem autogenen Training ebenbürtig ist. [80] Hierdurch könnte eine neue komplementär-psychotherapeutische Methode zur Behandlung von akutem psychischen Stress vorliegen. Sie kann auch dann zum Einsatz kommen wenn eine sprachliche Verständigung (z.B. aufgrund von Sprachbarrieren, verletzungsbedingt oder psychogen) schwer möglich ist.

Zur weiteren Untersuchung des Nutzens von SURE soll in einer kontrollierten prospektiven Studie in einem größeren Maßstab der Nutzen von SURE für besonders stressexponierte Studienpopulationen, wie Einsatzkräfte, untersucht werden. Hierzu muss ein großes Studienkollektiv rekrutiert werden. Den Studienteilnehmern soll per Videoanleitung die SURE Methode nahegebracht werden. Über die Erhebung von Fragebögen kann in Anlehnung an die Dissertation von Philipp Kutz aus dem Jahre 2010 der Erfolg der Methode im Verlauf gemessen werden. Die Studie wurde damals mit einem erheblichen (insbesondere personellen) Aufwand konventionell durchgeführt. [71] Die Studienteilnehmer erhielten postalisch ihre Studienmappe mit einem Begleitschreiben, Fragebögen, einem Anwendungskalender und einer DVD. Die DVD enthielt das Schulungsvideo, mit dessen Hilfe die Teilnehmer die SURE Methode erlernen konnten. Die Auswertung der Fragebögen erfolgte manuell nach Rücksendung der Studienunterlagen.

Abbildung 1: Studienunterlagen der SURE-Studie 2010 [71]

Für die jetzige Untersuchung sollen nun weitere Berufsgruppen herangezogen werden, die im beruflichen Alltag oder in bestimmten Situationen in besonderem Maße Stress ausgesetzt sind. In die Studienpopulation werden daher folgende vier Berufsgruppen einbezogen:

- Soldaten der Bundeswehr
- Mitarbeiter des Rettungsdienstes
- Polizeieinsatzkräfte
- Mitarbeiter in Pflegeeinrichtungen

Für jede Studienpopulation sollen so viele Teilnehmer wie möglich gewonnen werden. Ziel ist es, für jede Gruppe Teilnehmerzahlen von mindestens 500 Teilnehmern zu erreichen. Für das Studiendesign wird mit einer Gesamtteilnehmerzahl von 5.000 Probanden gerechnet. Während der Laufzeit von einem Jahr werden die Teilnehmer in regelmäßigen Abständen befragt. Die Studiengruppen sollen getrennt voneinander verwaltet und ausgewertet werden. Jedoch sieht das Studiendesign eine übergreifende Auswertung der Ergebnisse aller Gruppen vor. Die SURE-Studie beschäftigt sich unter Anderem mit der Fragestellung, in welchem Umfang das Verfahren zur Stressbewältigung beiträgt und dadurch als Präventivmaßnahme psychischen Folgeerkrankungen wie posttraumatischen Belastungsstörungen vorbeugen kann. Zum Ausschluss bereits bestehender psychischer Erkrankungen wird hierzu unter anderem der auf den internationalen Diagnosekriterien (DSM-III-R) basierende Test zur Diagnostik posttraumatischer Belastungsstörungen (PTSS-10) eingesetzt. [40] [83]

Weitere standardisierte Fragebögen wie der Erholungs-Belastungs-Fragebogen (EBF) oder der Einsatz von Eigenschaftswortlisten (EWL) sollen den Grad der beruflichen Stressbelas-

tung quantifizieren. [59] [63] Neben den Befragungs-Items, die allen Studiengruppen gemein sind (z.B. PTSS-10, EBF, EWL), müssen für die jeweilige Studiengruppe zusätzliche Items erstellt werden. Hierzu zählen gruppenspezifische Informationen wie zum Beispiel die Anzahl von Einsatzfahrten pro Woche (bei Rettungsdienstmitarbeitern) oder die Anzahl verstorbener Patienten pro Monat (bei Mitarbeitern aus der Pflege). Ein vollständiger Fragebogen enthält insgesamt zwischen 150 und 250 Items.

Angesichts des zu erwartenden Umfangs wurde beschlossen die Studie vollständig internetbasiert durchzuführen. Die Bereitstellung der Studienunterlagen inklusive des Schulungsvideos und Fragebögen soll über eine Onlineplattform erfolgen. Durch die zunehmende Nutzung und Akzeptanz des Internets, dem flächendeckenden Ausbau mobiler Datendienste und der guten IT-Ausstattung deutscher Haushalte stellen Onlinebefragungen technisch kein Hindernis mehr dar. [141] [142] [143] Bereits 92,8% aller berufstätigen Männer und Frauen verwenden das Internet laut ARD-ZDF Online Studie 2014 (n=1.434) zumindest gelegentlich, die Altersgruppe zwischen 18 und 40 Jahren sogar über 97%. Den stärksten Zuwachs verzeichnet die Nutzung mobiler Endgeräte und Videostreaming. Knapp 2 Stunden verbringt der Durchschnittsdeutsche im Internet und damit mehr als doppelt soviel Zeit mit digitalen Medien wie mit Zeitungen, Büchern und Zeitschriften zusammen. [37] Häufigste Anwendungen aller Altersklassen sind die Informationssuche oder Internetrecherche, Emailkorrespondenz, Abspielen von Videos und Teilnahme an Online-Communities. Dabei besitzt jeder Deutsche mindestens 2 internetfähige Geräte (Computer, Smartphone, Spielekonsole, Smart-TV, Tablet-PC oder eBook Reader) und schätzt seine Internetkenntnisse als „gut“ ein. [36] [37] Die in einer Studie verwendeten technischen Elemente müssen sich daher an dem Erfahrungshorizont der Studienpopulation orientieren. Dies bedeutet einen hohen Anspruch an Bedienbarkeit (Design und Ergonomie), Verfügbarkeit, Geschwindigkeit und Funktionen, die der Studienteilnehmer von seiner Erfahrung mit großen Onlineportalen gewohnt ist. [36] [14] In diesem Zusammenhang ist es wichtig, dass sich sowohl Fragebögen als auch das Videomaterial gewohnt komfortabel von jedem beliebigen Endgerät abrufen lassen. Inzwischen achtet die Mehrheit der Deutschen (76% laut Forsa Studie 2015) grundsätzlich darauf, wem sie welche Daten zur Verfügung stellen. [41] Vorsichtig gehen nach eigener Einschätzung 93% der deutschen Internetanwender mit ihren persönlichen Daten um und fordern gleiche Schutzregeln wie außerhalb des Internets ein. [37] Immerhin drei Viertel im Altersbereich der erwünschten Studienpopulation vertreten die Meinung, dass derzeit häufiger als noch vor 5 Jahren gegen Datenschutz verstoßen wird - eine Einschätzung, die zeitlich mit der Häufung medialer Präsenz diverser Datenschutzpannen und staatlicher Überwachungsprogramme korreliert. [43] [144] Aus diesem Grund soll bei der Durchführung der Studie Datenschutz und der transparente Umgang mit persönlichen Daten einen großen Stellenwert einnehmen.

In der Vergangenheit wurden in der Abteilung für Psychosomatische Medizin der Universität Regensburg zwar Studien (zumindest teilweise) auf elektronischem Wege durchgeführt. Es fehlte jedoch an einem einfach zu bedienenden und wiederverwendbaren Werkzeug. So besitzt die Universität Regensburg im Gegensatz zu anderen Universitäten kein eigenes Onlineportal oder Panel zur Teilnehmerrekrutierung und Durchführung von Onlinestudien. Für jede psychosomatische Studie mussten individuelle Programmierarbeiten durchgeführt werden, um eine Infrastruktur im Sinne von Servern sowie Webseiten und Datenbanken zu schaffen. Studienunterlagen wurden meist auf konventionellem Weg per Post an die Teilnehmer verschickt und manuell erfasst. Bislang fehlte eine Infrastruktur, die es Professoren, wissenschaftlichen Mitarbeitern oder Doktoranden ermöglichte, Studien zu entwerfen und im Internet bereitzustellen, ohne auf die Hilfe Externer (Informatiker, Webdesigner oder Rechenzentren) angewiesen zu sein.

Dass ein großer Bedarf an der Entwicklung von entsprechenden Onlineplattformen zu bestehen scheint, zeigt das dringende Gebot des deutschen Wissenschaftsrats zur „Entwicklung und Förderung von Informationsinfrastrukturen“ zur Durchführung großer sozial- und wirtschaftswissenschaftlicher Umfragestudien. [134] Insgesamt 99 deutsche Fachgesellschaften (bzw. deren Mitglieder) wurden nach existierenden Infrastrukturen und dem Bedarf nach Weiterentwicklung oder Neuanschaffungen befragt. Der Rat für Sozial- und Wirtschaftsdaten als Interessenvertreter der Sozialwissenschaften führte in diesem Zusammenhang eine Meinungsumfrage und Bedarfsanalyse unter seinen Mitgliedern durch. [45] [145] Von den 415 Teilnehmern der Onlinebefragung unterstützen 86% die Idee eines national geförderten „Online-Labors“ im Sinne einer Plattform zur Onlinebefragung, dem Bereitstellen von Video- und Audio -Materialien oder zum Durchführen von Experimenten wie Reaktionstests. Die gewünschte mittlere Teilnehmerzahl lag bei 500 Teilnehmern pro Studie. Als wichtigste Beweggründe wurden die Steigerung der Effizienz, methodische Gründe, sowie Chancengleichheit für Einrichtungen ohne Online-Panel angegeben. [45] In dem abschließenden Bericht kommt der Wissenschaftsrat zu dem Schluss, dass die Schaffung von Forschungsinfrastrukturen für den wissenschaftlichen Erkenntnisgewinn und zur internationalen Anschlussfähigkeit eine essentielle Bedeutung zukommt. In den Sozialwissenschaften ergäbe sich der Bedarf insbesondere unter dem Aspekt der praktischen Erleichterung der Forschung, zur Beantwortung wissenschaftlicher und gesellschaftlicher Fragen und zur Archivierung und Verfügbarmachung wertvoller Primärdaten. Er begrüßt den Aufbau von Forschungsdatenzentren direkt bei den Datenproduzenten und unterstreicht die Bedeutung von Engagement wissenschaftlichen Personals mit technischem Know-how zu Design- und Methodenentwicklung. [134]

Der Mangel lokaler wie auch nationaler Infrastrukturen zum Durchführen von Onlinebefragungen, der große Umfang der geplanten Studie, die limitierten finanziellen Ressourcen und die gestellten technischen Anforderungen, eröffnen die Frage, wie eine Onlineplattform aussehen müsste um den Ansprüchen psychosomatischer Forschungsprojekte gerecht zu werden. Diese Frage soll am Beispiel der SURE-Studie beantwortet werden. Die vorliegende Arbeit gibt hierzu einen Überblick über Probleme im Umgang mit internetbasierten Informationssystemen und zeigt hierzu Lösungsansätze auf. Die Arbeit gliedert sich in einen analytischen und einen konzeptionellen Teil. Zuletzt reflektiert ein funktionstüchtiger Prototyp zur Durchführung der SURE-Studie die zuvor angestellten Überlegungen im Sinne einer Best Practise Lösung.

3 Allgemeine Anforderungen an Onlineplattformen

Jedes Computersystem, das personenbezogene Daten speichert und mit Anwendern interagiert, muss drei grundsätzliche Kriterien erfüllen:

- Datenschutzanforderungen
- Datensicherheit
- Gebrauchstauglichkeit und Ergonomie

Datenschutz bezeichnet die Wahrung personenbezogener Daten (Alter, Geschlecht, Emailadresse, Antworten der Fragebögen usw.), wohingegen die Datensicherheit den Schutz der Daten vor Verlust (z.B. durch versehentliches Löschen) definiert.
Gebrauchstauglichkeit und Ergonomie beschreiben inwieweit das System die geforderten Aufgaben erfüllen kann und wie gut sich das System über eine grafische Benutzeroberfläche bedienen lässt. Die Anforderungen lassen sich durch technische oder organisatorische Maßnahmen bzw. personenbezogene Maßnahmen umsetzen.
Diese Kategorisierung hat ihren Ursprung im Arbeitsschutz. Im TOP Modell (technische sowie organisatorische und personenbezogene Maßnahmen) wird die Reichweite und Wirksamkeit der jeweiligen Maßnahme beschrieben. [125] Dieses Modell lässt sich gut auf elektronische Systeme übertragen. Technische Maßnahmen werden auf der Ebene von Hardware oder Software ergriffen, organisatorische und personenbezogene Maßnahmen sind abhängig von den beteiligten Anwendern des Systems.
Grundsätzlich sollte - sofern möglich - eine technische Maßnahme einer organisatorischen vorgezogen und bereits im Entwurf des Systems berücksichtigt werden.
Die folgende Grafik soll zusammenfassend die Maßnahmenhierarchie anhand von ausgewählten Beispielen verdeutlichen.

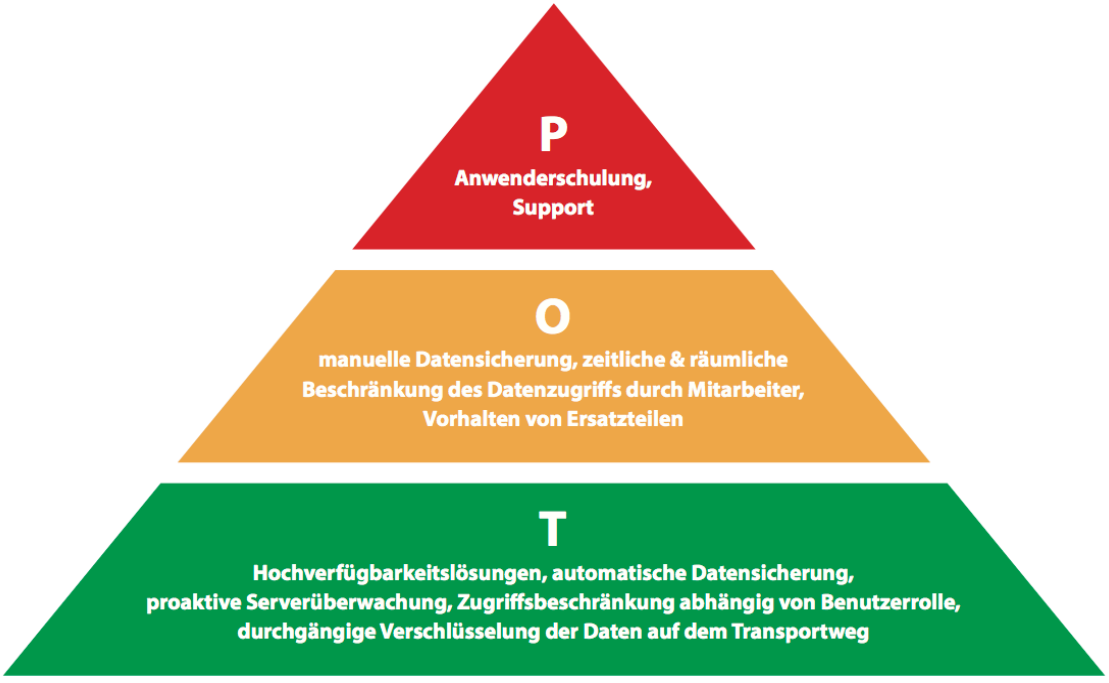


Abbildung 2: TOP Modell zur Maßnahmenhierarchie übertragen auf IT Systeme

3.1 Anforderungen an den Datenschutz

Datenschutz ist in Deutschland gesetzlich im Bundesdatenschutzgesetz BDSG § 9 verankert und fordert allgemein:

1. Zutrittskontrolle
2. Zugangskontrolle
3. Zugriffskontrolle
4. Weitergabekontrolle
5. Eingabekontrolle
6. Auftragskontrolle
7. Verfügbarkeitskontrolle
8. Getrennte Verarbeitung der zu unterschiedlichen Zwecken erhobenen Daten

Konkret bedeutet dies, dass die Server und Datenträger, auf denen die Daten der Studienteilnehmer gespeichert werden, vor unbefugtem Zugriff aber auch vor Beschädigung geschützt werden müssen. Punkt 1 betrifft das Rechenzentrum bzw. den Rechnerstandort: geschlossene Räume mit eingeschränktem Zugang zu Servern und Netzwerk, z.B. durch Code-Schlösser oder Fingerabdruckscanner, Videoüberwachung des Geländes, Protokollierung des Zugangs von technischem Personal zu Maschinenräumen usw. Die Punkte 2 bis 4 richten sich an das Softwaresystem. Ein Zugriff darf nur mittels einer digitalen Identifikation eines berechtigten Nutzers (Login mit Benutzername und Passwort) erfolgen (Punkt 2). Je nach Rolle des Nutzers muss der Zugriff auf die Daten beschränkt werden, die für den Nutzer relevant sind. Hierdurch kann verhindert werden, dass ein Studienteilnehmer die Studiendaten anderer Teilnehmer einsehen kann. Ein Studienbetreuer soll nur Einsicht in die Daten seiner Studie, nicht jedoch in die anderer Studien nehmen können (Punkt 3 und 6). Können Daten exportiert werden, sollten sie vor der Übertragung anonymisiert werden. Auch während der Datenübertragung über das Internet müssen die Daten geschützt werden. Dies kann durch verschlüsselte Datenübertragung realisiert werden (Punkt 4). Hierbei verweist das Datenschutzgesetz BDSG §9 ausdrücklich darauf, dass für Punkt 2 bis 4 „insbesondere die Verwendung von dem Stand der Technik entsprechenden Verschlüsselungsverfahren“ zu wählen sind. [146] Erfolgt ein Zugriff auf personenbezogene Daten, muss das System diese protokollieren und somit nachvollziehbar machen (Punkt 5). Werden mehrere Daten von Teilnehmern erhoben, dürfen nur die Daten durch das System angezeigt werden, die für die Studienauswertung relevant sind, also zum Beispiel die Antworten eines Fragebogens (Punkt 8). Mit Punkt 7 werden Datensicherheit und Datenintegrität angesprochen und im folgenden Kapitel erklärt.

3.2 Anforderungen an Datensicherheit

Ein Onlinesystem setzt sich aus vielen voneinander abhängigen Elementen zusammen. Nur als Ganzes ist ein sicheres Speichern von Studiendaten möglich.

Das Gesamtsystem besteht aus:

- Rechenzentrum und Anbindung an das Internet
- Server und weitere Hardware
- Software und Datenbank
- Anwender

In Analogie zu den Erkenntnissen der Arbeitssicherheit müssen Gefährdungsbeurteilungen und Risikobewertungen durchgeführt und abhängig von Häufigkeit und Schwere, entsprechende präventive Maßnahmen ergriffen werden.
Mit Hilfe einer Risikomatrix können die Risikofaktoren visualisiert werden.
Jedes häufige Ereignis sollte, sofern möglich, mit einer technischen Maßnahme abgefangen werden. Seltenen, aber schwerwiegenden Ereignissen muss abhängig vom notwendigen Aufwand mit technischen und/oder organisatorischen Maßnahmen begegnet werden.

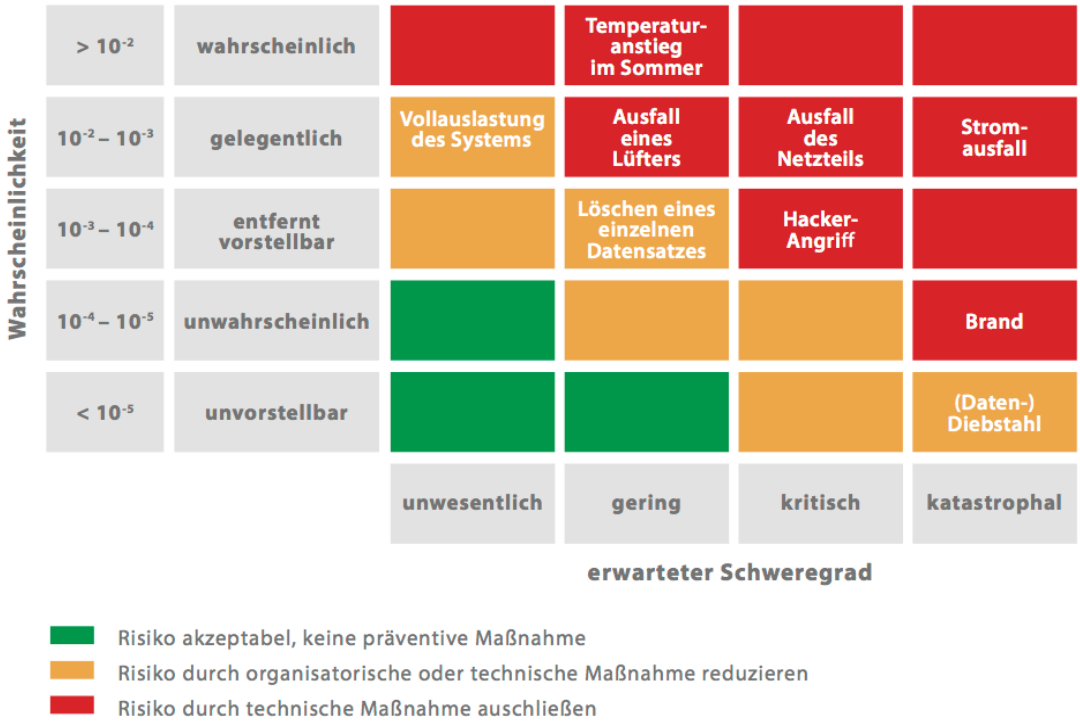


Abbildung 3: Beispiel einer Risikomatrix zur Datensicherheit

Ein Beispiel für ein schwerwiegendes aber vermutlich selten auftretendes Ereignis wäre ein Brand am Serverstandort. Mit Hilfe von Brandmelde- und Löschsystemen kann der hohe Schaden vermieden oder gemindert werden. Da sich ein Restrisiko nicht ausschließen lässt, wäre hier eine Kombination aus technischer und organisatorischer Maßnahme sinnvoll.
Eine sinnvolle technische Maßnahme wäre das Installieren von Branderkennungs- und Brandbekämpfungssystemen (Rauchmelder und CO_2 Löschsystemen) oder die Wahl eines entsprechend ausgestatteten Rechenzentrums. Die ergänzende organisatorische Maßnahme bestünde im Vorhalten einer Datensicherung an einem räumlich getrennten Standort.

Im Gegensatz zu diesem eher seltenen Fall sind Ausfälle von beweglichen Teilen in Rechnersystemen (z.B. Festplatten oder Lüfter) häufig und statistisch abschätzbar. Sie haben in der Regel einen Datenverlust oder den Ausfall des Systems zur Folge. Die adäquate technische Maßnahme, die sich aus hohem Risiko und Schadenschwere ableitet, besteht in der Schaffung von Redundanz z.B. durch Einsatz von gespiegelten Festplatten und mehreren Lüftern oder einer zusätzlichen Stromversorgung.

Für den Betrieb der Studienplattform ist entscheidend, dass das System rund um die Uhr verfügbar ist. Diese Anforderung setzt daher ein gut ausgestattetes Rechenzentrum voraus. Für eine optimale Erreichbarkeit des Portals muss für eine schnelle Anbindung an das Internet ebenso gesorgt sein, wie für eine unterbrechungsfreie Stromversorgung, eine Klimatisierung, Brandschutzanlagen und den Schutz vor Manipulation der Technik durch Unbefugte. Die eingesetzte Hardware muss redundant und fehlertolerant ausgelegt sein. Ein Ausfall einer Komponente darf nicht den Funktionszustand des Systems beeinflussen. Aufeinander abgestimmte Softwarekomponenten sollen den reibungsfreien Betrieb ermöglichen. Dies gilt insbesondere für das Datenbanksystem, das die wertvollen Studiendaten beherbergt. Eine Überprüfung der Datenintegrität muss vom System garantiert werden und Studiendaten müssen vor Manipulation oder Löschung geschützt werden. Kein Nutzer des Systems (außer dem jeweiligen Teilnehmer selbst) darf in der Lage sein, Studienergebnisse zu verändern oder zu verfälschen. Durch geeignete Sicherungsmaßnahmen muss der aktuelle Funktionszustand sowie der Datenbestand archiviert und jederzeit wiederhergestellt werden können.

3.3 Gebrauchstauglichkeit und Ergonomie

Eine weitere Anforderung an die Studienplattform ist die Gebrauchstauglichkeit. Die Plattform muss für jede Benutzergruppe alle Funktionen bereitstellen, die der Anwender für seine Aufgabe benötigt.

Studienteilnehmer müssen in erster Linie in der Lage sein, Fragebögen auszufüllen. Dieser Prozess muss möglichst intuitiv und in für den Nutzer gewohnter Weise erfolgen. Die grafische Oberfläche sollte demnach der Funktionsweise eines Papierfragebogens nachempfunden werden. Fragebögen sollen durch Ankreuzen der Auswahlmöglichkeiten ausgefüllt aber auch korrigiert und gegebenenfalls unterbrochen werden können. Studienbetreuer sollten auf die von Teilnehmern eingegebenen Daten ohne Programmierkenntnisse zugreifen und in ein Microsoft Excel oder SPSS Format überführen können. Zudem muss die Plattform die Studienbetreuer bei zeitintensiven Alltagsaufgaben und häufigen Anwendungsfällen (z.B. Erstellung von Fragebögen) so gut wie möglich unterstützen. [92] Darüber hinaus muss das System bei jeder Aktion mögliche Fehleingaben (z.B. Datumsangaben, Eingabe von Zahlen) oder Fehlbedienung durch Anwender (z.B. versehentliches Löschen von Fragebögen) vermeiden.

4 Marktüberblick

Entsprechend der beschriebenen Anforderungen der geplanten SURE- Studie wurde nach bereits verfügbaren Werkzeugen gesucht. Eine Übersichtsarbeit aus dem Jahr 2014 untersuchte die Art der Durchführung von Studien (n=101) im Bereich der Psychologie und Psychiatrie. Nur etwa 15% aller psychologischen und psychiatrischen Studien wurden vollständig online durchgeführt. Die Arbeit führt aus, dass Online-Studien bei korrekter Durchführung in Punkto Reliabilität mit Laborstudien vergleichbar seien und dass die methodische Qualität der Datenanalyse und Dateninterpretation durch elektronisch durchgeführte Studien deutlich verbessert werden könne. Weiterhin bestünde ein Benefit in der Wiederverwendung von Fragebögen, Softwareprogrammen, aber auch von Studiendaten. [13] Aus Ermangelung geeigneter Werkzeuge für Online-Fragebögen oder eines Online-Labors fordern die Autoren daher die Neuentwicklung eines national geförderten Online-Systems. Um einen Marktüberblick zu bekommen, wurden die in der Arbeit erwähnten Studien näher betrachtet und deren genutzte Software auf Tauglichkeit für den geplanten Einsatzzweck untersucht.

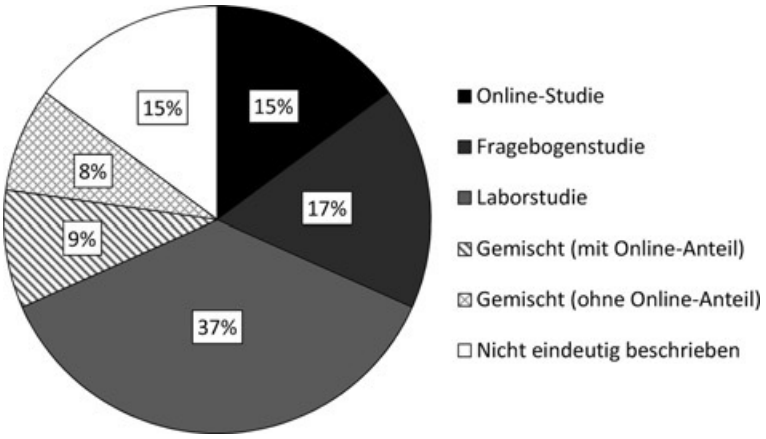


Abbildung 4: aktuelle Durchführung sozialwissenschaftlicher Studien, Bruder et. al. [13]

4.1 Grundlegendes zu Fragebogensoftware

Es findet sich eine Vielzahl unterschiedlicher Fragebogen Programme auf dem Markt. Grundlegend müssen zwei Gruppen unterschieden werden. Während die Einen auf den Servern des Anbieters lokalisiert sind, können die Anderen auf einem eigenen Server betrieben werden. Solche auch als Cloud bezeichnete Angebote sind in der Regel kostenpflichtig. Der Anbieter stellt die Infrastruktur und die Software zur Verfügung und verrechnet diese Leistungen entsprechend. Die Fragebögen werden über das Internet via Browser erstellt, ebenso werden Teilnehmerlisten via Webbrowser übermittelt. Die Teilnehmer können erstellte Fragebögen über eine spezifische Webadresse des Anbieters meist in der Form: *http://cloud-anbieter.com/index.php?fragebogen=mein-fragebogen* aufrufen.
Auf dem eigenen Server installierbare Software ist entweder als kostenfreie oder auch als kommerzielle Software verfügbar. Beiden gemeinsam ist die Notwendigkeit einer eigenen technischen und personellen Betreuung. Dafür können Fragebögen unter der Webadresse der Universität oder einer individuellen Webadresse, also zum Beispiel *https://sure.uni-regensburg.de* oder *https://studienportal.eu* abgerufen werden. Selbstbetriebene Softwarelösungen können, sofern sie quelloffen sind, beliebig erweitert und an eigene Bedürfnisse und Fragestellungen angepasst werden, was bei Cloud-basierten Lösungen nicht

möglich ist. Einige Universitäten betreiben eigene Studienportale. Diese sogenannten Panels dienen primär der Erfassung von Teilnehmern und der Organisation von Studien. Es handelt sich in der Regel um Eigenentwicklungen, die für die Bereitstellung von Studienunterlagen oder Fragebögen wiederum auf bestehende Fragebogensoftware zurückgreifen. Als kommerzielle Alternative haben sich etliche Anbieter am Markt etabliert, die für umfangreiche Studien geeignete Software bereitstellen oder die Durchführung von Studien übernehmen.

4.2 Fragebogensoftware Unipark

Unipark ist eine kommerzielle, online-basierte Software zur Erstellung von Fragebögen.
Für ihre Nutzung zahlt man abhängig von der Laufzeit des Projekts und der Anzahl der Teilnehmer. Der Preis für ein Jahresabonnement beträgt für 25 Projekte 250€ pro Jahr oder für 100 Projekte 700€ pro Jahr. In beiden Fällen ist die maximale Anzahl eingereichter Fragebögen pro Projekt auf 10.000 beschränkt. Die erhobenen Studiendaten lagern auf den Servern der Firma QuestBack GmbH in Bremen. (Stand 11/2014) Fragebögen können über eine intuitive Weboberfläche angelegt werden. Auch die Teilnehmerverwaltung erfolgt online. Die Software bietet viele unterschiedliche Fragetypen und eine webbasierte grafische Auswertung. Auch ein Export der Daten in SPSS ist möglich. Obwohl Unisys verschlüsselte Datenübertragung ermöglicht, fiel bei vielen überprüften Beispielen aus den Übersichtsarbeiten (selbst bei der Erhebung der Daten für die Übersichtsarbeit) auf, dass der Aufruf der Fragebögen unverschlüsselt via *http://unipark.de/uc/...* verlinkt wurde. [13] [45] [147] [126]
Ein Nachteil wie bei allen Cloud-Angeboten ist die Datenhaltung beim Anbieter. Zwar liegen die Daten in einem ISO 27001 zertifizierten Rechenzentrum, aber auf die eingesetzte Software und Technik (z.B. eingesetzte Verschlüsselung) oder auf einen möglichen Zugriff durch administratives Personal auf Anbieterseite, kann kein Einfluss genommen werden. Eine individuelle Erweiterung des Funktionsumfangs oder Skalierung für umfangreiche Studienpopulationen (über 10.000 eingereichte Fragebögen) ist nicht möglich. Neben der Geschäftsstelle in Deutschland verfügt der Anbieter über Büros, Service-Personal und technischer Infrastruktur für Befragungsprojekte in USA. Der Aufruf erstellter Fragebögen erfolgt über die Webadresse von Unipark. Der Teilnehmer verlässt hierdurch die Webseite der Universität und wechselt zu *http(s)://unipark.de/uc/....* oder auf amerikanische Server zu *http(s)://unipark.com/de/...*
Neben der LMU München, der ETH Zürich verwendet auch die Universität Freiburg und die Universität Mannheim, Basel und Ulm Unipark in ihren Studienportalen.

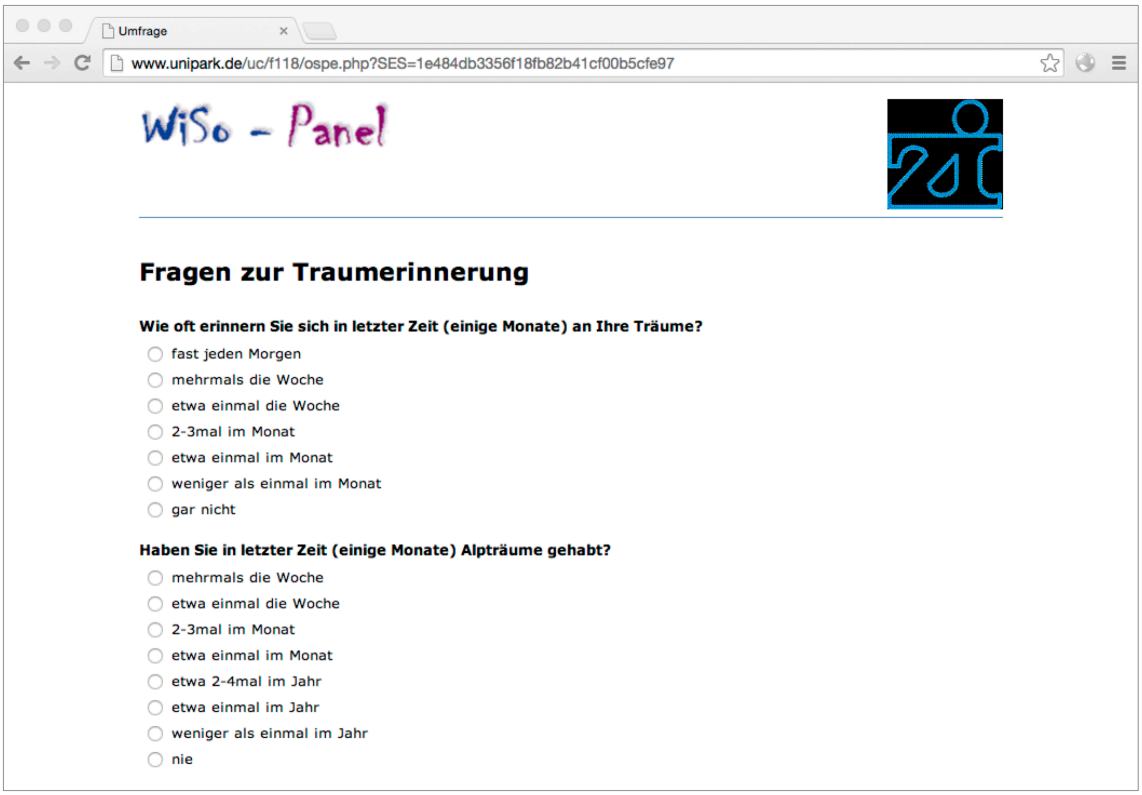


Abbildung 5: WiSo Panel der Universität Freiburg verwendet Unipark unverschlüsselt (Stand 11/2014)

„Forschung erleben“, das Studienportal der Uni Mannheim setzt ebenfalls Unipark ein.

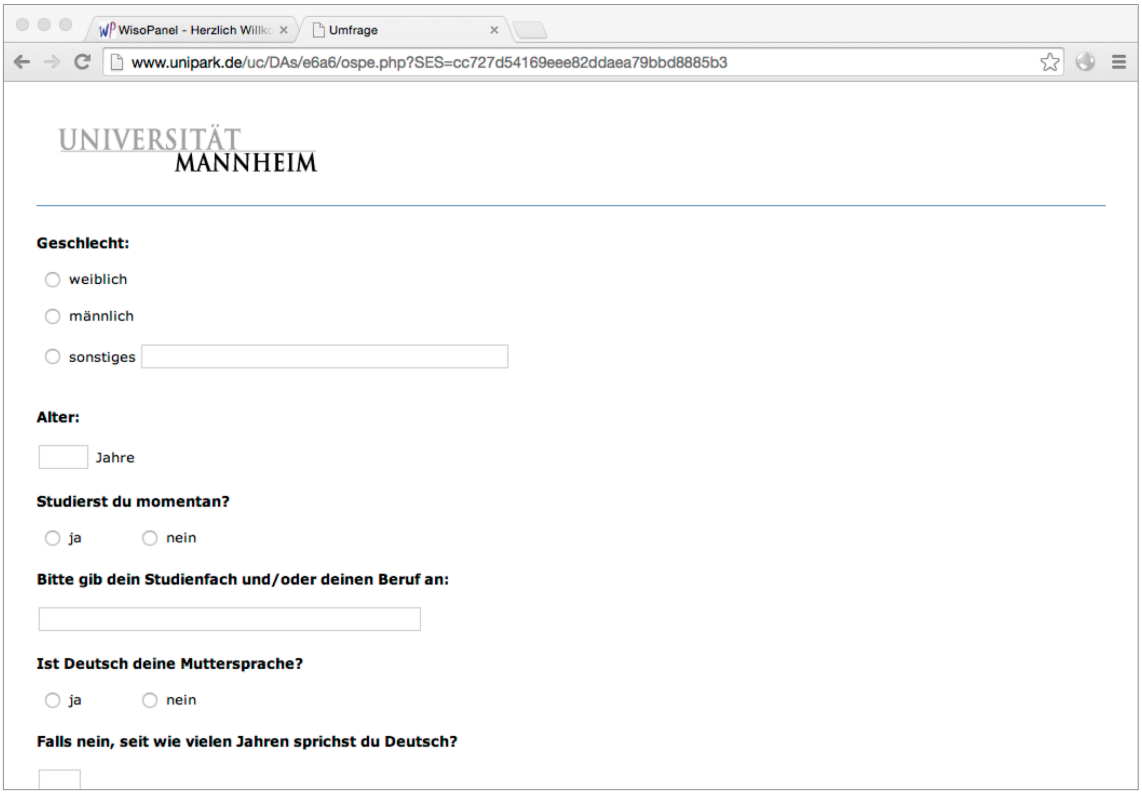


Abbildung 6: Studienportal der Universität Mannheim, Basel, Ulm nutzt Unipark unverschlüsselt (Stand 11/2014)

4.3 Surveymonkey

Ein ähnliches Fragebogen-System ist Surveymonkey. Es ist besonders im englischen Sprachraum verbreitet aber auch in deutscher Sprache verfügbar. Wie Unipark bietet es als kommerzielles Produkt die Möglichkeit der Fragebogengestaltung und Teilnehmerverwaltung über eine Weboberfläche.

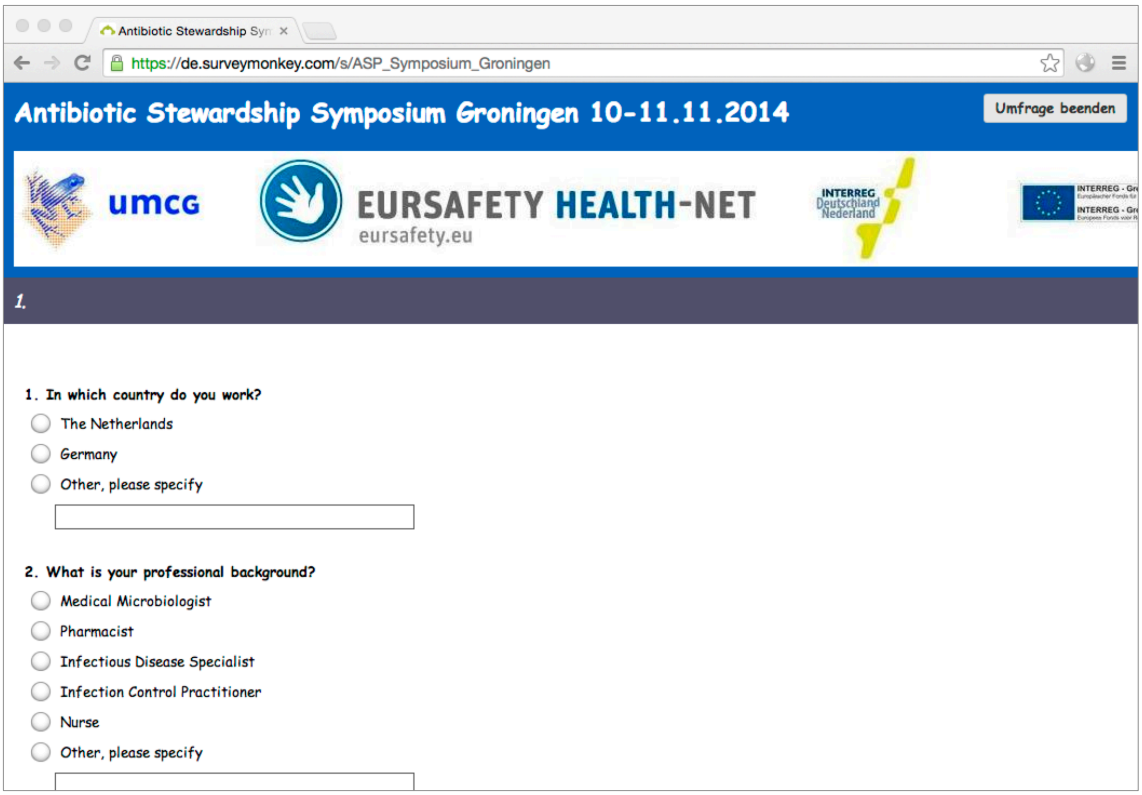


Abbildung 7: Surveymonkey, ein Onlinefragebogenwerkzeug mit Serverstandort USA

Das Tarifmodell sieht eine auf 100 Fragebögen beschränkte kostenfreie Mitgliedschaft vor. Kostenpflichtige Premium Mitgliedschaften erweitern den Funktionsumfang um verschlüsselte Datenübertragung sowie weitere Funktionen. Die Bezahlvarianten von Surveymonkey zeichnen sich durch unbegrenzte Anzahl einreichbarer Fragebögen aus. Das Plus-Paket kostet 250€/Jahr, der Gold-Tarif liegt bei 800€ jährlich und ermöglicht die Einbindung eigener Logos sowie rudimentäre Designanpassung (wie in obiger Abbildung gezeigt). [148] Wie Unipark erfolgt der Aufruf der erstellten Fragebögen über die Internetadresse von Surveymonkey. Die Server des Anbieters sind im Rechenzentrum von Amazon (Amazon EC2 Cloud) in San Jose Kalifornien USA untergebracht und unterliegen daher amerikanischer Gesetzgebung, was aus datenschutzrechtlichem Gesichtspunkt äußerst kritisch zu betrachten ist. Ebenso beunruhigend ist der Aufbau diverser Verbindungen zu Drittanbietern, die der Analyse des Benutzerverhaltens bzw. dem Anzeigen von Werbung dienen. Beim gezeigten Beispiel werden zu den Werbenetzwerken GoogleAds und Doubleklick, der Webseitenstatistik Google Analytics, Werkzeugen zur Erfassung von Nutzerverhalten (New Relic sowie Marinsm) und dem Webseitenmonitor Pingdom Verbindungen aufgebaut und ein *Cookie* hinterlegt. Es ist demnach davon auszugehen, dass neben Surveymonkey auch bei Drittanbietern Benutzerdaten gespeichert und ausgewertet werden, was in Konflikt mit dem deutschen Bundesdatenschutzgesetz steht. [146]

Abbildung 8: Cookies zur Analyse des Benutzerverhaltens

4.4 SoSci Survey

SoSci Survey ist eine Fragebogensoftware aus Deutschland, die speziell auf den wissenschaftlichen Einsatz im Bereich der Sozialforschung ausgelegt ist. Der Anbieter, die SoSci GmbH, bietet eine Cloud-basierte Lösung sowie die Möglichkeit der Installation der Software auf eigenen Windows oder Linux Servern an. Möchte man die Software im eigenen Haus betreiben werden Jahreslizenzkosten von 620€ fällig. Für Hochschulen wird eine kostenfreie Hochschullizenz zur Installation auf einem einzelnen Server vergeben. Keine der Studien darf dann jedoch in Kooperation oder im Auftrag gewinnorientierter Institutionen erfolgen.
Kann oder will man SoSci nicht auf eigener Hardware einsetzen, fallen für den Cloud-Dienst, abhängig von der Anzahl der Fragebögen, einmalige Kosten an. Die Preise orientieren sich an der Anzahl der Fragebögen und der Anzahl der Fragen. Eine Studie mit 5.000 Fragebögen mit mehr als 75 Fragen kostet 220€ (Stand 11/2014). Der Serverstandort ist ein zertifiziertes Rechenzentrum in München, das von der Firma m-Net betrieben wird. Für größere Fragebogenvolumina kann mit dem Anbieter ein individuelles Konzept verhandelt werden.
SoSci Survey ist in der Programmiersprache PHP geschrieben und bietet viele durchdachte Funktionen und die Möglichkeit durch individuelle Erweiterungen komplexe Befragungen zu gestalten. Die Programmgestaltung ist schlicht und entwicklerorientiert, was zusammen mit dem großen Funktionsumfang die Komplexität des Systems erhöht und etwas weniger intuitiv erscheint, wie die zuvor genannten Systeme.

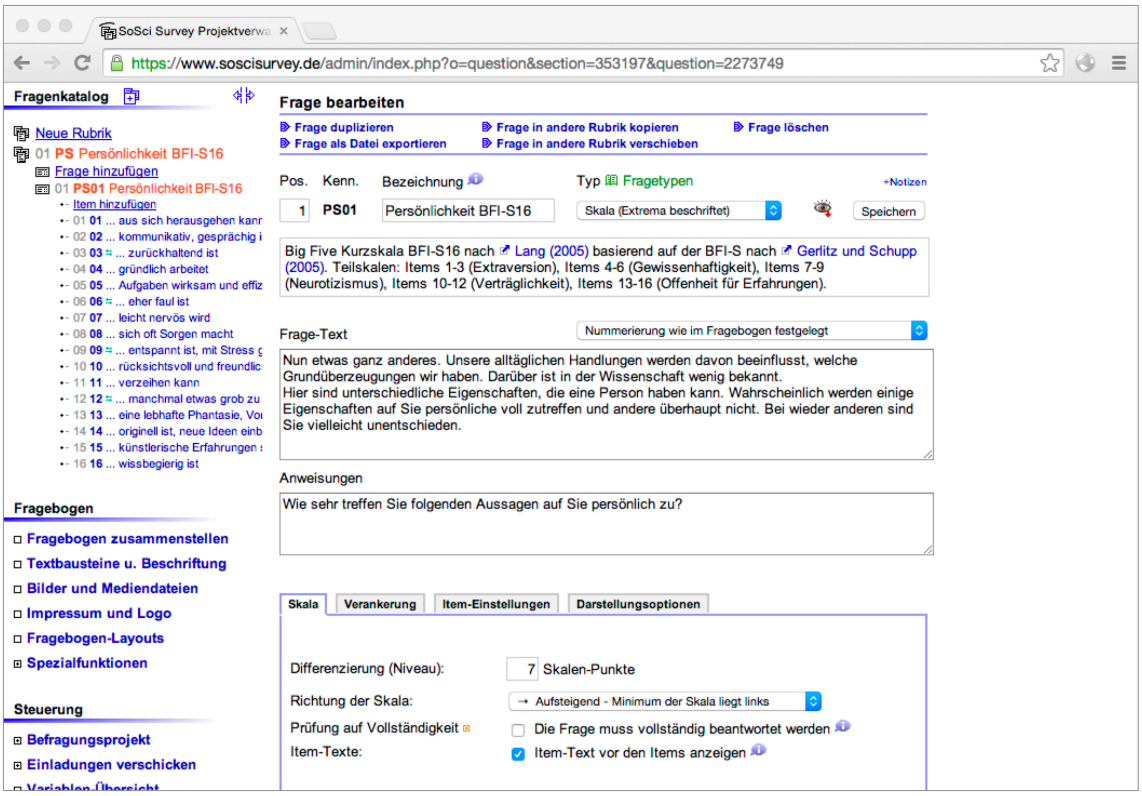


Abbildung 9: Erstellung eines Fragebogens mit SosSci Survey

Besonders hervorzuheben ist die Panel Funktion, die der Erfassung von Teilnehmern, deren Randomisierung sowie der Überwachung des Rücklaufs auf Rekrutierungsemails dient. Wie bei den anderen Anbietern können Listen mit Emailadressaten importiert werden. Die Sosci Survey Cloud-Version verschickt auf Wunsch zeitgesteuert Emails und erfasst neben unbeantworteten Anfragen auch fehlerhafte Emailadressen (sofern auf den Servern von Sosci Survey betrieben). Diese pfiffige Lösung hilft Rücklaufquoten zu ermitteln und ermöglicht es, Teilnehmer gezielt anzuschreiben oder an die Abgabe von Fragebögen zu erinnern.

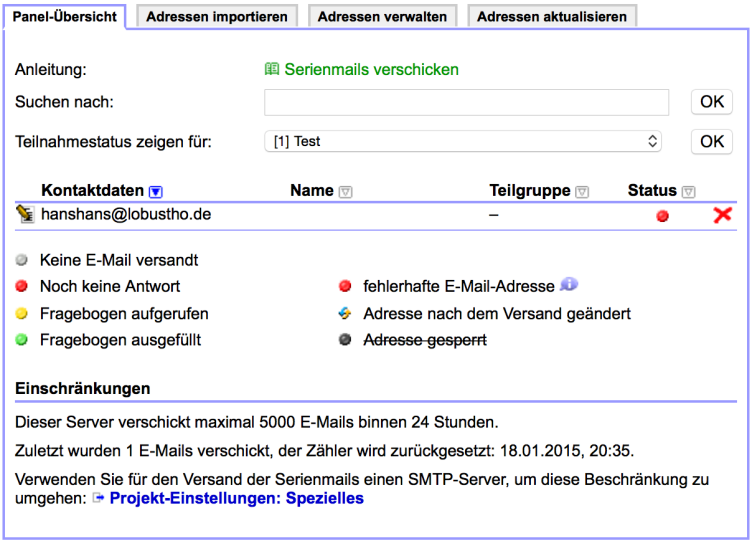


Abbildung 10: Panelfunktion mit Kontrolle des Rücklaufs auf Email-Rekrutierung

Laut Webseite des Herstellers ist SoSci Survey für größere Projekte ausgerichtet und setzt durch die Terminologie und den internen Aufbau Erfahrung oder eine gewisse Einarbeitungs-

zeit voraus. Lobenswert sind die umfangreiche (deutsche) Dokumentation und die FAQ Sektion. Der Funktionsumfang kann bei Bedarf durch eigene Programmteile erweitert werden, für spezifische Anpassungen des Designs sind Grundkenntnisse in HTML und CSS notwendig.

Neben der Software bietet der Anbieter umfangreiche Dienstleistungen von der Erstellung bis zur Auswertung von Fragebögen.

4.5 EvaSys

Evasys ist eine Software, die ursprünglich aus dem Bereich papierbasierter Fragebögen kommt. Der Hersteller konnte sich mit seinem Produkt besonders im universitären Umfeld z.B. mit der Erfassung und Auswertung studentischer Evaluationen etablieren. Mit Hilfe einer Windows Software können Fragebögen erzeugt werden. Diese werden mit einem Barcode versehen ausgedruckt, eingescannt und teil-automatisch ausgewertet. Seit einigen Jahren können auch Onlinebefragungen bzw. Hybrid-Lösungen (Papier und Onlinebefragungen wahlweise) mit EvaSys durchgeführt werden. Das Produkt rangiert im oberen Preissegment und wird nahezu in allen deutschen Fakultäten als Evaluationswerkzeug eingesetzt.
Es existieren mehrere Tarifmodelle. Das Cloud-Angebot, das primär für kleinere Umfragen konzipiert ist, setzt keine eigene Infrastruktur voraus. Die Daten werden auf Servern des Anbieters gespeichert. Aufgrund des geplanten Umfragevolumens ist dieses Modell nicht geeignet. Das für interne Befragungen (z.B. Evaluationen) von Universitäten eingesetzte Lizenzmodell zielt auf eine konstante Anzahl an Fragebogenrückläufern ab. Für den geplanten Einsatzzweck als Werkzeug für medizinische Studien passt die Lizenzvariante für Dienstleister besser. Hierbei wird berücksichtigt, dass zu Beginn einer Studie eine große Anzahl an Fragebogenrückläufern entsteht, die im Lauf der Zeit jedoch abnimmt. Für den Einsatz von EvaSys in beiden genannten Varianten muss eigene (oder beim Hersteller kostenpflichtig gemietete) Hardware und Infrastruktur vorgehalten werden. Die Installation erfolgt auf Windows Servern. Das in Frage kommende Lizenzmodell EvaSys HC (DL) ist eine Art Prepaid System. Es wird ein bestimmtes Kontingent an Fragebogenrückläufern erworben, das unabhängig von der Zeit bestehen bleibt. Ist dieses verbraucht, muss ein neues Paket hinzugekauft werden. Für das Beispiel SURE mit 5.000 Teilnehmern und mindestens 4 Rückläufern je Teilnehmer wären dies 20.000 Fragebögen. Laut einem individuellen Angebot von Electric Paper setzen sich die Gesamtkosten aus der Basislizenz (5.995,00€ zzgl. Steuer), dem Fragebogenpaket (3.995,00€ zzgl. Steuer für 10.000 Rückläufer) und der Remoteinstallation auf dem eigenen Server (995,00€ zzgl. Steuer) zusammen. Der 1-Jahres Supportvertrag (1 Jahr 1.199,00€ zzgl. Steuer) bietet die Möglichkeit Sicherheits- oder Programmupdates vom Hersteller sowie Hilfestellungen bei Problemen zu erhalten. Zusätzlich muss Personal sowie Hardware und Infrastruktur für den eigenen Server vorgehalten werden. Alternativ kann für jährlich 4.995,00€ netto ein Entry-Level Server (Intel Xeon E3-1245, 16 GB DDR3-ECC RAM, 2x 3 TB RAID 1, Windows Server 2008) bei Electric Paper gemietet werden. [149]
Die Betreuung des Betriebssystems mit Softwareupdates ist im Preis inbegriffen. Die angegebene Verfügbarkeit liegt gemäß Service-Level-Agreement bei 99% und lässt damit ungeplante Ausfallzeiten von drei Tagen und über fünfzehn Stunden pro Jahr vertraglich zu. [84] Das Rechenzentrum erfüllt laut Angaben des Herstellers alle Datenschutzbestimmungen. [150]
Für die umfangreiche Software bietet der Hersteller ein Schulungsangebot (1.995,00€ zzgl. Steuer für zwei Person) an. Ziel des Trainings ist primär die Erstellung, das Ausrollen und das Auswerten von Befragungen. In Summe ergeben sich (ohne Schulung und Hosting) einmalige Kosten von über 13.000€, laufende Kosten für den Supportvertrag sowie variable Kosten für weitere Fragebogenpakete.

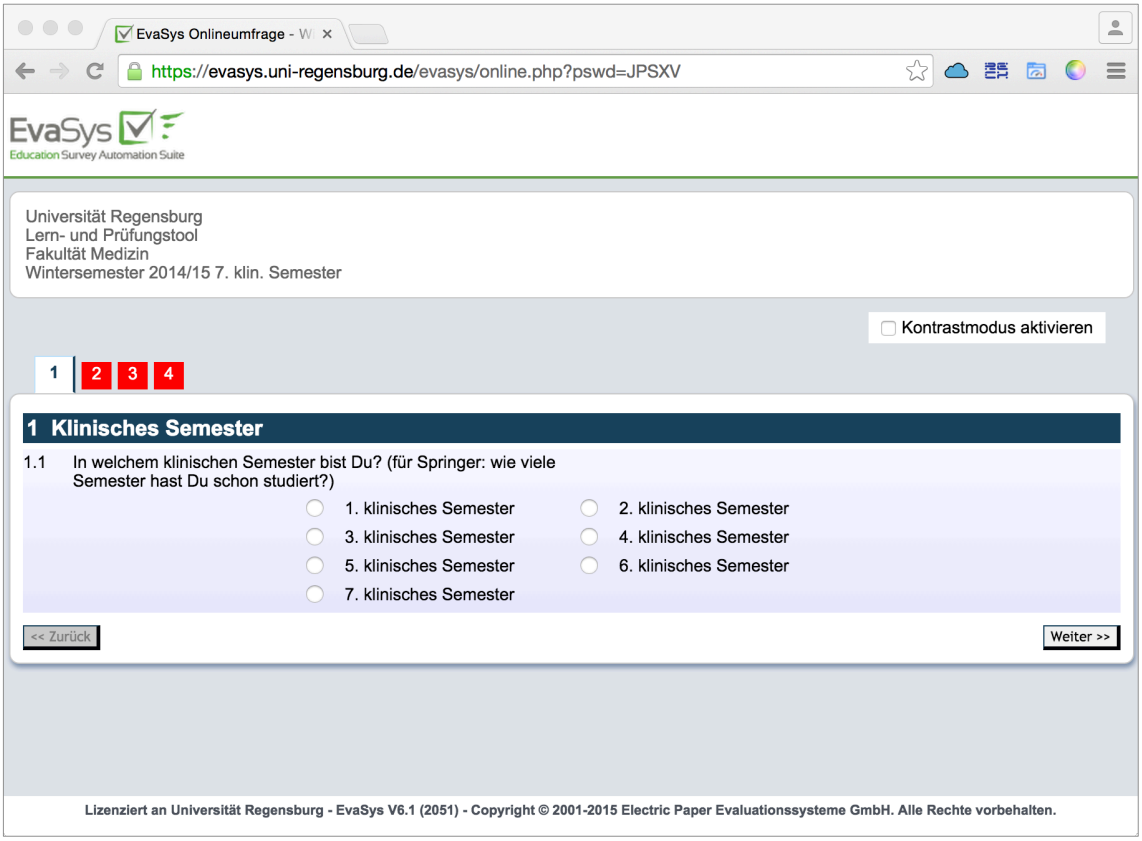


Abbildung 11: EvaSys, ein häufiges Werkzeug für universitätsinterne Umfragen

EvaSys bietet keine Möglichkeit der unterstützten Registrierung bzw. Rekrutierung (Panel-Funktion) von Studienteilnehmern. Alle Teilnehmer müssen im Vorfeld in der Datenbank erfasst sein und über eine Identifikationsmöglichkeit (z.B. Email-Adresse oder Benutzername) verfügen. Alternativ oder zusätzlich können nach dem Pin/Tan Verfahren Zugangscodes vergeben werden. Eine Unterstützung für Videos, die plattformunabhängig auf allen Endgeräten wiedergegeben werden können, ist in EvaSys nicht integriert. Sollen Videos dennoch integriert werden, müssten diese demnach von einem separaten System bereitgestellt werden. Allerdings bietet die Software eine gut dokumentierte und umfangreiche Schnittstelle zur Integration in eigene Softwareprojekte. Somit besteht technisch die Möglichkeit EvaSys in ein eigenes Portalsystem einzubetten und von der langjährigen Erfahrung und dem Funktionsumfang von EvaSys zu profitieren. Ebenfalls interessant in diesem Zusammenhang ist die Option bei Bedarf auch problemlos papierbasierte Befragungen durchzuführen.

Wird EvaSys auf eigener Infrastruktur verwendet gilt es zu berücksichtigen, dass das Maß der Verfügbarkeit, sowie der Datensicherheit und des Datenschutzes vom jeweiligen Systembetreuer und Rechenzentrum abhängt. Exemplarisch wurden drei Instanzen untersucht, die für die studentische Evaluation der Lehre eingesetzt werden. Allen gemeinsam waren unterschiedlich stark ausgeprägte Mängel (Stand 07/2015).

Die Installation auf dem Server der virtuellen Hochschule Bayern fiel in erster Linie durch mangelnde Unterstützung aktueller und empfohlener Verschlüsselungsprotokolle (*TLS 1.1* und höher) und einer alten Serversoftware (Windows Server 2003, IIS 6, PHP 5.3) auf. Für Windows 2003 wurde wie auch für Windows XP der Support eingestellt. Bestehende Installationen werden demnach nicht mehr mit Sicherheitsupdates versorgt. Im standardisierten SSL-

Test erhält die Installation nur eine mittelmäßige Bewertung von C nach amerikanischem Notensystem. (Bewertung des Marktüberblicks ☞ Kapitel 4.12) Bei Servern der Universität Regensburg kommt ein seit Jahren als unsicher eingestuftes Verschlüsselungsprotokoll (*SSLv3*) zum Einsatz, das die seit Anfang 2014 unter dem Namen „*Poodle*“ bekannte Sicherheitslücke eröffnet. [18] [90] Weiterhin wird nach wie vor ein Verschlüsselungsalgorithmus (*RC4*) eingesetzt, der seit Längerem als unsicher gilt und gemäß der technischen Richtlinie des Bundesamtes für Datenschutz seit 2013 nicht mehr eingesetzt werden darf. [18] [100] In der Testwertung ergibt sich ebenfalls die Note C. Das Ergebnis für das Äquivalent der Universität Göttingen liefert aufgrund veralteter und unsicherer Verschlüsselungskomponenten (*RC2*, *RC4, MD5*) sowie einiger Sicherheitslücken die schlechteste Note (Note F) im SSL-Test. Ein weiterer Grund für die schlechte Bewertung ist der Einsatz von „*anonymous Cipher Suites*“. [95] Die angesprochene Pseudo-Verschlüsselung stammt aus frühen Zeiten der Datenübertragung zu der man sich nur wenig Gedanken über Manipulation von Datenströmen bzw. Einblick in übertragene Informationen machte. Aus heutiger Sicht erlaubt die veraltete Technologie das einfache Abfangen der Datenübertragung, da die enthaltene Information faktisch unverschlüsselt übertragen wird. *Null Ciphers* sind daher seit vielen Jahren kontraindiziert.

4.6 LimeSurvey

LimeSurvey ist eine beliebte quelloffene Umsetzung einer ebenfalls in PHP geschriebenen Umfragesoftware. Es steht unter GPL Lizenz und kann (auch kommerziell) kostenfrei auf dem eigenen Server installiert werden. LimeSurvey verfügt wie die anderen Werkzeuge über eine webbasierte Fragen- und Teilnehmerverwaltung. Darüber hinaus können mehrere Benutzer und Benutzergruppen angelegt und mit unterschiedlichen Rechten versehen werden. Der Funktionsumfang ähnelt dem kommerzieller Cloud-Produkte. Neben Einfach- und Mehrfachauswahlfragen sind auch Freitext, Matrizen und Zahlenfragen möglich. Eine rudimentäre grafische Auswertung ist ebenso wie ein Datenexport in SPSS oder Microsoft Excel vorgesehen. Die grafische Benutzeroberfläche ist in einem frischen Design gehalten und verwendet viele Symbole. Die Programmlogik sowie die Benutzeroberfläche sind aus ergonomischem Aspekt jedoch noch verbesserungsfähig. Viele Funktionen erschließen sich erst durch Ausprobieren, da eine entsprechende Beschriftung der Knöpfe fehlt. Nachdem der Quellcode des Programms frei zur Verfügung steht kann die Software mit entsprechenden Programmierkenntnissen erweitert werden. Der modulare Aufbau ermöglicht das Einbinden eigener Module (z.B. eigene Fragetypen). Das Einbeziehen von Medien ist zwar prinzipiell möglich, jedoch umständlich. Bei Websuchen finden sich etliche weitere in PHP programmierte OpenSource Programme wie Formr Survey Framework oder phpESP, phpSurvey usw., die in wesentlichen Zügen dem Funktionsumfang von LimeSurvey entsprechen. LimeSurvey kann daher als Stellvertreter für alle übrigen Lösungen dieser Gattung angesehen werden.

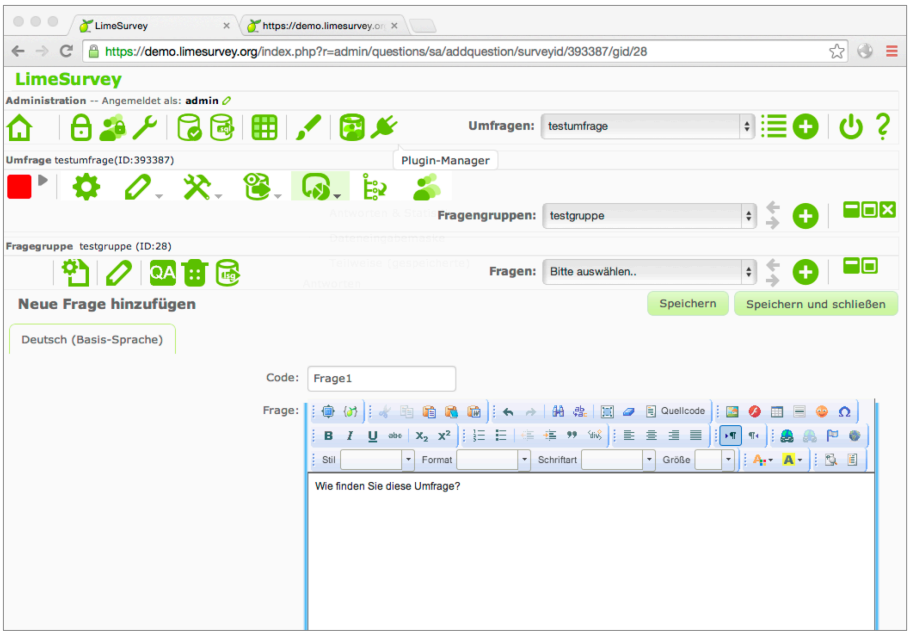


Abbildung 12: Fragebogenerstellung mit dem kostenlosen LimeSurvey

4.7 Google Forms

Auch Google bietet im Rahmen seiner kostenfreien webbasierten Office Lösung, Google Docs, ein Programm zur Erstellung von Onlineformularen an. Die Benutzeroberfläche besticht durch übersichtliche und intuitive Anordnung, die stark an die Optik von Microsoft Office angelehnt ist. Fokus ist nicht der wissenschaftliche Einsatz, sondern vielmehr die Erstellung von Umfragen im privaten oder geschäftlichen Umfeld. Daher beschränkt sich der Funktionsumfang auch auf grundlegende Fragetypen mit dem Ziel der Darstellung in Tabellenform oder als Balken/Tortendiagramm. Wie auch bei mit Google Docs erstellten Dokumenten liegen die Daten auf Servern von Google und können über die Google-Webseite verlinkt werden. Ein Export der Fragebögen in das Excel-Format ist ebenso möglich wie das Anschreiben von Teilnehmern via Email oder über Google Plus und Facebook.
Die Beschränkung der Komplexität der Umfragen beschreibt Google mit dem Speichern von Daten in 400.000 Zellen und 256 Spalten und einer maximalen Größe der Daten von 100MB. Diese Beschränkung entspricht somit der Tabellenkalkulations-Komponente von Google Docs. Bildet man einen Fragebogen in Tabellenform ab, so entspräche jeder Frage eine Spalte und jeder Zeile ein eingereichter Fragebogen. Die Anzahl der Fragen wäre dadurch auf 256 Fragen und 1562 Fragebögen beschränkt. Videos können nur via YouTube eingebunden werden (Stand 11/2014). Durch die Beschränkung auf die Anzahl der Fragen, aber auch aus datenschutzrechtlichen Gründen ist der Einsatz der Google Technik nicht empfehlenswert.

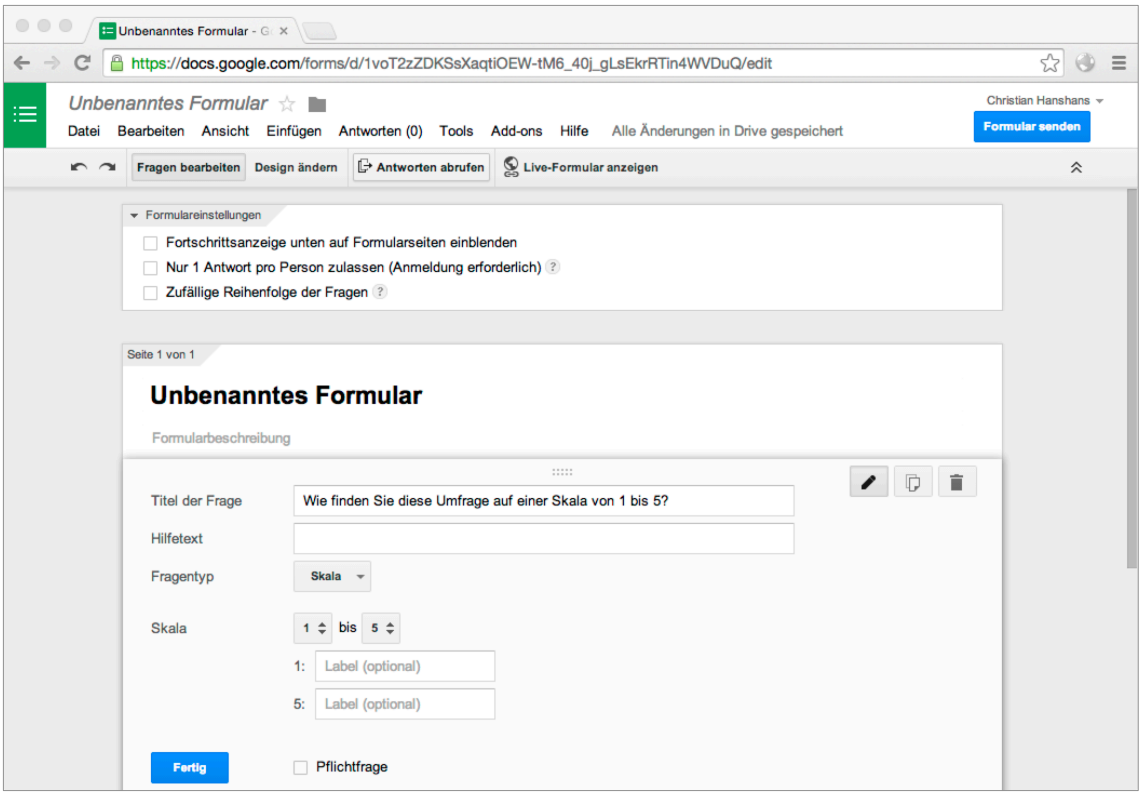


Abbildung 13: Onlinefragebögen mit Google Forms

4.8 Sphinx Survey

Mit 20 Jahren Erfahrung, gehört die Software aus Erding zu den umfangreicheren Softwarelösungen. Sphinx Desktop ist eine Windows Software zur Erstellung und Auswertung von Fragebögen in Papierform oder Online auf PDAs oder via Telefon. Die Grundversion kostet 1.995€ je PC und kann durch Module erweitert werden. So kann z.B. ein Scannermodul für das automatisierte Einlesen von Papierfragebögen für 995€ erworben werden. Für die Erstellung von Online-Fragebögen muss das Sphinx Onlinemodul für 190€ hinzugekauft werden. Möchte man Onlinestudien betreiben und die Installation auf den eigenen Server verlagern, wird zusätzlich die Serverkomponente (SphinxOnlineManager) benötigt. Sie ermöglicht, umfangreiche Studien, Mitarbeiter, Teilnehmer und Fragebögen online zu verwalten. Der Preis für die Serverlizenz beginnt ab 12.500€ (Stand 11/2014). Mit der Sphinx Cloud bietet Sphinx wahlweise ein webbasiertes Autorenwerkzeug zur Umfrageerstellung an. Ähnlich den anderen vorgestellten Lösungen erfolgt eine Preisstaffelung, abhängig vom Umfang der Fragebögen. Die Daten werden dann auf den Servern des Anbieters gespeichert.

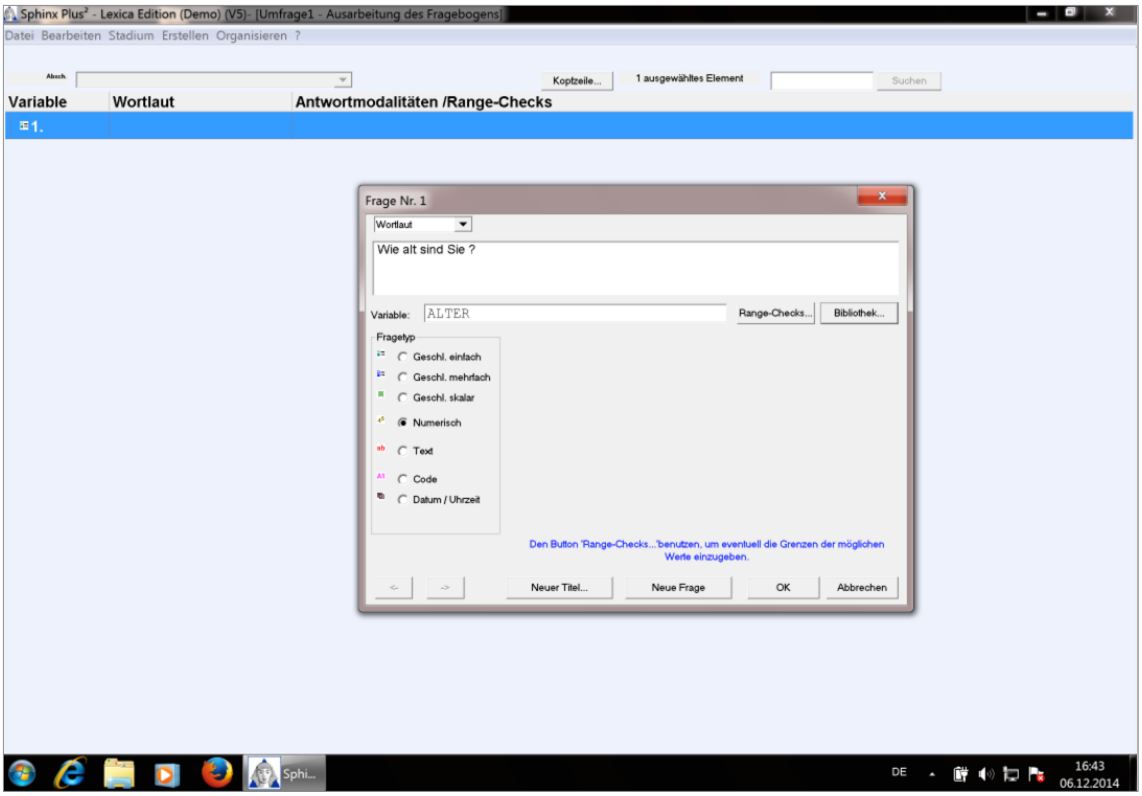


Abbildung 14: Sphinx Desktop Erstellung und Auswertung von Umfragen

4.9 NetQ

Der Anbieter NetQuestionaire bietet eine funktionsreiche Umfragesoftware. Die Umfragen werden über den Webbrowser erstellt und auf dem Server des Anbieters in den Niederlanden und der Schweiz vorgehalten. Die Jahreslizenzkosten sind abhängig von der Anzahl eingereichter Fragebögen. Für 10.000 Fragebögen pro Jahr fallen Kosten von 2.499,00€ an (Stand 11/2014). Zu den Besonderheiten des Angebots gehört die Möglichkeit der Teilnehmerrekrutierung via Webseite, Erinnerungsfunktionen und ein sogenanntes Bounce-Management. Hierunter versteht man den Umgang mit Emails, die ihren Adressaten nicht erreichen, ähnlich der Umsetzung bei Sosci Survey.

4.10 Marktforschungs-Software und Dienstleister

Für umfangreiche Datenerhebung, wie sie insbesondere für die Marktforschung oder pharmakologische Studien notwendig ist, haben sich einige Dienstleister mit spezialisierter Software am Markt etabliert. Sie bieten zusätzlich zur Software Dienste für die Umfrageerstellung sowie Beratungsleistung und Softwareschulungen an. Die Daten werden je nach Angebot auf Servern des Anbieters (oder dessen Subunternehmern) gespeichert. Die Kosten für die Dienstleistung und Bereitstellung der Technik bewegen sich je nach Anbieter jährlich im vier- bis sechsstelligen Bereich. Beispiele hierfür wären Interrogare (IRQuest), ConfirmIT (Confirmit), Amundis (2ask), Rogator (Rogator G4), Globalpark (Questback), sowie die kommerziellen Panelanbieter TNS Infratest, Respondi, Consumerfieldwork, GMI, Toluna und viele Weitere.

4.11 Eigenentwicklungen von Studienportalen

Neben kommerzieller und freier Software gibt es eine Reihe an Entwicklungen, die an unterschiedlichen Fakultäten entstanden sind. Ein Beispiel hierfür ist das WebLab des psychologischen Instituts der Universität Heidelberg. Hinter WebLab verbirgt sich ein Studienportal, über das sich Teilnehmer registrieren und Umfragen ausgerollt werden können. Registrierte Nutzer sind Teil des Systems und können für aktuelle oder zukünftige Studien rekrutiert werden. Die Heidelberger Forscher entwickeln ihre Software als *Weblab Toolkit* weiter und bieten Interessierten die Weitergabe des Quellcodes an. Das Bildschirmfoto zeigt eine Studie, die über das WebLab bereitgestellt wurde (Stand 11/2014). Leider werden personenbezogene Daten (Name, Vorname, Email) für Kontaktaufnahmen sowie die Teilnehmerregistrierung unverschlüsselt übertragen.

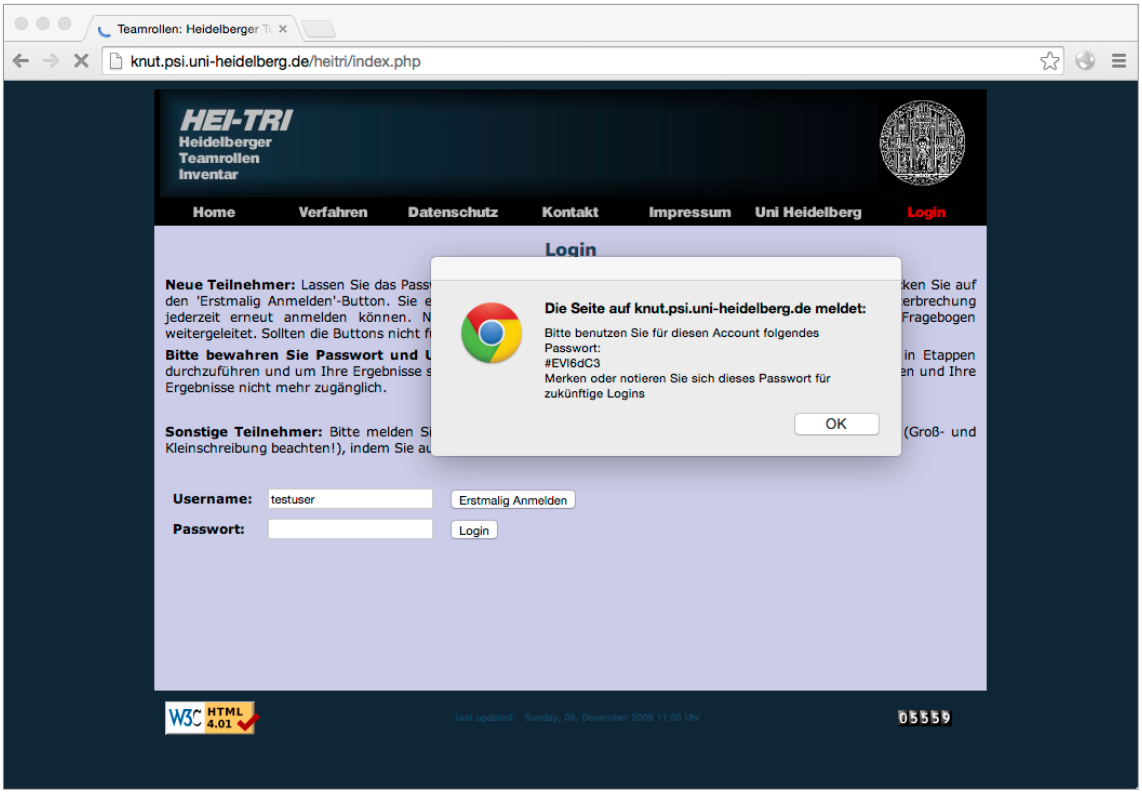


Abbildung 15: Universität Heidelberg verwendet die Eigenentwicklung WebLab

Ein anderes Beispiel ist das Freiburger WiSo Panel. Das wissenschaftliche Projekt der Universität Freiburg ist als Panel aufgebaut und dient damit primär der Gewinnung von Teilnehmern für nicht-kommerzielle Studien. Die Teilnahme an Studien wird mit kleinen Geschenken oder anderen Incentives belohnt. Die Einladung zur Teilnahme an einer geeigneten Studie erfolgt via Email. Beim Aufruf der Webadresse über die Google Suche wird die unverschlüsselt vorliegende Webadresse (*wiso-panel.net*) angezeigt. Dies hat zur Folge dass personenbezogene Daten (Name und Emailadresse) ungeschützt über das Internet übertragen werden. Die Plattform verfügt jedoch über eine weitere Webadresse (*wisopanel.net*), welche wiederum verschlüsselt angezeigt wird. Auf der Seite selbst fehlen die gesetzlich vorgeschriebene Impressumsangabe und eine Datenschutzerklärung. [19] Durch das Einbinden zweier externer Dienste (Facebook Connect Button) und eines Besuchertrackingsystems werden zudem Daten auf externen Servern (Facebook => USA, INFOnline GmbH => Bonn) erhoben, was unter Gesichtspunkten des Bundesdatenschutzgesetzes kritisch zu bewerten ist. [146] Ebenfalls

bedenklich ist, dass das Portal samt Quellcode und Datenbank mit vielen hundert anderen Internetseiten (darunter Websites auch mit erotischen Inhalten) auf einem gemeinsamen Server bei Strato untergebracht ist. [151]

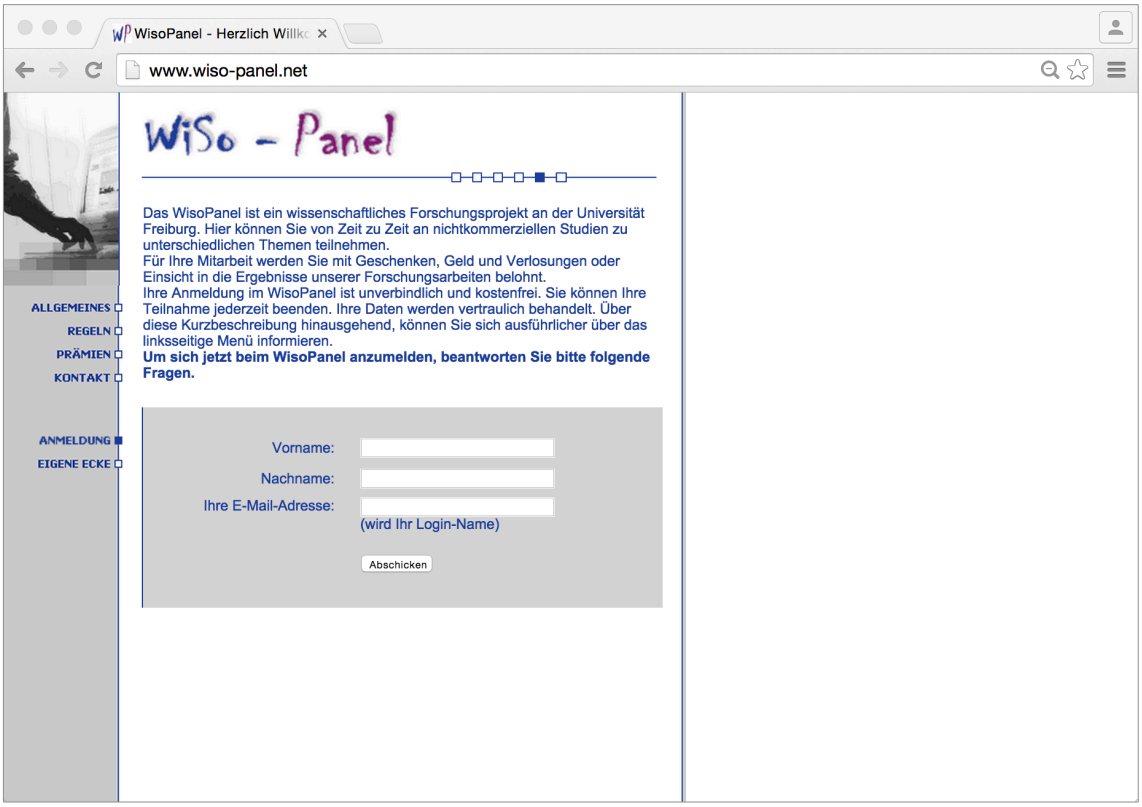


Abbildung 16: WiSo Panel - eine Eigenentwicklung der Universität Freiburg

Abbildung 17: Einladung zur Teilnahme an einer Studie des WiSo Panels mit Unipark Fragebögen

Das WiSo Panel beinhaltet selbst keine Werkzeuge für Onlinebefragungen, sondern verwendet die Cloud-basierten Fragebögen von Unipark. In den Einladungen zur Studienteilnahme, die via Email versandt werden, erfolgt die Verlinkung der Fragebögen wie im obigen Beispiel unverschlüsselt.

Hroot ist ein Akronym für **h**amburg **r**egistration and **o**rganisation **o**nline **t**ool. Die Entwicklung stammt aus dem WiSo Forschungslabor, einer zentralen Dienstleistungseinrichtung der Universität Hamburg und bildet die Grundlage des Hamburger WiSo Panels. Die Software dient der Online-Rekrutierung von Teilnehmern, der Randomisierung und dem Terminmanagement für die sozialwissenschaftlichen Testlabore. Das System bietet eine Kalenderfunktion und kann Erinnerungen an Teilnehmer via Email verschicken. Durch die ausgeklügelte Registrierungsprozedur können später Teilnehmer anhand spezifischer Merkmale gefiltert, in Experimente oder Studien eingeschlossen und verwaltet werden. Eine Erstellung von Fragebögen bietet die Software nicht an. Ihr Einsatzzweck ist primär die Organisation von Experimenten im Testlabor. [12] Für die Durchführung von Befragungen setzen die Hamburger Forscher Software von Survey System, LimeSurvey und Unipark ein. Da die Software als OpenSource -Anwendung frei zur Verfügung steht, nutzt zum Beispiel das psychologische Institut der Universität Heidelberg ebenfalls diese Softwarelösung für ihr Studienportal. [152]

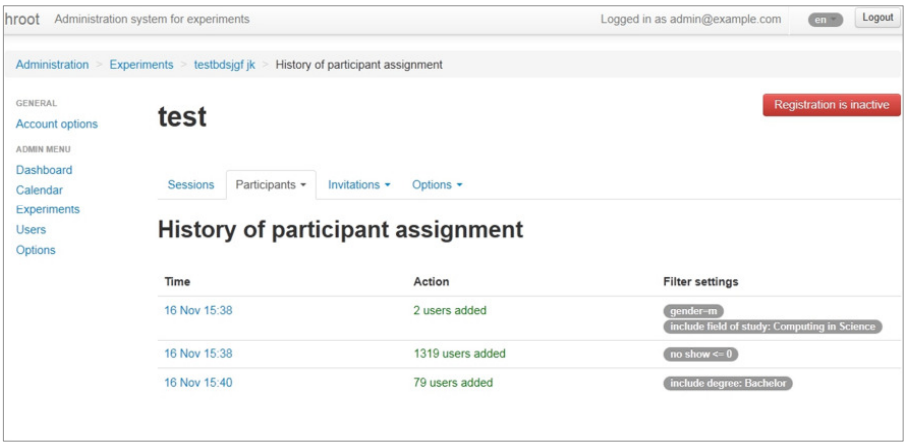


Abbildung 18: hroot ein Programm zur Rekrutierung und Verwaltung von Teilnehmern, Uni Hamburg [12]

4.12 Bewertung des Marktüberblicks

Es existieren bereits viele kommerzielle aber auch kostenfreie Werkzeuge zur Erstellung von Online-Fragebögen und zur Teilnehmerrekrutierung. Viele der genannten Programme enthalten sehr gute technische Lösungen von Problemen mit Onlinebefragungen. Jedoch vereinigt keines der untersuchten Programme alle wünschenswerten Funktionen unter einem Dach. Probleme ergeben sich zum Beispiel bei der Integration von Medien. In den meisten Fragebogensystemen können zwar Videos integriert werden, diese werden jedoch nur in einem Format ausgeliefert, was zu möglichen Inkompatibilitäten mit Browsern oder mobilen Endgeräten führen kann. Alle Cloud-basierten (reinen) Fragebogenprogramme setzen bereits bestehende Teilnehmerlisten oder Emailadressen voraus. Dies erfordert eine separate Rekrutierung von Teilnehmern via Internet und macht zusätzliche Arbeitsschritte erforderlich. Sollen potentielle Kandidaten für eine Studie über das Internet geworben werden, muss die Webseite der Studie ein Formular bereitstellen, das die Teilnehmer in einer Datenbank erfasst. Gleiches gilt für die Kommunikation mit den Teilnehmern, um sie zum Beispiel an die Abgabe von Fragebögen erinnern zu können oder um andere wichtige Mitteilungen zu versenden. Die Gestaltungsfreiheit von Fragebögen ist häufig durch die Vorgaben der Software eingeschränkt. Dies kann dazu führen, dass das Design der Onlineumfrage stark von der Webseite der Universität oder der Studienseite abweicht, insbesondere dann, wenn der Fragebogen auf eine externe Internetadresse verweist, wie dies bei Unipark, Surveymonkey oder ähnlichen

Angeboten der Fall ist. Ein mögliches Problem ergibt sich auch aus der Beschränkung auf eine bestimmte Anzahl von Fragebögen oder Datensätzen. Weiteres Manko einiger Angebote ist, dass die Möglichkeit fehlt, mehrere Benutzer mit unterschiedlichen Rechten auszustatten, so dass keine Verteilung von Verantwortlichkeiten möglich ist, bzw. dass Zugriffsrechte auf Studien oder Studienbereiche nicht reguliert werden können. Um die sichere Archivierung der Studiendaten muss sich der Betreiber beim Einsatz Cloud-basierter Lösungen für jedes Befragungsprojekt einzeln kümmern (z.B.:. in Form von CSV Dateien) oder aber den Anbieter weiterhin für die Speicherung der Daten bezahlen.

Nicht zuletzt spielen Datenschutz- und Sicherheitsaspekte eine wichtige Rolle. Die Datenhaltung sollte auf Servern der Universität oder auf Servern vertrauenswürdiger Kooperationspartner innerhalb Deutschlands erfolgen. Die Speicherung und Übertragung der Daten muss (gemäß des Bundesdatenschutzgesetzes) nach dem derzeitigen Stand der Technik verschlüsselt und vor Manipulation geschützt erfolgen.

Es stellte sich heraus, dass alle untersuchten Fragebogenprogramme oder Panels (sofern öffentlich im Internet verfügbar) Sicherheitsmängel aufwiesen. Bei vielen Angeboten werden nach wie vor unsichere Verschlüsselungsmethoden (z.B. *RC4* [18]) eingesetzt, schwache Schlüssel in Zertifikaten verwendet oder seit langem bekannte Sicherheitslücken schlichtweg nicht behoben. In einigen Fällen werden Fragebogenprogramme, Studiendatenbanken oder Emailkommunikation über Server großer Webhoster (1und1, Strato, GMX, Web.de, usw.) betrieben. Die günstigen Preise des *Shared-Hosting* sind verlockend, jedoch sicherheitstechnisch und datenschutztechnisch äußerst bedenklich, da auch hier wie bei Cloud-Angeboten kein Einfluss auf die Sicherheit des Systems und Zugriff auf die Daten von Dritten genommen werden kann. Hinzu kommt, dass große Anbieter bzw. große Serversysteme sehr beliebte Angriffsziele für Cyberkriminelle sind, da sich mit einem einmaligen Aufwand Zugang zu vielen Postfächern, Datenbanken oder Zugangsdaten verschaffen lässt. Im jüngsten Fall wurden vom WebHoster 1blue Passwörter aller Webhostingkunden erbeutet. [149] Aber auch Kunden anderer großer Anbieter (OVH, Hetzner, 1und1, Web.de, GMX,...) wurden in der Vergangenheit Opfer erfolgreicher Hackerangriffe. [154] [155] [156] [157]

Die Ergebnisse der Sicherheitsanalyse der untersuchten Angebote wurden vereinfacht in der folgenden Tabelle zusammengefasst. Eine vollständige und ausführliche Auflistung der Analyse befindet sich in elektronischer Form im Internbereich des Studienportals.

Werkzeug	Zertifikat	PFS	HSTS	Cookie Sicherheit	Rating	Bemerkung
Google Docs	+	~	-	✓	B	unsichere Verschlüsselung*, Sicherheitslücken
Surveymonkey	+	~	-	-	B	unsichere Verschlüsselung*
Unipark	+	~	-	-	B	unsichere Verschlüsselung*
LimeService	+	-	-	~	C	schwache Verschlüsselung, Sicherheitslücken
SoSciSurvey	++	~	✓	✓	C	Sicherheitslücken
Studienportal Uni Heidelberg	+	-	-	-	F	unsichere Verschlüsselung*, Sicherheitslücken, Hosting mit mdst. 7 weiteren Webseiten auf dem Server
WebLab Uni Heidelberg	-	-	-	-	F	unsichere Verschlüsselung* kein gültiges Zertifikat
WiSo Panel Uni Freiburg	+	-	-	-	F	fehlende Verschlüsselung fehlendes Impressum, fehlende Datenschutzerklärung SharedHosting mit über 80.000 weiteren Webseiten
WiSO Panel Uni Hamburg	++				F	unsichere Verschlüsselung*, diverse Sicherheitslücken

Tabelle 1: Sicherheitsanalyse aktuell verwendeter Fragebogensoftware, Stand 12/2014

Zertifikatstärke: +++ stärker 2048/256bit, ++ 2048/256bit, + 2048/128bit, - 1024/128bit
Umsetzungslevel: ✓ gute Umsetzung, ~ unvollständig umgesetzt, - nicht umgesetzt
PFS = Perfect Forward Secrecy
HSTS = HTTP Strict Transport Security (lückenlose Transportverschlüsselung)
DNSSec = verschlüsselte Übertragung der DNS-Information
DANE = DNS-based Authentication of Named Entities (Schutz vor Zertifikatfälschung)
Cookie Sicherheit = secure cookie flag, http only flag (Schutz von Login-Daten)
Qualys Rating = standardisierte Bewertung der sicheren Datenübertragung [153]
* gemäß der technischen Richtlinie des Bundesamtes für Sicherheit in der Informationstechnik (BSI) [18]

Die Untersuchung umfasste die wesentlichen Elemente für einen sicheren Transfer personenbezogener Daten. Hierzu zählen die sichere Zuordnung von Domain zu IP Adresse über das *DNS* System, der verschlüsselte Transportweg und einige grundlegende Sicherheitseinstellungen der Webanwendungen. Auf eine intensivere Analyse der jeweiligen Webanwendungen wie beim Studienportal (☞ Kapitel 7.2.2) durchgeführt, wurde nicht zuletzt aus rechtlichen Gründen verzichtet. Zur automatisierten und standardisierten Analyse und Bewertung der Transportverschlüsselung wurden die Testwerkzeuge SSL-Labs der Firma Qualys herangezogen. [158] Qualys hat sich auf Sicherheitsanalysen von Software spezialisiert und publiziert regelmäßig Empfehlungen für Konfiguration von Webservern. [57] Die von ihnen entwickelten Tests stellen einen Quasi-Standard zur Bewertung der verschlüsselten Datenübertragung von Webanwendungen dar. Gemäß des „SSL Server Rating Guide“ [158] werden Webanwendungen auf Mängel in der verschlüsselten Datenübertragung untersucht. Hierbei wird der Verbindungsaufbau mit unterschiedlichen Browsern simuliert und Sicherheitslücken, die verwendete Verschlüsselungsmethoden (☞ Kapitel 6.5.1) und viele weitere Kriterien erfasst. Jedes Kriterium wird gemessen und bewertet und geht gewichtet in die Gesamtwertung ein. Die Bewertung erfolgt analog zu amerikanischen Schulnoten mit A+ für die beste und F für die schlechteste Wertung.

Die Untersuchung zur verschlüsselten Übertagung von *DNS-Informationen* und dem Schutz vor Zertifikatfälschung wurde mit Testwerkzeugen von Verisign [159] und DNSViz [23] so-

wie über Konsolenprogramme durchgeführt. (☞ Kapitel 7.2.5) Die Analyse der Cookie Sicherheit erfolgte manuell durch Auswertung der Antwort der jeweiligen Server. Von besonderem Interesse war hierbei, ob die Webanwendung Vorkehrungen trifft, die den Diebstahl von Sitzungsinformationen (☞ Kapitel 6.5.7) oder das Einschleusen bösartiger Codes (☞ Kapitel 7.2.1.3) erschwert.

Aus den Mängeln der untersuchten Fragebogensysteme wird ersichtlich, dass die dringende Notwendigkeit besteht, sich mit dem Thema Datenschutz und Datensicherheit näher zu befassen und bei der Planung von elektronischen Studien von Beginn an zu berücksichtigen. Diesen beiden Themen wird insbesondere in den Kapiteln 6.5 und 6.6 Rechnung getragen.

5 Modell eines Studienportals

Fasst man alle bisher beschriebenen Anforderungen zusammen, so soll die für die SURE-Studie benötigte Plattform eine zentrale Anlaufstelle für Teilnehmer und Studienbetreuer darstellen, die bei der Akquise von Studieninteressenten behilflich ist und mit den Webseiten der Forschungsprojekte und Studienunterlagen inklusive Fragebögen und Medien bereitgestellt werden können. Zudem sollte die Umsetzung sowohl dem gesetzlich geforderten Datenschutz, als auch Datensicherheitsanforderungen genügen und die Möglichkeit bestehen, sich zukünftig schnell an veränderte Situationen (z.B. steigende Teilnehmerzahlen oder neue Sicherheitsanforderungen) anpassen zu können.

5.1 Personas

Mit Hilfe fiktiver Personen (*Personas*), die stellvertretend für eine Gruppe von Anwendern stehen, werden Erfahrung, Bedürfnisse und typische Alltagsaufgaben beschrieben. Die Anforderungen und Eigenschaften (z.B. EDV Kenntnisse), die diese Personen mitbringen, bilden die Grundlage für spätere Funktionen des Systems und Schwerpunkte bei der Entwicklung. *Personas* geben bei der Planung von Systemen einen schnellen Überblick über die Funktionalität eines Systems aus Anwendersicht.

1. Prof. Dr. med. Erwin Manger, 58 Jahre (Rolle: Manager)
   Prof. Manger ist Leiter der Abteilung der Psychosomatik an einem Universitätsklinikum. Er hat ein engagiertes Team junger Ärzte und Psychologen und betreut mehrere Doktoranden. Sein Hauptinteresse gilt der Patientenversorgung und Forschung. Prof. Manger überträgt seinen Mitarbeitern gerne Verantwortungsbereiche, möchte aber regelmäßig über Erfolg und Misserfolg informiert werden und den Überblick über den Stand der einzelnen Projekte behalten. Für konzeptionelle Fragen steht er jederzeit zur Verfügung. Bei der Durchführung und Ausarbeitung der Projekte lässt er seinen Mitarbeitern freie Hand. Prof. Manger ist ein technisch interessierter Mensch, dessen Hauptaugenmerk jedoch auf der Funktionalität liegt. Mit technischen Problemen möchte er sich nicht beschäftigen.
   Für Prof. Manger ist es wichtig, dass er für seine Mitarbeiter selbstständig neue Studien anlegen und Zuständigkeiten definieren kann. Berechtigungen für Teilbereiche möchte er für Mitarbeiter und Doktoranden einstellen können. Es ist ihm ein Anliegen, sich in laufende Projekte elektronisch einzuklinken, um den aktuellen Stand abfragen zu können.

2. Frau Dr. phil. Christina Elster 33 Jahre (Rolle: Studienbetreuer)
   Frau Elster ist Psychologin. Mit der Standardsoftware von Microsoft Office: Word, Excel und Power Point ist sie bestens vertraut. Sie besitzt ein Smartphone und nutzt bevorzugt Emails zur Kommunikation. Sie publiziert mehrmals jährlich in entsprechenden Fachzeitschriften und beherrscht die Statistiksoftware SPSS, mit deren Hilfe sie ihre Studiendaten analysiert.
   Ihr Interesse gilt neben der Patientenversorgung, der Organisation und Durchführung von klinischen Studien anhand von Befragungen. Die Rekrutierung von Studienteilnehmern erfolgt meist postalisch, durch Medien oder durch direkten Kontakt zur Studienpopulation. Die Erhebung der Studiendaten findet in der Regel in Papierform mit standardisierten Fragebögen statt. Digitalisiert werden die gesammelten Studiendaten mit Hilfe von Excel-Listen. Die Auswertung erfolgt meist mit SPSS. Frau Elster er-

hofft sich von einem elektronischen System eine schnellere Erfassung von Studienteilnehmern, bereits in elektronischer Form vorliegende Studiendaten und weniger Papier auf ihrem Schreibtisch.

3. Cand. Med. Christoph Dürr, 23 Jahre (Rolle: Studienassistent)
   Herr Dürr studiert Medizin im 7. Fachsemester. Er ist aktiver Nutzer von sozialen Netzen (Facebook, StudiVZ, LinkedIn). Auf seinem neuen iPhone nutzt er regelmäßig Apps zur Kommunikation, für Spiele und auch einige Medizin-Apps. In der WG verfügt er über eine schnelle Internetleitung und schaut gerne Videoclips auf Streaming-Portalen wie YouTube oder Vimeo. Die Nutzung des Internets zum Studium und in der Freizeit ist er gewohnt. Herrn Dürr interessiert die Ursache von psychosomatischen Erkrankungen. Daher hat er sich entschlossen seine Doktorarbeit in der Psychosomatik zu schreiben. Von Prof. Manger bekam er die Aufgabe „– Chronische Schmerzerkrankungen bei Patienten mit posttraumatischen Belastungsstörungen „– zu untersuchen. Zu diesem Zweck soll er aus Daten einer Studie von Frau Elster ein repräsentatives Kollektiv selektieren. Mit Hilfe von Vergleichsdaten aus Querschnittsstudien soll untersucht werden, ob ein Zusammenhang zwischen Auftreten von chronischen Schmerzen und posttraumatischen Belastungsstörungen besteht.
   Herr Dürr benötigt hierzu Zugriff auf Studiendaten von Frau Elster. Für seine Arbeit ist es nicht erforderlich und auch nicht gewünscht, Vollzugriff auf die Studie zu erhalten. Er soll jedoch die Daten aus der Studie anonymisiert exportieren. Erfahrung mit Statistiksoftware konnte er bisher nicht sammeln. Er möchte die Daten daher mit Microsoft Excel auswerten.

   Er bat Frau Elster, bei der Bearbeitung von Anfragen der Studienteilnehmer (per Email oder via Online-System) zu helfen.

4. Markus Teile, 53 Jahre (Rolle: Teilnehmer)
   Als langjähriger Rettungsdienstmitarbeiter kennt Herr Teile sein Geschäft. Er hat schon einige schwere Einsätze gefahren. Ein paar seiner Kollegen haben psychisch unter dem Erlebten gelitten. Er interessiert sich besonders dafür, wie er sich selbst schützen oder auch betroffenen Kollegen weiterhelfen kann. Daher hat er sich bereit erklärt, an der SURE- Studie teilzunehmen.
   Herr Teile besitzt zu Hause einen alten Computer, der hin und wieder spinnt. Wenn er etwas recherchieren möchte, tut er das in der Leitstelle. Dort steht ein neuer PC mit schnellem Internetanschluss. Kürzlich hat er sich einen Tablet Computer gekauft. Ein Gerät von Aldi, „irgendwas mit Android stand auf der Verpackung“. Damit möchte er in Zukunft unterwegs Zeitung lesen oder Bilder von seiner Digitalkamera ansehen. So richtig kenne er sich damit jedoch noch nicht aus. Herr Teile nutzt den Computer privat hauptsächlich um Emails zu beantworten oder im Internet beruflich zu recherchieren. Mit der Software auf der Arbeit kennt er sich aus. Er hat auch schon viele EDV Fortbildungen besucht. Einige seiner Kollegen fragen ihn, wenn sie mit dem System nicht klar kommen. Etwa zwei- bis dreimal pro Woche prüft er sein privates Email-Postfach. Häufig nutzt er Kartendienste zur Routenplanung oder die Fahrplanauskunft der Bahn.

5. Dipl. Inf. Alexander Rug, 27 (Rolle: Administrator)
   Herr Rug ist ein junger, motivierter Informatiker. Er hat eine halbe Stelle als wissenschaftlicher Mitarbeiter in der Psychosomatik und eine Promotionsstelle an der Hochschule Regensburg im Fachbereich Mathematik und Informatik.

Herr Rug unterstützt die Mitarbeiter und Mitarbeiterinnen der Psychosomatik, insbesondere, wenn es um das Exportieren von Daten aus Datenbanken geht bzw. bei der Auswertung der Daten mit Statistiksoftware. Neben dieser Tätigkeit pflegt er als Webmaster die Webseite des Instituts und entwickelt einfache Webformulare zur Erfassung von Umfragen via Internet.
Mit Einführung einer Studienplattform soll er die Funktion des Administrators übernehmen und für regelmäßige Sicherheitsupdates und Wartung der Datenbank sorgen. Für diese Arbeit benötigt er Zugang zum Betriebssystem des Servers. Zugriff auf Daten von Studienteilnehmern benötigt er jedoch nicht.

5.2 Anwendungsfälle (Use Cases)

Aus der Beschreibung der *Personas* und deren Anforderungen lassen sich Anwendungsfälle in Form von *Use Cases* ableiten. *Use Case* Diagramme sind als abstrakte Modellierungswerkzeuge in der Lage, unabhängig von der Software, die Interaktion eines Nutzers mit dem System bzw. anderen Akteuren des Systems zu visualisieren. Jeder Anwendungsfall ist ein komplexeres Szenario, das weiter im Detail analysiert werden muss.
Im Folgenden sollen für die notwendigen Rollen des Systems nur die wesentlichen *Use Cases* beschrieben werden.

Manager:
Zu den Hauptaufgaben des (Studien-)Managers gehört die Verwaltung von Studien und Mitarbeitern. Das Anlegen neuer Studien und die Vergabe der Berechtigung an Studienbetreuer bzw. Studienassistenten sind jeweils Anwendungsfälle. Zudem muss der Manager in der Lage sein, das Personal im System zu verwalten (z.B. anlegen oder löschen von Mitarbeitern) und auch den Verlauf aller Studien zu überwachen. Darüber hinaus verfügt er über alle Rechte, die den Rollen „Studienbetreuer“ und „Studienassistent“ zur Verfügung stehen.

Studienbetreuer:
Der Studienbetreuer ist für die Durchführung von Studien verantwortlich. Als Anwendungsfälle ergeben sich daraus das Bereitstellen von Studienunterlagen, Veröffentlichen von Medien, Anlegen von Fragebögen und die Betreuung der Studienteilnehmer. Hinzu kommt der Datenexport eingereichter Fragebögen sowie die Archivierung der Daten einer Studie. Die Auswertung der Daten erfolgt extern und ist daher nicht Teil des Systems.

Studienassistent:
Studienassistenten verfügen im Vergleich zum Studienbetreuer über eingeschränkte Berechtigung. Sie dürfen zwar auf die Daten einer Studie zugreifen und auch mit Teilnehmern interagieren, nicht jedoch Fragebögen erstellen oder Studienunterlagen veröffentlichen.

Administrator:
Für die Rolle des Administrators ist ein Zugriff auf Daten von Studienteilnehmern nicht notwendig. Seine Aufgabe bezieht sich in erster Linie auf die Wartung des Servers und der Plattform. Hierzu gehören auch die Sicherung und Optimierung der gesamten Datenbank, was einen Zugriff auf die Datenbank des Systems (nicht zwangsweise des Systems selbst) notwendig macht.

Als Konvention bei *Use Cases* gilt:

---------- := Beziehung
----------> := Abhängigkeit
---*extend*---> := Anwendungsfall kann ausgeführt werden (Sonderfall)
---*include*--> := Anwendungsfall muss ausgeführt

Eine einfache Linie (A------B) beschreibt eine Beziehung zwischen Akteur (A) und (B). Ein Pfeil (A------>B) zeigt eine Abhängigkeit zwischen zwei Akteuren (oder *Use Cases*) an.
Dies bedeutet, dass Anwendungsfälle, die A besitzt auf B übertragen werden. In der Informatik spricht man von *Vererbung*. B *erbt Use Cases* von A, ist also eine Erweiterung von A. Ein Manager besitzt demnach alle Anwendungsfälle eines Studienassistenten und Studienbetreuers.

Möchte man diese technische Schreibweise in Worte übersetzen, so würde aus einer einfach gestrichelten Linie (----------) ein *Darf* und aus ---*include*---> ein *Muss*.
Ein Studienteilnehmer darf an einer Studie teilnehmen, muss sich dazu aber zunächst registrieren und am System anmelden. Er darf einen Fragebogen einreichen. Dieser muss aber zwingend vom System vor dem Speichern anonymisiert werden.
Mit ---*extend*---> wird ein Sonderfall beschrieben, der eintreten kann, aber nicht zwangsweise muss. Ein Teilnehmer kann (muss aber nicht!) Fragebögen korrigieren. Die zwingende Erweiterung eines Anwendungsfalls wird als ---*include*---> bezeichnet.

Die folgende Grafik visualisiert die *Use Cases* stark vereinfacht.

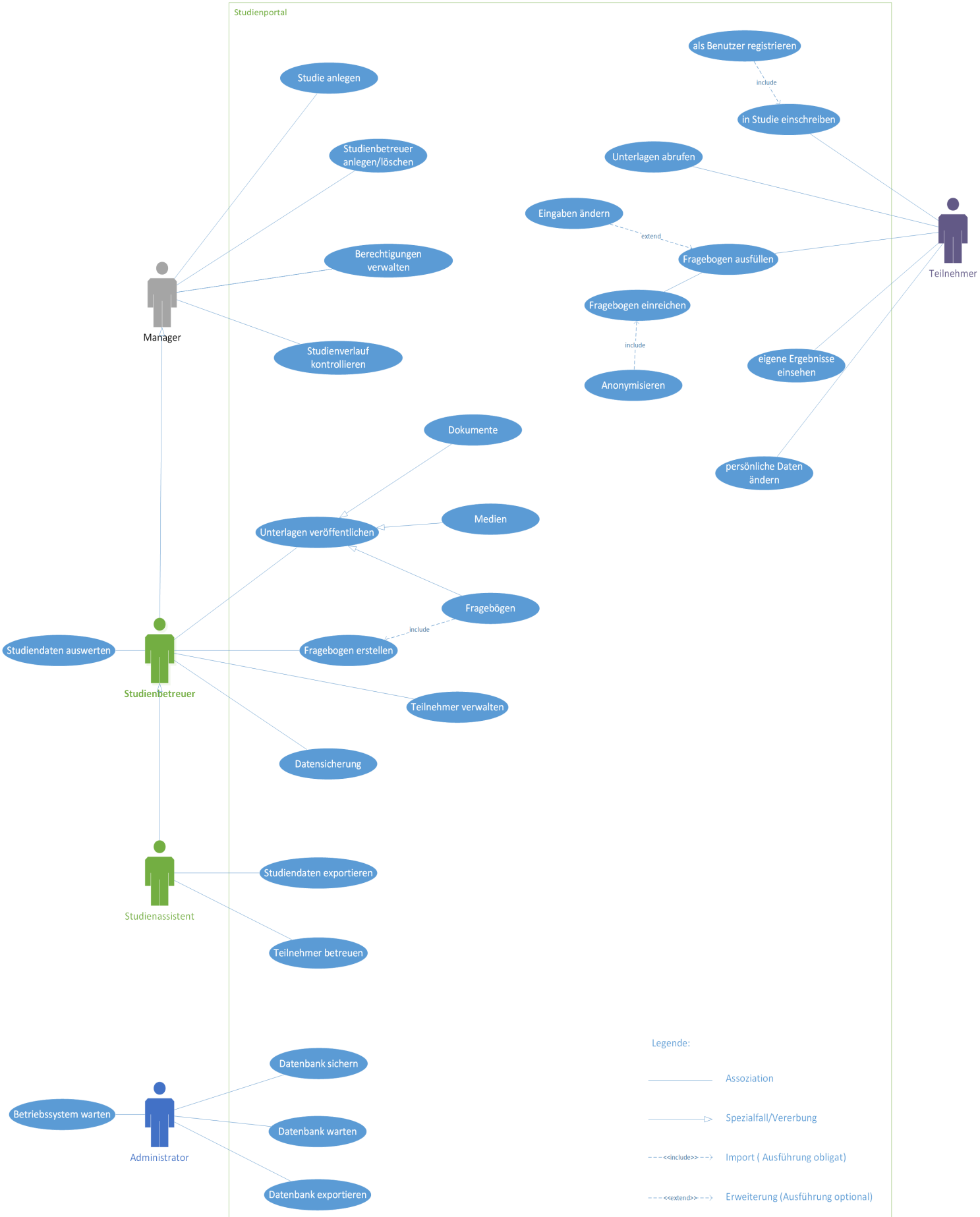


Abbildung 19: Vereinfachte UseCases des Studienportals im Überblick

5.3 Prozessmodelle

5.3.1 Registrierung der Teilnehmer

Die Teilnahme an einer Studie setzt zunächst voraus, dass der Teilnehmer ein registrierter Benutzer der Studienplattform ist und sich mit einem Benutzernamen und Passwort elektronisch ausweisen kann. In einem Registrierungsprozess, wie ihn die meisten Benutzer von anderen Internetportalen oder Webshops kennen dürften, wählt der Teilnehmer ein Pseudonym als Benutzernamen und ein Passwort aus. Für die Kommunikation des Systems und um Missbrauch auszuschließen, wird eine gültige Emailadresse benötigt. An diese Emailadresse wird eine Email mit der Bitte verschickt, einen Aktivierungslink zu bestätigen. Durch diese Bestätigung ist sichergestellt, dass der Benutzer unter der angegebenen Emailadresse erreichbar ist.

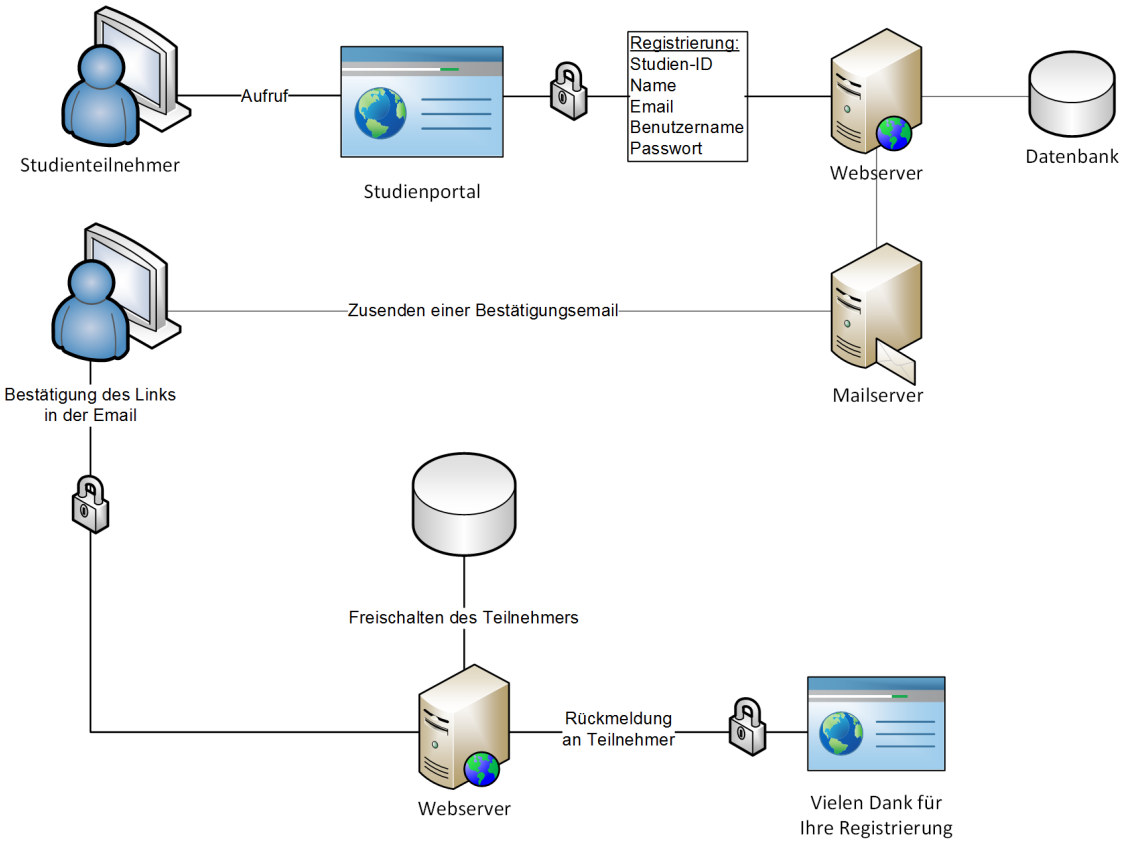


Abbildung 20: Schema des Registrierungsprozesses

5.3.2 Studienteilnahme

Ist ein Teilnehmer als Benutzer registriert und (manuell oder automatisiert) in eine Studie eingeschlossen worden, können die Studienunterlagen in Form von Hinweisen, Anleitungen, Datenschutzerklärungen, Videos, Audio-Dateien usw. abgerufen werden.
Abhängig von der gewählten Einstellung können Fragebögen einfach oder mehrfach bzw. zeitabhängig ausgefüllt werden.

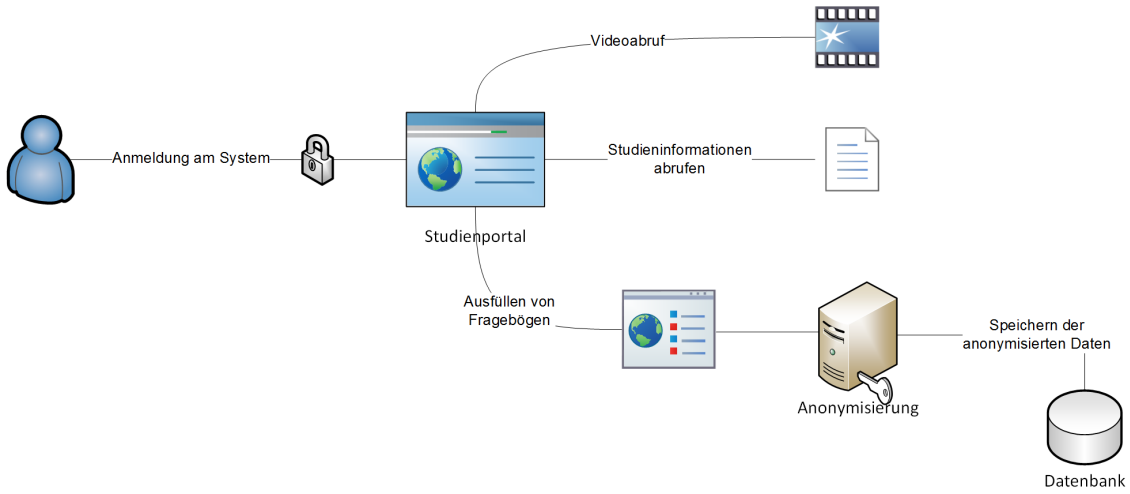


Abbildung 21: Schema des Abrufs von Studienunterlagen

5.3.3 Zugriff auf Studiendaten

Die Information der eingereichten Fragebögen wird neben zusätzlichen Metadaten wie einem Zeitstempel, der Bearbeitungsdauer, eventuellen Unterbrechungen als Datensatz in der Datenbank gespeichert. Einmal eingereichte Fragebögen können wie das analoge Vorbild nicht mehr verändert werden. Vollständig ausgefüllte und eingereichte Fragebögen stehen dem Studienmanager, den Studienbetreuern und Studienassistenten zur Auswertung zur Verfügung. Hierbei können die Daten der Fragebögen in Form einer Textdatei anonymisiert aus dem System exportiert werden. Die Datenbank selbst muss hierzu nicht geöffnet werden. Die Textdatei wird in Form einer Komma getrennten Liste (*CSV Datei*) ausgegeben und kann dadurch in vom Anwender bevorzugten Programmen weiterverarbeitet werden. Daten zur Nutzung des Systems, die Anzahl der Teilnehmer oder die Anzahl bereits eingereichter Fragebögen in Form von Tabellen, Listen oder Histogrammen, können mit Hilfe des Systems angezeigt und ausgedruckt werden.

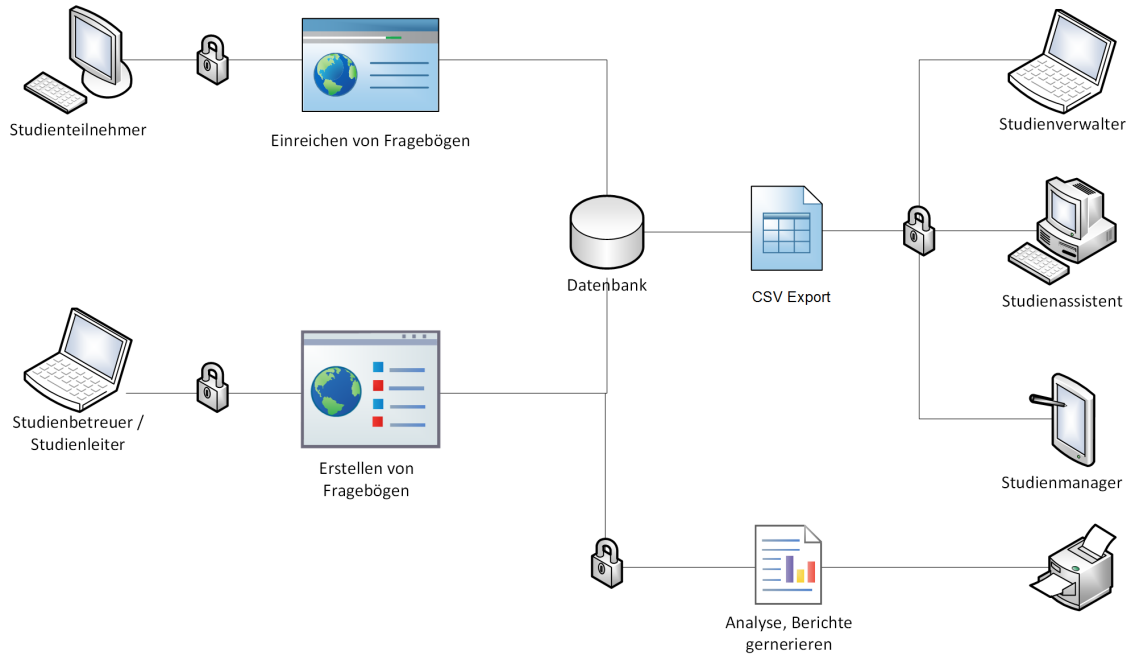


Abbildung 22: Schema Abruf und Auswertung von Studiendaten

6 Problemanalyse und Problemlösung

6.1 Rekrutierung großer Studienpopulationen

Das Erfassen großer Studienpopulationen ist für statistisch aussagekräftige Studien unabdingbar. Je nach Art und Umfang der Studie ist die Auswahl und Erfassung der Teilnehmer ein logistisches Problem. Mit Hilfe internetbasierter Systeme können diese Prozesse jedoch sehr gut automatisiert werden. Die Teilnehmer melden sich über ein elektronisches Formular an und werden automatisch in eine Studie eingeschrieben oder randomisiert in Studiengruppen eingeteilt. Neben der Logistik muss zunächst jedoch die Motivation für die Teilnahme an der Studie geschaffen werden. Hierfür kommen mehrere Kommunikationskanäle in Frage.

6.1.1 Email

Der Einsatz von Emails ist günstig und schnell. Innerhalb von Sekunden können Anschreiben an viele Tausende potentielle Studienkandidaten verschickt werden. In Zeiten von Spam und Massenmails gilt jedoch zu bedenken, dass Emails nicht selten als Spam eingeordnet und automatisch gelöscht oder in Spam-Ordner verschoben werden.
Auf dem Versandweg können Emails verlorengehen (z.B. durch volle Mailboxen). Es gibt keine Sicherheit, dass die versendete Nachricht tatsächlich beim Empfänger angekommen ist und gelesen wurde. Die von vielen Anwendern verwendete Lesebestätigung setzt ein Mailprogramm voraus das diese Funktion unterstützt. Viele Webmail-Postfächer, freie Emailprogramme (z.B. bei Linux Betriebssystemen) oder mobile Endgeräte unterstützen diese Funktion jedoch nicht.

6.1.2 Medien und Mundpropaganda

Medien fördern den Bekanntheitsgrad oder schaffen Neugierde. Gelingt es, über Radio, TV oder Internetkanäle Aufmerksamkeit zu erregen, kann dies dazu beitragen die Motivation für die Teilnahme zu fördern. Eine ähnliche, wenn nicht sogar noch wirkungsvollere Möglichkeit der Teilnehmerrekrutierung ist das Ausnutzen der Mundpropaganda. Ist ein Thema aktuell und interessant, verbreitet es sich insbesondere innerhalb gut vernetzter Kreise, wie dies bei den geplanten Studiengruppen (Einsatzkräfte) der Fall ist, schnell weiter. Es lohnt sich daher den Effekt klassischer Verbreitungswege auszunutzen, nicht zuletzt um auch weniger Internetinterent- und computer-affine Teilnehmer zu erreichen.

6.1.3 Soziale Netze

Eine sehr effiziente Möglichkeit Aufmerksamkeit zu erreichen ist der Einsatz von sozialen Netzen. Dieser Effekt ist auch der Werbeindustrie bekannt und wird zunehmend ausgenutzt. Über soziale Netze wie Facebook, Xing, oder Ähnliche verbreiten sich Informationen innerhalb der Nutzerkreise extrem schnell.
Mit einem mathematischen Modell konnten Forscher des Max-Plank-Instituts für Informatik nachweisen, dass sich Gerüchte innerhalb sozialer Netzwerke wie Facebook im Vergleich zu direkter Kommunikation zwischen Mitgliedern schneller als in jedem bekannten Computer-Netzwerk in sublogarithmischer Zeit verbreiten. [31] [30] Die Ursache hierfür ist ein Zusammenspiel zwischen beliebten und sehr gut vernetzten und gering vernetzten Personen. Diese

Erkenntnis macht sich auch *virales Marketing* zu Nutze. Hinter dem Begriff steckt das Prinzip der Mund-zu-Mund-Propaganda mit dem Unterschied, dass sich Nachrichten über soziale Netze viel effizienter und schneller transportieren lassen. Durch Funktionen wie *Like*, *ReTweet* oder *Share* von Artikeln, Blog-Einträgen, Webclips oder eigenständig verfassten Nachrichten, wird der digitale Freundeskreis über interessante Inhalte rasch informiert.

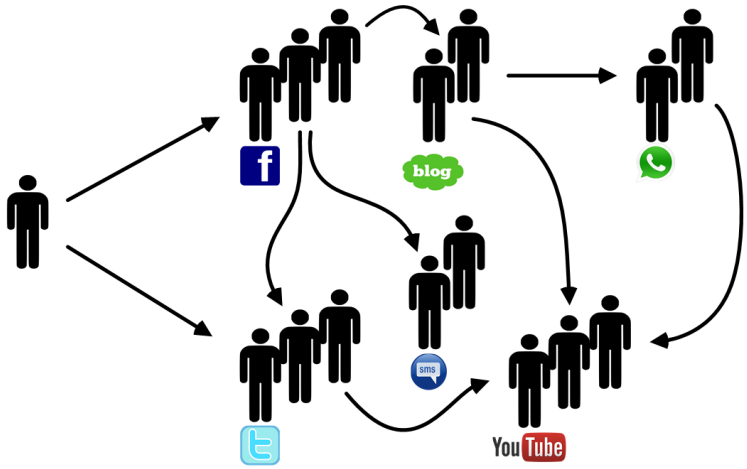


Abbildung 23: Funktionsweise der Informationsverbreitung über Social Media

Die JIM-Studie 2013 zeigte in einer repräsentativen Umfrage, dass 80% Prozent aller Jugendlichen zwischen 12 und 19 Jahren in sozialen Netzwerken aktiv sind. 77% ihrer Onlinezeit verbringen die Jugendlichen in sozialen Netzen wie Facebook, StudiVZ oder anderen Online-Communities. [39] Auch auf Smartphones sind entsprechende Apps in 61% aller Fälle anzutreffen. Jüngste demographische Daten (Januar 2014) aus den USA zeigen außerdem, dass soziale Netze in stark zunehmendem Maße auch für die Altersgruppe über 55 Jahre an Bedeutung gewinnen. Mit der höchsten Zuwachsrate aller Altersgruppen (80,4%) steigt dank zunehmendem Interesse der Generation „55+“ an Facebook die Anzahl der Facebook-Mitglieder auf über 28 Millionen. Auch wenn diese Daten von einer Werbeagentur erhoben wurden und sicher kritisch zu hinterfragen sind, zeichnet sich zumindest ein deutlicher Trend ab. [106]

Ähnlich einer Virusepidemie gibt es grundlegende Prinzipien, die für den Erfolg der Verbreitung von entscheidender Rolle sind. Schlüsselpersonen, die „infiziert“ sind und einen „engen Kontakt“ zu vielen anderen Personen pflegen. Das Internet ist hierbei ein schnelles und grenzüberschreitendes Fortbewegungsmittel. Besonders provokative, witzige oder skurrile Inhalte verbreiten sich schnell und weitreichend.
Soziale Netze können demnach für alle Altersgruppen ein effizientes Medium zur Gewinnung von Studienteilnehmern sein. Allerdings entscheidet die Relevanz (Interessen der Zielgruppe) und die Außergewöhnlichkeit für das jeweilige Kollektiv über den Grad der Verbreitung. Diese Erkenntnisse setzen als Konsequenz eine zielgruppengerechte Präsentation der Inhalte voraus. Videoclips oder Webseiten müssen potentielle Teilnehmer persönlich ansprechen.

6.1.4 Suchmaschine und eigene Webseite

Eine weitere Möglichkeit potentielle Studienteilnehmer zu rekrutieren ist die Webseite der Studie selbst. Um von Suchmaschinen besser gefunden zu werden, existieren diverse Verfahren zur Suchmaschinenoptimierung (*SEO = search engine optimization*). Jede Internetseite kann mit Metadaten und Schlagworten versehen werden. Suchmaschinen lesen diese Informa-

tion und verweisen bei Übereinstimmungen mit der Suchanfrage des Nutzers auf passende Seiten. Über den Rang, also die Position in der Trefferliste, entscheiden viele unterschiedliche Kriterien, die von Suchmaschine zu Suchmaschine unterschiedlich sind. Durch Optimierung der Schlagworte, aussagekräftige Seitentitel und passende Domainnamen, Barrierefreiheit und gute Verschlüsselung lässt sich bei der am häufigsten verwendeten Suchmaschine (Google) bereits ein guter Effekt erzielen. [44]

Suchmaschinenoptimierung ist ein komplexes Feld, das von der Optimierung auf die unterschiedlichen Browser und Endgeräte, über die Optimierung von Ladezeiten bis hin zur Integration sozialer Netzwerke und das Einbetten dynamischer Inhalte reicht.

6.1.5 Verlinkung auf externen Seiten und Forschungsdatenbanken

Im Internet existieren etliche Portal-Seiten in deren Datenbanken Studien eingetragen werden können. Auch auf diese Weise kann eine Studie von potentiellen Teilnehmern gefunden werden. Eines der größten Studienportale, das von der amerikanischen Regierung unterhalten wird, ist unter der Adresse: *http://clinicaltrials.gov* erreichbar. Ein weiteres internationales Portal, das von einem Verbund pharmazeutischer Firmen als non-profit Organisation betrieben wird, ist unter *http://clinicaltrials.ifpma.org* zu finden. Europäische Forschungsprojekte können in das europäische Forschungsregister unter *http://clinicaltrialsregister.eu* eingetragen werden. Für Deutschland gibt es eine neue unabhängig Plattform zur Erfassung von klinischen Studien:, das Deutsche Register Klinischer Studien. Das Portal ist unter den Webseiten des Universitätsklinikums Freiburg *http://drks-neu.uniklinik-freiburg.de/* aufrufbar.
Neben den bereits genannten Anlaufstellen, existieren fachspezifische Portale der jeweiligen Fachgesellschaften oder von Drittmittelorganisationen wie der DFG, BMBF sowie Studiendatenbanken der Universitäten.

6.2 Adhärenz der Studienteilnehmer

Ein allgemeines Problem bei der Durchführung von klinischen Studien ist die Adhärenz der Studienteilnehmer. Insbesondere bei Studien, die sich über einen längeren Zeitraum erstrecken, ist der Anteil der Studienabbrecher nicht unerheblich. Neben vorzeitigem Ausscheiden aus Studien kommt es häufig zu Fehlanwendungen bzw. zum Missachten der Studienanweisungen. Hierdurch wird das Studienergebnis verfälscht. Um Fehler durch unsachgemäße Anwendung zu verhindern und die Abbruchquote gering zu halten, ist es unerlässlich, technische Hürden zu minimieren und Studienanweisungen so intuitiv und einfach wie möglich zu gestalten. Darüber hinaus, müssen für internetbasierte Systeme möglichst alle Eigenschaften, die konventionelle Studien bieten in elektronischer Form abgebildet werden. Onlineplattformen bieten im Bereich des Medieneinsatzes Vorteile gegenüber konventionellen Studien. Diese Vorteile gilt es zu nutzen und technisch optimal umzusetzen.

6.2.1 Einsatz von individuellen Startseiten (Landing-Pages)

Der erste Kontakt des potentiellen Studienteilnehmers erfolgt über die Startseite.
Meist entscheidet der erste (visuelle) Eindruck darüber, ob ein Angebot als interessant oder uninteressant eingestuft wird. Große Onlineplattformen können jedoch nur selten auf die Bedürfnisse einzelner Zielgruppen eingehen, da sie auf die Bereitstellung eines umfangreichen

Angebots ausgelegt sind. Ein Studienportal, dessen Aufgabe es ist, eine oder mehrere Studien bereitzustellen, lässt sich daher gut mit Webinhalten großer Portale und Marktplätze vergleichen. Im Online-Marketing hat sich daher ein effektives Verfahren für den zielgruppenorientierten Erstkontakt etabliert. Hierbei wird der eigentlichen Portalseite eine zielgruppengerechte Seite vorgeschaltet, auf der die Benutzer bei Suchanfragen landen sollen.
Eine solche *Landing-Page* kann auch dem Studienportal vorgeschaltet werden. Der Teilnehmer kann sich über diese spezielle Internetseite über die Studie informieren und sich über ein elektronisches Formular zur Teilnahme an der Studie anmelden.
Der Vorteil der *Landing-Page* besteht darin, dass sie - im Gegensatz zum Studienportal - individuell gestaltet werden kann. Dies bedeutet, dass eine *Landing-Page* sowohl vom Design als auch vom Inhalt (z.B. Zusatzinformationen),) auf die Studienpopulation zugeschnitten werden kann. Im biostatistischen Sinne kann über eine *Landing-Page* eine Stratifizierung - also eine Schichtung der Studienpopulation realisiert werden. Im Kontext von SURE ermöglicht es eine Trennung der Studienpopulation in folgende Gruppen: Bundeswehr, Rettungsdienst, Pflege und Polizei. Jede der Gruppen weist ein anderes Tätigkeitsfeld mit unterschiedlichen Belastungsfaktoren auf und hat daher auch unterschiedliche Beweggründe für eine Studienteilnahme. Durch die zielgruppenorientierte Gestaltung kann auf die technischen Vorkenntnisse, Bedürfnisse und Erwartungen der Adressaten individuell eingegangen werden. Gut gestaltete *Landing-Pages* erhöhen das Vertrauen der Interessenten, verringern Vorbehalte gegenüber der Online-Nutzung und können den individuellen Nutzen durch die Teilnahme an der Studie transportieren.
Als weiteres förderliches Mittel können der Sprachjargon sowie die Verwendung von Beispielen aus dem Erfahrungsbereich der Zielgruppe, ansprechende Bilder, Symbole oder andere Medien dienen. So könnte auf der *Landing-Page* ein Video den Studienteilnehmer begrüßen und ihm den Zweck und persönlichen Nutzen der Studie erklären. [92] Der Einsatz von *HTML* und *CSS* Vorlagen zur Gestaltung der *Landing-Pages* ermöglicht es Webdesignern oder wissenschaftlichen Mitarbeitern auch ohne Programmierkenntnisse schnell neue *Landing-Pages* (z.B. mit Hilfe von Baukastensystemen) zu kreieren. Der Erfolg der vorgeschalteten Startseiten begründet sich auch auf einer besseren Möglichkeit der Verknüpfung mit sozialen Netzen und Suchmaschinenoptimierung z.B. durch den Einsatz gezielter Schlagworte (*Meta-Tags)*. Die Folge ist, dass die Webseite im Netz besser sichtbar ist und von potentiellen Interessenten besser gefunden werden kann. *Landing-Pages* wurden in zahlreichen Case Studies untersucht und werden aufgrund ihrer Wirksamkeit im Bereich des Online-Marketings intensiv genutzt. [160]
Dass Methoden des Online-Marketings nicht nur Webagenturen zur Steigerung des Absatzes von Onlineshops oder von kommerziellen Internetseiten vorbehalten sind zeigt ein Konzeptpapier der Technischen Universität Dresden. Das Projekt wird vom sächsischen Staatsministerium für Wissenschaft und Kunst unter der Kampagne „Pack dein Studium“ gefördert. Hierbei wurden seit 2009 Internetbefragungen unter Studierenden durchgeführt, um die Bedürfnisse und Interessen von Studieninteressenten und Studienanfängern herauszufinden. Aus den Befragungen ging hervor, dass 62% der Studienbewerber ihre Erstinformation zum Studium und Studienort aus dem Internet beziehen. Nur 40% bewerteten die Information des Universitätsportals als übersichtlich und informativ. 52,2% der Befragten wünschten sich eine spezielle Webseite für Erstsemester. Das Ergebnis der Umfrage führte zur Umsetzung einer *Landing-Page* für Erstsemester. Relevante Informationen dieser Zielgruppe wie Freizeitmöglichkeiten, Wohnungssuche, Erstsemesterveranstaltungen oder Lagepläne wurden in den Vordergrund gestellt. Durch sprachliche Gestaltung mit klaren Botschaften und Mottos konnte eine direkte Adressierung an junge Studierende umgesetzt werden. Innerhalb der Erstsemester-Webseite wurde auch die Anbindung an Fremdsysteme wie Facebook geplant. [32]

6.2.2 multimediale Unterstützung

Ein wesentlicher Vorteil internetbasierter Systeme ist die Möglichkeit multimediale Inhalte einsetzen zu können. Dies können Audioaufzeichnungen, Videos, Animationen oder Spiele sein. Im Vergleich zu konventionellen Studien stellt die Möglichkeit der Visualisierung einen klaren Vorteil dar. Medien können via Internet schnell und kostengünstig bereitgestellt und von den Teilnehmern einfach abgefragt werden. Die zeitunabhängige Zugriffsmöglichkeit bringt Flexibilität, die in traditionellen Studien nur schwer zu erreichen ist.
Analog zu präsenzbasierten Studien können interaktive Computerprogramme Tests ersetzen, für die üblicherweise Personal oder technische Ausrüstung vor Ort notwendig wären. So können Reaktionstests, Aufmerksamkeitstests oder neuropsychologische Überprüfungen mit Hilfe internetbasierter Programme am Bildschirm der Teilnehmer und somit ohne Bereitstellung von Räumlichkeiten, Personal oder Computern durchgeführt werden. Die Ergebnisse der Tests werden direkt in die Datenbank eingespielt und können sofort ausgewertet werden.

6.2.3 Erinnerungsmöglichkeit bei nicht eingereichten Fragebögen

Ist das mehrfache Ausfüllen von Fragebögen in bestimmten Zeitintervallen notwendig (Verlaufskontrolle oder Follow-up Studien), können Studienteilnehmer vom System automatisiert an die Abgabe des Fragebogens erinnert werden. Diese Erinnerungsfunktion kann zu einer besseren Adhärenz beitragen. Die notwendigen Erinnerungen, deren Anzahl bzw. die Zeit zwischen Erinnerung und Abgabe des Fragebogens können dabei interessante Zusatzinformationen liefern.

6.2.4 Fallstricke elektronischer Formulare

Füllt man Internetformulare aus oder loggt sich an einem Online -System ein, wird üblicherweise eine zufällige Sitzungsnummer erzeugt (*Session-ID*). Die *Session-ID* hat eine zeitlich begrenzte Gültigkeit. Bei längerer Inaktivität (keine Benutzereingabe oder Klicken von Schaltflächen auf der Webseite) kann es zu einem Zeitüberschreitungsfehler kommen.
Je nach Art der Internetseite, kann die *Session-ID* nur wenige Minuten bis Stunden, teilweise sogar Tage oder Wochen gültig sein. Nach Ablauf dieser Zeitperiode wird der Nutzer automatisch vom System abgemeldet. Die Eingaben in Formularfeldern können nun nicht mehr zum Server übertragen werden und gehen somit verloren. Es kommt zu einem Zeitüberschreitungsfehler (*Session Time-Out*).
Bei sicherheitskritischen Webseiten wie z.B. beim Onlinebanking haben *Session-IDs* eine vergleichbar kurze Gültigkeit von 1-5 Minuten. Auf anderen Webseiten kann man problemlos für Stunden oder gar Tage angemeldet bleiben. Findet ein Sitzungsabbruch bei einem umfangreichen Fragebogen statt, gehen alle getätigten Eingaben verloren. Die Bereitschaft der Nutzer den Fragebogen erneut auszufüllen, würde vermutlich sinken. Bei papierbasierten Fragebögen hingegen ist jederzeit eine Unterbrechung und Fortsetzung des Ausfüllens möglich. Um diese Lücke zu schließen, sollte der elektronische Fragebogen eine Pause-Funktion besitzen. Teilnehmer können das Ausfüllen des Fragebogens dadurch jederzeit unterbrechen und zu einem späteren Zeitpunkt wiederaufnehmen. Die bisherigen Eingaben werden gespeichert und bleiben somit erhalten.

6.2.5 Druck in Papierform

Sollte ein Teilnehmer mit dem Ausfüllen computerbasierter Fragebögen Schwierigkeiten haben, sollte das System den Druck in Papierform ermöglichen. Der Studienbetreuer muss Papierfragebögen über eine Druckfunktion erzeugen und den Teilnehmern auf Wunsch zur Verfügung stellen können. Der ausgedruckte Fragebogen kann dann per Hand ausgefüllt und an das Studienzentrum verschickt werden. Der ausgefüllte Fragebogen kann daraufhin vom Studienbetreuer entweder manuell in die Datenbank eingetragen oder nachträglich der exportierten Textdatei (*CSV Datei* des Datenbankexports*)* angefügt werden. Wünschenswert wäre auch, dass Teilnehmer selbstständig innerhalb des Systems einen leeren Papierfragebogen ausdrucken können, sofern ihnen dieses Medium lieber ist. Über eine Möglichkeit die erzeugten Fragebögen via Barcode mit Beleglesern automatisiert zu erfassen, wurde bereits nachgedacht. Die Implementierung würde jedoch sowohl die Zielsetzung als auch den Rahmen dieser Arbeit sprengen.

6.3 Management großer Studienpopulationen und Datenmengen

Betrachtet man zunächst konventionelle papierbasierte Studien, so setzen sich Zeit und Kostenaufwand für die Durchführung aus folgenden Teilprozessen zusammen:

- Rekrutierung von Studienteilnehmern
- Erstellung von Fragebögen
- Randomisierung und Einteilung in Gruppen
- Versand von Studienmaterial
- Betreuung der Studienteilnehmer (Kommunikation)
- Auswertung
- Archivierung

Wege zur elektronisch unterstützten Rekrutierung von Studienteilnehmern sowie die Einteilung in Studiengruppen wurden bereits angesprochen(☞ Kapitel 6.1). Eine Möglichkeit zur einfachen Erstellung von Fragebögen wird im Rahmen der Vorstellung des Prototypen ausführlich behandelt (☞ Kapitel 7.1.5). Dank elektronischer Fragebögen ist der Versand von Studienunterlagen obsolet. Das folgende Kapitel widmet sich den Möglichkeiten elektronischer Teilnehmerbetreuung.

6.3.1 Kommunikation mit Studienteilnehmern

Die Betreuung der Teilnehmer kann auf mehrere Wege erfolgen:

- Live-Chat
- Forum
- Interne Nachricht
- Email

Der Live-Chat stellt eine zeitsynchrone Form der Kommunikation dar. Dennoch können von einem Betreuenden mehrere Benutzer parallel betreut werden. Eine Chat-Kommunikation ist wie bei einem Telefongespräch zwischen Teilnehmer und Betreuer (ggf. Betreuern) möglich. Im IT-Sektor ist der Einsatz von Live-Chats bereits seit vielen Jahren etablierte Praxis, da

diese Methode im Vergleich zu telefonischer Betreuung oder Email günstiger und zeiteffizienter ist. Laut einer Studie von McKinsey kann durch den Einsatz von Foren und Live-Chats bis zu 25% der für Betreuung notwendigen Zeit eingespart werden. [161] Eine weitere Studie aus dem Bereich des Online-Marketings zeigt deutlich geringere Abbruchquoten bei Online-Einkäufen und längere Verweildauer auf den Webseiten, wenn eine Einkaufshilfe per Chat-Funktion angeboten wird. [139] Übertragen auf den Kontext klinischer Studien kann mit einer höheren Compliance und Adhärenz gerechnet werden.

Im Unterschied hierzu stellt das Forum eine asynchrone Kommunikationsform dar. Forenbeiträge sind für alle Teilnehmer (bzw. Teilnehmergruppen) sichtbar. Sie können moderiert oder nicht moderiert betrieben werden. Im moderierten Forum können Teilnehmer eine Frage stellen. Die Frage ist für andere Teilnehmer sichtbar und wird von einem Studienbetreuer oder Studienassistenten beantwortet. In nicht moderierten Foren können alle Teilnehmer einer Studie (oder Studiengruppe) auf Fragen antworten und sich untereinander austauschen. Hieraus ergeben sich zwei Vorteile. Zum einen sind Foren ein wichtiges Kommunikationsmedium. Sie finden breite Anwendung und fördern dank Interaktionsmöglichkeiten die Nutzung des Angebots. Zum anderen können Studienteilnehmer untereinander Themen diskutieren bzw. Fragen beantworten was wiederum den Betreuungsaufwand durch die Studienbetreuer reduziert. Eine repräsentative Untersuchung mit 499 Teilnehmern aus Deutschland und 502 in den USA aus dem Jahr 2007 ergab, dass der Hauptgrund für die Teilnahme an Foren (D: 64 %; USA: 71 %) der Austausch mit anderen Nutzern sei. [132]

Interne Nachrichten werden innerhalb einer Plattform zwischen Benutzern des Systems zugestellt. Sie können mit SMS Nachrichten verglichen werden. Im Unterschied zu Emails werden die Nachrichten jedoch nicht über fremde Server verschickt sondern bleiben innerhalb des Portals gespeichert. Dies bringt einen zusätzlichen Vorteil in puncto Sicherheit, da auf diese Weise der Zugriff auf die Nachrichten im Einflussbereich des Betreibers liegt und ihre Zustellung über eine kontrollierbare und gesicherte Verbindung erfolgt. Zudem kann kontrolliert und visualisiert werden, ob bzw. wann der Adressat die Nachricht gelesen hat. Für jede Studie können mehrere Studienbetreuer oder Studienassistenten eingesetzt werden, so dass die Betreuung der Studienteilnehmer auf mehrere Personen verteilt werden kann. Über ein Rollen- und Rechtemanagement können individuelle Berechtigungen für betreuende Personen gewährt werden. Zum Beispiel darf der Studienassistent A nur das Forum moderieren nicht aber Teilnehmer anschreiben wohingegen Studienassistent B sowohl ein Forum moderieren als auch Teilnehmer anschreiben darf.

Emails werden häufig noch immer über unverschlüsselte Datenverbindungen verschickt und in aller Regel unverschlüsselt beim Emailanbieter gespeichert wodurch sich Nachrichten jeglicher Kontrolle des Mail-Inhabers und der Studienleitung entziehen.
Um den Schutz persönlicher Daten zu verbessern, sollte daher auf ein internes Nachrichtensystem zurückgegriffen werden, da Mitteilungen nicht den Umweg über öffentliche Emailsysteme nehmen, sondern direkt zwischen Server und Teilnehmer ausgetauscht werden. Interne Nachrichten können von Teilnehmer zu Teilnehmer bzw. Teilnehmer zu Studienbetreuer verschickt werden. Analog zur SMS-Funktion aktueller Handys besteht die Möglichkeit einer 1:n Kommunikation also eine Nachricht an mehrere Personen gleichzeitig zu verschicken. Weiterer Vorteil eines internen Nachrichtensystems ist die Reduktion der übertragenen Daten. Jeder Datentransfer stellt einen Kosten- und Performancefaktor dar. Durch den Einsatz von Emails wird ein deutlich höheres Datentransfervolumen erzeugt, da jede Email neben den eigentlichen Nutzdaten (der Emailnachricht selbst) zusätzliche Daten überträgt. Jede Email beinhaltet einen sogenannten *Header*. Der *Header* enthält für die Zustellung der Email wichtige Infor-

mationen wie Absenderadresse, Zieladresse, Datum und Uhrzeit und weitere für den Versand wichtige Informationen. [127] Wird eine Email von einem Server empfangen, werden der Email weitere Informationen hinzugefügt. Diese Metadaten enthalten z.B. Informationen, die ein Spamfilter (Spam ja/nein?) oder ein Virenscanner (virenfrei ja/nein?) hinzufügt. Die Email wächst dadurch in der Dateigröße an. Dies kann je nach Nachricht über 30% ausmachen. [127] Für eine einzelne Nachricht betrachtet, sind dies zu vernachlässigende Datengrößen von wenigen Kilobytes. Für ein Studienportal mit mehreren tausend Nutzern kann dies zu einem deutlich höheren Datenvolumen führen und die Netzwerkauslastung erhöhen.

6.3.2 Ökonomische Aspekte

Der Einsatz einer online-basierten Studie kann insbesondere in Bereichen der Psychologie und Psychosomatik eine kostensparende Alternative zu konventionellen Studien darstellen. Eine große Rolle spielt die Zeit und Kostenersparnis durch die elektronische Bereitstellung der Studienunterlagen. Neben Papier und Portokosten entfallen Personalkosten für Sortierung und dem Zusammenstellen der Studienunterlagen. Die Einteilung der Studienteilnehmer in Gruppen, sowie die Pseudonymisierung können automatisiert werden und sparen weitere Arbeitszeit. Zudem eliminieren sich hierdurch potentielle Fehlerquellen.

Fragebögen können über eine grafische Oberfläche ohne Programmierkenntnisse erstellt und als Vorlage gespeichert werden. Mit Hilfe der Vorlagen ist eine Wiederverwendung bestehender Fragebögen möglich. Auch können bereits bestehende Fragebögen erweitert oder überarbeitet werden. Im Laufe der Zeit wächst so der Pool von standardisierten Fragebögen z.B. EBF, PTSS-10, EWL, usw. Das Zurückgreifen auf bereits existierende Fragebögen reduziert die Arbeitszeit, die zur Erstellung neuer Fragebögen notwendig ist.

Durch die direkte Speicherung der Benutzereingaben in einer Datenbank entfallen weiterhin Investitionen in Belegleser und Software zur Erzeugung und Erkennung von Antwortkreuzen (*OCR-Software*). Das maschinelle Einlesen und Digitalisieren von Papierfragebögen birgt zudem eine gewisse Fehlerrate und muss von geschultem Personal überwacht bzw. durchgeführt werden. Die Archivierung elektronischer Studiendaten wird dank der großen Kapazität aktueller Datenträger vereinfacht. Auf optischen Medien (z.B. DVD) oder auf magnetischen Speichern (Bandlaufwerken oder Festplatten) oder Flashspeichern (USB-Sticks) können viele hunderttausend Datensätze problemlos über Jahre platzsparend archiviert und schnell wieder abgerufen werden. In einer Metaanalyse wurden 16 papierbasierte und 11 elektronisch erfasste klinische Studien zwischen 2001 und 2011 eingeschlossen und die Kosten verglichen. Die Arbeitshypothese lautete: „Elektronische Studien vermeiden Fehler, verkürzen die Dauer und reduzieren die Kosten der Datenerhebung.“ [60]
Hierbei stellte sich heraus, dass elektronische Studien zwar in der Regel höhere Gesamtkosten verursachen, die Kosten pro Teilnehmer jedoch deutlich geringer ausfallen. Dies begründet sich in der Tatsache, dass elektronische Verfahren fast ausschließlich in größeren Multicenterstudien oder für Medikamentenstudien zum Einsatz kamen. Das Ergebnis der Studie zeigte einen geringfügigen zeitlichen Vorteil der elektronischen Verfahren. Für die Rekrutierung wurden in elektronisch durchgeführten Studien im Mittel 22,4± 9 Monate gegenüber den Papierbasierten mit 26,5± 13 Monate (p=0,34) benötigt.
Die Kosten pro Teilnehmer für papierbasierte Studien (pCRF) belief sich im Mittel auf 1.135€± 1.234 gegenüber elektronisch erhobenen Studien (eCRF) mit Kosten von 374€± 351. [60]

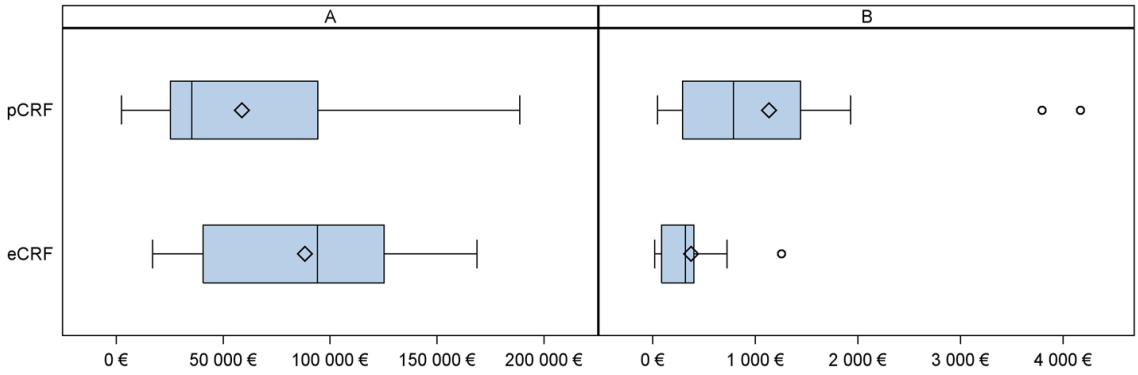


Abbildung 24: Metaanalyse Studienkosten links: Gesamtkosten, rechts: Kosten pro Teilnehmer [60]

Die gewonnenen Daten stammten aus Studien unterschiedlicher Fachbereiche mit teils stark variierender Teilnehmerzahl. Eine direkte Übertragung der Ergebnisse auf das Einsatzgebiet ist nicht möglich, da viele der eingeschlossenen Studien die Daten nicht online erhoben, sondern die Eingabe im Studienzentrum via Computer oder Tablet-PC erfolgte. Auffällig ist jedoch der deutliche Kostenunterschied pro Teilnehmer. Ein an dieser Studie besonders interessanter Ansatz ist die Befragung aller an den Studien beteiligter Personen. Für die Auftraggeber der Studien war der wichtigste Aspekt eine einfache Datenerfassung und das Gewinnen zuverlässiger Daten. Für die Studienverantwortlichen stand ebenfalls die Einfachheit der Dateneingabe im Vordergrund. An zweiter Stelle wurden barrierefreie Formulare genannt. Für die Datenanalysten waren vor allem die Reliabilität und weniger Datenbankabfragen sowie die Qualität der Datenbank wichtig. Über 40% der befragten Studienzentren (bzw. deren Mitarbeiter) bevorzugten die elektronische Erfassung der Daten. Je nach Funktion hatten zwischen 20% und 40% keine eindeutige Präferenz und nur zwischen 20% und 30% lehnten die elektronische Datenerfassung ab. [60]

Eine Simulationsstudie ging einen anderen Weg. Mit Werkzeugen der Informatik und Verfahrenstechnik wurde der Studienablauf papierbasierter Studien und elektronischer Datenerhebung modelliert. Ziel des mathematischen Modells war die Abschätzung von Kosten zukünftiger Studien. Hierzu wurden die Kosten der einzelnen Teilprozesse vergangener Studien analysiert. In die Berechnung gingen unter anderem die Größe der Studienpopulation, Anzahl der Befragungen, Materialkosten und Personalkosten ein. Ähnlich der bereits zuvor beschriebenen Studie konnte für die elektronische Datenerhebung ein deutlich geringerer Aufwand für Monitoring und Datenverwaltung nachgewiesen werden. Der Umfang des Sparpotentials korreliert mit der Größe der Studienpopulation. Auf der Grundlage des Modells können für eine Studie mit einer Laufzeit von 2 Jahren, 10.000 Fragebögen und 1.000 Teilnehmern die Kosten mit Hilfe elektronischer Datenerfassung um 55% gesenkt und die 1%-ige Fehlerrate durch die manuelle Dateneingabe vermieden werden. [98] Eine Übertragung der Daten auf die im Rahmen der SURE- Studie zu erwartenden Studienbedingungen scheint durchaus realistisch.

6.3.3 Auswertung

Durch die Eingabe der Daten über die Weboberfläche des geplanten Studienportals, ist im Gegensatz zu Papierstudien, eine grafische Auswertung und dadurch eine direkte Rückmeldung an die Teilnehmer möglich. Die Ergebnisse können grafisch präsentiert und im Verlauf beurteilt werden. Teilnehmer können ihre Ergebnisse (sofern dies von der Studienleitung erwünscht ist) mit denen des Kollektivs vergleichen.

Aus der Sicht der Studienleitung kann ein vorzeitiger Trend des Studienverlaufs direkt (ohne Zwischenauswertung) beobachtet werden.

6.3.4 Archivierung

Die Archivierung von papierbasierten Studien bedeutet Personalaufwand und bindet Lagerressourcen. Werden die papierbasierten Studiendaten (Fragebögen, Anwendungskalender usw.) digitalisiert, entstehen aus den Papierdokumenten Bild-Dateien, die in der Regel zusätzlich zu den Originaldokumenten über Jahre hinweg auf Datenträgern vorgehalten werden. Im Gegensatz zur Speicherung der Information in Bilddateien zeichnen sich Datenbankeinträge durch einen deutlich geringeren Speicherbedarf aus. Der Dateninhalt eines mehrseitigen Fragebogens mit ankreuzbaren Antwortmöglichkeiten lässt sich in wenigen Bytes erfassen. Am Beispiel des verkürzten Erholungs-Belastungsfragebogens (EBF) kann man sich das Größenverhältnis ganz gut veranschaulichen. Der genannte Fragebogen umfasst 25 Fragen mit je 7 Antwortmöglichkeiten. In gedruckter Form handelt es sich um ein 4-seitiges Papierdokument. Der Speicherbedarf für eine Tabelle mit 25 Fragen und je 7 Antwortmöglichkeiten beträgt lediglich 50 Byte (Datentyp *TinyInt* = 1Byte vorausgesetzt). Hiervon fallen an reinen Nutzdaten 25Byte und für die Zuordnung von Fragebogen zu Teilnehmern (*Primärschlüssel*) weitere 25Byte.

25 x 1Byte (Nutzdaten) + 25 x 1Byte (Schlüssel) = 50Byte

Digitalisiert man die Antwortbögen mit einem Scanner, so entsteht eine Datei, die aus Bildpunkten besteht. Die Datenmenge für die Datei setzt sich zusammen aus der Anzahl der Seiten, der Größe des Papiers (DinA4 = 11,7x8, 3 Zoll), der Auflösung in Punkten pro Zoll (100dpi) und der Anzahl der Farben (schwarz/weiß = 1bit).

4x 11,7inch x 100 dpi x 8,3inch x 100dpi x 1bit = 485.550Byte (≈ 0,5MB)

Durch geeignete Kompression können diese Bilddaten zwar weiter reduziert werden, jedoch ist dieser Vorgang mit Verlusten behaftet. Je nach gewähltem Dateityp und dem Kompressionsverfahren lassen sich Bilddaten des Beispiels auf bis zu 50KB reduzieren. [79]
Bei dieser vereinfachten Betrachtung wurde der zusätzliche Speicherverbrauch durch Metadaten, Codierungsverfahren, Farbräume ebenso wie der Overhead für die Verwaltungsschicht des Datenbankmanagementsystems vernachlässigt. Dennoch liegt der Größenunterschied zwischen der Datenhaltung in einer Datenbank und in Form von Bildmaterial bei einem Faktor von 1.000 bis 10.000.

6.4 Auswertung

6.4.1 Grafische Auswertung (Histogramm)

Die Auswertung der Studiendaten ist zwar mit Hilfe externer Werkzeuge vorgesehen. Trotzdem soll bereits über das Studienportal eine einfache grafische Darstellung der Studienergebnisse möglich sein. Hierzu sollen Histogramme zum Einsatz kommen. Mit Hilfe der Balkendarstellung können Studienbetreuer den aktuellen Verlauf der Studie abschätzen. Für Teilnehmer soll diese Funktion für eigene Fragebögen ebenfalls möglich sein. Für die Anwendung in der SURE- Studie dient diese Möglichkeit der Lernerfolgskontrolle und soll dazu

beitragen die Teilnehmer zu motivieren. Über die visuelle Rückkopplung soll ein Fortschritt bei der Stressbewältigung erkennbar werden.

6.4.2 Datenexport für externe Programme (Excel, SPSS)

Eine umfangreiche Auswertungsmöglichkeit innerhalb des Studienportals wäre programmiertechnisch zwar möglich, jedoch hängen die statistischen Mittel und gewünschten Grafiken stark von der Art der Studie und deren Fragestellung ab. Für komplexe statistische Analysen bieten externe Programme, die für die Berechnung und die visuelle Darstellung vorgesehen sind, ein hohes Maß an Flexibilität. Für den Datenexport wird daher ein Format gewählt, das von allen Statistikprogrammen importiert werden kann. Die Daten werden hierzu in ein Textdokument überführt und die Spalten durch Kommata getrennt. Jede Zeile entspricht einem Datensatz. Das Programm Microsoft Excel kann so gestaltete Dateien direkt öffnen und in einer Tabelle darstellen. Liegen die Daten in Form von Tabellen vor, kann der Anwender nach Wunsch statistische Funktionen darauf anwenden und die Daten visualisieren.

6.5 Datenschutz

Datenschutz ist ein essentieller Faktor, wenn es um den Umgang mit personenbezogenen Daten geht. Im psychosomatischen Bereich sind Patienten oder Studienteilnehmer besonders auf den Schutz ihrer persönlichen Daten angewiesen. Es gibt verschiedene Möglichkeiten durch die Zugriff auf persönliche Daten erlangt werden kann:

- Datenübertragung vom/zum Teilnehmer
- Webserver-Statistiken
- Zugriff auf die Datenbank
- Zugriff auf Daten über die Weboberfläche des Studienportals
- Zugriff auf das Dateisystem des Servers
- Cookies

Für die Sicherstellung des Datenschutzes, muss an alle Schwachstellen gedacht werden und es müssen technische oder organisatorische Maßnahmen getroffen werden, die den Zugriff auf persönliche Daten verhindern. (☞ Kapitel 3.1)

6.5.1 Verschlüsselte Datenübertragung via SSL/TLS

6.5.1.1 Grundlagen der Verschlüsselung

Bei der Eingabe einer Internetadresse wird eine Anfrage an den zuständigen Server gesendet. Dieser sendet die angefragte Information (Internetseite, Fragebogen, Videos...) an den Nutzer zurück. Der Browser stellt die empfangene Information entsprechend dar.
Meist findet diese Kommunikation zwischen Server und Browser unverschlüsselt statt. Dies hat zur Folge dass der Datenverkehr zwischen Server und Client abgefangen, gelesen oder sogar verändert werden kann. Sowohl Anfragen an den Server, als auch die zurück gesendeten Daten können jedoch sensibel sein. So können vom Studienteilnehmer gesendete Daten private Nachrichten oder auch Benutzername und Passwort enthalten.

Es gibt jedoch die Möglichkeit die Datenübertragung zwischen Browser und Server zu verschlüsseln. Die hierfür notwendige Software ist bereits in alle Betriebssysteme bzw. Browser integriert und muss vom Teilnehmer nicht zusätzlich installiert werden. Hinter *SSL* (*secure socket layer*) bzw. dessen Nachfolger *TLS* (*transport layer security*) verbirgt sich die angesprochene Technologie zur verschlüsselten Datenübertragung. Mit Hilfe eines elektronischen Ausweises, dem SSL-Zertifikat, kann sich der Server gegenüber dem Browser des Teilnehmers digital ausweisen und signalisieren, dass Daten verschlüsselt übertragen werden können. Die meisten Browser signalisieren dies mit einem Schloss-Symbol oder einer grünen Adressleiste.

Das SSL-Zertifikat besteht aus einem öffentlichen Schlüssel, *public key*, welcher der Öffentlichkeit bekannt gegeben wird und einem geheimen, *private key*, der nie preisgegeben wird. Über den öffentlichen Schlüssel werden der Name und die Adresse des Inhabers sowie Informationen zum Zertifikat-Aussteller (*CA = Certificate Authority*) bekanntgegeben. Eine Liste vertrauenswürdiger Zertifikat-Aussteller ist in Betriebssystemen bzw. Browsern vorinstalliert. Zertifikat-Aussteller können sein:

- Kommerzielle oder öffentliche Zertifizierungsstellen
- Firmen oder Privatpersonen (selbstsignierte Zertifikate)

Kommerzielle Zertifizierungsstellen sind Anbieter wie Verisign, Thawte oder GeoTrust. Ähnlich konventionellen Zertifizierungen (z.B. Qualitätsmanagement nach DIN/ISO 9001), übernehmen diese die Überprüfung des Antragsstellers und vergeben nach erfolgreicher Überprüfung ein Zertifikat. Selbstsignierte Zertifikate können von jeder Person oder Firma ausgestellt werden. Die ausstellende Person/Firma wird somit selbst zur Zertifizierungsstelle. Von kommerziellen Zertifizierungsstellen ausgestellte Zertifikate werden von Browsern automatisch als vertrauenswürdig eingestuft, selbstsignierte (oder ungültige) Zertifikate hingegen, werden als nicht vertrauenswürdig eingeschätzt. Vom Browser wird daher eine Warnmeldung ausgegeben. Der Benutzer kann dann entscheiden, ob er dem Aussteller des Zertifikats auf eigene Gefahr vertraut, oder nicht. Einige Browser nehmen dem Benutzer inzwischen bereits die Entscheidung ab und unterbinden den Aufruf der Webseite.

Der Einsatz von SSL-Zertifikaten dient nicht nur der Identifikation des Webseitenbetreibers sondern auch die Möglichkeit der verschlüsselten Datenübertragung. Das Verfahren, das hierbei eingesetzt wird, nennt sich *RSA*, ein auch als *public key Verfahren* bezeichnetes Verschlüsselungssystem, das von den drei Kryptographen **R**ivers, **S**hamir und **A**dleman 1977 veröffentlicht wurde. Es basiert auf einem Schlüsselpaar aus *public key* und *private key*, die miteinander in Verbindung stehen. Das Rückschließen vom öffentlichen auf den privaten Schlüssel wird hierbei durch ein mathematisches Codierverfahren verhindert. Da es mit entsprechendem Rechenaufwand möglich ist, den geheimen privaten Schlüssel aus dem frei zugänglichen öffentlichen Schlüssel zu dekodieren, werden Schlüssel mit entsprechend großer Länge benötigt. Die Länge der Schlüssel bestimmt die für das Knacken notwendige Rechenzeit. [104] Die Dauer für die Rückrechnung eines 700bit langen Schlüssels wurde zur Zeit der *RSA* Entwicklung auf mehr als 10^{15} Jahre geschätzt. 1990 wurden 512bit lange Schlüssel bereits als unsicher eingeschätzt und 1999 in einer Publikation der Beweis dafür erbracht. Im Jahr 2007 wurde in einem Experiment am Polytechnikum Lausanne (EPFL) versucht, einen 700bit langen *RSA* Schlüssel zu knacken. Hierzu wurden 400 handelsübliche Computer und Laptops zusammengeschaltet. Die Rechenzeit betrug lediglich 11 Monate. [6] Zwei Jahre später wurde ein 768bit Schlüssel mit Hilfe eines Clusters, bestehend aus 618 Servern in einer Rechenzeit von knapp 2 Jahren entschlüsselt. An diesem Projekt waren Mathematiker und

Kryptographen aus der Schweiz, Frankreich und Japan, Deutschland und den USA beteiligt. [114] Ein einzelner Rechner hätte hierfür allerdings 1700 Jahre benötigt. [66] In einem Übersichtsartikel aus dem Jahre 2001 zur Wahl geeigneter Schlüssellängen wurden für die nächsten 50 Jahre der Aufwand abgeschätzt der notwendig ist, um einen Schlüssel entsprechender Stärke zu brechen. Die Prognose deckt sich rückwirkend betrachtet mit bisherigen Veröffentlichungen. Gemäß der Prognose der beiden Autoren wird eine derzeit übliche Schlüssellänge von 2048bit bis maximal 2030 empfohlen. Die geschätzten Kosten für die Entschlüsselung des 2048bit Schlüssels binnen eines Tages würden sich dann auf $1{,}11 \times 10^9$ USD belaufen. Die Autoren betonen jedoch, dass die Grundlage ihrer Berechnungen und auch alle bis dato veröffentlichten Brüche der Verschlüsselungen keinesfalls den Stand der Technik darstellen, sondern dass das Brechen von Schlüsseln in erster Linie eine Frage des Aufwands sei. Vor allem spielen für die Sicherheit der Gesamtverschlüsselung nicht nur die Stärke des Schlüssels sondern auch alle anderen Teile des Computersystems eine wesentliche Rolle. [74] Dank stetig zunehmender Rechenleistung von Computern, aber auch durch die Möglichkeit verteilten Rechnens (*Grid Computing*) kann die Rechenzeit und der finanzielle Aufwand für die notwendigen Berechnungen drastisch reduziert werden. Ein Projekt namens BOINC, stellt Wissenschaftlern Rechenleistung für komplexe Aufgaben aus einem großen Pool von Computern freiwilliger Teilnehmer zur Verfügung. Derzeit (Stand 03/2015) besteht dieses *Grid* aus 677.938 Computern. Mit Hilfe dieses Supercomputers können diverse wissenschaftliche Projekte zur Klimasimulationen oder Molekularsimulationen unterstützt werden. Ein ähnlich umfangreiches Computernetz kann jedoch auch mit Hilfe vireninfizierter PC´s oder Smartphones aufgebaut werden. Die Wahl starker Verschlüsselungsmethoden ist daher besonders wichtig.

Durch die verschlüsselte Kommunikation zwischen Server und Client kommt es zu einer Zeitverzögerung. Verschlüsselte Informationen werden daher etwas langsamer übertragen als Unverschlüsselte. Für Server und Client entsteht dabei ein gewisser Rechenaufwand. Bei hoher Anzahl gleichzeitiger Zugriffe und bei Übertragung großer Datenmengen ist daher ein Server mit ausreichender Rechenleistung und Bandbreite notwendig. Insbesondere die Herstellung der Kommunikation zwischen Server und Teilnehmer während des Verbindungsaufbaus (*Handshakes*) ist mit einer erhöhten Belastung für den Server verbunden. [162] Durch kryptographische Zusatzinformation, die für die Übertragung notwendig ist, erhöht sich auch die Datenmenge, die notwendig ist, um eine Nachricht zu übertragen. Eine Nachricht kann dabei aus Text aber auch Bildern oder Videos bestehen. Sie wird zunächst nach einem festgelegten Protokoll in kleine Einheiten unterteilt und komprimiert. Soll eine Nachricht verschlüsselt werden, ist neben der komprimierten Information eine Zusatzinformation zur Datenintegritätsprüfung unumgänglich. Über den *MAC* (*Message Authentication Code*) wird vereinfacht eine Prüfsumme berechnet, mit der sichergestellt werden kann, dass eine Nachricht unverändert und vollständig übertragen wurde und vom richtigen Absender kommt. Durch diese Zusatzinformation wächst das Datenpaket (*TCP-Packet*), das dann über die Transportkanäle übertragen wird.

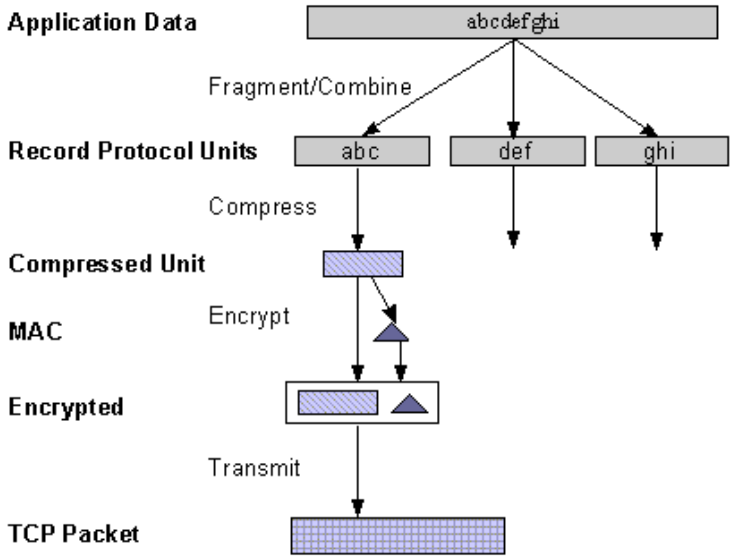


Abbildung 25: Schema verschlüsselter Datenübertragung, Apache.org [163]

6.5.1.2 Ablauf des Verbindungsaufbaus (Handshake)

Der Verbindungsaufbau zwischen dem Browser des Teilnehmers (Client) und dem Server wird mittels eines elektronischen Handschlags (*Handshake*) hergestellt. Beim *Handshake* handeln beide Parteien die Art der Datenkompression und der Verschlüsselungsmethode aus.

Der Client begrüßt den Server mit einer *client helo* Nachricht. Diese Nachricht enthält Informationen zu technischen Gegebenheiten, die der Computer des Clients mitbringt, z.B. bevorzugtes Kompressionsverfahren oder unterstützte Verschlüsselungsalgorithmen.
Der Server antwortet daraufhin mit einer *server helo* Nachricht. Die Antwort des Servers bestätigt die von beiden unterstützte Kompression und Verschlüsselung. Ist der *Handshake* abgeschlossen, überträgt der Server sein Zertifikat an den Client und sendet eine *hello done* Nachricht. Nun wartet er, während der Client das Server-Zertifikat überprüft. Ist der Aussteller des Zertifikats (*certification authority* oder *CA*) dem Betriebssystem des Client als vertrauenswürdig bekannt, überprüft der Client das Server-Zertifikat indem er beim Aussteller (*CA*) nachfragt, ob das Zertifikat gültig ist. Ist das Zertifikat gültig, sendet der Client eine Nachricht mit einer Zufallszahl (*pre master secret*), die mit dem (zuvor empfangenen) öffentlichen Schlüssel des Servers verschlüsselt wird. Den privaten Schlüssel des Zertifikates kennt nur der Server. Mit Hilfe des privaten Schlüssels (*private key*) kann der Server die vom Client gesendete Nachricht entschlüsseln. Der darauf folgende Prozess wird auch als Schlüsseltausch bezeichnet. Server und Client vereinbaren einen geheimen Schlüssel, den sie benutzen, um die übertragenen Daten zu verschlüsseln oder zu entschlüsseln. Hierzu wird aus dem *pre master secret* ein Schlüssel (*master secret)* generiert, den nur Server und Client kennen. Aus dem *master secret* wird für die Dauer der Verbindung ein symmetrischer Schlüssel *session key,* mit dem die Daten in Folge verschlüsselt werden. [29]

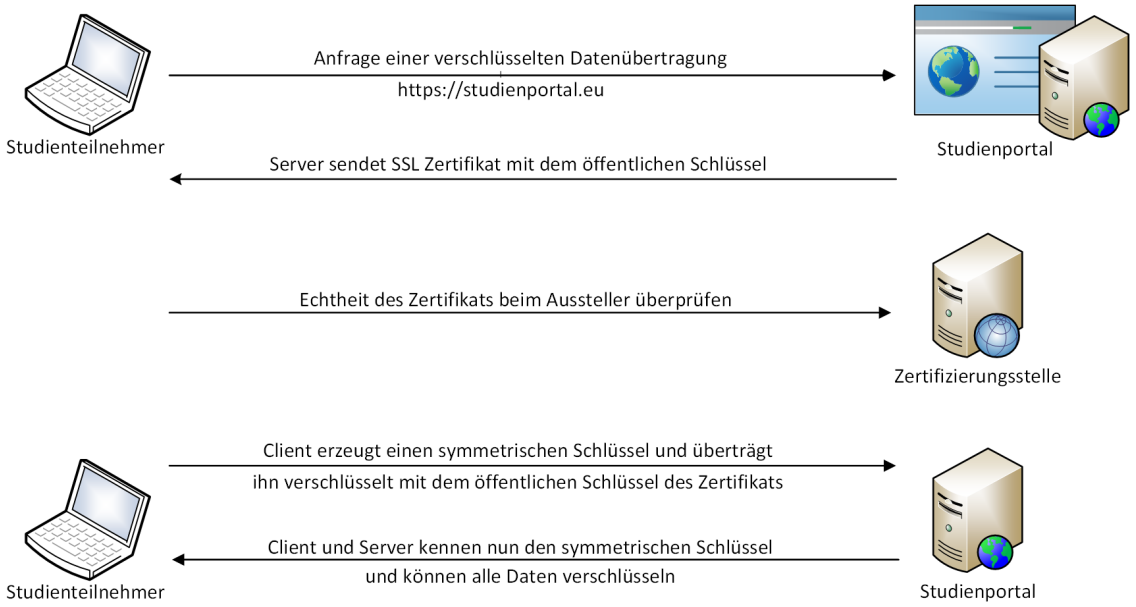


Abbildung 26: vereinfachter Handshakeprozess

Nach dem *Handshake* schließt sich die verschlüsselte Datenübertragung zwischen beiden Computern an. Mit welchem Algorithmus die symmetrische Verschlüsselung stattfindet, bestimmen die technischen Voraussetzungen des Clients und die Konfiguration des Servers. Üblicherweise wird für den Aufbau der Verbindung und dem *Handshake* das asymmetrische Verschlüsselungsverfahren *RSA* (*public key Verfahren*) verwendet.
Nachdem die Übertragung von Daten wegen relativ großer Schlüssellängen (1024 bis 4096bit) langsam und rechenintensiv ist, wird nach dem *Handshake* auf ein symmetrisches Verschlüsselungsverfahren mit kürzeren Schlüsseln (128 bis 384bit) gewechselt. Hierfür kommen unterschiedliche Algorithmen und Schlüssellängen zum Einsatz. *AES*, *3DES* und *RC4* sind gängige Verschlüsselungsalgorithmen. Um zu überprüfen, ob die Daten vollständig und korrekt empfangen wurden, werden mathematische Prüfverfahren (*Hashfunktionen*) verwendet. Die bekanntesten Vertreter sind *MD5* und *SHA*. Die Kombination aus asymmetrischem Verschlüsselungsverfahren, symmetrischem Verschlüsselungsalgorithmus und *Hashfunktion* wird als *Cipher Suite* bezeichnet. Ein Beispiel für eine schwache *Cipher Suite* wäre: TLS_RSA_WITH_RC4_40_MD5. Der Verbindungsaufbau wird hier über das asymmetrische *public key Verfahren* aufgebaut, symmetrisch mit einem 40bit unter Verwendung des *RC4* Algorithmus verschlüsselt, und die korrekte Datenübertragung mit der *Hashfunktion MD5* kontrolliert. Ein anderes Beispiel wäre die von der Engineering Task Force IETF als sicher eingestufte *Cipher Suite* TLS_DHE_RSA_WITH_AES_128_GCM_SHA256. [52] Der Verbindungsaufbau erfolgt hier nicht über einen asymmetrischen (wie zuvor beschrieben), sondern einen temporären symmetrischen Schlüssel, der mit Hilfe eines anderen Algorithmus (*Diffie-Hellman*) berechnet wird. Erst danach wird die Identität des Serverzertifikats wie beim klassischen *RSA* Verfahren überprüft, weitere Daten mit dem *AES* Algorithmus (128bit) verschlüsselt, und die Datenintegrität mittels *SHA-2* sichergestellt.

6.5.2 Perfect Forward Secrecy (PFS)

Die konventionelle Verschlüsselung mit asymmetrischem *public key* Verfahren besitzt eine Schwachstelle, die durch ein alternatives Verfahren beim Schlüsseltausch umgangen werden kann. *Perfect Forward Secrecy* berücksichtigt, dass durch Mitschneiden der verschlüsselten

Datenübertragungen und „Diebstahl“ des *private key* (z.B. durch Sicherheitslücken) ein potentieller Angreifer den *session key* knacken kann. Damit ist er in der Lage, nicht nur den Inhalt aller zukünftigen Datenübertragungen, sondern auch rückwirkend alle übertragenen Daten zu entschlüsseln. Auf diese Weise kann ein Angreifer in den Besitz von Zugangsdaten von Teilnehmern aber auch von privilegierten Nutzern wie dem Administrator oder dem Studienmanager gelangen. Mit Hilfe der Zugangsdaten können dann sehr einfach direkt Informationen aus dem gesamten System bezogen werden.

Beim *Perfect Forward Secrecy* Verfahren wird mit Hilfe eines symmetrischen Verschlüsselungsalgorithmus (*Diffi-Hellman Algorithmus*) [56] das *pre master secret* erzeugt. Es wird nicht über das Internet übertragen, sondern existiert nur im Arbeitsspeicher des Client und Servers. Im Gegensatz zum asymmetrischen RSA-Verfahren wird bei *Forward Secrecy* der *private key* des Servers nicht zur Erzeugung des *session key* sondern lediglich zum Nachweis der Serveridentität verwendet. Am Ende der Sitzung wird der gemeinsame Schlüssel von beiden Partnern zerstört und ist somit nicht rückwirkend mehr nachvollziehbar. Jeder neue *session key* wird unabhängig vom vorherigen Verbindungsschlüssel erzeugt, so dass vom Vorgänger nicht auf nachfolgende *session keys* geschlossen werden kann. Obwohl der *Diffie-Hellman* Schlüsselaustausch (*DHE*) bzw. dessen Spielart *ECDHE* (*Eliptic Curve Diffie-Hellman*) einen großen Sicherheitsgewinn bringt, wird er derzeit nur von wenigen Webseiten konsequent eingesetzt. [164] Selbst bei sicherheitskritischen Anwendungen wie etwa Homebanking wird dieses Verfahren kaum genutzt. Eine stichprobenhafte Überprüfung einiger Webseiten zeigt, dass trotz jüngster NSA Skandale und der damit verbundenen erhöhten Aufmerksamkeit der Öffentlichkeit kaum Konsequenzen gezogen wurden. (Stand 12/2014)

Grundvoraussetzung für *Forward Secrecy* ist die Unterstützung durch den Browser. Alte Betriebssysteme wie Windows XP ohne Browserupdates bleiben außen vor. Jedoch unterstützen alle aktuellen Browser diese Technologie.

- Mozilla Firefox (ab Version 21)
- Internet Explorer (ab Version 10)
- Safari (ab Version 5.19)
- Google Chrome (ab Version 27)
- iOS ab Version 6
- Android ab Version 4

6.5.2.1 Auswahl der geeigneten Cipher Suite

Je nach Zertifizierungsstelle (und Serverkonfiguration) werden 1024bit bis 4096bit lange Schlüssel für asymmetrische zertifikatbasierte Verfahren verwendet. Derzeit gängige Praxis ist der Einsatz von 2048bit *SSL Zertifikaten*. Die Verschlüsselung symmetrischer Verfahren liegt zwischen 40bit und 256bit langen Schlüsseln. Mit zunehmender Bit-Länge steigt die Sicherheit, jedoch auch der Rechenaufwand für Server und Client.

Die Belastung für den Prozessor (*CPU)* unterscheidet sich zwischen dem klassischen Schlüsselaustauschverfahren *RSA* und dem *Forward Secrecy Verfahren DHE* etwa um den Faktor drei. [120] In Vergleichstests von Vincent Bernat wurde der Rechenaufwand und die Zeitverzögerung quantifiziert. Die Messungen wurden auf einem Linux-Server mit einem 6-Kern-Prozessor und unter *OpenSSL* Version 1.0.1 durchgeführt. Mit Hilfe eines Simulationsprogramms führte ein Client 1000 *Handshakes* mit dem Server unter Verwendung unterschiedli-

cher *Cipher Suites* durch. Der für die *Handshakes* benötigte Rechenaufwand wird in der linken Grafik als Rechenzeit in Sekunden gemessen. Die rechte Grafik vergleicht das symmetrische *DHE* Verfahren (blaue Kurve) mit dem konventionellen Verfahren (rote Kurve) und misst die Antwort des Servers und die Anzahl der durchgeführten Transaktionen pro Sekunde.

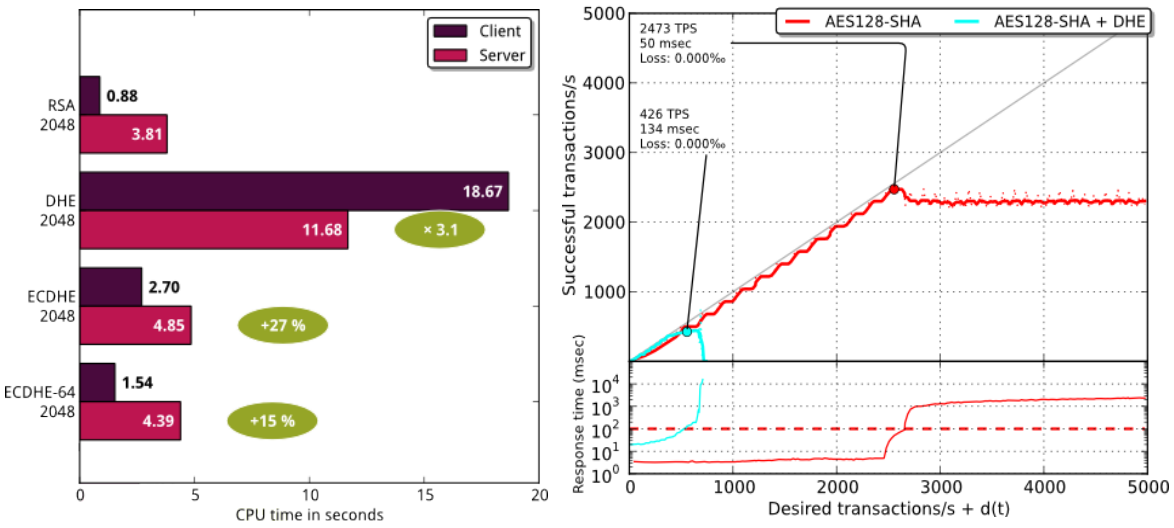


Abbildung 27: Prozessorlast und Rechenzeit verschiedener Schlüsseltauschverfahren, Vincent Bernat [120]

Das Testsystem erreicht unter Verwendung von *Forward Secrecy* mit *DHE* deutlich weniger (knapp 500) Transaktionen pro Sekunde. Mit *RSA* können etwa 2500 Transaktionen pro Sekunde durchgeführt werden. Nicht verwunderlich ist, dass die Antwortzeit des Systems von der Wahl des Schlüsseltauschverfahrens abhängig ist. Bereits bei geringer Last zeigt sich eine Differenz von etwa 70ms. Mit zunehmender Anzahl an Anfragen an den Server nimmt die Zeitverzögerung exponentiell zu. Der leistungsbegrenzende Faktor ist die Berechnung der Ellipsenfunktionen beim *DHE*- Verfahren. [64] *DHE* bietet zwar das höchste Maß an Sicherheit, ist damit jedoch auch gleichzeitig das rechenintensivste Verfahren. Ein guter Kompromiss ist der Einsatz von *ECDHE* für den Schlüsseltausch. Auf 64bit Plattformen kostet dieses Verfahren nur 15% mehr Rechenleistung. [120]
Für die Studienplattform ist jedoch nicht nur maximale Sicherheit, sondern auch die Abwärtskompatibilität zu älteren Browsern entscheidend. Neben den Verfahren zur *PFS* müssen daher auch möglichst sichere Verfahren eingesetzt werden die auch auf älteren Browsern unterstützt werden (vornehmlich für Microsoft Internetexplorer < Version 10).

Als bevorzugtes Schlüsselaustauschverfahren für das Studienportal werden die *ECDHE* Verfahren mit möglichst hoher Schlüsselstärke vom Server angeboten. Unterstützt ein Browser diese Technologie nicht, wird der aktuellen Empfehlung der IETF entsprechend das *DHE*-Verfahren genutzt. Dies trägt insbesondere Microsofts Internetexplorer 9 und 10 Rechnung. Älteren Browsern oder Smartphones wird das klassische *public key* Verfahren angeboten. Die Tabelle zeigt die Komponenten für die *Cipher Suites*, die für die Implementierung im Studienportal Verwendung finden.

Schlüsselaustausch	Verschlüsselung	Hashfunktion
ECDHE_RSA	AES256/AES 128/DES	SHA384/SHA256/SHA128
DHE_RSA	AES256/AES128/CAMELLIA/ DES	SHA384/SHA256/SHA128
RSA	AES256/CAMELLIA/DES	SHA384/SHA256/SHA128

Tabelle 2: Eingesetzte Verschlüsselungskomponenten

Die vom Webserver bevorzugten *Cipher Suites* sind diejenigen mit *Forward Secrecy* (*ECDHE*) und die mit möglichst hoher Performanz bei möglichst starker Verschlüsselung in absteigender Sicherheit.

1. ECDHE-RSA-AES256-GCM-SHA384
2. ECDHE-RSA-AES256-CBC-SHA384
3. ECDHE-RSA-AES256-CBC-SHA256

   ...

6.5.2.2 Grenzen der Verschlüsselung

An verschlüsselte Daten direkt heranzukommen ist nicht einfach. Unabhängig von der Stärke des Schlüssels oder des Algorithmus, existieren eine Reihe von Möglichkeiten, die es dennoch ermöglichen, auf verschlüsselte Daten zuzugreifen:

- Lücken in der Verschlüsselungssoftware
- Diebstahl des privaten Schlüssels
- Einbruch in das Betriebssystem
- Manipulation von Hard/Software (z.B. Zufallszahlengeneratoren)
- Kompromittierte Zertifizierungsstelle
- *Man in the Middle* Angriff

Die Linux-Distribution Debian berichtete erstmals 2006 in ihrer Mailingliste über schwere Lücken im *OpenSSL* Paket welches zur Verschlüsselung eingesetzt wird. Durch einen Fehler im Zufallszahlengenerator wurden die geheimen Schlüssel vorhersehbar. Damals wurde der Fehler öffentlich, da das Programm zur elektronischen Übermittlung der Steuererklärung (*ELSTER*) über betroffene Server ausgeliefert wurde. Bei *OpenSSL* handelt es sich um eine Software, mit deren Hilfe Serverdienste wie Webserver, Emailserver, FTP-Server, VPN, Fernwartungsanschlüsse oder Telefonie- Systeme verschlüsselt Daten austauschen können. Sie ist die Verschlüsselungssoftware, die weltweit am meisten genutzt wird. Die Konsequenzen von Mängeln an diesem Herzstück der Verschlüsselung sind daher besonders gravierend. Aus diesem Grund wurde die Anfang 2014 gefundene Schwachstelle unter dem Namen *Heartbleed* bekannt. Die Lücke bestand jedoch bereits seit zwei Jahren, bevor sie publik wurde. Betroffen hiervon waren nicht nur tausende Webserver, sondern auch viele andere Geräte wie DSL-Router, Internettelefone, VPN -Einwahlknoten, Firewalls, Hausautomatisierungs- oder Videoüberwachungssysteme. Besonders in die Öffentlichkeit rückte *Heartbleed*, da die Auswirkung mit einem Schlag alle Anwender betraf. Bei Bekanntwerden der Schwachstelle waren die meisten Webseiten davon betroffen. Auch 48 Stunden nach Bekanntwerden der Lücke waren 25-55% der beliebtesten amerikanischen Webseiten darunter Google, Yahoo, YouTube, Wikipedia und Flickr noch immer verwundbar. [33] Hierzulande waren nahezu alle Webseiten von Banken, Emailanbietern, Onlineshops bis hin zu Universitäten betroffen. Die meisten deutschen Banken haben erst Wochen bis Monate später reagiert und waren somit für potentielle Angriffe vulnerabel. [165] Selbst ein halbes Jahr nach Veröffentlichung und trotz der Medienpräsenz waren bzw. sind bis heute noch immer etliche Server anfällig für die Schwachstelle. [69] Auch Zertifizierungsinstanzen wie das Deutsche Forschungsnetz (DFN) waren davon betroffen. [166] Über die Infrastruktur des DFN läuft ein Großteil des wissenschaftlichen Datenverkehrs. Der Verein vernetzt mit seiner Infrastruktur deutsche Universitäten und Forschungseinrichtungen untereinander und stellt die Verbindung zu europäischen

Forschungsnetzen her. Über den DFN werden die SSL-Zertifikate fast aller Universitäten ausgestellt um damit Webmail, Lernplattformen, VPN-Einwahl oder Institutswebseiten abzusichern. Im Rahmen des *Heartbleed* Fehlers wurden die SSL-Zertifikate der Universität Regensburg erneuert und alle Benutzer dazu verpflichtet ihr Passwort zu ändern.

Ein weiteres Beispiel, das im Herbst 2014 für Aufmerksamkeit sorgte war ebenfalls eine mit *OpenSSL* zusammenhängende Sicherheitslücke mit dem Namen *Poodle*. Die meisten Server unterstützten ein bereits 18 Jahre altes unsicheres Protokoll *(SSLv3*) um Kompatibilität zu älteren Betriebssystemen wie Windows XP anbieten zu können. Über geschicktes Erzwingen des alten Protokolls konnte in die verschlüsselte Kommunikation eingegriffen werden. [90] Auch in diesem Fall sind noch immer viele Server verwundbar, die das alte Protokoll weiterhin unterstützen. Anwender können sich in diesem Fall zumindest teilweise schützen, indem sie in ihrem Browser (bzw. Mailprogramm) die Kommunikation über das Protokoll *SSLv3* verbieten. Ob der Browser in Bezug auf *Poodle* gefährdet ist, kann unter der folgenden Webadresse überprüft werden: *https://www.poodletest.com*

Eine Gefährdung kann auch durch den Verkauf oder den Austausch defekter Festplatten entstehen. In der Regel können die Daten (und darauf gespeicherte geheime private Schlüssel) mit entsprechenden Softwarewerkzeugen wieder rekonstruiert werden – selbst wenn sie zuvor formatiert wurden.

Die meisten Server verfügen über Fernwartungszugänge und sind permanent an das Internet angebunden. Nicht selten werden Standardpasswörter von Serversoftware weiter verwendet oder es werden zu schwache Passwörter gewählt (z.B. Benutzername=admin, Passwort=password). Somit kann es mittels Durchprobieren (*Bruteforce*) gelingen, administrative Zugangsdaten des Servers zu erlangen.

Daten können aber auch bereits vor der Verschlüsselung abgefangen werden. Pishing ist ein möglicher Weg, geheime Zugangsdaten von Anwendern zu erbeuten. Der Benutzer wird unter einem Vorwand (z.B. Ablauf eines Benutzerkontos oder Bestätigung einer Bestellung) auf eine gefälschte Webseite geleitet, um dort Zugangsdaten oder Kreditkarteninformationen einzugeben.

Es gibt jedoch auch Schadsoftware, die weitaus weniger offensichtlich operiert. Computerprogramme, die gratis zum Download angeboten werden können als Vektoren genutzt werden um sich auf dem Rechner des Opfers einzunisten. Diese *Backdoors* halten einem Eindringling eine Hintertür zum System offen über das ein Zugriff auf den Rechner erfolgen kann. Über Veränderungen von Systemeinstellungen kann auch der Datenverkehr gezielt umgeleitet oder Eingaben des Benutzers protokolliert werden. Dieses Verfahren kann auch bei Apps mobiler Endgeräte eingesetzt werden. Im Januar 2014 berichtete die New York Times im Rahmen der NSA Enthüllungen dass die Daten von Smartphone-Anwendern über die beliebte Spiele-App „Angry Bird“ ausgelesen werden könne. [72] Dem unbestätigten Originaldokument zu Folge, standen in besonderem Maß Kommunikationsdaten sozialer Netze und GPS Daten im Vordergrund, jedoch wurden ebenso besuchte Internetseiten und die verwendete Verschlüsselung erwähnt. [167]

Ebenfalls im Rahmen der Spionage Enthüllungen wurde in Berichten die Einflussnahme der Geheimdienste auf die Entwicklung von Kryptografie-Software untersucht. Hierbei ging es vor allem um Zufallszahlengeneratoren, dem Herzstück jeder Verschlüsselung. Mit ihrer Hilfe werden nicht vorhersehbare Schlüssel generiert. Eine Manipulation oder Fehler in der Erzeu-

gung der Zufallszahlen entspricht einem Hintertürchen in der Kryptografie. Einige wichtige Komponenten für eine wirksame Verschlüsselung wurden im Auftrag der Geheimdienste oder in enger Kooperation mit diesen entwickelt. Das National Institute of Standard and Technology NIST (vergleichbar dem DIN) empfahl in einem Standard den Zufallszahlengenerator *Dual_EC_DRBG* einzusetzen, dessen Entwicklung unter behördlichem Einfluss stand. [117] Für die Integration des umstrittenen Zufallszahlengenerators in die Softwarebibliothek BSAFE soll die NSA 10 Millionen Dollar an die Firma RSA Security gezahlt haben. [87] Die Bibliothek stellt die Grundlage für viele in C/C++ oder Java programmierte Softwareprodukte dar. Kryptografie-Experten befürchten auch die Beeinflussung von Hardwarebausteinen, die zur Beschleunigung von Verschlüsselung in Prozessoren integriert sind. Die meisten Betriebssysteme nutzen wegen des Geschwindigkeitsgewinns die Funktion von Hardware unterstützter Zufallszahlenerzeugung.

Das Prinzip der asymmetrischen Verschlüsselung basiert auf Zertifikatsketten. Das Betriebssystem oder der Browser verlässt sich darauf, dass die Zertifizierungsstelle und die Zertifikate, die diese ausstellt, vertrauenswürdig sind. Zertifizierungsinstanzen können jedoch kompromittiert werden und dadurch beliebige Webadressen (zu unrecht) als vertrauenswürdig eingestuft werden. Es gibt Zertifizierungsstellen, die (um die Erstellung des Zertifikats zu erleichtern) auch einen privaten Schlüssel für einen Server generieren und einen privaten Schlüssel gemeinsam mit dem öffentlichen Zertifikat in einem Dokument ausliefern. In diesem Fall hat die Zertifizierungsstelle Kenntnis über den sonst streng geheimen, privaten Schlüssel. Auch dies birgt bei der Kompromittierung der Zertifizierungsinstanz ein Sicherheitsrisiko.

Befindet sich ein potentieller Angreifer im lokalen Netzwerk oder im Rechenzentrum kann er den Datenverkehr zwischen Sender und Empfänger mitschneiden oder in den Datenstrom eingreifen. Als *Man-In-the-Middle* Attacke werden Angriffe bezeichnet, bei denen der Angreifer die Kommunikation zwischen Server und Client zu seinen Gunsten modifiziert. Der Angreifer kann eine Anfrage, die ursprünglich an den Server gestellt wurde, entgegennehmen und sich als Server ausgeben. Verfügt er dabei über ein gültiges Zertifikat, ist der Angriff nur schwer zu entdecken. Die Kommunikation würde statt mit dem Server, mit dem Angreifer ablaufen. Verfügt dieser über kein gültiges Zertifikat, kommt es im Browser zu einer Warnmeldung. Der Angriff wird dadurch erkennbar. Ein gültiges Zertifikat kann entweder durch Diebstahl des Originalzertifikats erlangt werden, oder über ein selbst ausgestelltes Zertifikat, das von einer vertrauenswürdigen Zertifizierungsstelle stammt. Es gibt noch eine Vielzahl anderer *Man-in-the-Middle* Angriffe die darauf ausgelegt sind, den Server zur Wahl schwacher *Cipher Suites* zu überreden oder Schwachstellen von Verschlüsselungsalgorithmen auszunutzen. [56] [133] [11] *Man in the Middle* Attacken sind allerdings nicht nur in lokalen Netzwerken (Universitäten, Firmen, Wohnheimen und WLANs) gefährlich, sondern können beim Datenverkehr im Internet erfolgen. So können zum Beispiel *DNS-Informationen* die in der Regel unverschlüsselt und unsigniert übertragen werden, manipuliert werden. Auf diese Weise kann ein Angreifer dem anfragenden Computer eine falsche *IP Adresse* unterschieben und auf eine gefälschte Webseite locken. (☞ Kapitel 7.2.5)

6.5.3 Webseiten-Statistik

Webserver protokollieren üblicherweise jeden Aufruf einer Webseite. Die Webseitenbetreiber können dadurch analysieren, welche Browser und Betriebssysteme die Besucher am häufigsten verwenden oder welche Seiten besonders beliebt sind um dadurch ihr Webangebot anpassen zu können. Den Administratoren dienen die Protokolle zur Kontrolle des monatlichen

Datentransfervolumens und ermöglichen es unerlaubte Zugriffsversuche zu erkennen und entsprechend darauf zu reagieren. Potentielle Angreifer können mit Hilfe des Zugriffsprotokolls lokalisiert und ausgesperrt werden.

Die Informationen, die hierbei üblicher Weise erhoben werden, sind:

- Aufgerufene Seite
- IP-Adresse (dadurch indirekt Standort)
- Datum und Uhrzeit des Zugriffs
- Angeforderte Ressourcen (URL)
- Betriebssystem des Benutzers
- Browser des Benutzers
- Bildschirmauflösung

Die gesammelten Informationen sind in der Regel nur Systemadministratoren zugänglich. Da für die Instandhaltung und Wartung des Servers nicht auf Administratoren verzichtet werden kann, muss durch organisatorische Maßnahmen sichergestellt werden, dass die Anzahl der zugriffsberechtigten Personen möglichst gering bleibt. Bei Zusammenarbeit mit externen Dienstleistungsunternehmen müssen Zugriffe auf den Server durch vertragliche Regelungen beschränkt werden. In einem Wartungsprotokoll muss der Grund und der zeitliche Umfang des Zugriffs dokumentiert werden. Ein Geheimhaltungsvertrag ist sinnvoll und sollte die Weitergabe von vertraulichen Informationen an Dritte explizit ausschließen.

6.5.4 Daten innerhalb des Studienportals

Während der Registrierung müssen nur wenige personenbezogene Daten erhoben werden. Hierzu zählen:

- Benutzername (frei wählbar)
- Passwort
- Emailadresse
- (ggf. Wohnort)

Benutzername und Passwort sind zwingend erforderlich, um sich als berechtigter Benutzer am Studienportal anzumelden. Das Passwort wird hierbei mit einer *Hashfunktion* codiert und in der Datenbank gespeichert. Die *Hashfunktion* lässt ohne entsprechenden Rechenaufwand keinen Rückschluss auf das Passwort des Benutzers zu, selbst wenn ein Zugriff auf die Datenbank besteht. [15] (☞ Kapitel 6.5.5) Dies bedeutet, dass auch Administratoren das Passwort der Benutzer nicht auslesen und sich auch nicht mit deren Kennung am System anmelden können. Innerhalb des Studienportals erfolgt die Kommunikation zwischen Studienbetreuern und Teilnehmern statt echter Name mit dem gewählten Benutzername (z.B. sani2014). Das Hinterlegen des richtigen Namens für die Funktion des Studienportals ist nicht zwingend erforderlich. Die Emailadresse dient dem System als eindeutiges Identifikationsmerkmal jedes Teilnehmers. Für Studienmanager und Studienbetreuer kann sie einsehbar gemacht werden, vor Teilnehmern bleibt sie verborgen. Eine Angabe des Wohnorts ist für die Nutzung des Systems nicht erforderlich, kann aber für die Gruppierung von Daten (z.B. zum Filtern nach Städten oder Regionen) hilfreich sein. Statt eine Freitext-Eingabe des Wohnorts zu erlauben kann es sinnvoll sein die Auswahl des Wohnorts durch vorgefertigte Auswahllisten zu beschränken. An Stelle einer freien Antwortmöglichkeit kann der Benutzer über ein Listenfeld die nächst größere Stadt (Regensburg, München, Nürnberg,...), eine Region (Oberpfalz,

Oberbayern, Oberfranken,...) oder das Studienzentrum (Uniklinikum Regensburg, LMU München, Uniklinik Nürnberg) auswählen.
Neben den 3 Pflichteingaben besteht die Möglichkeit während der Registrierung weitere Zusatzinformationen in die Datenbank einzuspeichern. Hierzu zählen: Anrede, Geburtsdatum, Geschlecht, usw.

6.5.5 Anonymisierung der Fragebögen

Einen weiteren wichtigen Punkt bei der Nutzung des Studienportals stellt die Zuordnung von Benutzern zu Fragebögen dar. Für das System muss eine 1:n Verknüpfung zwischen einer Person und einem oder mehreren Fragebögen möglich sein ohne von den Fragebogendaten auf den Teilnehmer rückschließen zu können. Die Lösung für dieses Problem ist eine Pseudonymisierung. Jeder eingereichte Fragebogen wird unter einem teilnehmerspezifischen Pseudonym statt dem tatsächlichen Benutzernamen gespeichert.
Um zu verhindern, dass ein Rückschluss von Pseudonym auf die Identität der Teilnehmer möglich ist, wird vor dem Abspeichern des Fragebogens mit Hilfe eines unidirektionalen Verfahrens ein Pseudonym erzeugt. Der Algorithmus generiert für den jeweiligen Teilnehmer immer das gleiche Pseudonym. Auf diese Weise können Fragebögen weiterhin vom System verarbeitet oder für externe Programme wie SPSS oder Excel exportiert werden ohne die Identität der Teilnehmer preiszugeben.

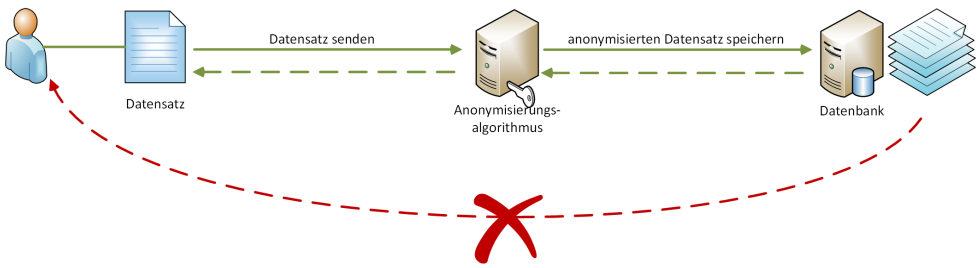


Abbildung 28: Prinzip der Pseudonymisierung

Beschreibung des Algorithmus:
In der Datenbank befinden sich viele konstante oder eindeutige Daten, die mit den Teilnehmern verknüpft sind: der Benutzername, der bei der Registrierung ausgewählt wurde, die Emailadresse, ggf. Vor- und Nachname und weitere Informationen wie Zeitpunkt der Registrierung, erster Zugriff auf das System, usw. Für einen fiktiven Benutzer: Max Mustermann, könnte ein Datensatz wie folgt aussehen:

Benutzername: sani2014
Emailadresse: max@mustermann.de
Name: Max Mustermann
Studie: SURE für den Rettungsdienst
Registrierungszeitpunkt: 01.01.2013, 15:24:23

Aus diesen Informationen werden nach bestimmten Kriterien Elemente ausgewählt und daraus eine Kennung erstellt.

Die Kriterien an unserem vereinfachten Beispiel könnten sein:

Erster Buchstabe des Vornamens: m
Zweiter Buchstabe des Nachnamens: u
Emailadresse: max@mustermann.de
Monat des Registrierzeitpunkts: 01
Stunde des Registrierzeitpunkts: 15

Kennung: mumax@mustermann.de0115

Die daraus erhaltene Kennung *mumax@mustermann.de0115* wird im zweiten Schritt mit Hilfe einer unidirektionalen Codierung zum Pseudonym.

Das Sicherheitskonzept besteht somit aus zwei Stufen:

a) Erzeugung einer Kennung
b) Kryptografische Umwandlung der Kennung

Grundlage der Umwandlung ist eine mathematische Funktion (*Hashfunktion*), die auf die Kennung angewendet wird, um daraus eine Zahlen/Buchstabenkombination zu erzeugen. Die bekannteste und recht überschaubare kryptografische *Hashfunktion* ist *MD5*. [103] Sie wird in vielen Content-Management-Systemen, Internet-Foren, Portalsystemen oder Online-Shops eingesetzt um Passwörter in Internet-Datenbanken zu speichern. Daher soll am Beispiel von *MD5* die prinzipielle Funktionsweise von *Hashfunktionen* veranschaulicht werden:

Die eigentliche Nachricht (in unserem Beispiel die erzeugte Kennung) wird zunächst von Buchstaben in Zahlen umgewandelt. Für jeden Buchstaben bzw. jedes Sonderzeichen existiert eine Zahl in einem Computer-Zeichensatz *(ASCII Zeichensatz*).
Aus mumax@mustermann.de0115 wird die Zahlenkombination:

109 117 109 97 120 64 109 117 115 116 101 114 109 97 110 110 46 100 101 48 49 49 53

MD5 simuliert einen 32bit Prozessor der 32bit lange Zeichen speichern kann. Dieser Prozessor besitzt vier Speicherbausteine (*Register*). So ergibt sich die Gesamtlänge von 4*32 = 128bit. Die Nachricht wird nun nach einer für *MD5* typischen Formel auf die Register verteilt. Innerhalb der Register finden weitere binäre Operationen statt, die den Inhalt des Registers bildlich gesprochen durchmischen. [103] Am Ende werden die Ergebnisse der Register wieder zu einer Nachricht zusammengefügt und ergeben das Pseudonym. Aus *mumax@mustermann.de0115* wird *347a5cef83f66828c365a5d36bfefb53*
Dieses Pseudonym kann nun nicht mehr zu *mumax@mustermann.de0115* zurückgerechnet werden. Das besondere an *Hashfunktionen* ist, dass eine Kennung immer das gleiche Pseudonym liefert. Um von *347a5cef83f66828c365a5d36bfefb53* auf *mumax@mustermann.de0115* rückschließen zu können, müsste man bei *MD5 Kodierung* mindestens 2^{128} Möglichkeiten durchspielen und würde eine Mindestanzahl von $3{,}40 \times 10^{38}$ Rateversuche benötigen. [128] Bereits recht früh wurde in unterschiedlichen Ansätzen beschrieben, wie man *MD5* brechen könnte. [105] Durch den Einsatz moderner Hardware ist die benötigte Rechenzeit hierfür deutlich reduziert. Mit handelsüblicher PC-Hardware (z.B. Grafikkarten) können die notwendigen Berechnungen in sehr kurzer Zeit ausgeführt werden. Eine entsprechende Bauanleitung wurde im Computermagazin c´t in Ausgabe 6/2009 veröffentlicht. Im Internet kursieren Webseiten die *MD5* Entschlüsselungen anbieten. Diese arbeiten dabei mit umfangreichen Datenbanken (*Rainbow Tables)* mit bekannter Zuordnung von Passwort zu *Hash*. Dank dieser Lis-

ten lässt sich der Passwort-Hash *16d7a4fca7442dda3ad93c9a726597e4* binnen Sekunden zu *test1234* dechiffrieren. Neben *MD5* gibt es eine Reihe alternativer *Hashverfahren*, die mit größerer Schlüsselstärke und anderer Durchmischung arbeiten und daher als sicherer eingestuft werden. Typische Vertreter sind zum Beispiel *mcrypt* (128bit), *SHA-1* (160bit), *RIPMED-160* (160bit), *SHA-256* (256bit) und *SHA-512* (512bit). [15]

Durch den Einsatz von Kennung und unidirektionaler Codierung ist für Außenstehende (z.B. Studienbetreuer) selbst bei bekannten Informationen wie Benutzername, Name und Emailadresse ein Errechnen des Pseudonyms nicht möglich. Die einzige Gefahr besteht darin, dass durch Zugriff auf den Quellcode (z.B. durch Administratoren oder Hacker) der Algorithmus zur Erzeugung der Kennung und die verwendete *Hashfunktion* bekannt wird. In diesem Fall kann mit Hilfe des Algorithmus für jeden in der Datenbank vorhandenen Teilnehmer ein Pseudonym generiert und in Listenform gespeichert werden. Diese Liste kann dann mit den eingereichten Fragebögen abgeglichen werden. Somit wäre eine Zuordnung von Fragebogen zu Teilnehmer möglich. Aus Sicherheitsgründen wird daher der genaue Algorithmus, der die Kennung und das Pseudonym erzeugt, nicht in der schriftlichen Arbeit erläutert. Der Algorithmus befindet sich kommentiert im Quellcode des Studienportals.

6.5.6 Vollständige Anonymität – eine Gefahr?

Eine vollständig anonyme Nutzung des Studienportals ist technisch durchaus denkbar. Mit Hilfe von „Einweg-Emailadressen“ kann die Registrierung am Studienportal erfolgen. Es existieren Emaildienste (z.B. *www.10minutemail.com*), die eine Emailadresse für die Dauer von wenigen Minuten bereitstellen. Für diese Dienste ist keine Registrierung erforderlich. Nach Ablauf der Zeit zerstört sich die Mailadresse selbständig. Mit Hilfe einer solchen temporären Emailadresse kann die Registrierung am Studienportal erfolgen. Als bestätigter Benutzer hat der Teilnehmer nun die Möglichkeit das System zu nutzen, ohne über seine Emailadresse identifizierbar zu sein. Um seine IP-Adresse zu verschleiern, kann er den Datenverkehr über anonyme Proxyserver oder ein VPN umleiten. Die wahre Identität und Herkunft des Anwenders ist nun kaum mehr nachvollziehbar. Diese Möglichkeit ist nicht unproblematisch. Teilnehmer können nicht mehr vom System oder von Studienbetreuern per Mail benachrichtigt werden um z. B. an die Abgabe der Fragebögen erinnert zu werden. Die Funktion „Passwort vergessen“ mit deren Hilfe das gewählte Passwort zurückgesetzt werden kann, erfordert ebenfalls eine gültige Emailadresse. Fraglich ist auch wie genannte Dienste mit den Daten der temporären Mailadressen umgehen. Es bestünde über die Bestätigungsemail oder Passwort-Rücksetzfunktion theoretisch die Möglichkeit Zugang zum Portal zu erlangen. Auch aus sicherheitstechnischer Sicht gilt es zu bedenken, dass ein anonymer Nutzer innerhalb des Systems ein potentielles Risiko darstellt. Unbekannte Schwachstellen im Quellcode des Portals könnten ausgenutzt werden um auf die Datenbank zuzugreifen. Es liegt daher in der Verantwortung des Studienmanagers zu erwägen, ob echte Emailadressen Voraussetzung für die Registrierung sein müssen. Es können für das Portal zulässige Emailadressen definiert und unerwünschte Emailadressen ausgeschlossen werden. So können für die Studiengruppe „SURE für den Rettungsdienst“ nur Nutzer mit Emailadressen *@brk.de, *@asb.de, *@johanniter.de usw. zugelassen werden. Auch eine manuelle Freischaltung von Teilnehmern durch den Studienbetreuer ist technisch möglich.

6.5.7 Cookies

Der Einsatz von Cookies ist ein weiterer datenschutzrelevanter Punkt. Die kleinen Textdokumente mit maximal 4.000 Zeichen werden bei (fast) jedem Webseitenaufruf vom Browser erzeugt und verbleiben auf dem Computer des Anwenders. Sie werden verwendet um temporär Daten zu speichern, die von den Erzeugerwebseiten zu einem späteren Zeitpunkt wieder abgefragt werden. Sie können je nach Programmierung unterschiedliche Informationen enthalten. Mit Hilfe der Cookies merkt sich ein Webshop zum Beispiel Warenkörbe oder zuletzt betrachtete Produkte. Cookies werden jedoch auch verwendet, um Sitzungsinformationen kurzfristig zu speichern. In Homebanking Anwendungen ist die Gültigkeit einer Sitzung meist recht kurz. Nach wenigen Minuten verliert das Cookie seine Gültigkeit. Die Sitzung wird dadurch automatisch beendet und der Benutzer ausgeloggt. Cookies sind standardisiert und enthalten folgende Information: [101]

- Name
- Inhalt (*value*)
- Herkunft (*Domain, Path*)
- Gültigkeit (*Expires, Max-Age*)
- Sicherheitseinstellungen (*Secure, HttpOnly, Discard*)

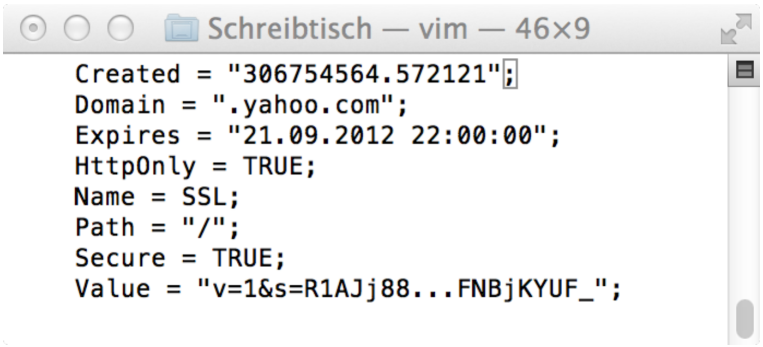


Abbildung 29: Beispiel eines Session Cookie von Yahoo.com

Bei der Anmeldung an Online-Systemen passiert in der Regel Folgendes:
Der Anwender meldet sich an der Webseite mit seinem Benutzernamen und Passwort an. Der Webserver überprüft die Zugangsdaten und stellt eine Anfrage an den Browser mit der Bitte, ein Cookie abspeichern zu dürfen. Ist dies in den Browsereinstellungen verboten (Cookies nicht erlaubt), kommt es zur Fehlermeldung. Sind Cookies erlaubt, was der Grundeinstellung der meisten Browser entspricht, wird ein Cookie mit einer Sitzungsnummer auf dem Computer des Benutzers hinterlegt. Für die Dauer der Sitzung bleibt die Sitzungsnummer erhalten. Meldet sich der Teilnehmer vom System ab verliert die Sitzungsnummer ihre Gültigkeit. Ein Zugriff auf Inhalte des Studienportals ist ohne gültige Sitzungsnummer nicht mehr möglich. Meldet sich der Nutzer nicht ordnungsgemäß vom System ab bleibt das Cookie so lange gültig wie es in der Gültigkeitsdauer vom Programmierer definiert wurde.

Da dem Studienportal mit Hilfe des *Session Cookies* mitgeteilt wird, dass der Teilnehmer berechtigt ist auf den internen Bereich zuzugreifen, birgt dies einige Risiken. Bei einem gemeinsam genutzten Computer kann das *Session Cookie* in falsche Hände geraten. Behält das Cookie seine Gültigkeit, kann ein anderer Benutzer dies ausnutzen und auf den internen Bereich zugreifen, ohne sich einloggen zu müssen. Ein Problem besteht auch dann, wenn ein Cookie über eine unverschlüsselte Verbindung übertragen wird. Findet die Übertragung innerhalb eines lokalen Netzes statt, können andere Nutzer des Netzwerks den Datenverkehr belauschen

und so den Inhalt des Cookies (mit der sensiblen Sitzungsnummer) lesen und sich so, ohne das Passwort kennen zu müssen am Studienportal anmelden.

Der Einsatz von Cookies kann jedoch nach *RFC 6265* Standard durch die Festlegung von Sicherheitseinstellungen sicherer gemacht werden. Für die Nutzung von Cookies im Studienportal werden daher folgende Maßnahmen ergriffen:

Die Cookie-Einstellung *Secure=1* erzwingt eine *SSL* verschlüsselte Übertragung des Cookies und stellt somit sicher, dass ihr Inhalt nicht durch Lauschangriffe aus dem Netzwerk gelesen werden kann. Mit der Funktion *Discard* wird der Browser angewiesen, das Cookie nach beendeter Sitzung sofort zu löschen. Auf diese Weise können alte Cookies und die darin enthaltenen alten Sitzungsdaten nicht missbraucht werden. Die Dauer der Sitzung ist auf eine definierbare Zeitspanne (Minuten bis wenige Stunden) begrenzbar. Mit Ablauf der Frist verliert das Cookie seine Gültigkeit und wird gelöscht. Mit der Einschränkung *session.cookie_httponly=1* wird verhindert, dass das *Session-Cookie* von anderen Programmen (z.B. JavaScript) abgefragt werden darf. Somit kann das Cookie nur vom Webserver der jeweiligen Webseite gelesen werden. Zusätzlich muss durch entsprechende Einstellung am Webserver verhindert werden, dass während einer Sitzung die Verbindung von einer verschlüsselten (*https*) auf eine unverschlüsselte (*http*) Verbindung gewechselt werden kann.

6.6 Datensicherheit

Im Vergleich zum Datenschutz dient die Datensicherheit nicht dem Schutz persönlicher Daten vor ungewolltem Zugriff sondern dem Schutz gespeicherter Daten vor Verlust.
Es gibt zwei grundsätzliche Maßnahmentypen bei der Datenwiederherstellung:

- Präventivmaßnahmen
- Notfallwiederherstellung (*disaster recovery*)

Bei Onlinestudien mit einer großen Anzahl an Benutzern und hohen Datenmengen ist der Verlust von Daten unbedingt zu vermeiden. Daher müssen bereits im Vorfeld technische und organisatorische Maßnahmen getroffen werden um den Verlust von Daten zu verhindern. Auch die Verfügbarkeit der Internetplattform ist von großer Bedeutung. Ist sie nicht verfügbar, können keine Studienergebnisse gesammelt werden. Langsame Geschwindigkeit oder Nicht-Verfügbarkeit des Systems kann sich negativ auf die Compliance der Teilnehmer auswirken. Das folgende Kapitel beschäftigt sich daher mit technischen Konzepten, die eine maximale Verfügbarkeit des Studienportals und Sicherheit der Daten gewährleisten sollen.

Präventivmaßnahmen sind in erster Line eine solide Konzeption und Implementierung auf Software- wie Hardwareebene. Notfallwiederherstellung kann durch unterschiedliche Backup-Strategien umgesetzt werden. Die folgende Abbildung visualisiert die Wertigkeit und die Abhängigkeiten der im folgenden Kapitel beschriebenen Elemente.

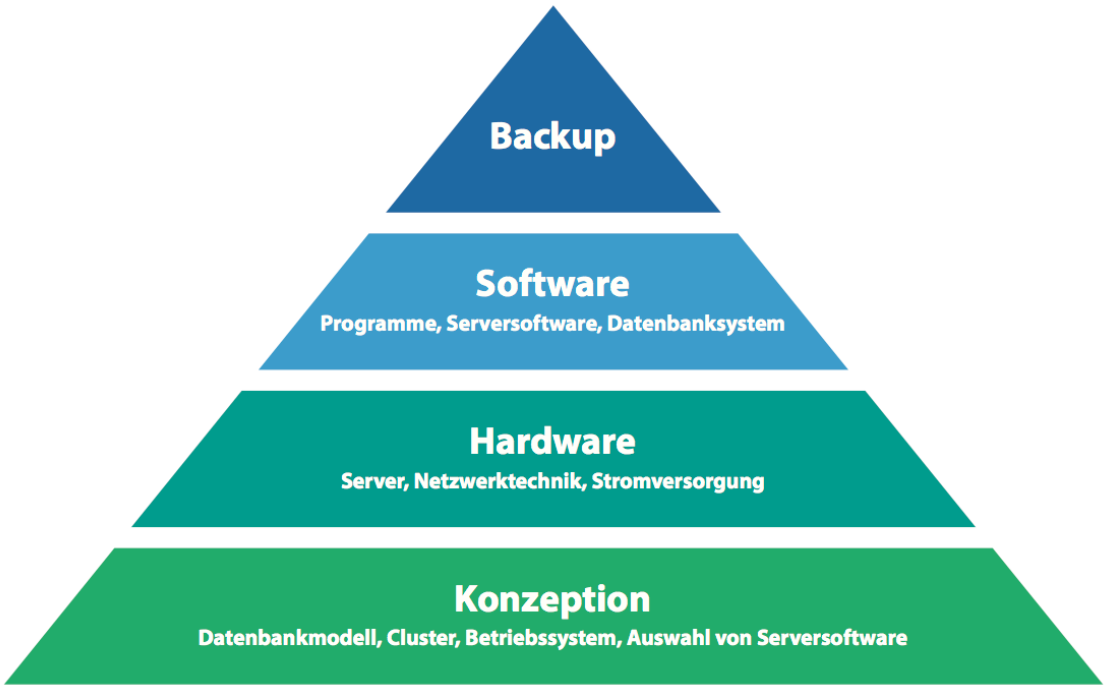


Abbildung 30: Hierarchie der Einflussfaktoren zur Datensicherheit

6.6.1 Server und Hardware

Um die Verfügbarkeit der Plattform zu gewährleisten, ist eine Redundanz des Servers und all seiner Komponenten zu gewährleisten.
Hierzu zählen:

- Festplatte
- Netzteil
- Mainboard und Prozessor
- Arbeitsspeicher
- Netzwerk/Außenanbindung
- Strom
- Anbindung an das Internet

Besonders häufig von Ausfällen betroffen sind mechanisch bewegliche Elemente des Systems. Festplatten haben aufgrund thermischer Belastung und des mechanischen Verschleißes (Schreib-Leseköpfe sowie rotierende Magnetscheiben) eine begrenzte Haltbarkeit. Das Risiko für Festplattendefekte beträgt innerhalb der ersten 3 Monate etwa 3% und steigt bei einer Betriebsdauer von zwei bis drei Jahren auf 8,6%. [99] Aus diesem Grund werden in der Praxis Verfahren eingesetzt, die Daten auf mehreren Festplatten vorhalten. Ein grundlegendes Verständnis von Redundanzansätzen ist wichtig, da diese Konzepte auf andere Komponenten des Gesamtsystems (Server, Netzwerk, Internetanbindung) übertragen werden können. Das Verteilen von Daten auf mehreren Festplatten wird *Raid* (***R**edundant **A**rray of **I**ndependent **D**isks*) genannt. Werden sämtliche Daten auf einer zweiten Festplatte bereitgehalten, spricht man von gespiegelten Festplatten oder *Raid 1*. Eine andere Möglichkeit ist das Verteilen von Datenelementen auf mehreren Festplatten (*Raid 5*). Die mehrfache Datenhaltung geht in beiden Fällen auf Kosten der Speicherkapazität. Bei der Spiegelung (*Raid 1*) wird die gewünschte

Festplattenkapazität ein weiteres Mal benötigt, bei der Verteilung der Daten (*Raid 5*) ist der Verlust an Speicherkapazität durch Redundanz mit 33% im Vergleich zu *Raid 1* (50%) geringer, es werden jedoch mindestens drei Festplatten benötigt. Neben den beiden genannten klassischen *Raid* Methoden existieren Mischformen, auf die nicht weiter eingegangen werden soll. Während *Raid 1* durch das gleichzeitige Lesen von zwei Festplatten die Geschwindigkeit steigert und beim Ausfall einer Festplatte kaum nennenswerte Geschwindigkeitseinbußen zur Folge hat, geht bei *Raid 5* im Falle eines Ausfalls, Geschwindigkeit durch die Rekonstruktion der Daten verloren. *Raid 5* ist ein guter Kompromiss zwischen Geschwindigkeit und Speichereinbußen durch Redundanz. [168] Durch die Preisentwicklung bei Speichermedien und die Verfügbarkeit großer Festplatten erfreut sich *Raid 10* zunehmender Beliebtheit. *Raid 10* ist eine Kombination aus Spiegeln (*Raid 1*) und Aneinanderhängen *Raid 0* von Festplatten. Durch das Verketten (*Raid 0*) wird aus zwei Festplatten eine große Festplatte. Die zu speichernden Daten (Datenblöcke) werden auf die beiden Festplatten verteilt. Hierdurch wird sowohl Lese- als auch Schreibgeschwindigkeit deutlich erhöht. Jedoch hätte der Ausfall einer Festplatte einen kompletten Datenverlust zur Folge. Daher spiegelt man das *Raid 0* System mit Hilfe von *Raid 1*. Das Spiegeln stellt die nötige Sicherheit zur Verfügung. Für *Raid 10* Systeme sind mindestens 4 Festplatten notwendig.
Neben Festplatten zählen Netzteile ebenfalls zu den ausfallträchtigen Komponenten eines Servers. Moderne Server besitzen daher meist zwei Netzteile und stellen somit eine redundante Stromversorgung bereit. Es gibt jedoch Hardwarebausteine, die nicht doppelt in einem Server vorgehalten werden können. Die Hauptplatine (Mainboard) ist ein solches Bauteil. Ein Defekt führt zum Ausfall des gesamten Systems. In der IT-Welt haben sich daher Konzepte etabliert, mit denen gesamte Server redundant gehalten werden können. Solche Systeme setzen sich aus einem Netzwerk von mehreren Servern zusammen und bilden einen *Cluster*. Die Server, die Teil eines *Clusters* sind, heißen Knoten (*Nodes*). Analog zu den Redundanzprinzipien bei Festplatten gibt es das Konzept der redundanten Daten- oder Aufgabenverteilung auch auf Serverebene.

Beinhalten zwei (oder mehrere Server) den gleichen Datenbestand und führen die gleiche Aufgabe aus, so kann beim Ausfall eines Servers der Ausfall kompensiert werden (*Mirroring Prinzip*). Um die Daten zu replizieren ist ein permanenter Datenabgleich zwischen allen Knoten notwendig, wodurch der Netzwerkverkehr steigt. [51] Für *Mirroring* ist daher in besonderem Maße eine stabile und schnelle Netzwerkverbindung zwischen den Knoten notwendig. Es kommen daher oft besonders schnelle (aber auch teurere) Netzwerkkomponenten mit Glasfaserleitungen (*FibreChannel*) oder Drahtleitungen (*InfiniBand*) zum Einsatz.

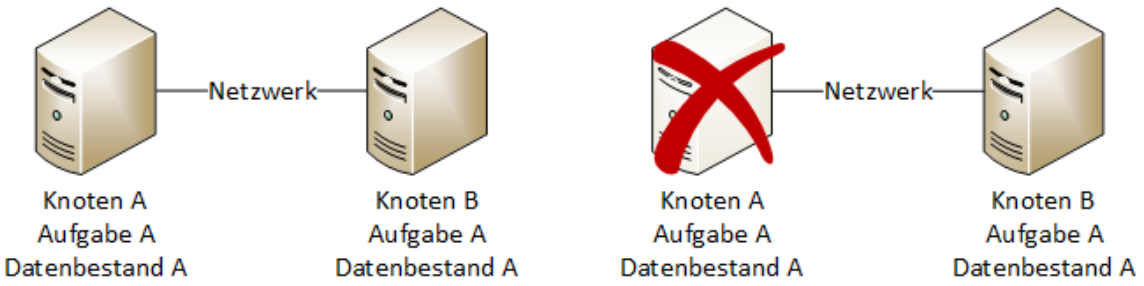


Abbildung 31: Mirroring Prinzip von Nodes entspricht Raid 1

Werden die Daten von zwei (oder mehreren) Servern gleichzeitig angeboten, kann dies einen Geschwindigkeitsvorteil durch Lastverteilung bringen. Eine weitere Komponente (*Load-Balancer*) muss jedoch den Datenverkehr nach außen regeln. Ein defekter Knoten verursacht

auch in diesem Fall keine Ausfallzeit. Jedoch steigt die Last auf den oder die verbliebenen Knoten. Dieser *aktiv/aktiv Konfiguration* genannte Ansatz ist weniger flexibel und besitzt eine höhere Komplexität, da die Serversoftware eine Unterstützung für diese Konfiguration mitbringen muss. Häufiger werden daher *aktiv/passiv Konfigurationen* eingesetzt. Hierbei stellt nur der Hauptknoten die Daten zur Verfügung (aktiv). Der oder die weiteren Knoten bleiben als Reserve im Hintergrund (passiv). Fällt der Hauptknoten aus, übernimmt ein Reserveknoten die Aufgabe und wird so zum Hauptknoten. Wenn die Reserve abrufbereit auf den Einsatz wartet, wird diese Konfiguration auch *hot standby* bezeichnet. Die Konfiguration ist einfacher, da sie nicht auf Unterstützung der Serversoftware angewiesen ist. Dafür entsteht für die Zeit des „Umschaltens“ auf die Reserve eine geringe Ausfallzeit. [51] Nachteil beider genannter Konfigurationen ist, dass die Rechenleistung der Knoten für keine andere Aufgabe zur Verfügung steht.

Kommt ein zentraler Datenspeicher zum Einsatz auf den die Knoten zugreifen können, spricht man vom *Shared Prinzip*. Hierbei unterscheidet man zwei Varianten:

Beim *Shared All Prinzip* konkurrieren alle Knoten um den gemeinsamen Datenspeicher. Fällt ein Knoten aus, tritt ein anderer Knoten des *Clusters* an dessen Stelle und greift dabei auf die gemeinsamen Daten zurück. Beim *Shared All* Prinzip in aktiv/aktiv Konfiguration erzeugt ein Knotenausfall keine Ausfallzeit. Es besteht jedoch die Gefahr der Datenkorruption, wenn mehrerer Knoten auf den gemeinsamen Speicher schreiben dürfen. Um dies zu vermeiden, muss von der Serversoftware ein Mechanismus bereitgestellt werden, der sicherstellt, dass nur ein Knoten auf den Speicher schreiben darf (*Distributed Lock Manager*).

Häufiger verwendet wird das *Shared Nothing* Prinzip.
Hierbei besitzt jeder Knoten einen eigenen gekapselten Speicherbereich auf dem gemeinsamen Speicher. Im Bedarfsfall kann ein Knoten den Speicher- und Aufgabenbereich eines ausgefallenen Knotens übernehmen. Der Vorteil des *Shared Nothing* Prinzips besteht darin, dass jeder Knoten im Cluster eigenen Aufgaben nachgehen kann und nur im Bedarfsfall einspringen muss. Für Redundanz geht im Regelbetrieb keine Rechenleistung verloren. [77]

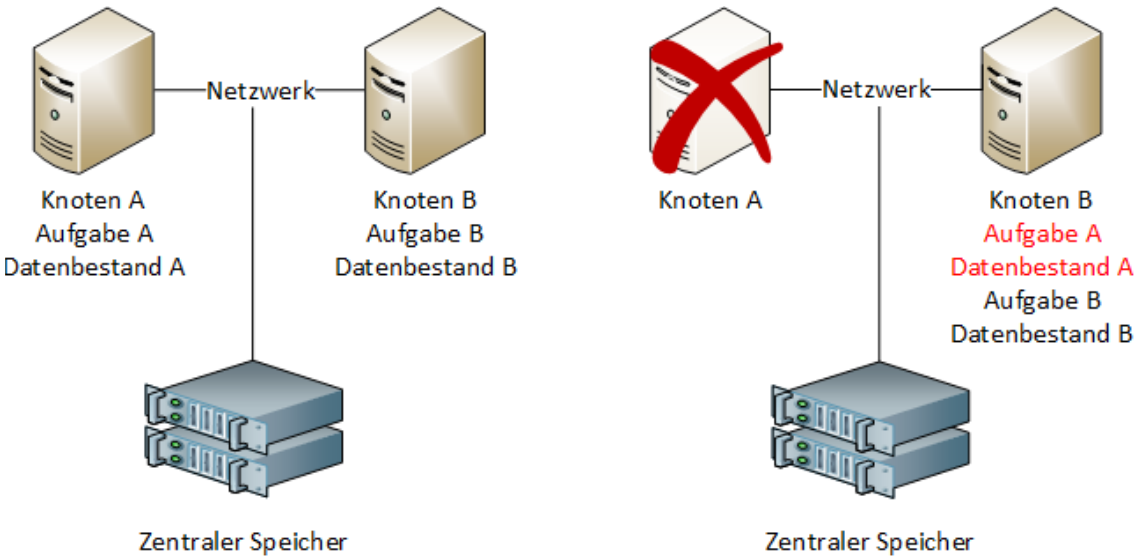


Abbildung 32: Shared All links Regelbetrieb, rechts Ausfall von Knoten A

Ein anderer und zunehmend populärer Ansatz verteilt die Daten (ähnlich *Raid 5*) auf mehrere Knoten des *Clusters*. Es entsteht ein Speicher, der sich aus vielen einzelnen Datenelementen zusammensetzt. Die Sicherheit und die Geschwindigkeit steigen mit der Anzahl der Knoten im Cluster. Eine Erweiterung des Speichers erfolgt durch Hinzufügen weiterer Knoten. Da dieses Konzept eine Mindestanzahl an drei Knoten voraussetzt, wird es meist in größeren Clustern (> 10 Knoten) eingesetzt. Als Vorteil gilt neben der einfachen Skalierbarkeit der Leistungsgewinn durch parallelen Zugriff auf Datenelemente, die auf unterschiedlichen Knoten vorgehalten werden. Allerdings ist hier ein schnelles und zuverlässiges Clusternetzwerk Grundvoraussetzung, weshalb dieses Setup nur dann Sinn macht, wenn teurere Netzwerktechnik (z.B. 10Gbit Ethernet oder Infiniband) zum Einsatz kommt. Werden Daten über einen *Cluster* verteilt, spricht man von einem *Cluster-Dateisystem*.

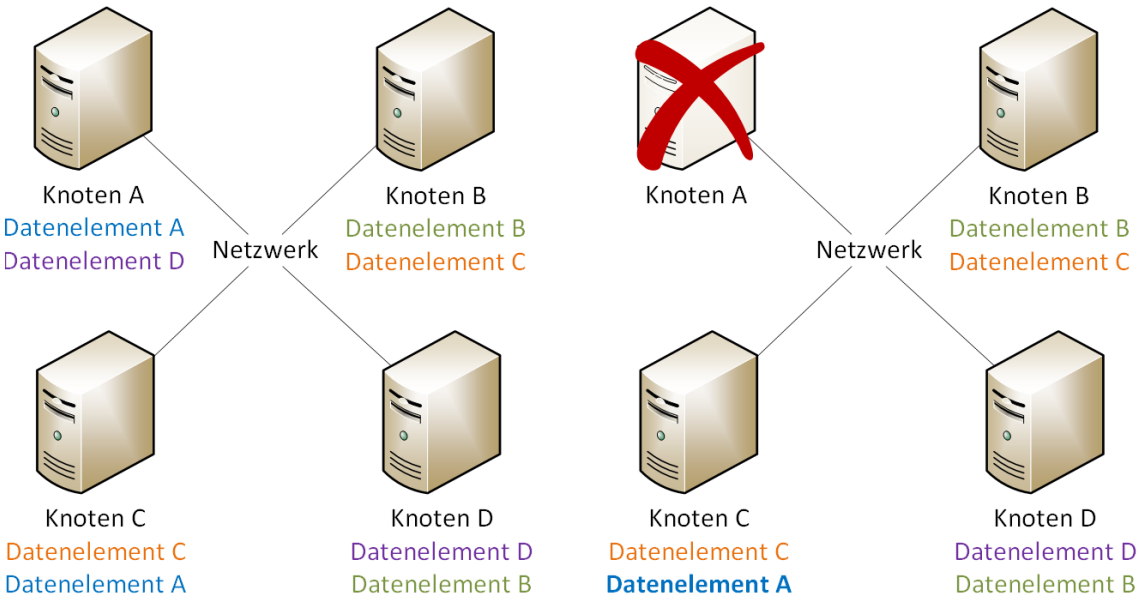


Abbildung 33: Cluster-Dateisystem links: Regelbetrieb, rechts: Ausfall von Knoten A

Da alle genannten Prinzipien der Ausfallsicherheit eines Gesamtsystems dienen, werden sie auch als *HA-Cluster* (*High Availability Cluster)* bezeichnet.

Für ausfallsichere Systeme müssen neben den Servern auch alle zum Betrieb notwendigen Komponenten berücksichtigt werden. Man spricht in diesem Zusammenhang vom *Single Point of Failure* - also einem Glied in der Kette das bei seinem Ausfall das gesamte System lahmlegt. Daher spielen in Hochverfügbarkeitsanwendungen auch Gegebenheiten des Serverstandorts eine wichtige Rolle. Hierzu zählen redundante Netzwerktechnik, redundante und lastverteilte Anbindung an große Internetknotenpunkte, unterbrechungsfreie Stromversorgung (*USV*), redundante Klimatechnik, usw. [77] Die Wahl des richtigen Rechenzentrums ist ebenso wichtig, wie die Wahl eines geeigneten Cluster-Konzepts, da der Grad der Verfügbarkeit von allen potentiellen Fehlerquellen abhängt. [169] Durch Redundanz auf allen Ebenen können Verfügbarkeiten bis zu 99,99999% erreicht werden, was einer ungeplanten Ausfallzeit von 3 Sekunden pro Jahr entspricht. Ein vom Kosten/Nutzenverhältnis vertretbares Maß an Aufwand ist eine Verfügbarkeit von 99,99% (53 Minuten ungeplante Ausfallzeit pro Jahr). [169]

Verfügbarkeit	Ausfallzeit/Jahr	Klasse	Einsatzzweck
99,00%	3d 15h	1	Verfügbarkeit nicht essentiell. (z.B. private Webseite)
99,90%	8h 46min	2	Betrieb darf nur minimal unterbrochen werden. (z.B. normale Webseiten)
99,95%	4h 23 min	3	Betrieb muss während der Hauptgeschäftszeit gewährleistet sein. (z.B. eLearning Plattformen)
99,99%	53 min	4	Hochverfügbarkeit Dauerhafter Betrieb muss rund um die Uhr gewährleistet sein. (z.B. Online Shops, KIS)
99,999%	5 min	5	Hochverfügbarkeit Dienste müssen unter allen Umständen ununterbrochen verfügbar sein. (z.B. Rettungsleitstellen)

Tabelle 3: Verfügbarkeit von IT-Diensten und deren Einsatzgebiete [84] [77]

6.6.1.1 Betriebssystem

Die Grundlage aller Softwarekomponenten bildet das Betriebssystem. Auf heimischen Computern kommt nach wie vor meist Windows zum Einsatz. [170] Für den Einsatz in Serverumgebungen gibt es jedoch deutlich mehr Alternativen. Neben der Servervariante von Microsoft Windows stehen eine Reihe von Linux und BSD basierten Serverbetriebssystemen zur Auswahl. Für die Wahl des Betriebssystems gibt es folgende Punkte zu berücksichtigen:

- Lizenzkosten
- Wartungsaufwand
- Systemanforderungen
- Sicherheit

Windows basierte Serverbetriebssysteme unterscheiden sich von Linux bzw. BSD basierten Betriebssystemen und Unix durch ihr Lizenzmodell. BSD (Barkley Software Distribution) Betriebssysteme sind den Linux Systemen lizenzrechtlich artverwandt, werden jedoch aufgrund ihrer Herkunft den Unix-Systemen zugerechnet. Unix Systeme haben ihre Wurzeln in den siebziger Jahren. Dabei handelt es sich um Betriebssysteme, die für spezielle Aufgaben (z.B. Telekommunikationsaufgaben) oder für spezielle Hardware konzipiert wurden. Die beiden bekanntesten Vertreter sind Solaris (Sun Microsystems) oder HP-UX (Hewlett Packard). Beide Hersteller optimieren ihre Systeme auf leistungsfähige eigene Hardware. Obwohl sich Linux und BSD recht ähnlich sind, unterliegen sie unterschiedlichen Lizenzmodellen. Sie unterscheiden sich aus Sicht des Administrators primär in der Systemarchitektur (*Kernel*) und der technischen Dokumentation. [48]
Windows Server werden als kommerzielle Produkte angeboten und sind entsprechend der Version, dem Funktionsumfang und Hardwareeinsatz (z.B. Anzahl der Prozessoren oder Arbeitsspeicher) kostenpflichtig. Die meisten Linux und BSD Varianten sind als freie Software kostenlos verfügbar. Windows Systeme werden von ihrem Hersteller in vorgefertigter Form ausgeliefert (*Closed Source*). Ein Einblick in den der Software zu Grunde liegenden Quellcode ist nicht möglich. Somit können Entwickler das Betriebssystem nicht verändern. Im Gegensatz hierzu legen Linux-Systeme den Quellcode offen. Jeder interessierte Anwender kann das Betriebssystem genau studieren und im Bedarfsfall modifizieren. [58] Unterhalten und

weiterentwickelt wird der Linux-Kern (*Kernel*) von über 10.000 freien Entwicklern und über 1.000 Firmen weltweit. [62] Das *OpenSource* Prinzip führte dazu, dass sich eine Vielzahl unterschiedlicher Linux-Varianten entwickeln konnte. Die Distributionen (z.B. Debian, Ubuntu, Redhat, Suse Linux,...) unterscheiden sich hauptsächlich in der Zusammenstellung der Softwarepakete, der Dokumentation bzw. den unterschiedlichen Anwendungsgebieten. So gibt es Distributionen, deren Haupteinsatzgebiet Serverdienste darstellen, wohingegen andere auf Endanwender und deren Bedürfnisse abgestimmt sind. Linux Systeme, die für den Servereinsatz konzipiert wurden, verfügen in der Regel über keine grafische Benutzeroberfläche, sondern werden über eine Kommandozeile (*Konsole*) administriert. Windows Server Systeme setzen auf benutzerfreundliche Oberflächen, die für die Verwaltungsaufgaben entsprechende Programme und Masken anbieten. Da die Darstellung grafischer Benutzeroberflächen Ressourcen (Prozessor und Arbeitsspeicher) benötigt, haben Linux -basierte Systeme in der Regel geringere Hardwareanforderungen. Die Verwaltung dieser Systeme erfordert jedoch tiefer greifende Kenntnisse. Dennoch sind Linux Systeme für den Einsatz als Serverbetriebssysteme sehr attraktiv. Sie haben den Ruf stabiler und schneller zu sein. Dies liegt unter Anderem auch daran, dass der *Kernel* auf die jeweilige Hardware zugeschnitten werden kann. Die Aufgabe des *Kernels* besteht darin, die vom Computer bereitgestellten Hardwareressourcen (Prozessor, Arbeitsspeicher, Festplatten und Peripherie) zu verwalten. Hierfür benötigte Schnittstellen für das Ansprechen spezifischer Hardwarekomponenten (z.B. einer Netzwerkkarte) können entweder fest integriert oder als Erweiterungen bei Bedarf geladen werden. [91] Die Flexibilität, das Betriebssystem auf die jeweilige Hardware optimieren zu können, ist der Hauptgrund dafür, dass 476 der 500 größten Supercomputer unter Linux betrieben werden. [76] Einen weiteren Vorteil bildet die Sicherheit von Linux. Sowohl *Kernel* als auch Serversoftware unterliegen einem permanenten *peer-review* Prozess. Gegebenenfalls kann der Anwender selbst Sicherheitslücken schließen. Durch die Vielfalt der Distributionen erhöht sich der Aufwand bei der Entwicklung von Schadsoftware für Hacker. Anwender von proprietäreren Betriebssystemen sind stets auf den Hersteller angewiesen. Wird eine Sicherheitslücke bekannt, ist die Zeit bis zum Schließen der Lücke von der Reaktionszeit des Herstellers abhängig. Während dieser Zeit ist das System besonders vulnerabel für Angriffe (*Zero-day-Exploits*). Aus Kosten- und Flexibilitätsgründen nimmt die Verbreitung von *Linux* als Serverbetriebssystem, insbesondere für Einsätze im *Cloud-Computing* in den letzten Jahren stetig zu. [42] Häufigster Grund für den Wechsel von Windows oder Unix auf Linux ist die Kostenreduktion (22%), gefolgt von Stabilität (14%) und Geschwindigkeit (13%). [42]

6.6.1.2 Datenbanksystem

Für die elektronische Abbildung und Speicherung von Informationen wird ein *Datenbanksystem* benötigt. Ein *Datenbanksystem* besteht aus den eigentlichen Daten und einer Software für deren Verwaltung (*Datenbankmanagementsystem*). Umgangssprachlich wird Datenbank häufig synonym für *Datenbankmanagementsystem* verwendet. In aller Regel stehen Daten in Beziehungen zueinander. *Relationale Datenbanken* versuchen diese Beziehung durch in Verbindung stehende Tabellen abzubilden. [111] Es gibt eine Reihe kommerzieller und freier *Datenbankmanagementsysteme* mit unterschiedlichem Funktionsumfang, die für die Speicherung in Frage kommen. Wichtige Entscheidungskriterien für die Auswahl des *Datenbankmanagementsystems (DBMS*) sind:

- Datensicherheit
- Datenschutz
- Integritätssicherung

- Mehrbenutzerfähigkeit
- Datensicherungsmöglichkeiten
- Geschwindigkeit
- Clusterfähigkeit
- Unterstützte Betriebssysteme
- Lizenz / Lizenzkosten
- Datenbank-Verwaltung

Weit verbreitete *Datenbankmanagementsysteme* sind: Oracle, Microsoft SQL Server, PostgreSQL und MySQL. Sowohl Oracle als auch Microsoft bieten ihre Datenbankserver als kommerzielle Software an. Oracle unterscheidet für den kommerziellen (nicht-privaten) Einsatz nach Anzahl der Prozessoren oder Anzahl der Datenbankbenutzer. Datenbankbenutzer im Verständnis des Herstellers können auch Programme sein (wie z.B. ein Webserver), welche auf die Daten zugreifen. Da die Anzahl der Benutzer, die über das Internet Zugriff auf die Datenbank hat, schlecht kalkulierbar ist, kommt für den Einsatz in einem Onlineportal nur das Lizenzmodell nach Anzahl der Prozessoren in Frage. Oracle Server sind lizenzabhängig clusterfähig und können auf Windows, Linux und Unix-Servern betrieben werden. Der Preis für die Einsteigerlizenz (Oracle Standard Edition One) liegt für Server mit einem Prozessor bei etwa 4.578.00€ (Stand August 2013). Sie ist jedoch in dieser Version nicht clusterfähig. Die entsprechend clusterfähige Lizenz (Oracle Database Standard Edition) kostet für einen Prozessor 13.813€ (Stand August 2013). [171] Ebenfalls kostenpflichtig ist der Microsoft *SQL* Server. Lizenziert wird entweder nach Anzahl von Prozessorkernen oder nach Anzahl der auf die Datenbank zugreifenden Geräte bzw. Benutzer. Es gibt für jedes Lizenzmodell gestaffelte Preise für Unternehmen, Bildungseinrichtungen, Schüler und Studenten sowie Entwickler. [172] Für den Einsatz als *Datenbankmanagementsystem* für Internetportale kommt analog zu Oracle nur eine Lizenz nach Anzahl an Prozessorkernen (Core-License) in Frage. Die benutzer- oder gerätebasierte Lizenzierung ist für den Einsatz in lokalen Netzwerken mit Datenbankanwendungen konzipiert, die direkt auf den Computern der Benutzer laufen und direkt auf die Datenbank zugreifen (z.B. ein Warenwirtschaftssystem). Aktuelle Server besitzen zwischen 4 und 16 Kerne. Beim Microsoft *SQL* Server mit Core-License müssen alle Kerne des Servers lizenziert werden. Eine Core-License deckt zwei Kerne ab. Es müssen pro Server jedoch mindestens 4 Kerne lizenziert werden. Lizenzen werden von unterschiedlichen Anbietern verkauft und unterscheiden sich daher im Preis. Für Bildungseinrichtungen liegt der Preis für eine Core-License bei etwa 1.130€. Geschäftskunden zahlen knapp 4.000€. Gibt es noch keinen Softwarevertrag mit dem Hersteller (Microsoft OPEN-Vertrag), besteht eine Mindestabnahme von 5 Core Lizenzen. Microsoft SQL Server sind nur unter Microsoft Betriebssystemen lauffähig und können nicht unter Linux/Unix betrieben werden.

MySQL und PostgreSQL sind Vertreter von OpenSource *Datenbankmanagementsystemen.* MySQL galt lange Zeit als besonders einfach zu installierendes und schnelles *Datenbankmanagementsystem*. Es hat sich daher binnen kurzer Zeit zum beliebtesten *Datenbankmanagementsystem* für Internetanwendungen entwickelt. Allerdings fehlten ihm zumindest anfangs einige Funktionen kommerzieller Datenbanken. PostgreSQL hingegen hatte aufgrund seines großen Funktionsumfangs und SQL- konformer Umsetzung den Ruf die kostenfreie Alternative zu Oracle zu sein, konnte sich aber wegen komplizierterer Installation und geringerer Geschwindigkeit gegenüber MySQL nicht durchsetzen. [75]
Inzwischen bestehen in Funktion und Geschwindigkeit beider Systeme kaum mehr Unterschiede, die für den Einsatz als Datenbank für Webseiten oder Internetportale relevant wären. Aufgrund der freien Verfügbarkeit, der schnellen Entwicklung und der guten Integration in

Webserver ist MySQL das derzeit am weitesten verbreitete *Datenbankmanagementsystem* bei Internetportalen wie Wikipedia, Facebook, YouTube, Google und Twitter. [173]

DBMS	Lizenzkosten	Betriebssysteme	Clusterfähig	Software	Replikation
Oracle	ab 4.500€	Windows, Linux, Unix	lizenzabh.	ja	ja
MS SQL	ab 2.200€	Windows	Ja	ja	ja
Postgres	-	Windows, Linux, Unix	Ja	Ja	ja
MySQL	-	Windows, Linux, Unix	Ja	ja	ja

Tabelle 4: Kosten laut Hersteller Microsoft, Oracle, MySQL, PostgreSQL Stand 03/2014 [174] [175] [71] [177]

6.6.1.3 Webserver

Webserver liefern den Browsern (Hyper-)Text, Bild und Medien aus. Neben dem Bereitstellen von statischen Inhalten, haben sie die Aufgabe dynamische Inhalte aus Datenbanken zu erzeugen. Für die Erzeugung dynamischer Inhalte gibt es unterschiedliche Skriptsprachen, von denen PHP sicher die Bekannteste ist. Der Webservermarkt ist recht übersichtlich. Die beiden häufigsten Vertreter sind der Apache Webserver und der Microsoft Internet Information Server (IIS). Während Apache kostenfrei und quelloffen für Windows, Linux und Unix zur Verfügung steht, ist der IIS Teil der Windows Serverbetriebssysteme. Neben den beiden Hauptvertretern hat Nginx als ebenfalls plattformunabhängiger und quelloffener Webserver aufgrund seiner Geschwindigkeitsvorteile insbesondere bei statischen Inhalten (Videos, Photos usw.) an Verbreitung zugenommen. Außer den drei genannten, gibt es weitere Webserver, die auf spezielle Anwendungen oder Geräte ausgelegt sind (wie z.B. Internetrouter, Smart TV*s*, Entertainment-Systeme).

Apache ist mit 51% Marktanteil der am häufigsten eingesetzte Webserver gefolgt von Microsofts IIS 19%. Nginx hat einen Anteil von 15% und ebenso wie der Apache Webserver ein positives Wachstum zu verzeichnen (Stand April 2013). [138] Die Auswahl des passenden Webservers ist schwierig, erfüllen doch alle drei voranstehend Genannten die Kriterien von:

- Modularität/Erweiterbarkeit
- Integration der Datensysteme Oracle, MSSQL, MySQL, PostgreSQL
- Skriptsprachenunterstützung (PHP, ASP, JSP, Perl)
- Sichere Datenübertragung (*HTTPS*)
- Zugriffsbeschränkung (passwortgeschützte Bereiche)
- Protokollierung der Zugriffe
- Puffern häufiger Inhalte (*Caching*)

Wichtig sind die sicherheitsrelevanten Aspekte eines Webservers, denn jeder Webserver ist prinzipiell angreifbar. Durch den offenen Quellcode sind *OpenSource* Lösungen tendenziell im Vorteil. Allerdings entstehen die meisten Sicherheitslücken und Schwachstellen nicht durch Fehler der Software, sondern vielmehr durch Fehlkonfiguration und Programmierfehler, Schnittstellen und Softwaremodule. [48] Da ein Webserver nie isoliert zu betrachten ist, sondern immer im Kontext mit dem Betriebssystem, dem *Datenbankmanagementsystem* und den

verwendeten Skriptsprachen, gibt es Best Practise Lösungen in Form von aufeinander abgestimmten Softwarepaketen.
Eines dieser Pakete wird *LAMP* abgekürzt und enthält **L**inux als Betriebssystem, den **A**pache Webserver, eine **M**ySQL Datenbank und die Skriptsprache **P**HP. Für jede Linux Variante gibt es eine entsprechende Empfehlung zur sicheren Implementierung. Für den IIS wird die Kombination aus MSSQL und ASP empfohlen und in entsprechenden Whitepapers beschrieben, wie diese sicher betrieben werden kann. [119]

6.6.2 Backup

Datensicherungen zählen zu den essentiellen Maßnahmen zur Gewährleistung von Datensicherheit. Backups sind keine prophylaktische Maßnahme, sondern ein Notfallplan zur Wiederherstellung eines Systemzustands oder Datenbestands zu einem bestimmten Zeitpunkt in der Vergangenheit. Je nach Anforderung und Ressourcen ist die Planung eines individuellen Backup-Szenarios notwendig. Es gibt unterschiedliche Backup-Ebenen:

- Datenbank-Backup
- Datei-Backup
- Server-Backup
- Live-Snapshot/Image

Soll der Datenbestand eines Datenbanksystems gesichert werden, ist ein Datenbank-Backup sinnvoll. Jedes *Datenbankmanagementsystem* bringt Hilfsprogramme mit, mit deren Hilfe eine Sicherung der Datenbankstruktur und der Datensätze erzeugt werden kann. Datenbank-Backups sind in der Regel schnell durchgeführt und können im laufenden Betrieb stattfinden. Je nach Bedarf können Sicherungen auch in kurzen Intervallen durchgeführt werden. Datenbanken für Internetanwendungen sind klein, d.h. sie besitzen Dateigrößen von einigen Kilobytes bis zu wenigen hundert Megabytes. Nur selten (und nur in großen und komplexen Anwendungen) erreichen sie Größen im Gigabytebereich. Daher benötigen Datenbank-Backups nur wenig Zeit für die Erstellung und wenig Speicherplatz. Sie eignen sich daher zum engmaschigen Schutz des Datenbestands.

Eine weitere Strategie ist das Sichern von Dateien. Ein Datei-Backup umfasst die Sicherung von Ordnern oder einzelnen Dateien. Es können Dokumente, Bilder, Videos aber auch Programme und Programmeinstellungen gesichert werden. Datei-Backups können ebenfalls im laufenden Betrieb erzeugt werden. Durch das Lesen vieler verteilter Dateien bremst die Datensicherung jedoch den Zugriff anderer Prozesse auf die Festplatte aus. Hintergrund hierfür ist, dass für das Positionieren des Lese/Schreibkopfes der Festplatte verhältnismäßig viel Zeit benötigt wird und während der Positionierung nicht gelesen werden kann. Datei-Backups werden meist zu einer komprimierten Archivdatei (z.B. Windows => *zip* oder Linux => *tar.gz*) zusammengefasst. Die Kompression von Dateien kostet während der Erstellung Rechenleistung. Wird ein großes Archiv erstellt, kann die Belastung des Prozessors nicht unerheblich sein, was wiederum eingeplant werden muss. Eine Möglichkeit den Speicherbedarf für Backups zu reduzieren sind inkrementelle Backups. Bei fortlaufenden Backups werden lediglich neue oder veränderte Dateien gesichert. Der Vergleich aller eingeschlossenen Dateien mit der Vorgängerversion benötigt jedoch Rechenleistung. Inkrementelle Backups sind besonders dann interessant, wenn entweder der verfügbare Speicherplatz auf dem Backup Medium oder die Kapazität der Netzwerkverbindung limitierende Faktoren sind.

Eine Erweiterung des Datei-Backups ist das Server-Backup. Hierbei werden alle Daten der Anwender, Datenbanken, Konfigurationsdateien, Programme und das Betriebssystem gesichert. Ein solches Backup ermöglicht die Wiederherstellung des Betriebssystems inklusive aller Daten, erzeugt aber auch entsprechend große Archiv-Dateien. Wie bei Datei-Backups kommt es hier zur Verlangsamung des Systems durch Lese- und Komprimierungsprozesse. Ein Server-Backup kann etliche Stunden in Anspruch nehmen. Da für diesen Zeitraum die Leistungsfähigkeit des Servers beeinträchtigt ist, empfiehlt es sich, das Backup zu einer Zeit mit minimalem Benutzerzugriff (z.B. morgens zwischen 2:00 – 6:00 Uhr) zu verlegen. Um den Speicherbedarf für die Archivdateien zu minimieren kann auch hier inkrementell gesichert werden. Der Rechenaufwand im Falle einer Rücksicherung ist im Vergleich zu sequentiellen (nicht inkrementellen) Backups erhöht. Server Backups werden in niedrigerer Frequenz durchgeführt. Übliche Frequenzen liegen zwischen täglich und monatlich. Ihr primäres Ziel ist es, das Betriebssystem und die Softwareinstallation zu archivieren.

Snapshots erzeugen eine Momentaufnahme des Betriebssystems. Sie können auf Dateisystemebene oder auf Blockebene erzeugt werden. Dateisysteme, wie das Windows- Dateisystem *NTFS* (sofern die Funktion *Volumenschattenkopie* aktiviert ist), legen bei einem *Snapshot* eine Kopie jeder geänderten oder neuen Datei an und können dadurch bis zu 64 Versionsstände im laufenden Betrieb speichern. [178] *ZFS* als Unix- Vertreter beherrscht ebenfalls *Snapshots* auf Dateisystemebene. Ähnlich dem Microsoft-Prinzip, wird dabei in einem *copy-on-write* Verfahren genannten Vorgang der aktuelle Stand eingefroren und nur die im Vergleich zum vorherigen Systemstand geänderten Dateien gespeichert. *Snapshots* können im laufenden Betrieb schnell und ohne großen Leistungsverlust erstellt und zurückgespielt werden. Ausfallzeit entsteht in der Regel nicht. Fällt jedoch im laufenden Betrieb eine (nicht redundante) Festplatte aus, ist eine direkte Wiederherstellung auf die neue Festplatte nicht möglich. Es muss zunächst das Betriebssystem installiert werden, um einen *Snapshot* wiederherstellen zu können.

Images (Festplattenabbilder) arbeiten auf Blockebene und gehen einen anderen Weg.
Sie greifen direkt auf die Datenblöcke der Festplatte zu und erzeugen eine exakte Kopie der Festplatte (inklusive Betriebssystem und Dateisystem). Der Vorteil gegenüber *Snapshots* ist, dass das Wiederherstellen unabhängig von einem funktionstüchtigen Betriebssystem erfolgen kann. Mit Hilfe des *Images* kann die Datensicherung direkt auf eine neue Festplatte übertragen werden. Images können auch dazu verwendet werden Server zu klonen. Das Erzeugen von *Images* bei physikalischen Servern erfordert in der Regel eine längere Ausfallzeit. Der Server muss für die Zeitdauer der Sicherung heruntergefahren werden. Mit spezieller Software bzw. in virtualisierten Umgebungen können *Images* zwar im laufenden Betrieb erfolgen spätestens für das Zurückspielen ist jedoch ein Neustart des Systems nötig.

Eine hybride Form aus *Snapshot* und *Image* hat sich unter Linux etabliert. Mit Hilfe des *LVM* (*Logical Volume Manager*) wird eine Verwaltungsebene erzeugt, die zwischen den Festplatten und dem Dateisystem vermittelt. Hierzu stellt der *LVM* virtuelle Festplatten (*logical volumes*) bereit. Das logische Laufwerk ist unabhängig von physikalischen Festplatten und kann sich auf mehrere Festplatten verteilen oder auch nur ein Teil einer Festplatte sein. Innerhalb eines LVM Volumes können (wie bei echten Festplatten) das Betriebssystem und das Dateisystem installiert werden. *Snapshots* können von *logical volumes* im laufenden Betrieb erzeugt werden. Sie sichern den aktuellen Zustand der virtuellen Festplatte ähnlich einem *Image* (also inkl. Betriebs- und Dateisystem). Jedoch werden bei jedem *LVM Snapshot* lediglich die Änderungen gesichert, wie es bei *Snapshots* auf Dateisystemebene üblich ist. Die Archivdateien sind daher verhältnismäßig klein. Zudem kann sowohl Sicherung als auch Wiederher-

stellung im laufenden Betrieb in sehr kurzer Zeit erfolgen. [4] *LVM-Snapshots* sind sehr flexibel. Ähnlich wie bei *Images* können Festplatten und somit Serverinstallationen geklont werden. Wichtige Anwendungsfelder hierfür sind Testumgebungen für Entwickler oder Wiederherstellungspunkte vor Software-Updates.

Bei einigen Backup-Methoden gilt zu berücksichtigen, dass eine Rücksicherung den bisherigen Datenbestand unselektiv überschreibt. Wird z.B. ein fünf Tage altes *Image* zurückgespielt, gehen alle Daten der letzten fünf Tage verloren. Kommt es durch ein misslungenes System-Update zu einem schweren Fehler, ist ein reines Datenbank-Backup nicht ausreichend um den funktionstüchtigen Zustand wiederherzustellen. Es muss zunächst die dafür notwendige Umgebung (Betriebssystem und Datenbankmanagementsystem, ggf. noch weitere Software) installiert werden, bevor die Datenbanksicherung eingespielt werden kann. Wie beide Beispiele verdeutlichen, ist es durchaus sinnvoll, Backup-Optionen zu kombinieren, und einen Kompromiss aus Leistungseinbuße und Flexibilität einzugehen.

Eine sinnvolle Backup-Strategie wäre zum Beispiel:

Datenbank-Backup: 7:00 bis 23:00 stündlich
Datei-Backup: Montags bis Samstags zwischen 00:00 und 2:00
Server-Backup: Sonntags zwischen 1:00 und 5:00
Snapshot: vor jeder Wartung und jedem Update
Image: nach Installation und Konfiguration des Betriebssystems und der Software

Die folgende Tabelle fasst die Backupmöglichkeiten und ihre Funktion zusammen.

	Datenbank Backup	Datei-Backup	Server-Backup	Snapshot	Image
Datensätze	Ja	-	Ja	Ja	Ja
Dateien	-	Ja	Ja	ja	Ja
Betriebssystem	-	-	Ja	Ja	Ja
Betriebszustand	-	-	-	ja	-

Tabelle 5: Übersicht Datensicherungsmöglichkeiten

6.6.3 Monitoring

Um mögliche Störungen oder Probleme mit Soft- oder Hardware früh zu erkennen und darauf reagieren zu können ist es notwendig, Serverhardware und Serverdienste zu überwachen. Häufig kündigen sich Probleme bereits im Vorfeld an und können behoben werden bevor es zu einem Zwischenfall kommt. Eine volle Festplatte kann ebenso zu schwerwiegenden Problemen führen, wie der Ausfall eines Lüfters. Es ist daher sinnvoll mit Hilfe von Monitoring-Software alle kritischen Komponenten zu überwachen.

Hierzu zählen:

- Webserver
- Datenbank
- Mailserver
- Betriebssystem Sicherheitsupdates

- Gültigkeit des Sicherheitszertifikats
- Prozessorlast
- Speicherauslastung
- Netzwerkinterfaces (Bandbreite, Anzahl Verbindungen, Datenvolumen)
- Festplatten (*S.M.A.R.T.* Status)
- Temperatur
- Netzteil, Lüfter
- Switches
- Notstromversorgung (*USV*)

Dienste (wie Webserver oder Mailserver), die direkt über das Internet erreichbar sind, lassen sich mit relativ einfachen Methoden überwachen. Sie kommunizieren unter der *IP-Adresse* des Servers auf spezifischen Kommunikationskanälen (*Ports*). Jeder Dienst hat hierbei einen für ihn typischen *Port*. Der Dienst „antwortet" sofern eine Netzwerkanfrage auf seinem *Port* gestellt wird. Über die Antwort des Dienstes lässt sich ein Rückschluss auf dessen Funktionszustand ziehen. Die Überwachung öffentlich zugänglicher Serverdienste von außen (*externes Monitoring*) erfordert keine zusätzliche Maßnahme am Server. *Externes Monitoring* ist daher sicherheitstechnisch betrachtet unkritisch und einfach zu implementieren. Zwei einfache Beispiele zeigen wie externes Monitoring prinzipiell funktioniert. Hierzu wird versucht über den Kommunikationskanal des Webservers (*Port 80*) und des Postausgangsservers (*Port 25*) eine Verbindung aufzubauen. Nach erfolgreicher Verbindung zum Webserver (Abbildung links) bestätigt der Client (in unserem Beispiel das Konsolenprogramm *Telnet*) dies mit einer kurzen Statusnachricht. Der Mailserver (Abbildung rechts) reagiert nicht auf die gestellte Anfrage. Auf diese Weise lässt sich ohne Eingriff am Server die Erreichbarkeit von öffentlich zugänglichen Serverdiensten feststellen. Die Überprüfung in regelmäßigen Intervallen kann mit Hilfe entsprechender Software leicht automatisiert werden.

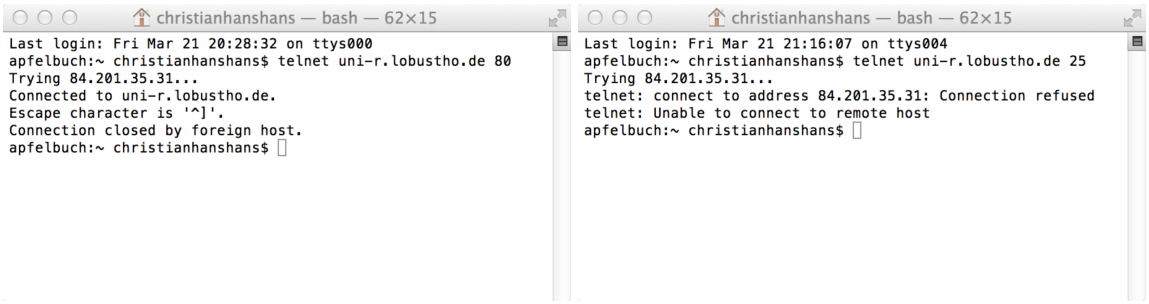


Abbildung 34: Funktionsprinzip externen Monitorings (Webserver, Mailserver)

Schwieriger sind Dienste zu überwachen, die nicht direkt mit dem Internet verbunden sind. Aus Sicherheitsgründen sollten einige Dienste (z.B. der Datenbankserver) nicht über das Internet erreichbar sein. Andere Informationen wie z.B. Temperatur oder Festplattenzustand können nicht direkt aus der Ferne abgefragt werden. Auf dem Server muss entweder eine Software ausgeführt werden, die Vitalparameter des Systems sammelt und an die Überwachungsinstanz meldet (*passives Monitoring*), oder das *Monitoring* System muss den zu überwachenden Server kontaktieren und die gewünschten Informationen abfragen (*aktives Monitoring*). Beide Wege bergen Sicherheitsrisiken in sich, da Systeminformationen permanent mit einem entfernten Rechner ausgetauscht werden. In beiden Fällen sind Hilfsprogramme von Nöten, die Zugriff auf den Server bzw. dessen Software haben. Es ist daher wichtig, dass auch hier Daten verschlüsselt (z.B. über einem *VPN-Tunnel*) übertragen werden.
Wichtig ist insbesondere die Früherkennung von Hardwaredefekten. Aktuelle Hardware verfügt über eine Vielzahl an Sensoren, die ausgelesen und überwacht werden können. Moderne

Festplatten verfügen über *S.M.A.R.T.* (*Self Monitoring Analysis and Reporting Technology*). Mit integrierten Mikrochips wird der Gesundheits- und Betriebszustand (z.B. Temperatur, Drehzahl, Alter, fehlerhafte Lese/Schreiboperationen) permanent protokolliert. Aus Untersuchungen von Festplattenausfällen großer Rechenzentren weiß man, dass auffällige *S.M.A.R.T.* Informationen ein guter Prädiktor für drohende Ausfälle sind. Warnzeichen sind insbesondere das vermehrte Auftreten von fehlerhaften Lese/Schreibzugriffen und eine dauerhaft erhöhte Betriebstemperatur. [5] [99]

Eine einfache und daher beliebte Überwachungsmethode ist das *aktive Monitoring* via *SSH* (Linux) oder *WMI* (Windows). [140] [85] Hierbei baut das Überwachungssystem eine Fernwartungsverbindung zum Server auf und meldet sich als Benutzer am System an. Mit entsprechender Berechtigung können nun Systembefehle ausgeführt und das Ergebnis ausgewertet werden. Ein einfacher Befehl wäre z.B. freien Festplattenspeicher anzeigen.
Der Vorteil hierbei ist, dass in der Regel keine spezielle Software auf dem Server installiert werden muss. Die Einrichtung der Überwachung ist daher sehr einfach. Das Risiko besteht jedoch darin, dass direkter Zugriff bzw. Kontrolle über den zu überwachenden Server ausgeübt werden kann. Für einige Befehle sind erweiterte (administrator-äquivalente) Rechte notwendig. Ein gekapertes Monitor-System stellt dann eine große Gefahr für die überwachten Server dar. Dennoch wird dieses Verfahren häufig eingesetzt und nicht selten mit den Zugangsdaten des Administrators der Server überwacht. Aus Sicherheitsgründen sollte dieses Monitoring Verfahren nicht verwendet werden.

Eine andere Möglichkeit besteht darin, auf dem Server ein Hilfsprogramm (*Daemon* oder *Dienst)* mit beschränkten Rechten zu installieren. Das Programm hat die Aufgabe alle gewünschten Gesundheitsinformationen zu sammeln und lediglich diese Informationen dem Monitor-System mitzuteilen. Das Monitor-System hat keine Möglichkeit auf andere Daten des überwachten Servers zuzugreifen oder Programme auszuführen. *SNMP* (*simple network management protocol*) ist ein Beispiel einer solchen Implementierung. Es steht für Windows wie Linux ein entsprechender *Dienst* bzw. *Daemon* bereit. Die Auswahl, welche Informationen des Systems übertragen werden dürfen, lässt sich in einer Konfigurationsdatei festlegen. *SNMP* verfügt über mehrere Protokollversionen. In Version 3 (*SNMPv3*) ist zusätzlich eine Authentifizierung mit Benutzername/Passwort enthalten. [113] Beschränkt man den Zugriff auf eine *IP*-Adresse (nämlich die des Monitor-Systems) und überträgt die Daten über eine verschlüsselte Verbindung, lässt sich ein vernünftiges Maß an Sicherheit mit überschaubarem Aufwand erreichen.

Für die Überwachung ist ein unabhängiges Serversystem notwendig, das sich sinnvollerweise in einem anderen Rechenzentrum befindet. Monitoring ist ein weites Feld und kann daher im Rahmen dieser Arbeit nicht detailliert ausgeführt werden. Es gibt eine Vielzahl kommerzieller und freier Monitoring-Lösungen, die auf eigenen Servern installiert werden können. Es gibt auch Anbieter, die sich als Dienstleister auf die Überwachung von Servern spezialisiert haben und die Infrastruktur sowie die Software bereitstellen und warten. Hierbei gilt es jedoch Sicherheits- und Datenschutzgrundlagen zu beachten. Die Auswahl des richtigen Werkzeugs hängt nicht zuletzt vom Budget und dem gewünschten Betriebssystem ab.

7 Implementierung des Prototypen

7.1 Weboberfläche und Benutzerverwaltung

Eine komfortable Verwaltung des Studienportals über das Internet macht den Einsatz einer intuitiven Weboberfläche notwendig. Das Studienportal muss zudem zwischen unterschiedlichen Rollen und deren Rechte im System unterscheiden. Es existieren viele Softwarewerkzeuge, die sowohl die Verwaltung von Inhalten als auch die Benutzerverwaltung beinhalten. Es bietet sich daher an, auf entsprechende Programme aufzubauen und die benötigten Funktionen (wie Fragebogenerstellung) zu ergänzen.
Zur Erstellung von Internetseiten mit dynamischen Inhalten werden üblicher Weise *Content-Management-Systeme* (*CMS*) eingesetzt. Jedoch sind nicht alle *Content-Management-Systeme* quelloffen. Für die gestellten Anforderungen muss zudem ein besonderes Augenmerk auf Sicherheit und Stabilität gelegt werden. Das System sollte zudem über eine gute Dokumentation verfügen, damit Erweiterungen einfach in die Software eingebaut werden können. Die wichtigste Anforderung ist ein geschützter Bereich, der nur für eine bestimmte Benutzergruppe (die Studienpopulation oder das Studienpersonal) sichtbar sein soll. Innerhalb dieses Bereiches sollen Videos angezeigt und Fragebögen ausgefüllt werden. Als Kandidaten kämen die freien *Content-Management-Systeme* wie Typo3, Joomla, Drupal oder Wordpress in Frage. In der Grundausstattung erfüllt jedoch keines der genannten *CMS* diese Anforderung. Häufig finden sich Plugins oder Module die zusätzlich installiert werden können um die gewünschte Funktion zu ergänzen. Plugins sind meist nicht vom Entwickler-Team des *CMS*, sondern von anderen Entwicklern programmiert. Ihr Einsatz muss kritisch betrachtet werden, da sie datensicherheits- und datenschutztechnische Gefahren in sich bergen können. So können kostenfreie Module oder Templates Sicherheitslücken oder sogar Schadcodes enthalten. Der Schädling *CryptoPHP* infizierte auf diese Weise zehntausende *Content-Management-Systeme* und baute mit den gekaperten Servern ein *Botnetz* auf. [67]

Lernmanagementsysteme (*LMS*) sind in der Grundausstattung den gestellten Anforderungen etwas näher. eLearning-Plattformen bringen bereits eine hierarchische Benutzergliederung mit. Ihr Fokus ist auf Interaktion zwischen Lernenden und Lehrenden abgestimmt. Sie bringen Unterstützung für Medien mit und sind an Hochschulen mit tausenden Benutzern im Einsatz. Die Langzeiterfahrung mit großen Benutzerzahlen universitärer Einrichtungen machen die gängigen Lernmanagement-Systeme zu einem zuverlässigen Unterbau. Die notwendigen Erweiterungen beschränken sich auf die Erstellung und Auswertung von Fragebögen.
Es gibt wie bei *Content-Management-Systeme* eine Vielzahl an unterschiedlichen eLearning-Plattformen. In der deutschen Hochschullandschaft dominieren Moodle, Blackboard und Ilias. Bei Blackboard handelt es sich um eine kommerzielle und sehr umfangreiche Lernplattform. Moodle und Ilias sind kostenfrei und quelloffen. Moodle ist in Deutschland das am weitesten verbreitete eLearning-System. Es sind in Deutschland 3352 (Stand 19.19.2013) öffentliche Moodle-Installationen registriert. [179] Unter anderem setzen die Universität Regensburg, Würzburg, TU München, Heidelberg oder die virtuelle Hochschule Bayern Moodle zur Bereitstellung von Lerninhalten ein.

7.1.1 Moodle

In einer umfangreichen Übersichtsarbeit aus dem Jahre 2011 wurden wesentliche Aspekte zur Wahl des geeigneten *Lernmanagementsystems* (*LMS*) gesammelt. Da sich bis dato nur wenige Publikationen mit ergonomischen Aspekten von *LMS* befassten, wurde die Gebrauchstaug-

lichkeit der Hauptvertreter dieser Systeme im Rahmen einer empirischen Studie genauer betrachtet. Das Ergebnis der Studie spricht für Moodle als sichere, verlässliche und anwenderfreundliche Plattform. [116]
Moodle als Basis des Studienportals zu nutzen, hat zudem viele praktische Vorteile. Der modulare Aufbau, die gute Dokumentation, ein durchdachtes Datenbankdesign mit der Möglichkeit einer granularen Rechtebeschränkung sind eine gute Ausgangssituation für eigene Entwicklungen. Die Möglichkeit die Oberfläche durch Sprachdateien an die gewünschte Terminologie (z.B. Kurs =>Studie, Kursbetreuer => Studienbetreuer,...) anzupassen, ist ein weiterer Pluspunkt. Die weite Verbreitung des Systems spielt aus zwei Gründen eine besondere Rolle. Zum einen wird der Einstieg für wissenschaftliche Mitarbeiter erleichtert die für die Verwaltung von Studien verantwortlich sind. Sie kennen in der Regel das System bereits aus dem Lehrbetrieb. Zum anderen kennen auch potentielle Teilnehmer Moodle bereits, da Moodle nicht nur in Schulen und Hochschulen, sondern auch in privaten Bildungseinrichtungen (wie in Volkshochschulen) oder in Unternehmen zur Mitarbeiterschulung eingesetzt wird. Dieser Bonus kann dazu beitragen, dass Vorbehalte und Bedienungsschwierigkeiten als initiale Hürde verringert werden. Anwenderschulungen für Moodle werden nahezu an allen Fakultäten angeboten, die Moodle einsetzen. Die Schulung der Studienbetreuer in den Grundfunktionen von Moodle kann daher über bestehende Schulungsangebote der Universität oder privater Bildungseinrichtungen erfolgen. Eine spezifische Schulung für die zusätzlichen Funktionen des Studienportals kann somit mit deutlich reduziertem Zeitaufwand erfolgen.

7.1.2 Debian Linux

Als Basis aller Server wurde das Linux Betriebssystem Debian 7 gewählt. Bei Debian handelt es sich um eine freie Linux Distribution, die für die Wahl als Serverbetriebssystem aus mehreren Gründen besonders gut geeignet ist. Im Gegensatz zu RedHat oder Suse Linux aber auch zu Microsoft Betriebssystemen sind keine Abonnements oder Supportverträge notwendig um das System zu nutzen. Ubuntu, ein Abkömmling von Debian, wurde mit dem Gedanken auf Anwenderfreundlichkeit entwickelt und hat daher seine besonderen Stärken als Desktop Betriebssystem. Obwohl es auch eine Serverversion ohne grafische Oberfläche gibt, ist diese in der Minimalinstallation (1.2 Gigabyte) vom Speicherbedarf her deutlich umfangreicher. Debian lässt sich im Vergleich zu seinen Hauptkonkurrenten mit deutlich geringerem Speicherbedarf installieren. Die Minimalinstallation des Betriebssystems belegt etwa 450 Megabyte, die Installation eines vollständigen Webservers (☞ Kapitel 7.1.3) knapp 1 Gigabyte der Festplatte. Die schlanke Installation spart Speicherplatz und erweist sich insbesondere im Hinblick auf Datensicherungen als vorteilhaft. Kopien der Festplatteninhalte können schneller erstellt und wegen der kleineren Archivgröße schneller zum Backup-Server übertragen werden. Netzwerk und Server sind durch die Datensicherungen kürzer beeinträchtigt. Das Vorhalten von Backups benötigt ebenfalls weniger Speicherplatz.
Debian wird von einem demokratisch organisierten und nicht kommerziellen Verein freier Entwickler unterhalten und weiterentwickelt. In einem Gesellschaftsvertrag sind die Regeln des Projekts festgeschrieben, nämlich die zu 100% freie und unabhängige Software und der transparente Umgang mit Sicherheitslücken und Fehlern. [180]

7.1.3 Webserver (LAMP)

Voraussetzung für das Betreiben von Moodle ist ein Webserver, der die Programmiersprache PHP unterstützt. Die Moodle Dokumentation empfiehlt Linux als Betriebssystem, den Apache Webserver und MySQL als Datenbankmanagementsystem zu verwenden. [181] Für die Installation des Webservers wurde daher eine klassische *LAMP* Installation bestehend aus Debian 7, Apache 2.2, MySQL 5.5 und PHP 5.5 installiert.
Da PHP 5.5 nicht Teil des Debian Paketmanagements ist, musste es aus dem Quellcode kompiliert und manuell installiert werden. Es bietet im Vergleich zum von Debian ausgelieferten PHP 5.4 einige sicherheits- und performancetechnische Vorteile. [182] Durch Kompilieren mit der Erweiterung *OpCache* konnte eine deutliche Leistungssteigerung beobachtet werden. Durch das Zwischenspeichern häufig verwendeter Teile des Quellcodes als vorkompilierten Bytecode, können Anfragen deutlich schneller bearbeitet werden. Zwar vermögen auch andere Webserver (Nginx, LightHttpd und Microsoft IIS) PHP zu verarbeiten, dennoch hat sich in der Praxis der Apache Webserver lange Zeit bewährt und wird daher von der Moodle Gemeinde präferiert. Das Apache Konzept ermöglicht es, gewünschte Funktionen durch die Installation entsprechender Module zu integrieren. Der Webserver kann dadurch an die Bedürfnisse der Webanwendung angepasst werden. So erfolgt z.B. die Integration der Programmiersprache PHP über das Modul *mod_php5,* die verschlüsselte Datenübertragung via *SSL* mit *mod_ssl* usw. MySQL kann von PHP über einen Konnektor (*php5-myql*) angesprochen werden. Als Speicherverwaltungssystem der Datenbank wurde *InnoDB* verwendet, das im Vergleich zu *MyISAM* Fremdschlüssel sowie Transaktionen unterstützt, vor allem aber auf die Verarbeitung großer Datenmengen ausgelegt ist. [183] [184] Die Speichereinstellungen wurden mit Hilfe des Werkzeugs MySQLTuner an den zur Verfügung stehenden Arbeitsspeicher angepasst. [47]

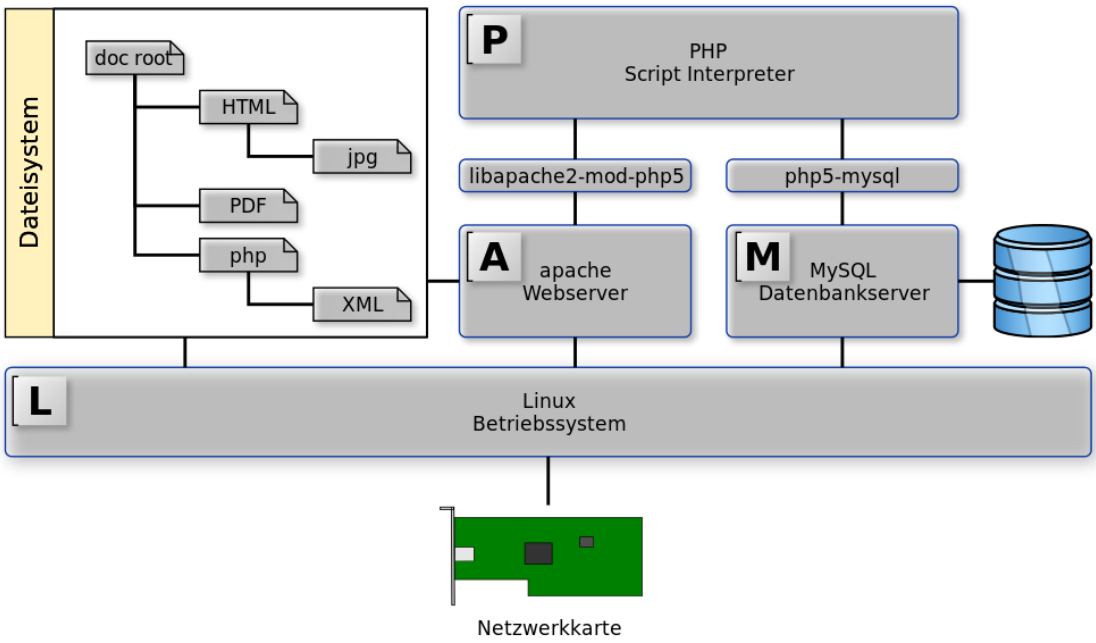


Abbildung 35: Schematischer Aufbau eines LAMP Servers, Karsten Adam [185]

Der Webserver Nginx kann nicht nur als Apache-Ersatz, sondern auch in Kombination mit Apache betrieben werden. Nginx ist ressourcensparender und kann dem Apache *als Proxy* vorgeschaltet werden. So könnte er als *Reverse Proxy* dem Ausliefern statischer Inhalte oder der Lastverteilung dienen. Besonders für eine Hochlastumgebung hat sich diese Symbiose

bewährt und wird daher von vielen großen Webseiten (z.B. Wordpress oder Wikimedia) eingesetzt. Um die Komplexität des Systems nicht zu erhöhen, wird für den Prototyp auf diese Möglichkeit wissentlich verzichtet. Für den Aufbau einer Lastverteilung beim Einsatz mehrerer Webserverinstanzen wäre der Einsatz von Nginx jedoch durchaus sinnvoll.

7.1.4 Benutzeroberfläche

7.1.4.1 Startseite

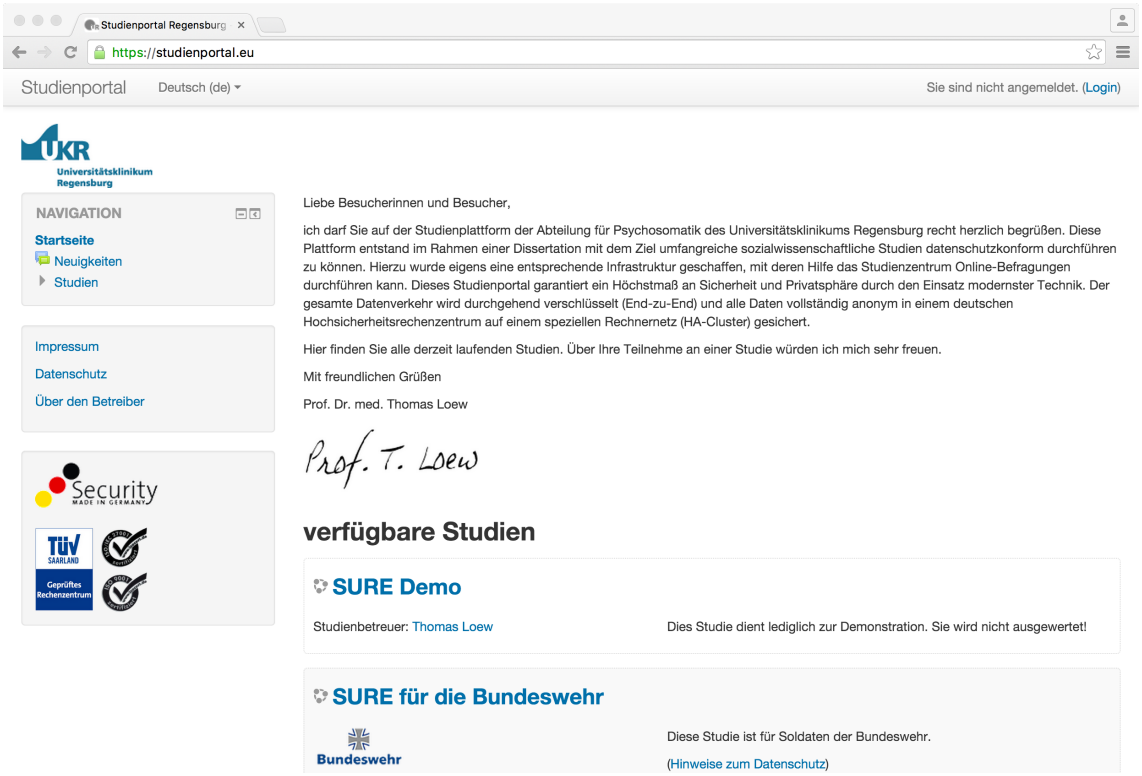


Abbildung 36: Startseite des Studienportals

Die Startseite des Studienportals orientiert sich grafisch an der Webseite der Abteilung für Psychosomatik der Uniklinik Regensburg. Die Farbgestaltung, die Einbindung des Logos, sowie des *Favicons* soll die Zugehörigkeit des Portals zum Institut wiederspiegeln. Eine persönliche Begrüßung (wie hier exemplarisch mit der Unterschrift des Studienmanagers) soll Vertrauen beim potentiellen Studienteilnehmer schaffen. Grundlage der grafischen Oberfläche ist ein bestehendes Moodle-Template (afterburner). Für den Einsatz im Studienportal wurden die *CSS Dateien* sowie Bilder angepasst. Das verwendete Template wird für alle Studien verwendet. Es besteht jedoch auch die Möglichkeit, für jede Studie ein eigenes Design (Farbe, Anordnung, Logo) zu verwenden. Dadurch kann der Studienbereich an das Look and Feel der *Landing-Page* angepasst werden.

Um die Nutzung auf mobilen Endgeräten zu ermöglichen wurde des Weiteren ein *Responsive Design* für die Geräteklassen Smartphone und Tablet-PC implementiert.

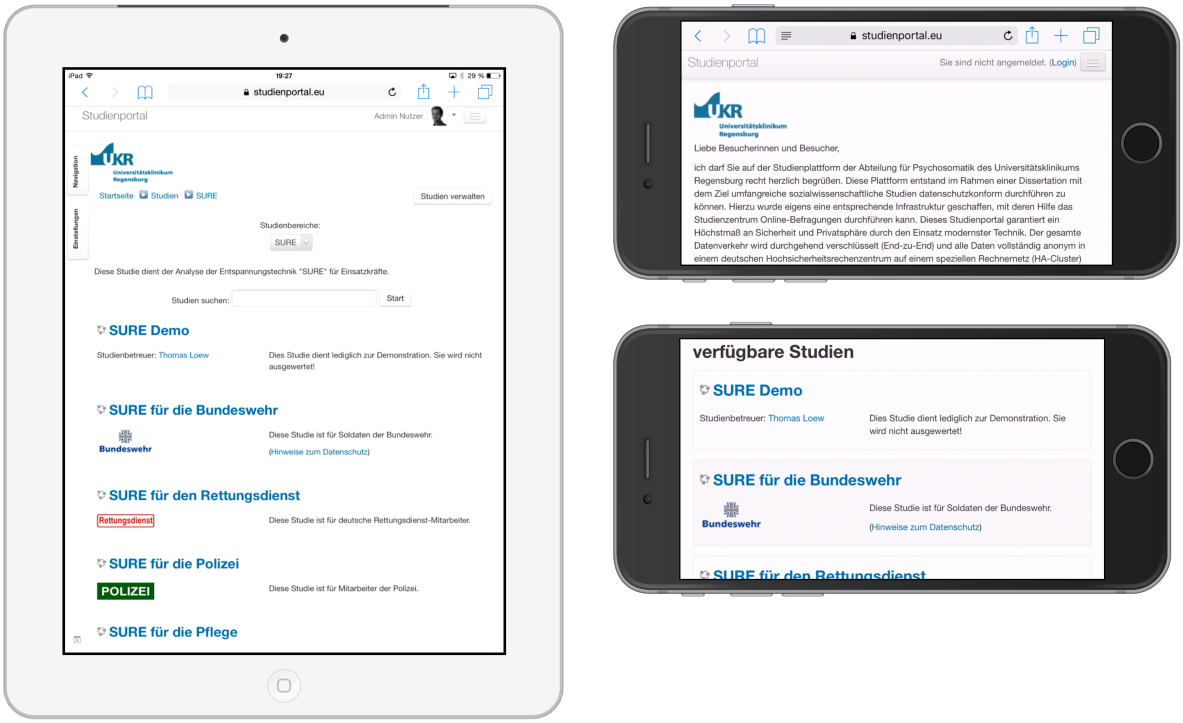


Abbildung 37: Optimierung für die Nutzung von mobilen Endgeräten (hier iPad und iPhone)

7.1.4.2 Login

Die automatische Registrierung der Teilnehmer ist über ein Registrierungsformular innerhalb einer *Landing-Page* möglich. Nach erfolgreicher Registrierung können sich die Teilnehmer entweder über die *Landing-Page* am System anmelden oder über das Studienportal.
Die Anmeldung via *Landing-Page* setzt die Implementierung eines Formularfelds und die Verknüpfung der *Landing-Page* mit dem Studienportal voraus.

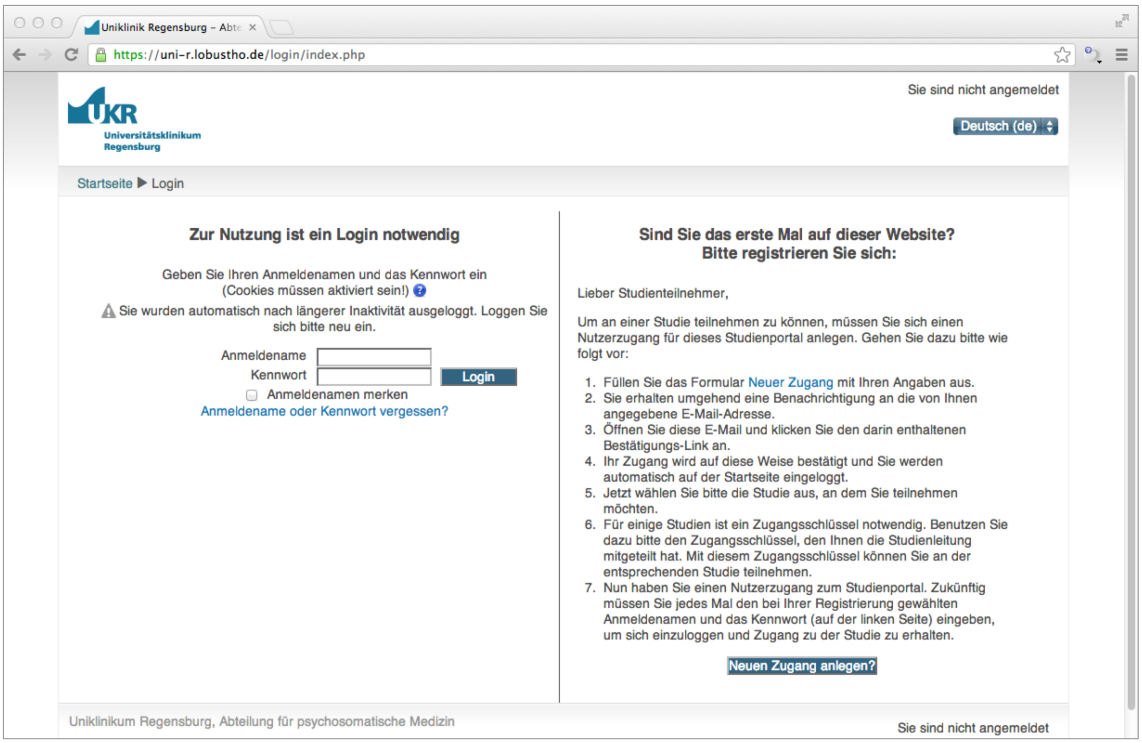


Abbildung 38: Anmeldemaske für registrierte Benutzer

7.1.4.3 Terminologie

Um den Studienteilnehmern und Betreuern den Umgang mit dem Studienportal zu erleichtern, wurde die Terminologie von Moodle an den Kontext klinischer Studien angepasst. Moodle arbeitet mit einem Kurskonzept. Ein Kurs entspricht einer Studie, der Kursbetreuer dem Studienbetreuer, der Student einem Studienteilnehmer usw. Für die Anpassung an unterschiedliche Sprachen verfügt Moodle über die Möglichkeit, individuelle Sprachpakete in die Datenbank aufzunehmen. Für jedes Moodle-Modul und für jeden Teil von Moodle gibt es Variablen, die mit einer Beschriftung versehen werden können. Sie bilden das Sprachpaket. Im Prototyp wurde ein Sprachpaket in deutscher Sprache mit entsprechender Terminologie erstellt und das Standardsprachpaket ersetzt. Für weitere Sprachen muss die Terminologie analog angepasst werden. Vor einem Update empfiehlt sich unbedingt die Sicherung des Sprachpakets.

Abbildung 39: Einfache Anpassung der Terminologie

7.1.4.4 Das Fragebogen-Modul

Moodle bringt bereits eine beachtliche Anzahl an Modulen mit. Drei davon sind in der Lage Onlineformulare oder Fragebögen zu erzeugen. Jedoch wurden bei Keinem davon alle benötigten Funktionen befriedigend abgedeckt. Das Quizz-Modul (*mod_quiz*) ist für den Einsatz als Prüfungswerkzeug, für Tests oder elektronische Klausuren konzipiert, nicht aber auf anonyme Bearbeitung ausgelegt. Neben individuellen Bewertungen werden Zeitstempel und wei-

tere detaillierte Bearbeitungsmerkmale erhoben. Es ist sehr komplex und schien daher für eine Anpassung nicht geeignet.
Das Modul *mod_feedback* zeigte sich zunächst vielversprechend. Es ist ebenfalls ein Standardmodul und dafür vorgesehen, Umfragen unter Teilnehmern durchzuführen. Es bietet neben einer Anonymisierung auch eine grafische Auswertung und eine Export-Funktion. Allerdings wurde bei anonymer Bearbeitung nicht pseudonymisiert, so dass mehrfach ausgefüllte Fragebögen nicht zu einem (anonymen) Teilnehmer zugeordnet werden können.
Ein weiteres Modul, das für die Erstellung von Fragebögen in aktuellen Moodle Versionen zur Verfügung steht ist *mod_survey*. Das Modul ist sehr umfangreich, erlaubt anonyme Befragungen, Export-Funktion und Zeitsteuerung. Ähnlich *mod_feedback* ist *mod_survey* auf Meinungsumfragen ausgelegt. Es ist jedoch deutlich komplexer und umfangreicher und könnte für spätere Implementierungen eine gute Grundlage bilden. Aufgrund der Komplexität und der vielen nicht benötigten Funktionen, wurde das Modul jedoch nicht genutzt. Ein weiteres wichtiges Argument gegen die Modifikation von Standardmodulen ist, dass diese bei Updates von Moodle verlorengehen würden. Ein eigenständiges Moodle-Modul bietet die Möglichkeit, flexibel auf die Anforderungen einzugehen und für zukünftige Versionen offen zu bleiben.

Für das Fragebogenmodul wurde daher ein eigenes Modul implementiert, das seinen Ursprung in phpESP [186] hat. phpESP ist ein schlankes in PHP geschriebenes Fragebogenskript, das bereits viele der benötigten Fragetypen beherrscht. Es wurde daher in der Vergangenheit bereits in einige *Content-Management-Systeme* integriert, wird jedoch seit 2010 nicht mehr aktiv weiterentwickelt. Aus der Moodle Community gab es bereits Ansätze, den alten Code in ein Moodle-Modul einzubinden. Diese Codebasis wurde für das Fragebogenmodul herangezogen. Das Fragebogenmodul befindet sich im Modulordner als *mod_questionaire*.

7.1.5 Fragebogen erstellen

Über die Oberfläche ist das Erstellen neuer Fragebögen durch das Hinzufügen einer neuen *Aktivität* möglich.

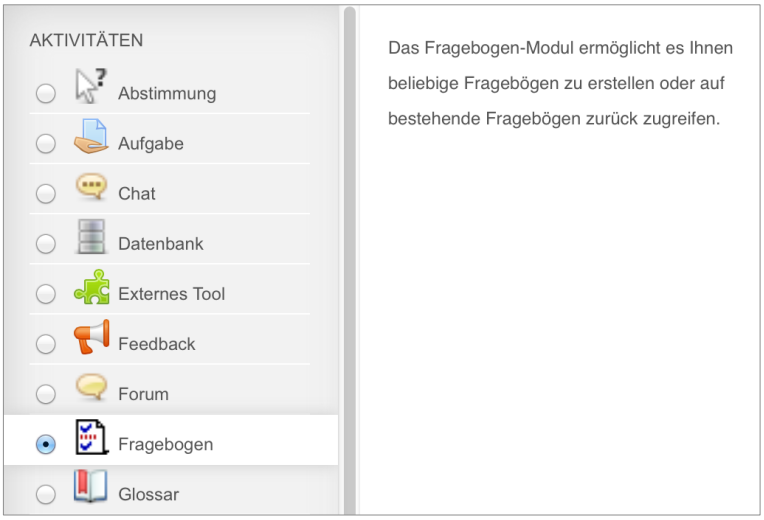


Abbildung 40: Fragebögen können als Aktivität hinzugefügt werden

Der Anzeigename und eine Beschreibung des Fragebogens (z.B. mit Hinweisen zur Bearbeitung), können über einen *Rich-Text-Editor* wie von Textverarbeitungssoftware gewohnt, formatiert und eingegeben werden.

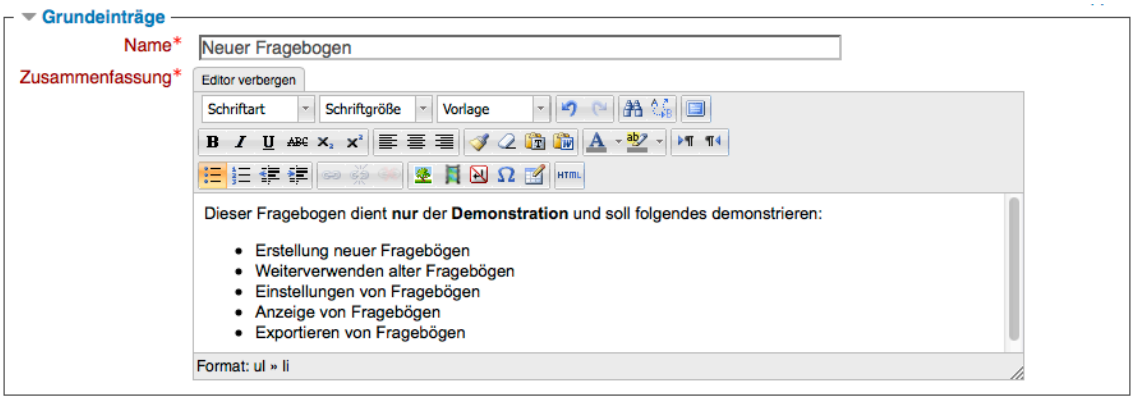


Abbildung 41: Fragebogenbeschreibung

Über eine Zeitsteuerungsfunktion kann die Bearbeitung des Fragebogens auf einen Zeitraum beschränkt werden.

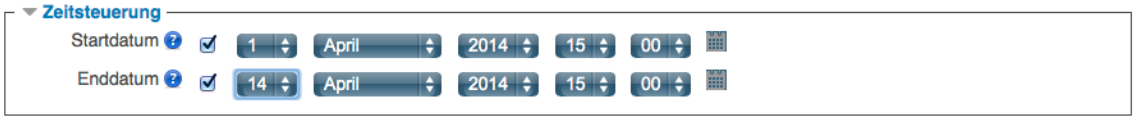


Abbildung 42: Zeitsteuerung der Fragebogenbearbeitung

Über eine weitere Einstellmöglichkeit lässt sich die Anzahl der Fragebögen definieren, die von einem Teilnehmer maximal eingereicht werden dürfen. Die derzeitige Voreinstellung sieht folgende Modi vor:

- Einmaliges Einreichen des Fragebogens
- Mehrmaliges Einreichen eines Fragebogens (zeitunabhängig)
- Tägliches Einreichen
- Wöchentliches Einreichen
- Monatliches Einreichen

Mit einer Auswahlmöglichkeit kann zwischen anonymer und personenbezogener Speicherung ausgewählt werden. Die anonyme Speicherung erfolgt gemäß dem bereits beschriebenen Algorithmus. (☞ Kapitel 6.5.5) Nach dem Einreichen des Fragebogens kann der ausgefüllte Fragebogen wahlweise angezeigt werden. Mit der Option *Speichern* bzw. *Fortsetzen* kann der Teilnehmer das Ausfüllen des Fragebogens unterbrechen und zu einem späteren Zeitpunkt fortsetzen.

Abbildung 43: Abgabeeinstellungen der Fragebögen

Zuletzt kann entschieden werden, ob der geplante Fragebogen aus einer Vorlage oder von Grund auf neu erstellt werden soll. Wurde ein Fragebogen zuvor von seinem ursprünglichen Ersteller für die Wiederverwendung freigegeben, kann dieser Fragebogen von anderen Studienbetreuern direkt übernommen und erweitert werden.

Abbildung 44: Fragebogen aus Vorlage erstellen

Wurde der Fragebogen gespeichert, kann der Fragebogen angeklickt werden. Es öffnet sich die Registerkarte: *Erweiterte Einstellungen*. Nun kann die Nutzung des Fragebogens eingestellt werden. Hierbei gibt es drei Möglichkeiten:

- *Privat*: kann nur innerhalb der jeweiligen Studie genutzt werden
- *Vorlage:* kann in anderen Studien genutzt und verändert werden
- *Öffentlich*: kann in anderen Studien genutzt, aber nicht verändert werden

Abbildung 45: Freigabe eines Fragebogens als Vorlage für den Fragebogen-Pool

Über die Abgabe-Einstellungen kann bei Abgabe des Fragebogens entweder eine Seite der *Landing-Page* (z.B. *danke.html*) aufgerufen oder alternativ ein Text innerhalb des Studienportals angezeigt werden. Wird eine Emailadresse angegeben, so wird eine Benachrichtigung über den Eingang eines neuen Fragebogens an die hinterlegte Adresse geschickt.

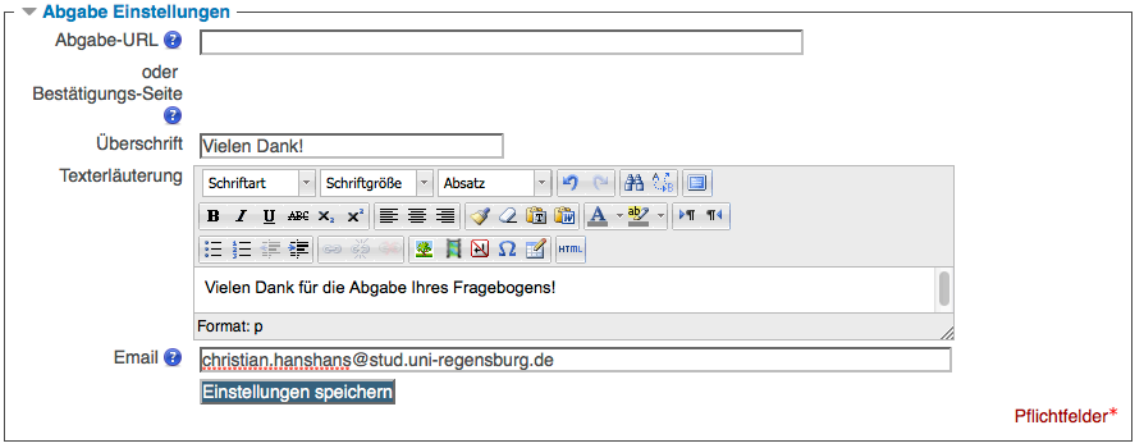


Abbildung 46: Verhalten bei eingereichten Fragebögen

Das Register *Fragen* ermöglicht die Bearbeitung von Fragebögen oder das Hinzufügen von neuen Fragen. Folgende Frageelemente sind implementiert:

- Einfachauswahl (Listenfeld)
- Einfachauswahl (Radiobuttons)
- Ja/Nein Antwort
- Lickert Skala
- Mehrfachauswahl (Checkboxen)
- Eingabefeld (Zahl)
- Eingabefeld (Datum)
- Eingabefeld (Text)

Dazu kommen noch die beiden Gestaltungselemente Beschriftung und Seitenumbruch.
Die Eingabe der Fragen ist intuitiv. Die Fragebezeichnung ist eine optionale Angabe und kann z.B. die Item- Nummer eines standardisierten Fragebogens sein.
Das folgende Beispiel zeigt die Eingabe einer Frage nach der Lickert-Skala. Alle anderen unterstützen Fragetypen können in ähnlicher Form eingegeben werden. Die Anordnung der Antwortmöglichkeiten kann hierbei vertikal oder horizontal erfolgen. Der Fragetext kann wie in einer Textverarbeitungssoftware (Schriftart, Schriftfarbe, Ausrichtung, Aufzählungen, Einfügen von Bildern oder Links usw.) frei formatiert werden. Der Benutzer hat die Möglichkeit mit der Tabulatortaste von Eingabefeld zu Eingabefeld zu springen. Jede Antwortmöglichkeit wird durch eine neue Zeile (Eingabetaste) getrennt. Über das Kontextmenü *Kopieren* und *Einfügen* (bzw. Tastaturbefehle *Strg+C* und *Strg+V*) können Antwortmöglichkeiten in die nächste Frage übertragen werden. Dies ist besonders hilfreich bei Fragebögen mit numerischen oder ordinalen Skalen, die über mehrere Fragen hinweg gleich bleiben. Mit einem Klick können alle Antwortmöglichkeiten übertragen werden. Nur die Frage muss eingetippt werden. Auf diese Weise können auch umfangreiche Fragebögen sehr schnell und mit wenig Mouse-Interaktion per Tastatur angelegt werden.
Soll der Teilnehmer die Möglichkeit bekommen eine nicht genannte Antwortmöglichkeit als Freitext einzugeben, kann dies durch die Angabe von *!other* als Antwortmöglichkeit geschehen. Dem Benutzer wird die Antwortoption *andere* angezeigt und beim Klick ein Eingabefeld eingeblendet.

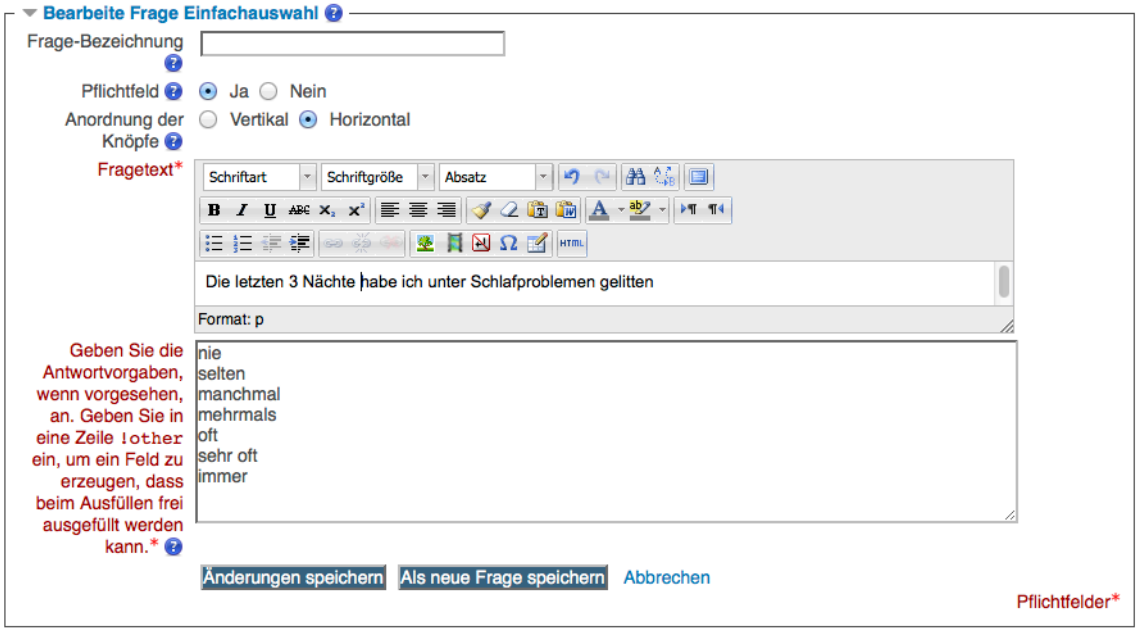


Abbildung 47: Frage und Antwortmöglichkeiten eingeben

Nach vollständiger Eingabe kann die Frage mit *Änderungen speichern* gesichert werden. Es wird danach die Fragenübersicht angezeigt. Mit Hilfe von Pfeilsymbolen können Fragen umsortiert, mit *x* gelöscht und mit dem Stiftsymbol bearbeitet werden.
Mit Hilfe der Funktion *Als neue Frage sichern* können zuvor eingegebene Fragen als Vorlage herangezogen und geklont werden. Dies lohnt insbesondere, wenn nur kleine Änderungen an der Frage gewünscht sind und eine Frage besonders aufwändig formatiert wurde.
Der Reiter *Vorschau* zeigt den aktuellen Fragebogen aus Teilnehmersicht. Er kann testweise ausgefüllt, nicht jedoch abgeschickt werden. Aus Anwendersicht ist das Fragebogenmenü gezielt schlicht gehalten. Es gibt drei Registerkarten:

- *Fragebogen ausfüllen*
- *eigene Fragebögen*
- *alle Fragebögen*

Das Register *Fragebogen ausfüllen* zeigt den Fragebogen an.

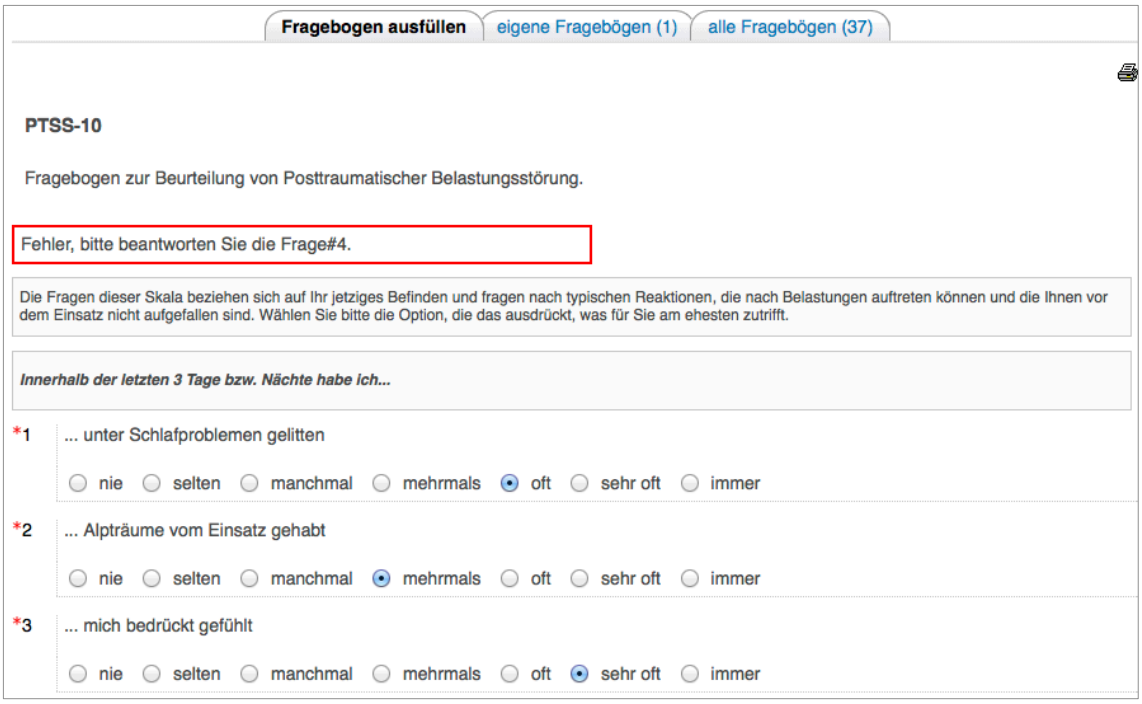


Abbildung 48: Fragebogen aus Teilnehmersicht

Am Ende jedes Fragebogens gibt es zwei Knöpfe. *Fragebogen einreichen* führt zur Abgabe des ausgefüllten Fragebogens. Sollten Pflichtfelder nicht ausgefüllt sein, wird der Benutzer auf die fehlende Frage mit einem eindeutigen Warnhinweis aufmerksam gemacht und der Fragebogen nicht angenommen. Erst wenn alle Pflichtfelder ausgefüllt wurden kann der Fragebogen eingereicht werden. Bei der Erstellung des Fragebogens muss sich der Ersteller gut überlegen, ob und in welchem Umfang Pflichtfelder eingesetzt werden. *Speichern* speichert die bisherige Bearbeitung. Der Teilnehmer wird darüber informiert, dass er jederzeit den Fragebogen fortsetzen kann. Beim nächsten Einloggen kann der Fragebogen fortgesetzt werden. Teilnehmer können den Link zum Fragebogen als Favorit im Browser oder als Verknüpfung auf dem Desktop abspeichern.

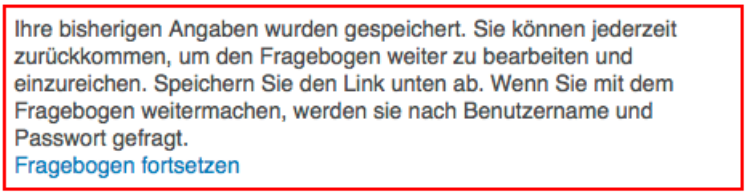


Abbildung 49: Pausieren der Bearbeitung

Nach Abgabe des Fragebogens kann den Studienteilnehmern das eigene Ergebnis im Vergleich zum Studienkollektiv angezeigt werden. In der Standardeinstellung ist dies erlaubt. Der Studienmanager kann diese Funktion auf Wunsch jedoch deaktivieren

7.1.6 Datenauswertung und Datenexport

Die eingereichten Fragebögen können, wie im vorigen Kapitel angesprochen und visualisiert werden. Hierbei wird in einem Histogramm die Häufigkeitsverteilung aller eingereichter Fragebögen dargestellt. Die Abbildung zeigt die Auswahlmöglichkeiten aus Sicht des Studien-

managers. Ob Studienbetreuer Fragebögen (zu Testzwecken) einreichen dürfen, kann vom Studienmanager festgelegt werden. Im Gegensatz zu Teilnehmern werden für Studienbetreuer, Studienassistenten und Studienmanager keine Pseudonyme erzeugt, so dass Test-Einträge beim Exportieren leicht herausgefiltert werden können.

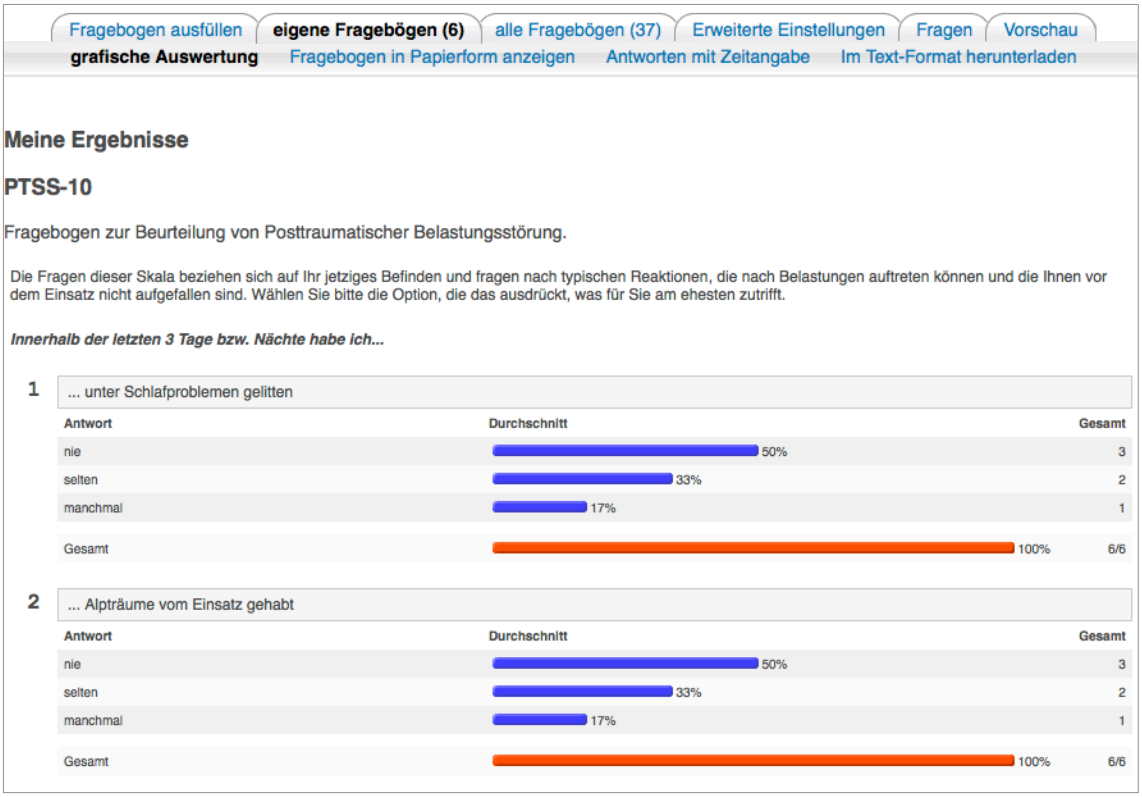


Abbildung 50: Anzeigen eigener Fragebögen im Verlauf

Die gesammelten Studiendaten können von berechtigten Studienbetreuern oder Studienassistenten über die Schaltfläche *Im Textformat herunterladen* aus der Datenbank exportiert werden. Es kann entschieden werden, ob die Antwortnummer und/oder der Antworttext exportiert werden sollen. Die Antwortnummer (*Antwort-ID*) entspricht der Antwortreihenfolge. Für die externe Auswertung kann es vorteilhafter sein mit Zahlen statt Text zu arbeiten.

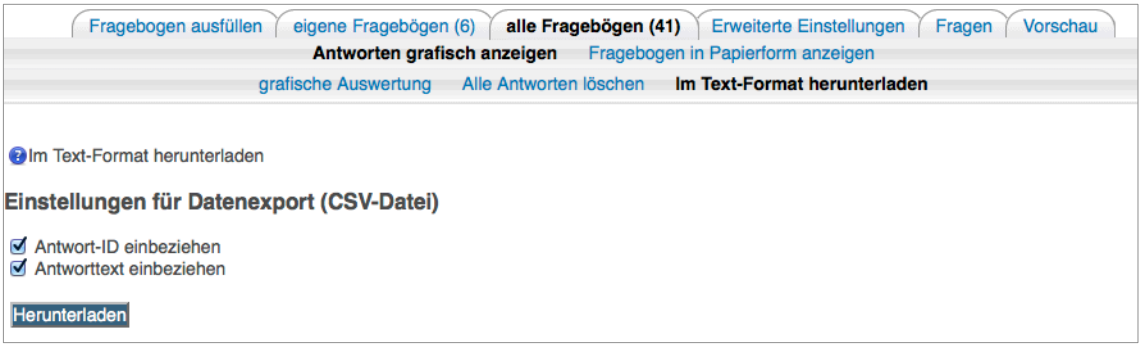
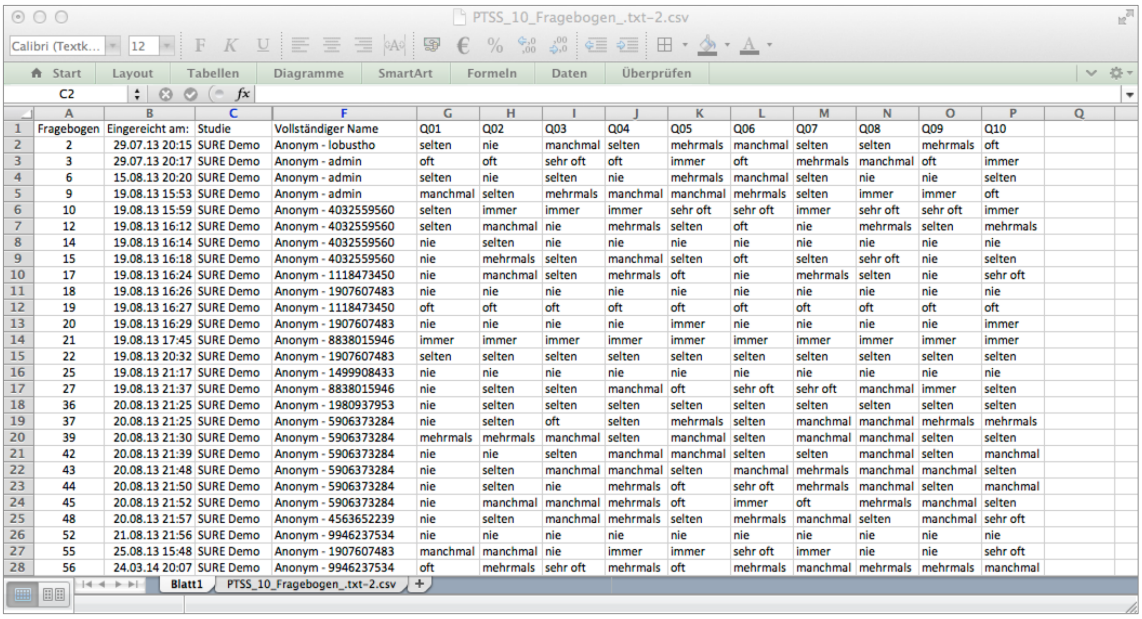


Abbildung 51: Datenbank Export der Studiendaten

Nach Klick auf *Herunterladen* wird eine *.csv* Datei zum Download angeboten.
Diese Datei kann von Statistikprogrammen importiert und dort weiter verarbeitet werden.



PTSS_10_Fragebogen_.txt-2.csv

	A	B	C	F	G	H	I	J	K	L	M	N	O	P	Q
1	Fragebogen	Eingereicht am:	Studie	Vollständiger Name	Q01	Q02	Q03	Q04	Q05	Q06	Q07	Q08	Q09	Q10	
2	2	29.07.13 20:15	SURE Demo	Anonym - lobustho	selten	nie	manchmal	selten	mehrmals	manchmal	selten	selten	mehrmals	oft	
3	3	29.07.13 20:17	SURE Demo	Anonym - admin	oft	oft	sehr oft	oft	immer	oft	mehrmals	manchmal	oft	immer	
4	6	15.08.13 20:20	SURE Demo	Anonym - admin	selten	nie	selten	nie	mehrmals	manchmal	selten	nie	nie	selten	
5	9	19.08.13 15:53	SURE Demo	Anonym - admin	manchmal	selten	mehrmals	manchmal	manchmal	mehrmals	selten	immer	immer	oft	
6	10	19.08.13 15:59	SURE Demo	Anonym - 4032559560	selten	immer	immer	immer	sehr oft	sehr oft	immer	sehr oft	sehr oft	immer	
7	12	19.08.13 16:12	SURE Demo	Anonym - 4032559560	selten	manchmal	nie	mehrmals	selten	oft	nie	mehrmals	selten	mehrmals	
8	14	19.08.13 16:14	SURE Demo	Anonym - 4032559560	nie	selten	nie	nie	nie	nie	nie	nie	nie	nie	
9	15	19.08.13 16:18	SURE Demo	Anonym - 4032559560	nie	mehrmals	selten	manchmal	selten	oft	selten	sehr oft	nie	selten	
10	17	19.08.13 16:24	SURE Demo	Anonym - 1118473450	nie	manchmal	selten	mehrmals	oft	nie	mehrmals	selten	nie	sehr oft	
11	18	19.08.13 16:26	SURE Demo	Anonym - 1907607483	nie	nie	nie	nie	nie	nie	nie	nie	nie	nie	
12	19	19.08.13 16:27	SURE Demo	Anonym - 1118473450	oft	oft	oft	oft	oft	oft	oft	oft	oft	oft	
13	20	19.08.13 16:29	SURE Demo	Anonym - 1907607483	nie	nie	nie	nie	immer	nie	nie	nie	nie	immer	
14	21	19.08.13 17:45	SURE Demo	Anonym - 8838015946	immer	immer	immer	immer	immer	immer	immer	immer	immer	immer	
15	22	19.08.13 20:32	SURE Demo	Anonym - 1907607483	selten	selten	selten	selten	selten	selten	selten	selten	selten	selten	
16	25	19.08.13 21:17	SURE Demo	Anonym - 1499908433	nie	nie	nie	nie	nie	nie	nie	nie	nie	nie	
17	27	19.08.13 21:37	SURE Demo	Anonym - 8838015946	nie	selten	selten	manchmal	oft	sehr oft	sehr oft	manchmal	immer	selten	
18	36	20.08.13 21:25	SURE Demo	Anonym - 1980937953	nie	selten	selten	selten	selten	selten	selten	selten	selten	selten	
19	37	20.08.13 21:25	SURE Demo	Anonym - 5906373284	nie	selten	oft	selten	mehrmals	selten	manchmal	manchmal	mehrmals	mehrmals	
20	39	20.08.13 21:30	SURE Demo	Anonym - 5906373284	mehrmals	mehrmals	manchmal	selten	manchmal	selten	manchmal	manchmal	selten	selten	
21	42	20.08.13 21:39	SURE Demo	Anonym - 5906373284	nie	nie	selten	manchmal	manchmal	selten	selten	manchmal	selten	manchmal	
22	43	20.08.13 21:48	SURE Demo	Anonym - 5906373284	nie	selten	manchmal	manchmal	selten	manchmal	mehrmals	manchmal	manchmal	selten	
23	44	20.08.13 21:50	SURE Demo	Anonym - 5906373284	nie	selten	nie	mehrmals	oft	sehr oft	mehrmals	manchmal	selten	manchmal	
24	45	20.08.13 21:52	SURE Demo	Anonym - 5906373284	nie	manchmal	manchmal	mehrmals	oft	immer	oft	mehrmals	manchmal	selten	
25	48	20.08.13 21:57	SURE Demo	Anonym - 4563652239	nie	selten	manchmal	mehrmals	selten	mehrmals	manchmal	selten	manchmal	sehr oft	
26	52	21.08.13 21:56	SURE Demo	Anonym - 9946237534	nie	nie	nie	nie	nie	nie	nie	nie	nie	nie	
27	55	25.08.13 15:48	SURE Demo	Anonym - 1907607483	manchmal	manchmal	nie	immer	immer	sehr oft	immer	nie	nie	sehr oft	
28	56	24.03.14 20:07	SURE Demo	Anonym - 9946237534	oft	mehrmals	sehr oft	mehrmals	oft	mehrmals	manchmal	mehrmals	mehrmals	manchmal	

Abbildung 52: Fragebogendaten in Microsoft Excel

7.1.7 Videounterstützung (Browserweiche)

Moodle unterstützt von Haus aus die Integration von lokal auf dem Server gespeicherten Videos, bzw. Videos, die auf den öffentlichen Videoportalen YouTube und Vimeo gespeichert sind. Werden Videos auf Videoportale hochgeladen, können sie über das integrierte Video Modul verlinkt werden. Hierauf wird automatisch im Quelltext ein Videoplayer (z.B.: YouTube. Youtube Player) eingebettet, der die Videos vom Videoportal bezieht und sich um die Auswahl des passenden Videoformats kümmert. Leider kommt die Bereitstellung von Videos über öffentliche Videoportale aus datenschutzrechtlichen Gründen nicht in Frage, da hier personenbezogene Daten unkontrollierbar auf fremden Servern erhoben und verarbeitet werden.
Möchte man auf dem eigenen Server gespeicherte Videos verwenden (Stand Moodle 2.5) kann man diese über eine Maske hochladen. Allerdings beschränkt sich die Wiedergabe auf das Videoformat der hochgeladenen Daten oder auf Adobe Flash. Geräte ohne installierten Flashplayer (z.B. Smartphones und Tablets) können Videos nicht im Flashformat anzeigen. Mobile Endgeräte bevorzugen *MP4* oder ähnliche Medienformate. Auf der anderen Seite unterstützen nicht alle Browser bzw. Betriebssysteme *MP4*. Der mit Windows 7 ausgelieferte Internetexplorer unterstützt als prominentes Beispiel von Haus aus kein *MP4*. Um sicher zu gehen, dass die Videos auf allen gängigen Betriebssystemen und Browsern angezeigt werden können, muss auf die Besonderheiten jedes Endgeräts, Betriebssystems sowie dessen Browser eingegangen werden. Anhand der technischen Ausstattung des Studienteilnehmers müssen Videos im passenden Format abgespielt werden.
Auf großen Videoportalen erledigen spezielle Streaming Server diese Aufgabe. Sie liefern das gewünschte Video automatisch in einem passenden Videoformat aus. Streaming Server sind meist kommerzielle Produkte. Die Kosten für den beliebtesten Streaming Server (Adobe Media Server Professional 5) belaufen sich auf ca. 5.800€ (Stand 2013). Eine OpenSource Alternative stellt Red5 dar. Red5 ist kostenfrei, quelloffen und wird bereits von einigen großen Videoportalen eingesetzt. Allerdings ist hierfür Programmierung und Anpassung notwendig. Testweise wurde versucht Red5 auf einem Linux Server zu installieren, zu konfigurieren und in Moodle zu integrieren. Der zeitliche und technische Aufwand, insbesondere beim Hinzufü-

gen weiterer Videos stellte sich als zu umfangreich für den geplanten Einsatzzweck dar. Daher wurde eine eigene Lösung programmiert, welche die Aufgabe eines Streaming-Mediaservers übernimmt, jedoch keine zusätzliche Serversoftware benötigt. Hierbei wird das gewünschte Video in unterschiedlichen Formaten auf dem Server gespeichert und abhängig von den technischen Gegebenheiten des verwendeten Endgerätes im passenden Format ausgeliefert (Browserweiche). Zum Anzeigen der Videos wird der quelloffene Videoplayer Flowplayer eingesetzt. Er kommt auch im Moodle-eigenen Video Modul zum Einsatz. Überprüft werden die Browser Internet Explorer, Safari, Opera, Firefox. Für mobile Endgeräte wird zwischen Android, iOS, Blackberry und Windows Mobile unterschieden.

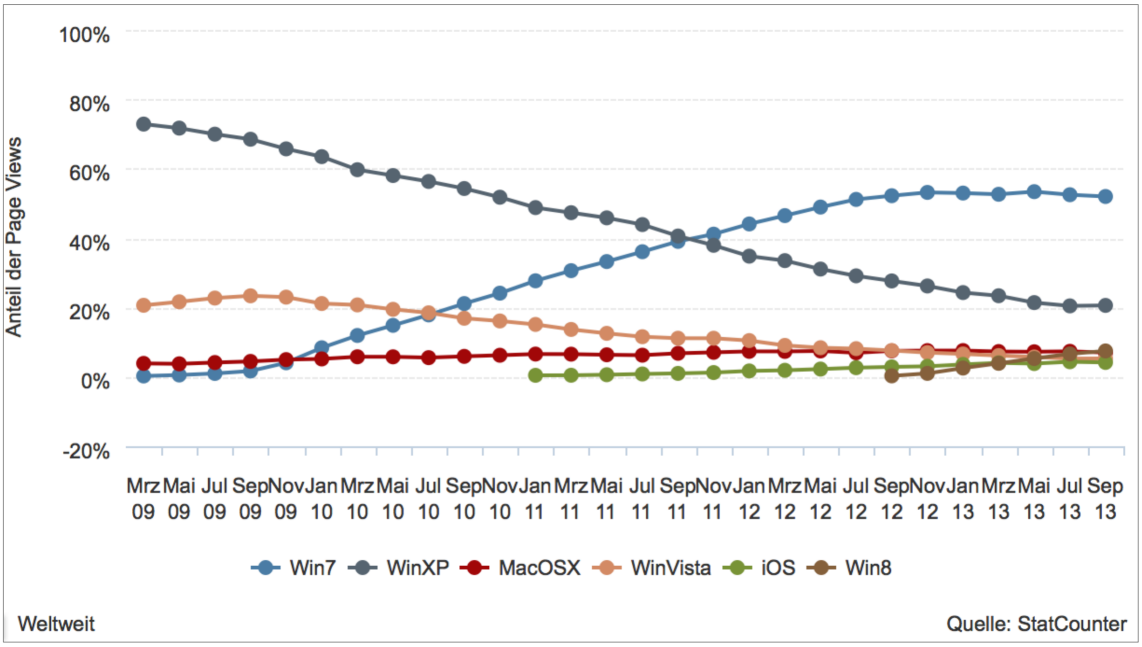


Abbildung 53: Prävalenz von Betriebssystemen weltweit, statista [187]

Die Verbreitung der Browser in Deutschland (Stand Oktober 2013) wird in der folgenden Übersichtstabelle zusammengefasst. Sie enthält darüber hinaus eine Einteilung der Browser in der Fähigkeit Videos via *HTML5* oder *Flash* darzustellen. Videoinhalte können entweder über *HTML5* oder als Flashvideo an PCs oder Smartphones ausgeliefert werden. Die meisten mobilen Endgeräte unterstützen das *HTML5* Format. Bei PCs hingegen unterstützen nur die aktuellsten Browser *HTML5* vollständig. Aus der Tabelle wird ersichtlich, dass weder Flash noch *HTML5* von allen Browsern unterstützt wird.

Browser/Gerät	Verbreitung	HTML5 Video	Flash Video
Firefox ab Version 3.5	41,7%	ja	ja
Chrome > 3.0	22,5%	ja	ja
Internet Explorer 9, 10, 11	15,7%	ja	ja
Internet Explorer 6, 7, 8	7,3%	nein	ja
Safari > 3	5,4%	ja	ja
iOS > 3	3%	ja	nein
Opera > 10.6	2,0%	ja	ja
Sonstige	1,5%	?	?
Android > 2.3	0,9%	ja	ja/nein
Firefox < 3	0,1%	nein	ja

Tabelle 6: Übersichtstabelle Browserkompatibilität und Browserverbreitung, Verbreitung: statscounter 10/2013 [51]

In einem Unterverzeichnis der Moodle-Installation (*var/www/moodle/videoplayer*) befinden sich die Browserweiche sowie die Programmdateien des Videoplayers. Die Datei *index.php* enthält den Algorithmus der Browserweiche. Sie überprüft den Browser des Anwenders und liefert das für die Browserversion am besten geeignete Videoformat. Im Gegensatz zu Flash wird für *HTML5*-Videos keine zusätzliche Software benötigt. Jedoch bevorzugen die Browser je nach Generation unterschiedliche Videocontainer. Um dies abzufangen, werden den Browsern die drei gängigsten Videocontainer (*mp4, WebM, OGV*) zur Verfügung gestellt. Die vereinfachte Funktionsweise der Videoauswahl wird an Hand des Schaubildes verdeutlicht.

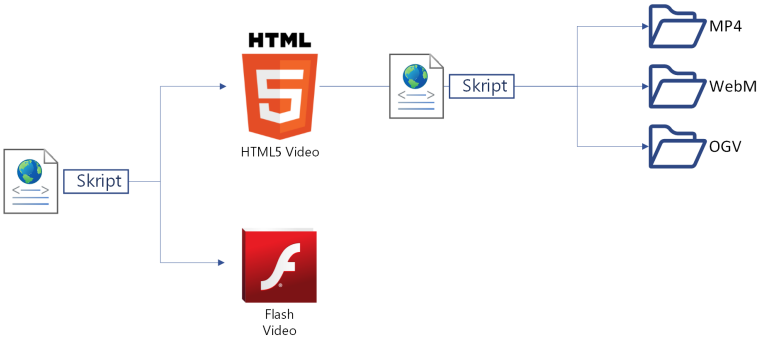


Abbildung 54: Funktionsprinzip der Browserweiche, Logos: wikimedia.org CC-BY

Die in den Containern enthaltenen Videos können wiederum mit unterschiedlichen Codecs komprimiert sein. Jedoch haben sich Quasi-Standards, also sehr häufig verwendete Codec-Kombinationen etabliert. Die Tabelle fasst die Wahl der Codecs für die jeweiligen Videocontainer zusammen.

Container	Video Codec	Audio Codec
Container	Video Codec	Audio Codec
mp4	H.264	AAC
WEBM	VP8	Vorbis
OGV	Theora	Vorbis

Tabelle 7: Wiedergabe von Videos über das Studienportal, flowplayer.org [188]

mp4 via *HTML5* wird sich vermutlich langfristig als Standard durchsetzen und ist daher die Präferenz mit der versucht wird Videos auszuliefern. Mit Ausnahme von Opera und älteren

Versionen des Mozilla Firefox unterstützen aktuelle Browser und Smartphones *mp4* via *HTML5*. Für ältere Versionen des Internet Explorers, die standardmäßig mit Windows XP, Windows Vista und Windows 7 ausgeliefert wurden, wird Flash als Videoformat ausgewählt. *OGV* dient als abwärtskompatible Lösung für ältere oder exotische Browser und mobile Endgeräte. Die folgende Tabelle gibt einen Überblick über *HTML5*-fähige Browser.

Browser	Ogg	mp4	Webm
Internet Explorer	-	9.0	-
Firefox	3.5	21.0	4.0
Chrome	3.0	3.0	6.0
Safari	-	3.1	-
Opera	10.5	-	10.6
iPhone/iPad	-	3.1	-
Android	4.0	2.3	2.3

Tabelle 8: Browserkompatibilität HTML5

Es wird schnell klar, dass die heterogene Browserlandschaft und eine Vielzahl an kursierenden Browserversionen eine Fallunterscheidung notwendig macht, um sicherzustellen, dass jeder Benutzer Videos problemlos wiedergeben kann. Daher wird zunächst überprüft, ob es sich um ein mobiles Endgerät handelt. Die beiden häufigsten mobilen Betriebssysteme, Android und iOS unterstützen *mp4/HTML5* und werden daher direkt damit versorgt. Bei seltene-ren Geräten wie Windows-Phones, Blackberry, Palm, usw. wird zunächst versucht, *mp4* auszuliefern. Scheitert dies, wird *WebM* und *OGV* ausprobiert. Bleibt auch dies erfolglos, wird versucht, das Video über Flash abzuspielen, bevor der Benutzer darüber informiert wird, dass sein Gerät nicht unterstützt wird.

Wird das Video von einem PC oder Mac angefordert, werden zunächst Betriebssystem und Browser, sowie dessen Version bestimmt. Anhand der Browserversion wird (auf Basis obiger Tabelle) entschieden, ob ein Browser *HTML5*-fähig ist, oder nicht. Ältere Browser, wie Internet Explorer 6 bis 8, aber auch ältere Firefox Versionen < 3.5 beherrschen keine *HTML5* Videos und müssen daher mit Flash versorgt werden. Hierfür wandelt der Flowplayer die *mp4*-Version des Videos in ein Flashvideo um. Ist kein Flash-Plugin mit Version 9 oder neuer installiert, wird der Benutzer darauf hingewiesen, dass das Video nur mit aktuellem Flash-Player wiedergeben kann und auf die Webseite von Adobe verwiesen. Der Algorithmus wird anhand Flussdiagramms auf der nächsten Seite dargestellt:

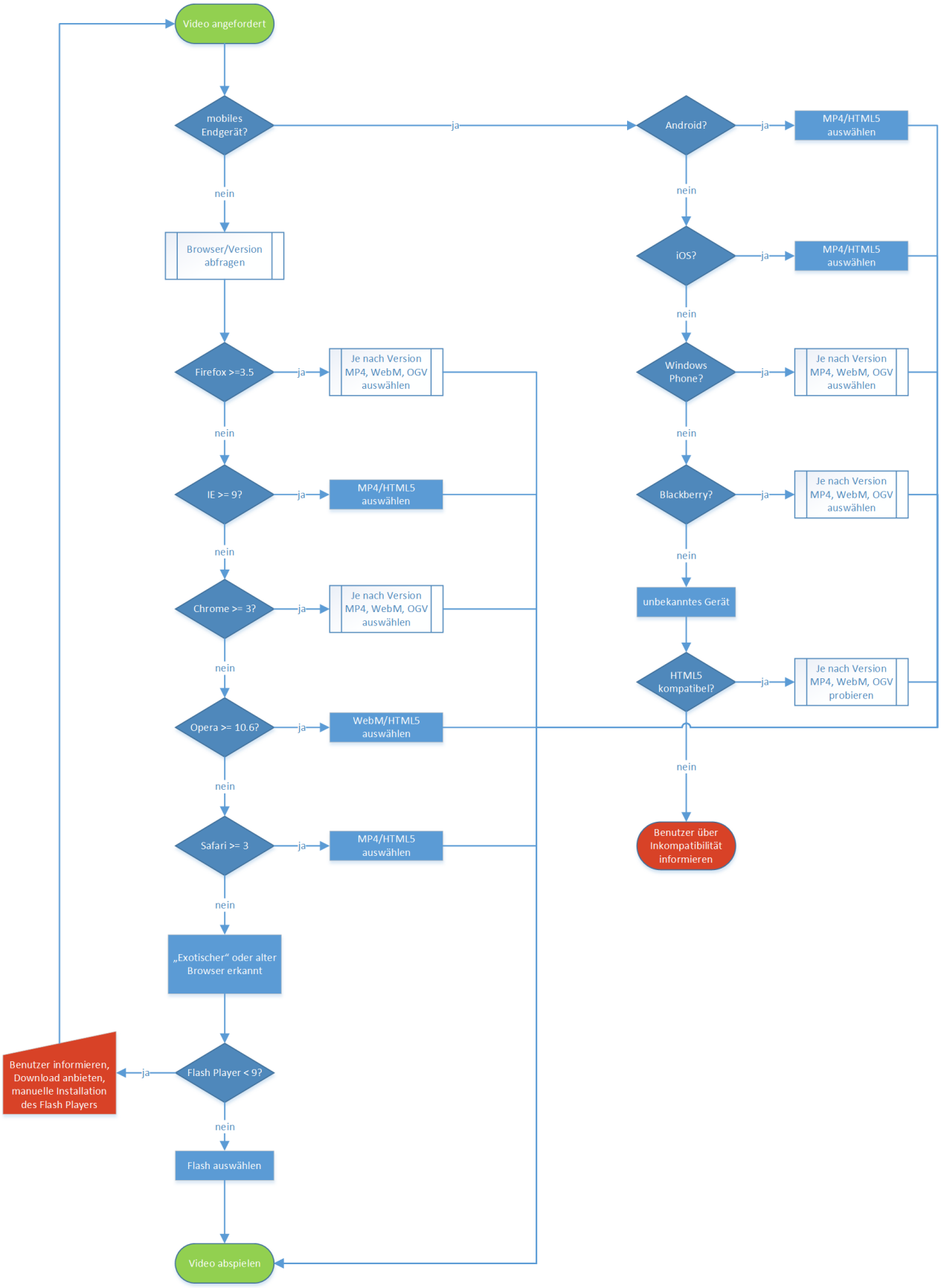


Abbildung 55: Flussdiagramm mit Algorithmus der Browserweiche

Soll ein Video über die Browserweiche bereitgestellt werden, muss zuvor für jeden Container eine passende Videodatei erzeugt werden. Konvertierungssoftware zur Umwandlung von Videos in die benötigten Containerformate gibt es für jedes Betriebssystem. Für die Konvertierung der SURE -Videos von *MPEG-2* nach *MP4*, *WebM* und *OGV* wurde der kostenfreie MiroVideoConverter verwendet. Er steht als OpenSource Software für Windows, Linux und Mac unter *www.mirovideoconverter.com* zur Verfügung.
Der Ordner, in dem die Videos abgespeichert werden, sowie der Dateiname der drei Formate, können angepasst werden. Auf diese Weise können mehrere unterschiedliche Videos an unterschiedlichen Speicherorten verwendet werden. Es müssen jedoch die zu einem Video gehörenden Videodateien immer im selben Unterverzeichnis liegen und den gleichen Dateinamen (bei unterschiedlicher Dateiendung) besitzen. Prinzipiell können die Videos auch auf einem separaten Server gespeichert sein, sofern der Videoordner mit den Videodateien über eine URL erreicht werden kann. Auf diese Art und Weise kann auch eine Lastverteilung realisiert werden.
Im folgenden Beispiel sollen zwei Videos *sure* und *other_video* im Ordner *video* gespeichert werden. Die notwendigen Dateien für das SURE -Video sind *sure.mp4,* sure.*webm* und *sure.ogv*. Die Videodateien im Beispiel müssen demnach folgende Struktur aufweisen:

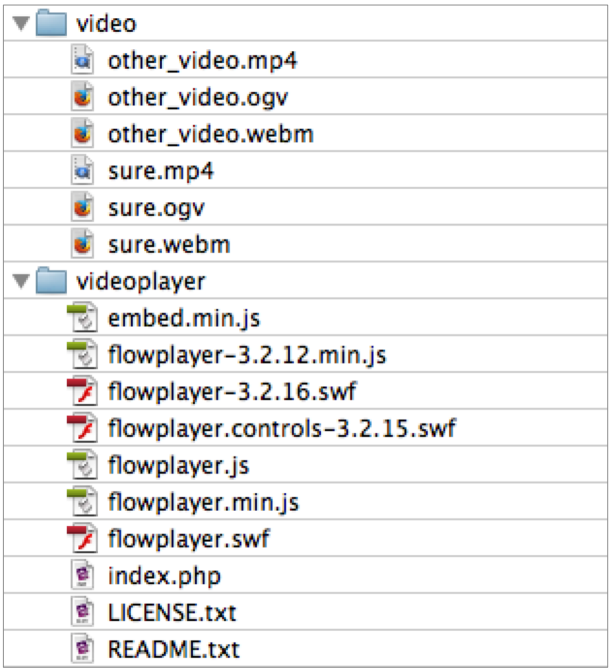


Abbildung 56: Übersicht des Videoverzeichnisses

Die Integration in Moodle erfolgt über einen eingebetteten Link. Der Link enthält den *absoluten Pfad* zur Browserweiche (*index.php*), sowie zwei Parameter für den Speicherort der Videos (Ordner *vf=**video***) und den Dateinamen (Dateiname *vs=**sure***).

Für das erste Video:

*https://uni-r.lobustho.de/videoplayer/index.php?vf=**video**&vs=**sure***

Abbildung 57: Einbinden eines Videos in Moodle über die Browserweiche

und das zweite Video:
*https://uni-r.lobustho.de/videoplayer/index.php?vf=**video**&vs=**other_video***

Für den Aufruf von Videos über eine sichere Leitung muss bei Eingabe der URL auf *https* geachtet werden. Wird das Skript auf einem separaten Server eingesetzt, muss dieser über ein gültiges *SSL Zertifikat* verfügen.

Es gilt zu beachten, dass bei Kenntnis der URL eine Wiedergabe möglich wäre, ohne als Benutzer am System angemeldet zu sein. Durch Überprüfung des Moodle Cookies innerhalb der Browserweiche kann der Zugriff auf den Personenkreis angemeldeter Benutzer reduziert werden. Das Downloaden des Videos oder das Abfilmen des Bildschirms durch angemeldete Benutzer kann jedoch auch durch höhere Sicherheitsmaßnahmen nicht verhindert werden.
Im Prototyp wurde auf die Implementierung von Schutzmaßnahmen der Videos verzichtet.

7.1.8 Registrierung der Studienteilnehmer

Die Registrierung der Teilnehmer, sowie deren automatische Einschreibung in die Studien, erfolgt über studienspezifische *Landing-Pages*. Der Moodle-Datei */login/signup.php*, die für die Registrierung neuer Benutzer zuständig ist, können Variablen übergeben werden. Mit Hilfe individueller Formularfelder können zum Zeitpunkt der Registrierung beliebige Daten an Moodle weitergereicht werden. Dies kann dazu benutzt werden, um Moodle mitzuteilen, in welcher Studie der Teilnehmer eingeschrieben werden soll.
Um das Registrierungsformular an eine beliebige Webseite anzufügen, wurde die Datei *landing-page.php* programmiert. Die Datei greift hierbei auf die Moodle-eigene Registrierungsfunktion zurück und erweitert diese. Eine selbst erstellte *Landing-Page* kann somit einfach um das Registrierungsformular erweitert werden, ohne Kenntnisse von den Programmiersprachen HTML und PHP zu besitzen. Die Datei *landing-page.php* enthält fünf Variablen, die angepasst werden können. Der übrige Quellcode muss nicht bearbeitet werden. Er implementiert das Standard-Moodle- Registrierungsformular, erzwingt eine verschlüsselte Verbindung, speichert und verschlüsselt den Benutzernamen, das Passwort sowie die Studie und die Rolle Teilnehmer in der Moodle- Datenbank.

```
define("studie", "sure-bundeswehr");
define("user_rolle_id", 5); // Rolle 5 = Teilnehmer/in
define("path_landingpage_text", "landingpage_text.html");
define("path_to_login_folder", "../login");
define("path_to_file", "/landingpage/landing-page.php");
```

Abbildung 58: Variablen für die Integration des Registrierungsformulars in Landing-Pages

Die wichtigste Variable ist hierbei die Variable *studie.* Sie muss die von Moodle vergebene *Studien-ID* enthalten. Die *Studien-ID* kann beim Anlegen neuer Studien über die Grundeinstellung frei gewählt werden. Sie sollte aber keine Umlaute, Sonder- und Leerzeichen enthalten. Es muss unbedingt sichergestellt werden, dass *Studien-ID* und der Wert der Variablen übereinstimmen, da hierüber die Verknüpfung zwischen *Landing-Page* und Studie hergestellt wird. Nur so können Teilnehmer automatisch in die richtige Studie eingeschrieben werden.

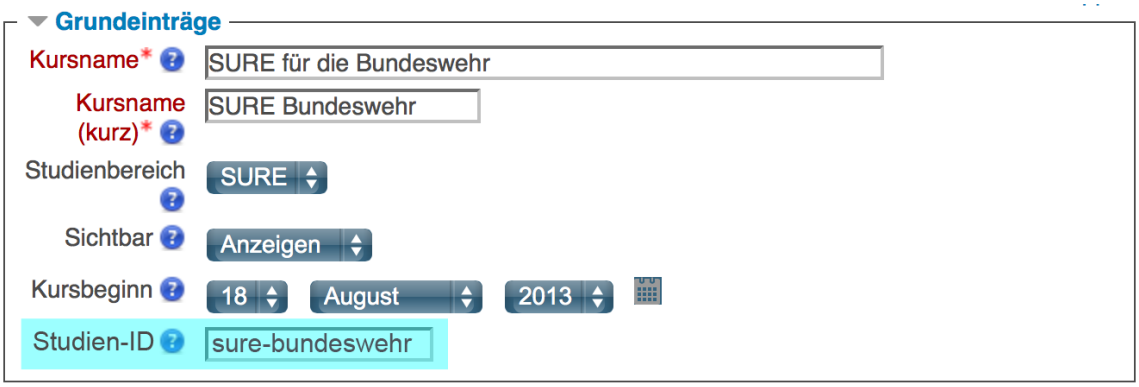


Abbildung 59: Verknüpfung von Landing-Page und Studie

Erläuterung der Variablen:

user_rolle_id
Über *user_rolle_id* wird die Rolle definiert, die der neue Benutzer im System haben wird. Teilnehmer haben die *Rollen-ID* 5. Diese Zahl sollte nicht geändert werden.

path_landingpage_text
Mit Hilfe der Variable *path_landingpage_text* kann eine beliebige *HTML-basierte* Webseite eingebunden werden. Das Registrierungsformular wird automatisch an den Fuß der Seite angefügt. Die Variable enthält den Pfad zur gewünschten *HTML-Datei* ausgehend von der Datei *landing-page.php*.

path_to_login_folder
path_to_login_folder enthält den Pfad zur Login-Seite von Moodle.

path_to_file
path_to_file enthält den Pfad und den Dateinamen der *landing_page.php*.
Sollten Landing-Page und Studienportal nicht auf dem gleichen Server liegen, muss die Datei *landing_page.php* in *index.php* umbenannt werden. Die Pfadangaben aller Variablen müssen dann entsprechend angepasst werden und auf die Internetadresse des Studienportal-Servers verweisen. Liegen *Landing-Page* und Studienportal auf dem gleichen Server, können relative Pfade, wie im obigen Beispiel gezeigt, verwendet werden.

Die Integration des Registrierungsformulars in bestehende Webseiten via *landing-page.php* ist zwar wie gezeigt recht einfach, jedoch ist die gestalterische Freiheit hierdurch eingeschränkt. Um eine individuelle Formulargestaltung umzusetzen, sind HTML und PHP Kenntnisse erforderlich. Es müssen Formularfelder erzeugt und mit Hilfe der URL an Moodle übergeben werden. In diesem Fall muss jedoch auch die Plausibilitätskontrolle der Eingaben selbst programmiert werden, da nicht auf die Moodle-Funktion zurückgegriffen werden kann.
Die nicht individuell gestaltete Registrierungsmaske *via landing-page.php* wurde im Rahmen des Prototypen nicht weiter auf Optik optimiert. Sie dient primär als Beispiel dafür, wie prinzipiell eine Integration des Formulars mit geringem technischen Aufwand möglich wäre.
Für größere Studien sollte erwogen werden die Registrierungsmaske dem Seitendesign anzupassen und an entsprechender Stelle in der *Landing-Page* zu platzieren.
Für die Erstellung eines individuellen Formulars sind rudimentäre Kenntnisse der Programmiersprache PHP, CSS und HTML notwendig.

Die folgenden beiden Beispiele sollen den Unterschied zwischen dem Einsatz der *landing-page.php* - Methode und einem individuellen Formulardesign verdeutlichen.

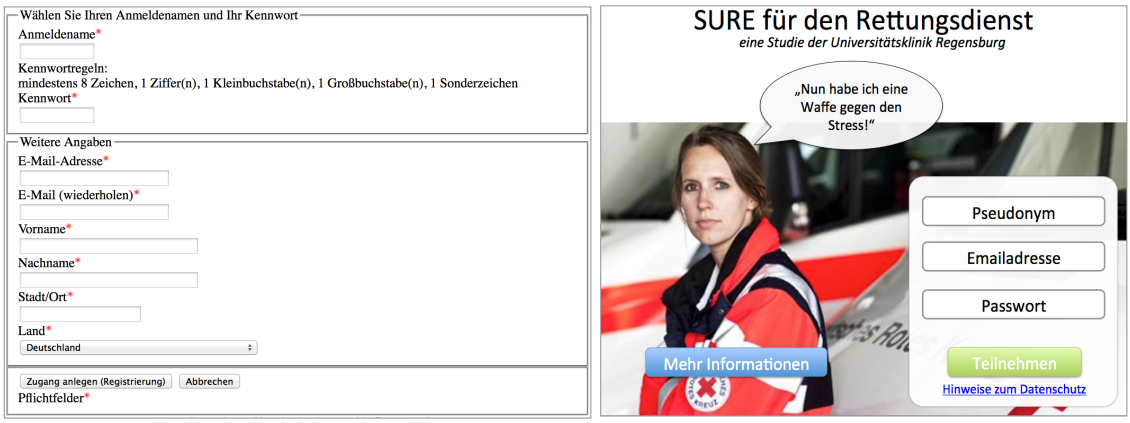


Abbildung 60: allgemeines Registrierungsformular v.s. individuelles Registrierungsformular in einer Landing-Page

Da besonders während des Zeitpunkts der Registrierung sensible Daten wie Benutzername und Passwort übertragen werden, muss unbedingt darauf geachtet werden, dass die *Landing-Pages* mit einem *SSL Zertifikat* abgesichert werden und ausschließlich verschlüsselt aufgerufen werden können.

7.1.9 Studienverwaltung

Moodle bietet die Möglichkeit Kursanträge zu erstellen. Diese Funktion wurde übernommen und auf den Studienkontext übertragen. So können beispielsweise Doktoranden, die im System die Rolle Studienbetreuer oder Studienassistent besitzen auf elektronischem Weg eine neue Studie beantragen. Der Manager wird darüber per Email informiert und kann per Knopfdruck zustimmen oder den Antrag ablehnen. Eine andere Möglichkeit ist das manuelle Anlegen einer Studie durch den Administrator oder Studienmanager. Jede Studie enthält einige zusätzliche Informationen. Neben bereits besprochenen *Studiennamen* und der *Studien-ID* kann ein *Studienbeginn* definiert werden. Die Studie wird dadurch erst zum eingestellten Zeitpunkt für Teilnehmer sichtbar.
Studien können in Bereiche zusammengefasst werden. Dies ermöglicht es, Studien im Fall der vier geplanten Bereiche (SURE spezifiziert für die Bundeswehr, den Rettungsdienst, die Pflege und die Polizei), zu einem Studienbereich mit dem Namen SURE zusammenzufassen. Berechtigungen und viele allgemeine Einstellungen können in einem Studienbereich angewendet werden und übertragen sich automatisch auf alle zugehörigen Studien. Die *Studienbeschreibung* kann frei gestaltet werden. Es können zusätzlich Medien (Bilder, Videos usw.) und Dokumente (z.B. eine Datenschutzerklärung) angefügt werden. Für jede Studie können Studiengruppen gebildet werden. Alle Elemente der Studie (z.B. Fragebögen, Medien, Dokumente usw.) können selektiv für alle Teilnehmer oder nur für eine spezielle Gruppe von Teilnehmern sichtbar gemacht werden. Die Einteilung der Benutzer in Gruppen muss derzeit manuell vorgenommen werden. (z.B. vom Studienbetreuer oder Studienassistenten) Es wäre jedoch technisch möglich, Gruppen (z.B. bei der Registrierung) automatisch randomisiert aufin Studiengruppen zu verteilen. Diese Funktion ist allerdings in der aktuellen Fassung des Prototyps nicht implementiert.

Abbildung 61: Manuelles Anlegen einer neuen Studie

7.1.10 Rollen und Benutzerrechte

Moodle besitzt bereits ein feingliedriges Rechtekonzept. Benutzer können zu Gruppen zusammengefasst und diese mit spezifischen Berechtigungen ausgestattet werden. Benutzergruppen kann Zugriff auf gesamte Module oder Teilfunktionen von Modulen gegeben werden. [38] Es können neue Rollen hinzugefügt oder bestehende Rollen dupliziert und die Kopie modifiziert werden. Auf diese Weise lassen sich mit einem Klick Berechtigungen für mehrere Benutzer gleichzeitig verändern. Natürlich können auch für einzelne Benutzer Rechte vergeben werden, die über die Rechte ihrer Gruppe hinausgehen.
Hilfreiche Symbole signalisieren mögliche Gefahren bei Vergabe der jeweiligen Berechtigung. Symbolhaft werden die Möglichkeiten - Schadcode in Eingabefelder einzufügen (⚠), - Gefahren für die Privatsphäre durch Einblick in Nutzerdaten (⚠), - Gefahr des Löschens von Datensätzen (⚠) oder die Möglichkeit an Nutzer -Emails zu versenden (⚠) dargestellt.

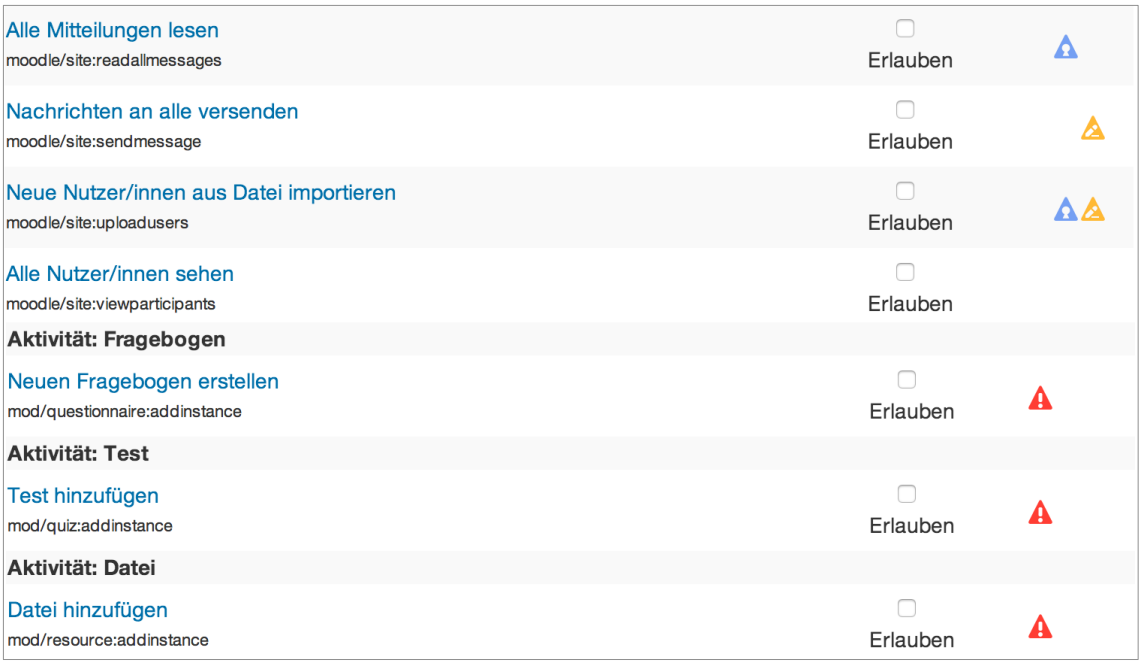


Abbildung 62: Beispiel der Vergabe von Berechtigungen für die Rolle des Studienassistenten

Auch für selbst programmierte Module kann die Rollenverwaltung und Rechtevergabe angewendet werden. Ist man sich unsicher, ob ein Benutzer oder eine Gruppe eine bestimmte Berechtigung besitzt, kann ein Bericht generiert werden, der abhängig von der Auswahl, alle Funktionen von Moodle oder auch nur die einzelner Module umfasst.

	Manager/in	Studienersteller/in	Studienbetreuer	Studienassistent	Teilnehmer/in
Umfrage hinzufügen mod/survey:addinstance	**Erlauben**	Nicht gesetzt	**Erlauben**	Nicht gesetzt	Nicht gesetzt
Fragebogen-Vorlage kopieren mod/questionnaire:copysurveys	**Erlauben**	**Erlauben**	**Erlauben**	Nicht gesetzt	Nicht gesetzt
Antworten herunterladen mod/survey:download	**Erlauben**	Nicht gesetzt	**Erlauben**	**Erlauben**	Nicht gesetzt
Umfrage beantworten mod/survey:participate	**Erlauben**	Nicht gesetzt	**Erlauben**	**Erlauben**	**Erlauben**
Antworten ansehen mod/survey:readresponses	**Erlauben**	Nicht gesetzt	**Erlauben**	**Erlauben**	Nicht gesetzt

Abbildung 63: Bericht der Berechtigungen des Fragebogenmoduls

7.2 Sicherheitsimplementierungen

Das Bundesamt für Sicherheit in der Informationstechnik (BSI) ist die oberste deutsche Bundesbehörde in Fragen der IT-Sicherheit. Ihr Auftrag ist die Analyse von Bedrohungen und Beratung in Fragen der IT-Sicherheit. Die vom BSI veröffentlichte Publikation zur „Sicherheit von Webanwendungen - Maßnahmenkatalog und Best Practices“ wurde daher als Orientierungshilfe für die Sicherheitsanalyse und Implementierung der Studienplattform herangezogen. Das Dokument richtet sich speziell an Projektleiter und Softwareentwickler, die Webanwendungen planen oder entwickeln. [17] Entsprechend dem Leitfaden wurden folgende Aspekte untersucht bzw. nach aktuellem Stand der Technik (12/2014) implementiert:

- Transportwege verschlüsseln
- sichere Passwörter erzwingen
- Zugangskontrollen einrichten
- Server sicher konfigurieren
- Sicherheitsüberprüfung (Audit)
- Systemüberwachung

Für Internetanwendungen gibt es eine Vielzahl an potentiellen Eintrittspforten. Die folgende Grafik zeigt einen Überblick über die unterschiedlichen Angriffsmöglichkeiten. Es sollte jeder Bedrohung mit einer Lösung auf technischer Ebene begegnet werden. Ist dies nicht umsetzbar, muss zumindest der Aufwand für potentielle Angreifer so weit wie möglich erhöht werden. Das folgende Kapitel beschäftigt sich mit der Analyse von Bedrohungen. Es zeigt wie diese, sofern im Einflussbereich des Servers bzw. des Betreibers liegend im Prototypen des Studienportals gelöst wurden.

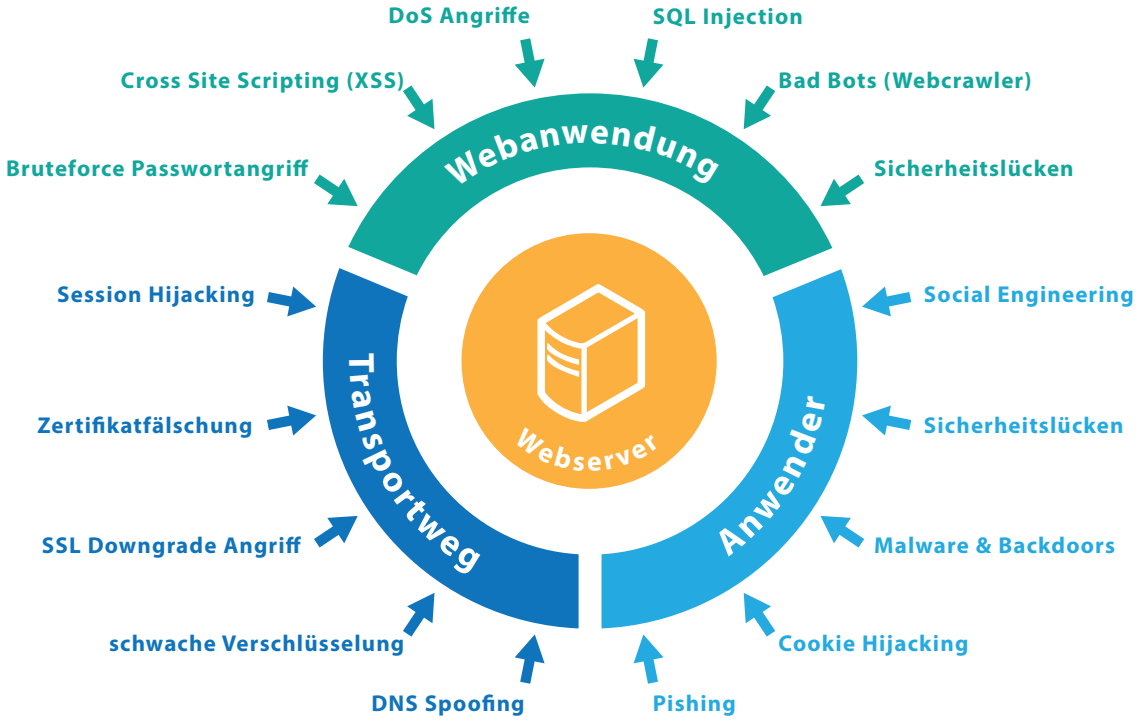


Abbildung 64: mögliche Angriffspunkte für Webanwendungen

7.2.1 Moodle Sicherheitseinstellungen

7.2.1.1 sichere Passwörter erzwingen

Um die Sicherheit des Studienportals zu gewährleisten ist es wichtig, dass die Benutzer, die Zugang zum System haben sichere Passwörter verwenden. Dies gilt in besonderem Maße für privilegierte Benutzer wie Studienbetreuer, Studienmanager oder Administratoren. Gelingt es einem Angreifer das Passwort eines privilegierten Benutzers zu erraten, können dessen Sonderrechte im System missbräuchlich verwendet werden, um z.B. Studienergebnisse einzusehen, im Studienbereich beliebige Inhalte zu platzieren oder mit Teilnehmern zu kommunizieren. Moodle bietet die Möglichkeit, Regeln für den Umgang mit Benutzernamen und Passwörtern zu definieren. Für die Wahl sicherer Passwörter empfiehlt das Bundesamt für Sicherheit in der Informationstechnik die Wahl einer Mindestlänge von 8 Zeichen, die sowohl eine Zahl als auch einen Groß- und einen Kleinbuchstaben enthalten. [17] Diese Empfehlung lässt sich dank einer von Moodle bereitgestellten Oberfläche leicht umsetzen.

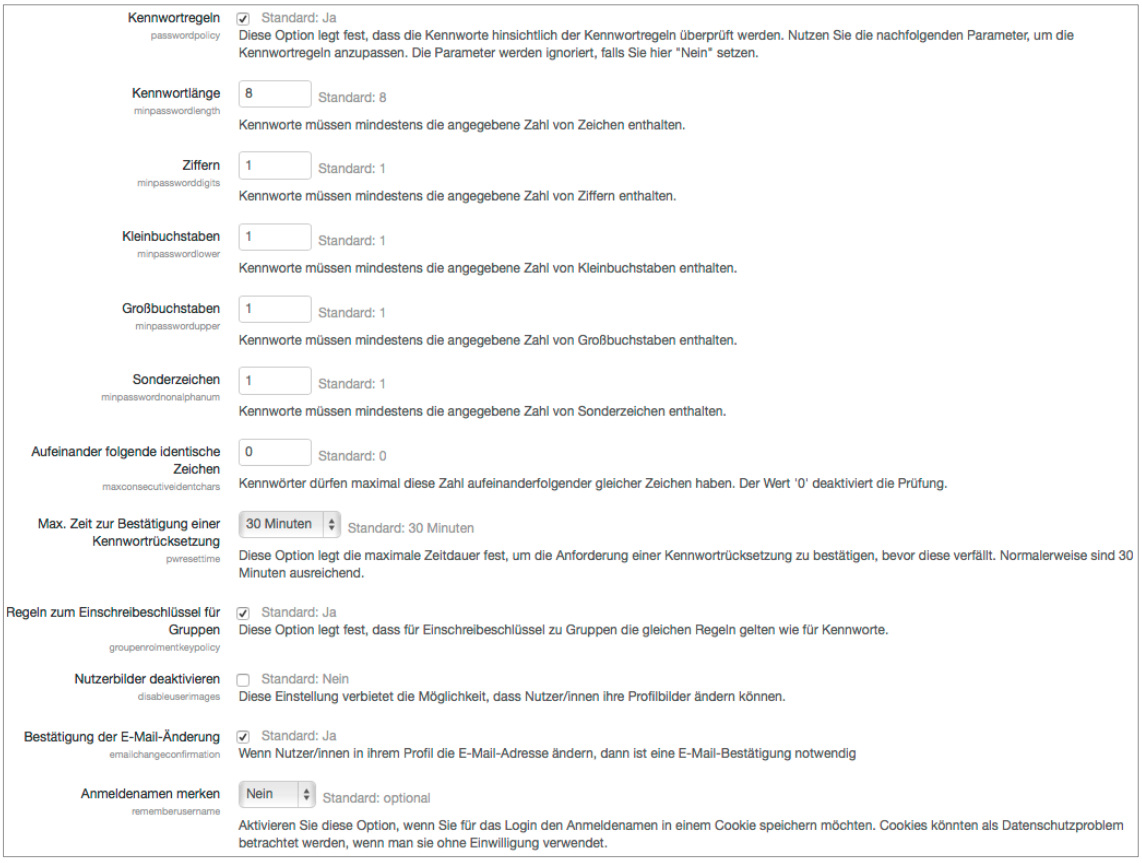


Abbildung 65: Passwörter müssen ein Groß- und Kleinbuchstaben und ein Sonderzeichen enthalten

Um zu verhindern, dass ein Angreifer durch einen *BruteForce-Angriff* ein Passwort erraten kann, besitzt Moodle die Möglichkeit eine Zeitsperre einzurichten. Während dieser frei definierbaren Zeit ist eine Passworteingabe für den gesperrten Benutzer nicht möglich. Automatisiertes Durchprobieren von Passwörtern kann hierdurch unterbunden oder zumindest stark verzögert werden.

Abbildung 66: Der Benutzer wird nach 5 Fehleingaben für 30 Minuten gesperrt

7.2.1.2 Cookie- und Session Hijacking unterbinden

Die größte Gefahr durch *SessionHijacking* geht von einem unverschlüsselten Transportweg aus. [17] [135] Moodle bietet die Möglichkeit einige wichtige Sicherheitseinstellungen anzupassen. Sie sind in der Grundinstallation jedoch nicht aktiviert. Zudem sind sie selbst in aktiviertem Zustand für einen sicheren Betrieb nicht ausreichend. [135] [89] Die Funktion *Login* über *HTTPS* schwenkt während der Übertragung von Benutzername und Passwort auf eine gesicherte Verbindung um. Allerdings kann nach dem Login der unverschlüsselte Datenverkehr weiterhin mitgelesen und manipuliert werden. Die Einstellung *Nur sichere Cookies* erzwingt den sicheren Übertragungsweg von Cookies. Die Funktion aktiviert die bereits beschriebene *secure cookie* Funktion. (☞ Kapitel 6.5.7) Dem Cookie wird hierzu der Eintrag *Secure=1* hinzugefügt. Wird diese Einstellung verwendet, muss serverseitig dafür gesorgt werden, dass die Verbindung durchgängig verschlüsselt erfolgt, da anderenfalls eine Anmeldung am System nicht mehr möglich ist.

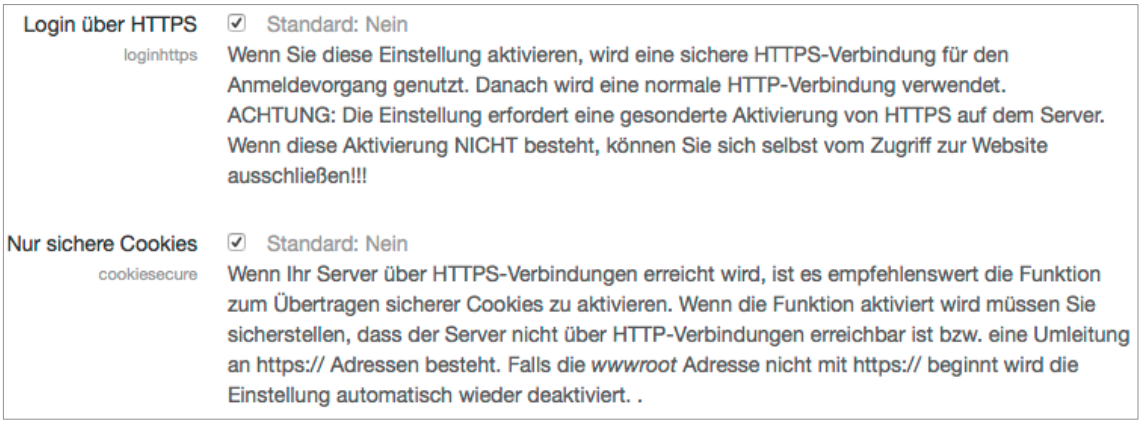


Abbildung 67: Notwendige Moodle HTTP Sicherheitseinstellungen

Nur durch das Erzwingen einer durchgängig verschlüsselten Verbindung kann das Erbeuten der *Session-ID* auf dem Übertragungsweg am sichersten verhindert werden. [89] [82]

In den meisten Browsern kann man die Cookie-Einstellung einsehen. Im Chrome Browser lässt sie sich besonders einfach durch einen Klick auf das Schlosssymbol in der Adressleiste anzeigen.

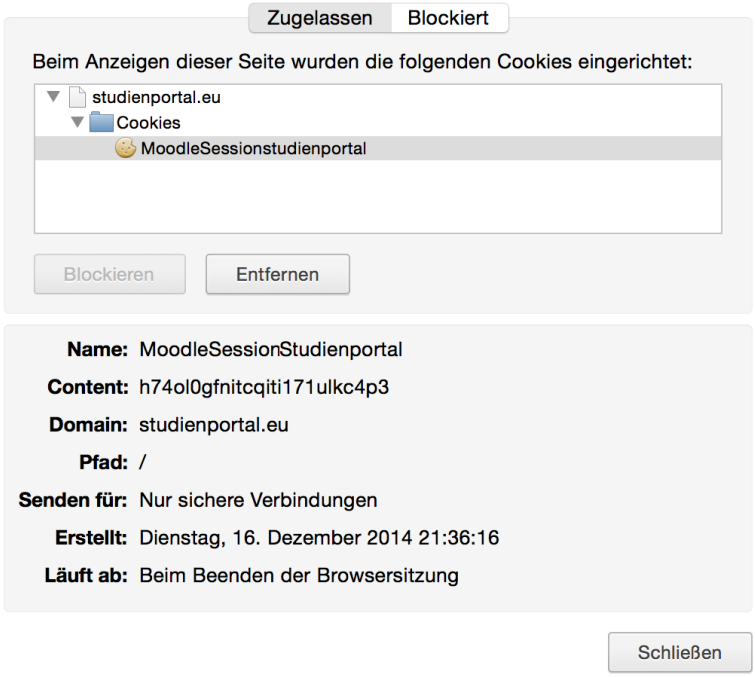


Abbildung 68: Cookie wird nur über https übertragen

Eine Möglichkeit trotz verschlüsselter Datenübertragung an die *Session-ID* zu gelangen, ist der Einsatz von Skripten, die vom Browser des Benutzers ausgeführt werden. Durch JavaScript kann auf Cookies zugegriffen werden, selbst wenn sie zuvor sicher übertragen wurden. [82] Um zu verhindern, dass sie von Skripten gelesen werden, kann der Zugriff auf Cookies beschränkt werden. Diese Funktion kann in Moodle über *Nur HTTP-Cookies* aktiviert werden. Dies setzt im Cookie die Einstellung *http-only=1*. [89] Nun darf nur noch der Webserver, nicht jedoch ein Skript das Cookie lesen.

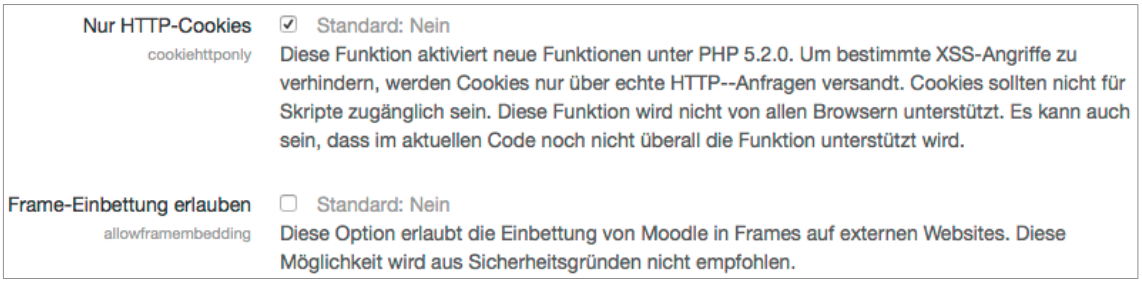


Abbildung 69: Zugriff auf Cookies durch JavaScript verbieten

7.2.1.3 CrossSite-Scripting (XSS) und SQL Injects (SQLi)

Für die eingesetzte Grundlage des Studienportals (Moodle 2.5) sind in der *CVE* Datenbank (*CVE = common vulnerabilities and exposures*) 63 Sicherheitslücken bekannt. Den Löwenanteil hiervon macht *XSS* aus.

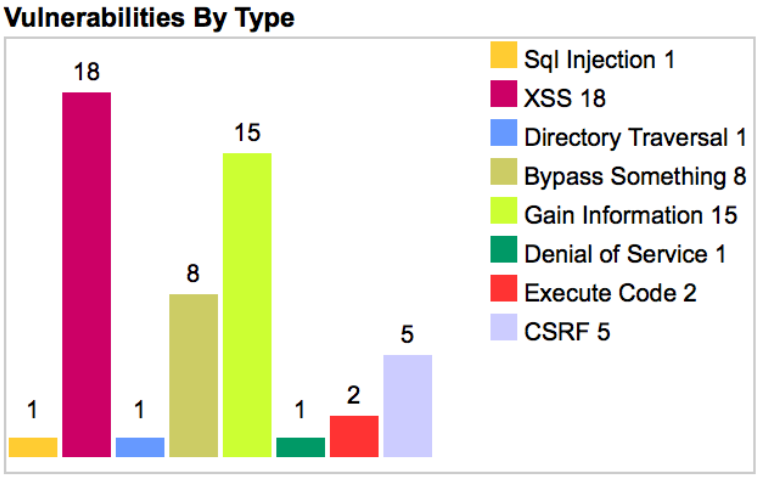


Abbildung 70: Bekannte Sicherheitslücken in Moodle 2.5, cvedetails.com [96]

Unter *CrossSite Scripting* versteht man das Einschleusen von bösartigem Code in Webanwendungen. *XSS* ist überall dort möglich, wo eine Benutzereingabe oder eine Interaktion des Systems mit dem Benutzer (z.B. Warnmeldungen) stattfindet. Da *XSS*-Angriffe sowohl über den gespeicherten Code auf dem Server, als auch über die URL durchgeführt werden können, ist faktisch jede Internetseite für diese Art des Angriffs verwundbar. [17] Ziel von *XSS*-Angriffen ist das Ausführen von JavaScript im Browser des Benutzers. Das eingeschleuste Skript kann wiederum genutzt werden, um einen Zugriff auf den Computer des Anwenders zu erlangen, um dort sensible Daten zu stehlen oder weitere Angriffe vom Computer des Opfers auszuführen. *XSS* Lücken in Webseiten sind beliebt und werden in großen Datenbanken veröffentlicht oder sogar verkauft. Unter *http://xssed.com* findet sich eine Datenbank mit über 45.000 eingetragenen Webseiten und den dazugehörigen *XSS* Schwachstellen. Darunter sind prominente Vertreter wie Google, Facebook, YouTube, PayPal, eBay und Portale vieler deutscher Banken und Emailanbieter [189]. Es ist daher besonders wichtig die Webseite auf XSS- und *SQLi*-Angreifbarkeit zu untersuchen und vor vermeidbaren Gefahren zu schützen. [17] Mit Hilfe einiger Einstellungen in Moodle können Eintrittspforten verringert werden. Hierzu müssen alle Moodle Funktionen deaktiviert bzw. eingeschränkt werden, die es einem Benutzer ermöglichen, *HTML*- oder *JavaScript-Code* in Moodle einzubauen. Die Funktion *Frame-Einbettung erlauben* (siehe vorherige Abbildung) sollte daher ebenso deaktiviert sein wie die Funktion *OBJECT/EMBED erlauben*. *Anmeldenamen merken* sollte deaktiviert und *Automatische Vervollständigung von Kennworten verhindern* aktiviert werden, um zu verhindern, dass Skripte die im Browser hinterlegten Zugangsdaten auslesen können.

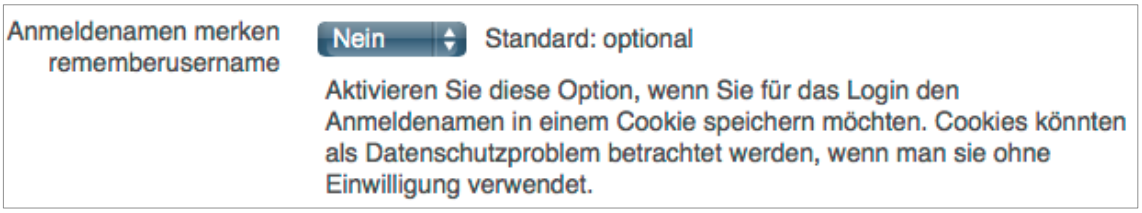


Abbildung 71: Anmeldenamen sollten nicht im Browser gespeichert werden

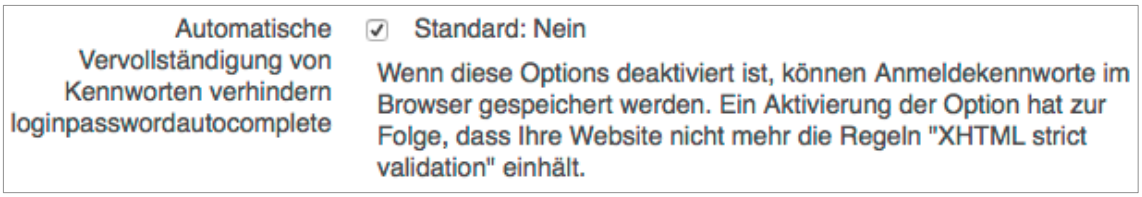


Abbildung 72: Speichern des Passworts im Browser unterbinden

7.2.2 Sicherheitsanalyse von Moodle

Um *XSS* zu vermeiden, müssen technisch betrachtet, alle Benutzereingaben überprüft werden, bevor sie angezeigt oder in der Datenbank der Webanwendung gespeichert werden. Über eine Liste unerlaubter Zeichen erfolgt ein Herausfiltern oder Umwandeln von Zeichen, die für das Einbringen von Schadcodes geeignet wären. Erlaubte Zeichen hingegen sollen gespeichert oder angezeigt werden. Diese Kontrollaufgabe muss die Webanwendung übernehmen. Wird diese Überprüfung und Umwandlung nicht korrekt durchgeführt, kann es gelingen Code einzuschleusen, der beim Aufruf der Webseite vom Browser des Benutzers ausgeführt wird. Besonders gefährliche Zeichen sind < oder >, da über sie JavaScript eingebaut werden kann. [129] Gelänge es zum Beispiel über ein Suchfeld folgenden Code auszuführen:

<script>alert(document.cookie)</script>

würde das aktuelle Cookie (mit der darin enthaltenen *Session-ID*) ausgegeben.
Aus diesem Grund wurde der interne Bereich von Moodle mit sogenannten *Pentest* Werkzeugen auf *XSS* Schwachstellen untersucht. Mit Hilfe der ExploitMe Werkzeuge wurde zunächst die Filterfunktion von Moodle bei Benutzereingaben kontrolliert. [112] Mit einem entsprechenden Regelwerk wurden im nächsten Schritt typische *XSS* Angriffe simuliert.
Die Überprüfung umfasste folgende Sonderzeichen: *; \ / < > " ' = ,* sowie Skripten, die über Bilder, *iFrames*, *StyleSheets*, *XML*, *Body*, *Meta* oder *Header* sowie als *Inline-Script* eingebracht werden können. Folgende Elemente des Studienportals (Moodle) wurden untersucht:

- Anmeldemaske
- Studienseite
- Suchfunktion
- Fragebogen-Modul
- Videoweiche
- Forum
- Nachrichten

Die meisten Module enthalten versteckte oder sichtbare Eingabefelder, über die Zeichen unencodiert angenommen werden. Die durchgeführten *XSS*- Angriffe waren jedoch bei keinem der Module erfolgreich. Da nur eine Auswahl klassischer Angriffsszenarien überprüft wurde, ist ein erfolgreicher *XSS*- Angriff dennoch nicht auszuschließen. Besonders durch das Forum Modul und die Suchfunktion wäre ein erfolgreiches Einschleusen von Schadcode wegen der unzureichend gefilterten Sonderzeichen denkbar.

Modul	Anzahl Tests	Angriff erfolgreich	Bemerkung
Login	154	-	unencodiertes Zeichen im Feld Benutzername
Studienübersicht	184	-	-
Suchfunktion	31	-	mehrere unencodierte Zeichen im Suchfeld
Fragebogen	308	-	-
Videoweiche	308	-	-
Forum	208	-	unencodierte Zeichen in Betreff, Nachrichtentext und Dateianhang möglich
Nachrichten	368	-	-
Studien	308	-	-

Tabelle 9: XSS Analyse mit XSS Heuristik mit ExploitMe

OWASP (Open Web Application Security Project) ist eine unabhängige non-profit- Organisation mit einer sehr umfangreichen Datenbank für bekannte Softwareschwachstellen. Das Ziel der Organisation ist es, Entwicklern und Betreibern von Webseiten kostenfrei Wissen und Werkzeuge für Systemaudits an die Hand zu geben. Eines dieser Werkzeuge ist das Programm OWASP ZAP. In der Literatur wurde Moodle von 3 Sicherheitsanalysten der US Air Force mit Hilfe dieses Werkzeugs analysiert und Schwachstellen aufgedeckt. In Anlehnung an die Sicherheitsanalysen von Floyd, Schultz und Fulton wurde Moodle entsprechenden Analysen unterzogen. [135] Hierbei wurde die Angreifbarkeit des gesamten Servers und der Moodle Installation überprüft.

7.2.3 Sicherheitsmängel von Moodle

Beim Vergleich der Moodle Version 2.5 mit aus Version 2.1 bekannten Schwachstellen konnten die zuvor beschriebenen *XSS* Vulnerabilitäten auch in der aktuellen Version wiedergefunden werden. Ein orientierender Scan ermittelte etliche als leicht bis mittelschwer eingestufte Risiken.

Risiko-Level	Risiken
Hoch	1
Mittel	8
Leicht	304
Hinweis	14

Tabelle 10: Ergebnis des Sicherheitsscans nach der Grundinstallation

Hauptverantwortlich für diese Meldungen sind ausnutzbare Schwachstellen, von denen sich zumindest einige durch Anpassungen des Webservers reduzieren lassen. Einer der Schwachpunkte ist die Auskunft über das verwendete Betriebssystem und die Version des eingesetzten Webservers. Simuliert man einen Webseitenaufruf in der Standardkonfiguration des Webservers, so wird folgende Information übertragen:

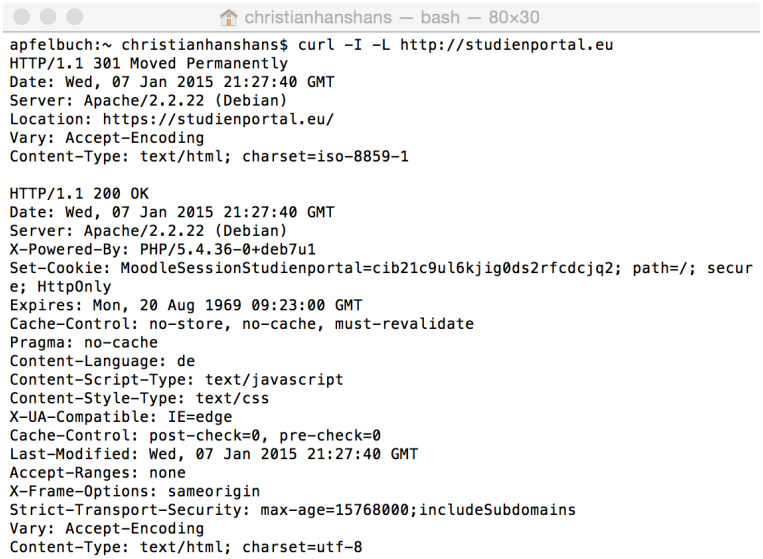


Abbildung 73: HTTP Header des Webservers vor Anpassung

Der abgebildete *HTTP-Header* ist wie eine Art Wasserzeichen, die jeder angezeigten Übertragung angehängt wird. Wie in der Abbildung zu erkennen ist, steht hier neben dem Betriebssystem (Debian) auch die verwendete Version des Webservers und die PHP-Version. Führt man diese Simulation z.B. bei der Lernplattform der Universität Regensburg durch, erfährt man, dass der Apache Webserver mit der *PHP 5.3* betrieben wird. Diese Information erlaubt eine schnelle Suche nach Sicherheitslücken der entsprechenden Software und sollte daher abgeschaltet werden.
Ein weiterer wichtiger Punkt ist die Einschränkung der Ausführbarkeit von Skripten. Durch eine entsprechende Konfiguration des Webservers wird der Browser, der den *HTTP Header* liest, instruiert, wie mit Skripten umzugehen ist. Hierzu muss serverseitig das Apache Modul *mod_header* installiert und aktiviert sein. Nicht jeder Browser kann die Sicherheitsanweisungen korrekt umsetzen. Die meisten Anwender sollten jedoch davon profitieren. Für ein sicheres Ausliefern von Webseiten sollte der Webserver folgende *Header* Informationen senden. Die notwendigen Einstellungen wurden daher entsprechend gesetzt:

Header Einstellung	**Erklärung**
Strict-Transport Security	Alle Webseiteninhalte werden ausschließlich via *HTTPS* übertragen.
X-Frame-Options 'sameorigin'	Skripten dürfen nur in Frames ausgeführt werden wenn sie auf demselben Server liegen.
X-XSS-Protection '1'	Aktiviert den browsereigenen *XSS*-Schutz.
X-Content-Type-Options 'nosniff'	Schützt Internetexplorer und Chrome vor *MIME Sniffing*
Content-Security-Policy	Bestimmte Inhalte dürfen nur von bestimmten Servern geladen werden.
Caching 'no-cache, no-store, must-revalidate'	Der Browser soll die Webinhalte immer vom Server direkt beziehen um zu verhindern, dass über den Browser-Zwischenspeicher Skripten oder Dateien ausgetauscht werden.

Tabelle 11: Apache Webserver Headereinstellungen

Insbesondere mit Hilfe der *Content-Security-Policy* (*CSP*) kann die Sicherheit deutlich erhöht werden. Der Vorteil von *CSP* besteht darin, dass individuell eingeschränkt werden kann, aus welchen Quellen Skripten, Bilder, *CSS Dateien*, Audio/Video Dateien, Frames oder eingebettete Objekte stammen dürfen. Zum Beispiel kann eine Regelung vorsehen, dass JavaScript

und eingebettete Objekte nur auf dem eigenen Server liegen dürfen - Video-Dateien und Stylesheets jedoch von einem ausgewählten Server geladen werden dürfen. [124]

Die Anpassung der beschriebenen *Header* Einstellungen kann in den Apache Sicherheitseinstellungen (*/etc/apache2/security*) vorgenommen werden.

Abbildung 74: Anpassung der Apache Sicherheitseinstellungen

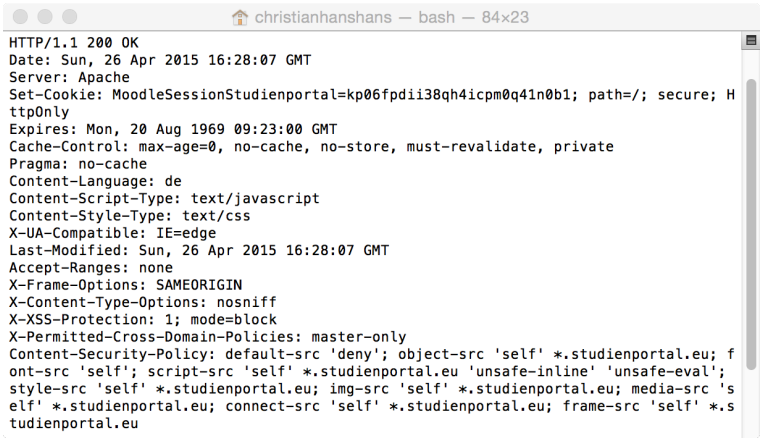


Abbildung 75: HTTP Header nach der Anpassung

Durch die getroffenen Maßnahmen konnten die von OWASP erkannten Schwachstellen behoben werden. Dies spiegelt sich auch in einem Sicherheitsaudit wieder. In einer erneuten Untersuchung konnte nur noch die als hoch eingestufte Schwachstelle gefunden werden. Hinter dieser verbirgt sich ein Fehler in der Suchfunktion, über die es möglich ist, ein nicht erlaubtes Sonderzeichen als Parameter zu übergeben, was prinzipiell einen *XSS* Angriff ermöglichen könnte. Da diese vermeidliche Eintrittspforte jedoch tief in Moodle integriert ist und auch bei manuellen Versuchen nicht ausgenutzt werden konnte, wurde im Rahmen dieser Arbeit nicht versucht dies zu korrigieren.

Risiko-Level	Risiken
Hoch	1
Mittel	0
Leicht	0
Hinweis	0

Tabelle 12: Ergebnis des Sicherheitsscans nach Sicherheitsoptimierung

7.2.3.1 SQL Injection

SQL Injects sind Angriffe auf das Datenbanksystem. Sie nutzen ähnlich wie bei *XSS* nicht gefilterte Benutzereingaben oder andere Schlupflöcher im Quellcode um am eigentlichen Programmcode vorbei mit der Datenbank zu kommunizieren. Dies kann dazu dienen, fremde Daten in die Datenbank einzuschleusen (z.B. Javascript) oder Daten aus der Datenbank zu extrahieren. Einige *SQL Injects* legen es auch darauf an, das Datenbankmanagementsystem gezielt zu überlasten oder zum Absturz zu bringen. Moodle gilt diesbezüglich als sehr sicher. *SQL Injects* konnten für Moodle weder in der Literatur noch in den durchgeführten Sicherheitsaudits gefunden werden. [89] [135]

7.2.4 Verschlüsselung des Transportwegs

7.2.4.1 SSL-Zertifikat

Für die sichere Datenübertragung zwischen Studienteilnehmern bzw. Studienbetreuern und Server wurde ein *SSL Zertifikat* mit einem 4096bit starken Schlüssel eingesetzt. Die verwendete Schlüsselstärke liegt deutlich über dem derzeitig (12/2014) üblichen 2048bit. Die Wahl von 4096bit wurde in Verbindung mit einer verhältnismäßig kurzen Gültigkeit von einem Jahr aus Sicherheitsüberlegungen gezielt gewählt. Nach Ablauf eines Jahres muss das Zertifikat neu ausgestellt werden. Häufiger Zertifikatswechsel (öffentlicher und privater Schlüssel) und lange Schlüssel gewähren die bestmögliche Sicherheit, erfordern aber auch einen entsprechenden administrativen Aufwand bei der Neuausstellung und Integration der Zertifikate. Es wurde darauf geachtet, dass die Zertifikatskette über die gleiche Schlüssellänge verfügt und dabei keine Schwachstelle entsteht. Aus Kostengründen wurde für den Prototyp ein Klasse 1 Zertifikat genutzt. Das Zertifikat deckt keine Subdomains ab, sondern nur:

https://studienportal.eu
https://www.studienportal.eu

Bei Klasse-1- Zertifikaten wird nur die Gültigkeit der Email-Adresse des Webmasters überprüft. Sie sind daher schnell und unkompliziert auszustellen und zudem kostengünstig.
Für ein Produktivsystem sollte der Einsatz von Klasse 2 oder besser noch Klasse 2 EV (= extended validation) Zertifikaten erwogen werden. Klasse 2 Zertifikate beinhalten zusätzliche Organisationsprüfungen und sind daher vertrauenswürdiger. Das Zertifikat enthält neben Domain und Emailadresse des Webmasters eine zusätzliche Überprüfung des Unternehmens oder der Organisation (z.B. Adresse, Telefonnummer, Handelsregisterauszug,...). Mit dem EV Zusatz wird dem Benutzer eine grüne Adresszeile im Browser angezeigt und der gesicherte Übertragungsweg sowie die Identität des Seitenbetreibers in der Adressleiste des Browsers präsentiert.

7.2.4.2 Verschlüsselten Seitenaufruf erzwingen (Rewrite Regel)

Um sicherzustellen, dass das Studienportal nur über eine sichere Verbindung aufgerufen wird, kann auf Ebene des Webservers eine sogenannte *Rewrite-Regel* definiert werden. Seitenaufrufe, die via *http://...* erfolgen (also unverschlüsselt auf Port 80), können hierdurch automatisch auf *https://...* weitergeleitet werden (also verschlüsselt auf Port 443). Für den Seitenbesucher geschieht diese Umleitung unbemerkt im Hintergrund. Sie stellt jedoch sicher, dass das Pass-

wort oder andere Informationen nicht ungesichert übertragen werden. Die Umleitung wird in der Konfiguration des Webservers oder in einer *.htaccess-Datei* definiert:

RewriteRule ^(.)$ https://%{HTTP_HOST}$1 [redirect=301]*

Ruft ein Teilnehmer das Studienportal über *http://studienportal.eu/study/view.php?id=2* auf, wird die zuvor eingegebene URL (inklusive aller Parameter) automatisch in *https://studienportal.eu/study/view.php?id=2* überführt. Suchmaschinen und Browsern wird dabei die Zusatzinformation übergeben, dass die via http aufgerufene Seite dauerhaft auf *https* (redirect=301) verschoben wurde und unter *http://* nicht erreichbar ist. Dies führt dazu, dass die URLs von Suchmaschinen-Treffern in Zukunft als *https* statt *http* angeboten werden.

7.2.4.3 Aufrechterhalten der Transportverschlüsselung (HSTS)

Ist ein Benutzer bereits über eine verschlüsselte Verbindung am Studienportal angemeldet besteht noch immer die Gefahr, dass während der aktuellen Sitzung von einer verschlüsselten Verbindung auf eine unverschlüsselte umgeschaltet wird. Dies wiederum könnte von potentiellen Angreifern genutzt werden, um übertragene Daten mitzulesen oder zu verändern. Um dies zu verhindern, kann ***H**TTP **S**trict **T**ransport **S**ecurity* (*HSTS*) eingesetzt werden. [57] Hierbei wird dem Browser vom Server über eine Zusatzinformation (im *http-Header*) signalisiert, dass eine Webseite nur verschlüsselt abrufbar sein soll. Der Browser speichert diese Information für einen festgelegten Zeitraum und überträgt innerhalb dieser Zeit keine unverschlüsselten Daten von oder zu dieser Webseite. [49] *HSTS* muss hierzu vom Webserver unterstützt und explizit aktiviert werden.

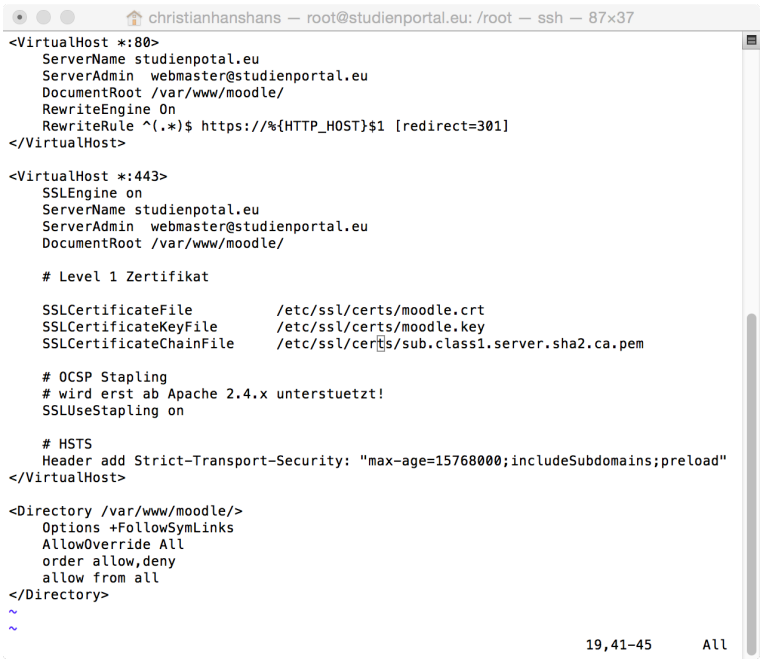


Abbildung 76: Erzwingen verschlüsselter Verbindungen beim Apache Webserver

HSTS ist noch relativ unbekannt und wird laut einer Studie aus 2012 nur von 0,2% aller deutschen Webseiten genutzt. Selbst bei sicherheitsrelevanten Anwendungen wie Homebanking findet es nur selten Anwendung. [108] Mit Ausnahme von SosciSurvey setzt keiner der kommerziellen Fragebogen-Software-Anbieter *HSTS* ein. (Stand 12/2014)

Für den Seitenbesucher gibt es derzeit leider kein Merkmal, an dem man die sichere *HSTS* Übertragung erkennen könnte. Um eine Webseite auf *HSTS* zu überprüfen gibt es kostenfreie Online-Tests (z.B. *www.ssllabs.com/ssltest/*). Alternativ können die *HTTP Header* mit dem Konsolenprogramm *curl* abgerufen werden, welche unter Linux und Apple Betriebssystemen standardmäßig installiert ist.

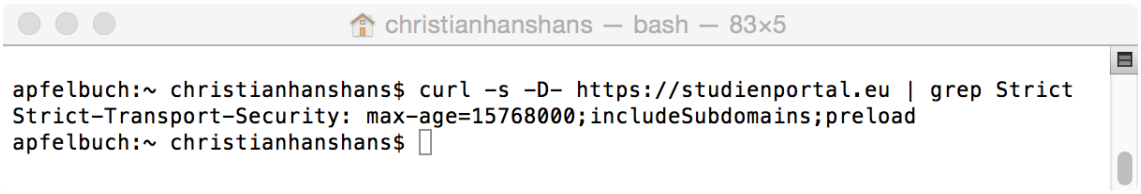


Abbildung 77: Abfrage von HSTS Header via Linux oder Mac Terminal

HSTS ist auf die Zusammenarbeit der Browser angewiesen. Nicht jeder Browser kann die übertragenen Header-Informationen auch sinnvoll verwerten. Eine Unterstützung von strikter Transportsicherheit ist jedoch mit Ausnahme des Microsoft Internetexplorer mit allen aktuellen Browsern und Smartphones oder Tablets gewährleistet. Mit *HSTS* Unterstützung ist laut Microsoft im Internetexplorer 12 (Windows 10) zu rechnen. Eine Liste unterstützter Browser ist in der folgenden Tabelle zusammengefasst:

Browser	**Transportsicherheit ab Version**
Internetexplorer IE	-
Firefox	31
Chrome	31
Safari	7 (OSX 10.9)
Opera	26
Android	4.4
iOS	8

Abbildung 78: HSTS fähige Browser Stand 12/2014, caniuse.com [27]

Google pflegt für Chrome eine Liste vertrauenswürdiger Domains, welche die *HSTS* Anforderungen erfüllen und integriert sie fest (daher *Preloaded HSTS*) in den Quellcode des Browsers. [190] Mit der Installation des Browsers ist der Seitenbesucher (sofern die Webseite gelistet ist) sofort vor Missbrauch geschützt. Auch andere Browser wie Firefox oder Safari bedienen sich dieser Liste. Ein Antrag auf Aufnahme in die *Preloaded HSTS* Liste kann unter *https://hstspreload.appspot.com* gestellt werden. Nach erfolgreicher Überprüfung wird die Webseite in die nächste Version des Browsers integriert. Für das Studienportal wurde *HSTS* mit einer Gültigkeitsdauer von einem halben Jahr implementiert. Jede Subdomain wurde in die strikte *SSL* Übertragung einbezogen und die Domain *studienportal.eu* erfolgreich in die *HSTS-Preload* Liste eingetragen.

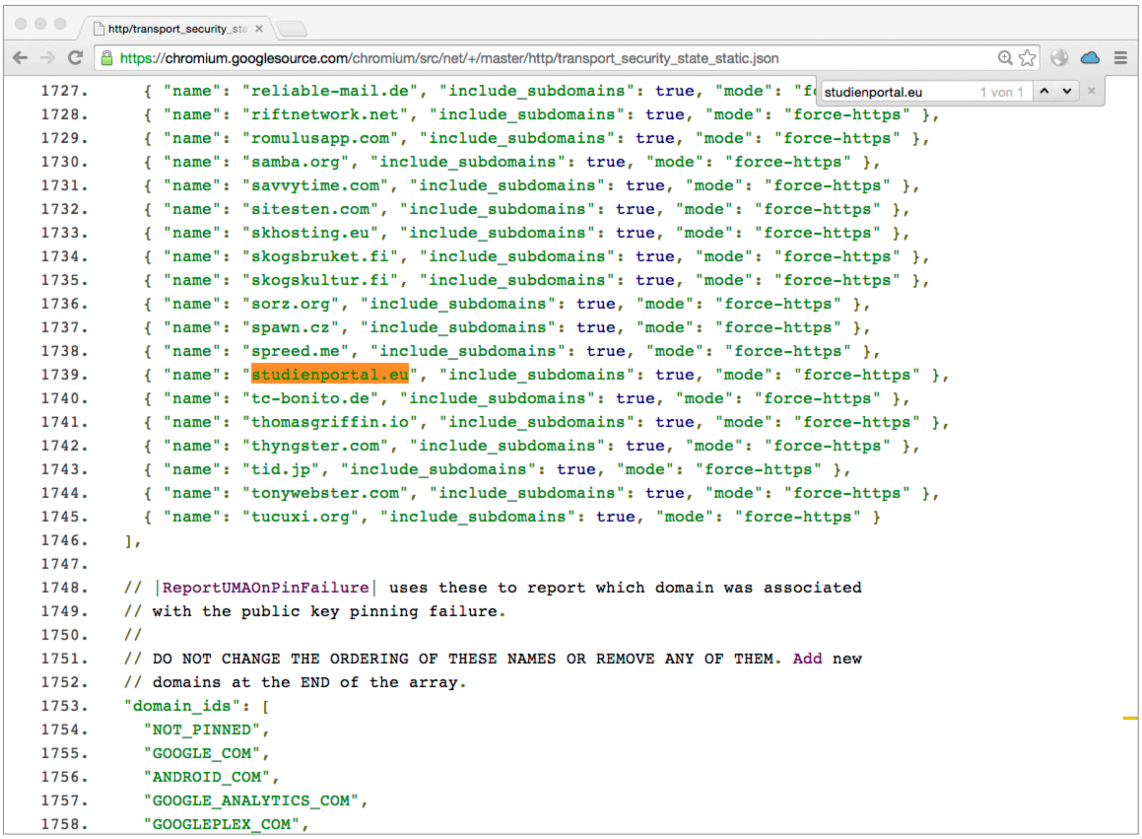


Abbildung 79: Studienportal.eu ist fest im Quellcode von Google Chrome integriert

Die Umsetzung von *HSTS* ist sehr elegant und mit geringem Aufwand möglich. Sie birgt jedoch die Gefahr, dass die Webseite nicht mehr erreichbar ist, sollte das *SSL Zertifikat* seine Gültigkeit verlieren. Die Gültigkeit des Zertifikats muss daher unbedingt gewährleistet sein und die Neuausstellung von Zertifikaten vor Ablauf der Gültigkeit erfolgen, um ungeplante Ausfallzeiten zu verhindern. Alle vom Webserver ausgelieferten Daten müssen vollständig via *SSL* erreichbar sein. Durch die Aktivierung von *HSTS* werden über *http://* eingebundene Ressourcen wie Links, Bilder oder Videos nicht mehr angezeigt. Dies gilt auch für die Subdomains. Sollen diese auf dem Server betrieben werden (z.B. *https://sure.studienportal.eu),* müssen auch sie vollständig via *https://* erreichbar sein und über ein gültiges *SSL Zertifikat* verfügen.

Berücksichtigt man bisher beschriebene Maßnahmen und Servereinstellungen, kann bereits ein hohes Maß an Sicherheit für die Datenübertragung zwischen Teilnehmern und Studienportal erreicht werden. Führt man die Sicherheitsanalyse der Plattform analog der Marktanalyse (☞ Kapitel 4.12) durch, erzielt das Studienportal folgende Ergebnisse:

Werkzeug	Zertifikat	PFS	HSTS	Cookie Sicherheit	Rating	Bemerkung
Studienportal.eu	+++	✓	✓	✓	A+	sichere Verschlüsselung*

Tabelle 13: Sicherheitsanalyse des Studienportals analog der Marktanalyse, Stand 12/2014

Qualys Rating = standardisierte Bewertung der sicheren Datenübertragung [153]
* gemäß der technischen Richtlinie des Bundesamt für Sicherheit in der Informationstechnik (BSI) [18]

7.2.5 DNSSec

DNS steht für ***D**omain **N**ame **S**ervice* und ist der Dienst, der zwischen IP-Adressen und Domainnamen vermittelt. *DNS* Anfragen werden üblicherweise unverschlüsselt übertragen und können daher manipuliert werden. [122] Den Ablauf einer normalen *DNS* Anfrage zeigt die folgende Schemazeichnung.

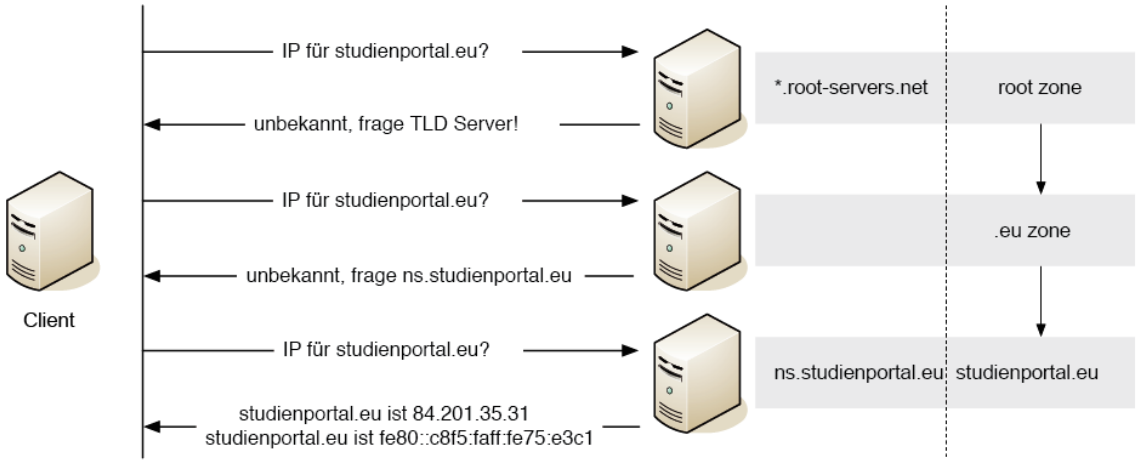


Abbildung 80: Ablauf einer DNS-Anfrage, Bildvorlage Rijswijk [24]

Der Anfragende hat keine Möglichkeit zu überprüfen, ob die angefragte Webseite tatsächlich unter der gelieferten IP-Adresse erreichbar ist. Auf diese Art können *Man-in-the-Middle* Angriffe erfolgreich durchgeführt werden. Hierbei wird entweder ein *DNS Cache* (*Cache Poisoning*) auf dem Weg zum Client (z.B. Router) mit falschen *DNS-Informationen* gefüttert, oder eine gefälschte Antwort an den Client geschickt (*DNS Spoofing*). Als prominentes Beispiel dient das Zensurinstrumentarium Chinas. Es bedient sich unter anderem der Manipulation von *DNS-Informationen*, um die Internetzensur durchzusetzen. [81] Auf diese Weise können Webseiten blockiert oder Informationen umgeleitet werden. Verfahren wie *DNS Spoofing* oder *DNS Hijacking* sind in jedem Netz möglich (z.B. Uni oder Firmennetz, WiFi Hotspots, UMTS/LTE, Heim-DSL) und stellen ein nicht zu unterschätzendes Risiko dar.

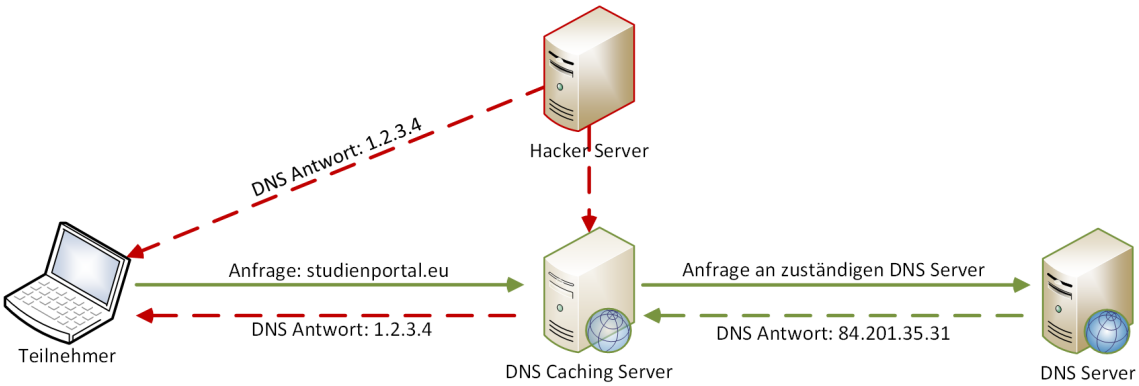


Abbildung 81: Man-in-the-Middle Angriffe via DNS-Manipulation

DNSSec schließt diese Lücke in dem die Adresse des Servers und die Echtheit der *DNS-Informationen* durch eine Signatur überprüfbar wird. Der Transportweg der übertragenen *DNS* Daten zwischen Client und Server wird durch Verschlüsselung und die digitale Unterschrift vor unbemerkter Manipulation geschützt. Der Client hat dadurch die Gewissheit, dass

er mit dem richtigen Server verbunden ist und die *DNS-Informationen* unverfälscht übertragen wurden. Hierzu ist auf der Client Seite ein *DNSSec*-fähiger *Resolver* notwendig. Ein *Resolver* kann an einer zentralen Stelle im lokalen Netzwerk der Universität oder Firma bereitgestellt werden und alle angeschlossenen Endgeräte schützen. Aber auch ein Browser oder das Betriebssystem eines einzelnen Computers kann einen *Resolver* enthalten. [130] Mit dem Ausbau von *DNSSec* in öffentlichen und lokalen Netzwerken wird in Zukunft die Anzahl von geschützten Clients wachsen. Internetdienstanbieter könnten an zentraler Stelle für das Filtern ungültiger *DNS-Informationen* sorgen. Entsprechende Umsetzungsmöglichkeiten sind bereits seit einiger Zeit veröffentlicht und von einigen Anbietern erfolgreich im Einsatz. [24] Für heimische Windows, Mac- und Linux-Computer gibt es einen quelloffenen *Resolver*, mit dem die fehlende *DNSSec* Unterstützung des Betriebssystems nachgerüstet werden kann. [191] Alternativ können alle gängigen Browser mit einem Plugin (*DNSSec-Validator)* erweitert werden. Das Plugin signalisiert die verschlüsselte Übertragung von *DNS* Daten (Schlüsselsymbol), sowie die Echtheit des *SSL Zertifikats* (Schlosssymbol). [192]

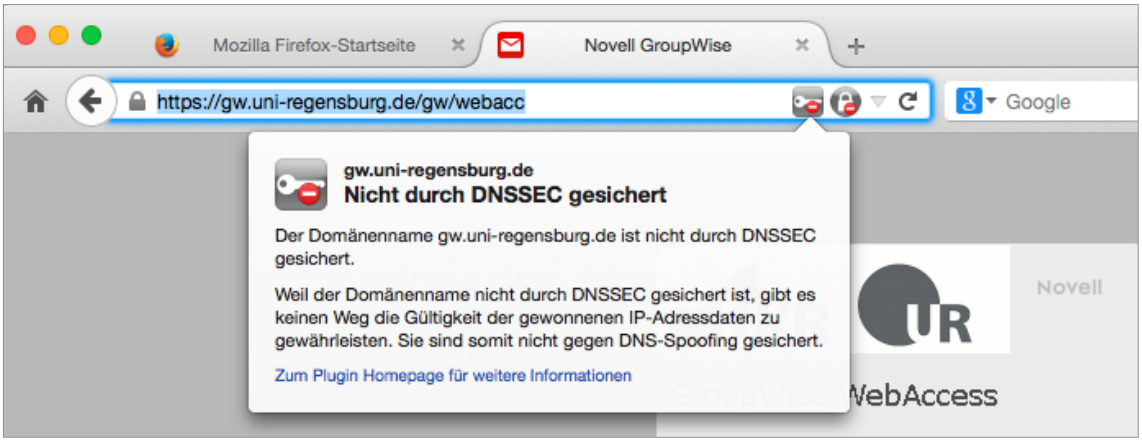


Abbildung 82: Browser Plugin überprüft DNSSec und TLSA

Einer Studie der Universität Duisburg zu Folge, überprüfen derzeit nur 3,8% der (deutschen) untersuchten Clients aus dem IT-Umfeld die *DNS-Informationen* auf Gültigkeit. [122] Obwohl nur via *DNSSec* die Vertrauenswürdigkeit von *DNS-Informationen* sichergestellt werden kann, findet sich serverseitige Unterstützung nur sehr selten in der Praxis. In einer aktuellen Stichprobe Ende 2014 (n=500) verfügten nur 0,6% der untersuchten Webseiten über die verschlüsselte *DNS* Übertragung. [193] Ein Grund für die geringe Verbreitung ist vermutlich der Tatsache geschuldet, dass es derzeit in Deutschland kaum Domainregistrare gibt, die ohne zusätzlichen Aufwand *DNSSec* unterstützen.(Stand 12/2014) Stattdessen muss derzeit entweder auf kostenpflichtige Dienstleister (*managed DNS Server*) zurückgegriffen oder ein eigener Namensserver betrieben werden. [130] Dennoch ist *DNSSec* das Mittel der Wahl, da es die Grundlage einer vollständigen Transportverschlüsselung bildet und entsprechend der Forderung des Bundesdatenschutzgesetzes dem aktuellen Stand der Technik entspricht.

Für das Studienportal wurde daher eine eigene *DNS* Infrastruktur bestehend aus zwei Namensservern (*Primary/Secondary*) aufgebaut. Für die Domain *studienportal.eu* wurde zunächst ein Schlüsselpaar analog zum *public-key Verfahren* erzeugt. Der öffentliche Anteil des Schlüssels wird allen *DNS Server* bekannt gegeben. Mit Hilfe des privaten Schlüssels, der Gültigkeitsdauer (*TTL*) und dem Domainnamen wird für jeden *DNS-Eintrag* eine Signatur (*RRSIG*) erstellt und auf dem für das Studienportal zuständigen *DNS Server* gespeichert. Ruft ein Teilnehmer nun *https://studienportal.eu* auf, wird eine Anfrage an die oberste Hierarchieebene des *DNS Systems* gestellt (*root Zone*). Von dort werden solange Anfragen an die

tieferen Ebenen gestellt, bis der zuständige Namensserver gefunden wurde, der die Anfrage beantworten kann. Der zuständige *DNS-Server* wiederum muss allen übergeordneten Namensservern bekannt sein und einen *DS-Record* (= *Hash* des öffentlichen Schlüssels) an alle übergeordnete *DNS-Server* bekannt gegeben haben, damit die Vertrauenskette geschlossen ist.

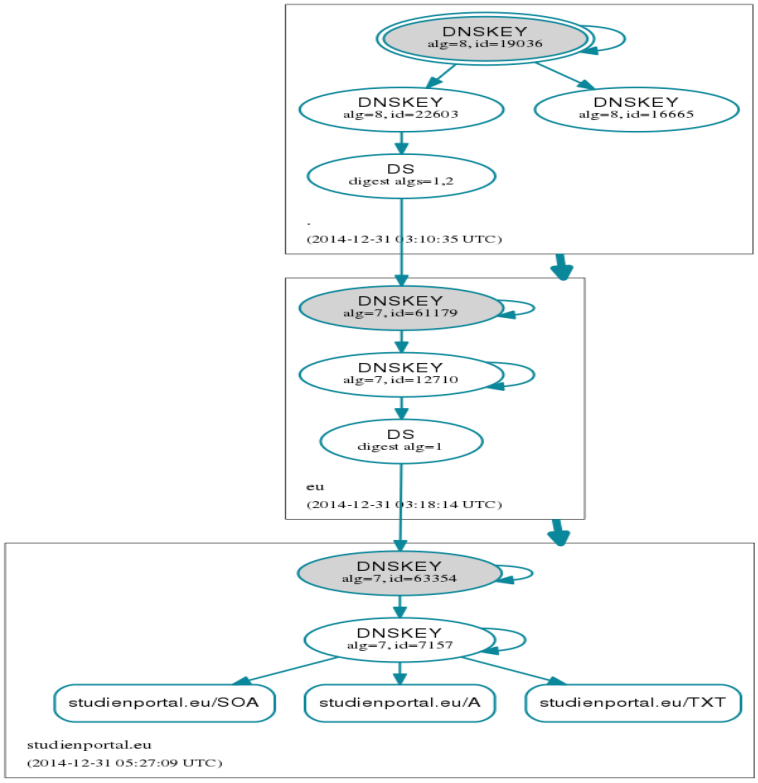


Abbildung 83: Überprüfung und grafische Darstellung der DNS-Kette mit dnsviz.net

Der Flexibilität und Kostenüberlegung geschuldet, befinden sich die beiden Namensserver im gleichen Rechenzentrum und im gleichen Netzsegment. Aus Gründen der Redundanz und Performanz sowie zum Schutz vor *DoS* Angriffen sollten für den Produktivbetrieb mehrere *DNS-Server* an unterschiedlichen Serverstandorten zum Einsatz kommen. Die Implementierung von *DNSSec* bleibt unabhängig von Software-Updates des Betriebssystems, des Webservers und von Moodle bestehen.

7.2.6 DANE/TLSA

SSL Zertifikate dienen als elektronische Unterschrift, die dem Anwender die Echtheit der ausgelieferten Daten bestätigen soll. Das *public-key* -Verfahren basiert auf einer Vertrauenskette von der Zertifizierungsstelle (*CA*) über Zwischeninstanzen (*Intermediate CA*) bis zum tatsächlichen Serverzertifikat. Hierbei wird den öffentlichen *CAs* und deren Zwischeninstanzen automatisch vertraut. Der Serverbetreiber hat keine Möglichkeit einzuschränken, welche Zertifizierungsstelle Zertifikate für die eigene Webseite ausstellen darf. Seitenbesucher wiederum können kaum überprüfen, ob das angezeigte Zertifikat tatsächlich vom Serverbetreiber ausgestellt oder von Dritten ausgetauscht wurde. Wird eine *CA*, wie in der Vergangenheit bereits häufiger geschehen, unterwandert, stellt dies ein großes Sicherheitsproblem dar, da die Browser den ausgelieferten Zertifikaten blind vertrauen. [194] [195]
Folgendes Beispiel zeigt einen *Man-in-the-Middle* Angriff. Hierzu wurde im lokalen Netzwerk über einen Proxy Server der Zugang zum Internet hergestellt. Dieser liefert für ver-

schlüsselte Verbindungen ein gültiges *SSL Zertifikat*. Der Browser signalisiert eine sichere Verbindung (Schlosssymbol), obwohl der gesamte Datenverkehr mitgelesen werden kann. Der Aussteller des Zertifikats für die besuchte Webseite ist nicht korrekterweise die Universität Regensburg (*Uni Regensburg CA*), sondern in unserem Beispiel eine fremde Zertifizierungsinstanz. Ein Verdacht könnte angesichts der unterschiedlichen Zertifizierungsinstanzen aufkommen. Allerdings müsste man sich hierzu die Zertifikateigenschaften mit Hilfe des Browsers anzeigen lassen. Wird ein Zertifikat, das ursprünglich von einer Zertifizierungsinstanz z.B. von Verisign ausgestellt wurde, durch ein anderes gültiges Verisign-Zertifikat ersetzt, ist die Fälschung kaum noch zu erkennen. Im beschriebenen Beispiel konnten mit Hilfe des *Man-in-the-middle* Angriffs die Zugangsdaten zur Lernplattform der Uni Regensburg (rote Markierung) unbemerkt ausgelesen werden.

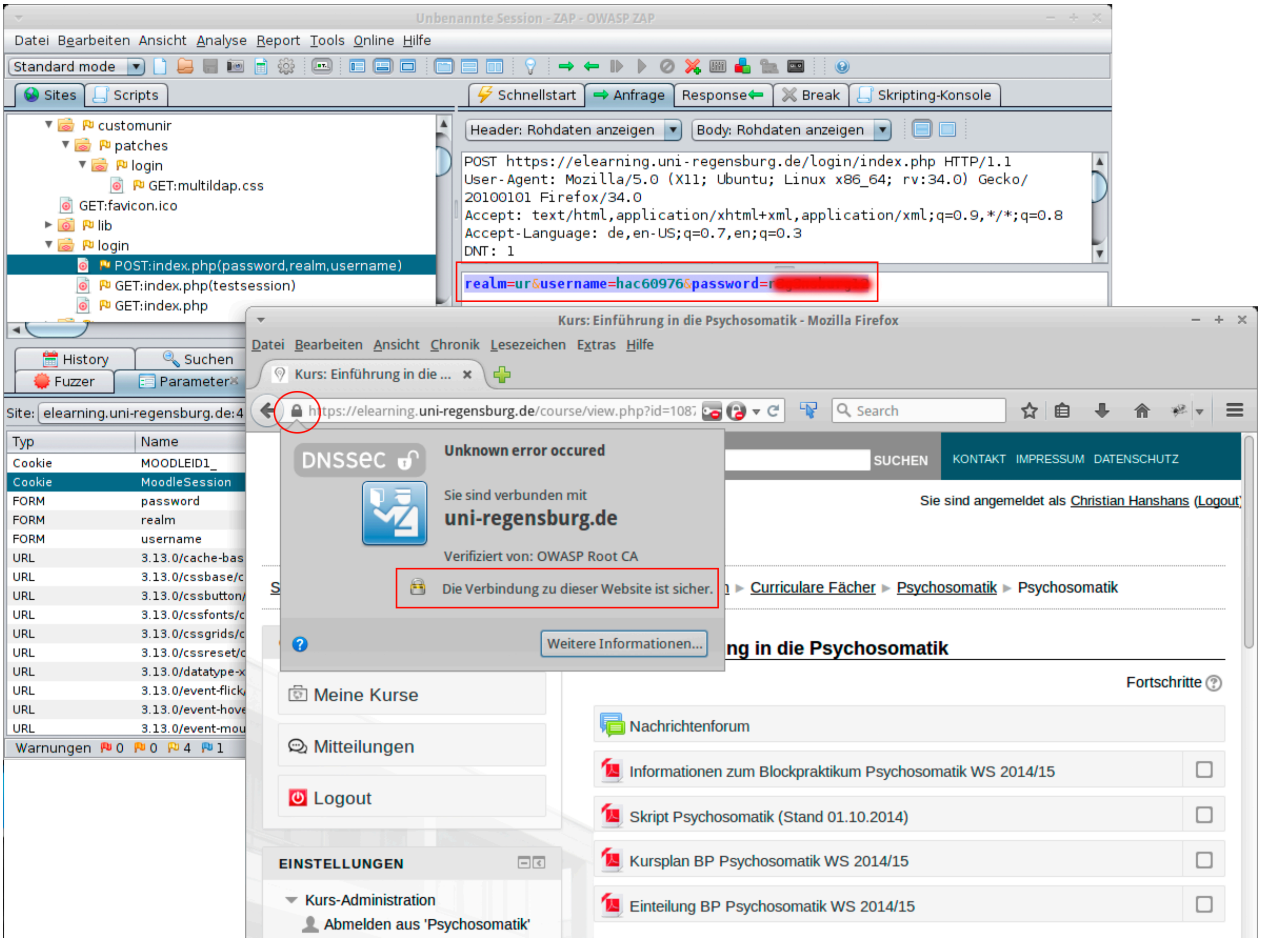


Abbildung 84: Erbeuten der Zugangsdaten durch Man-in-the-Middle Angriff

Dass dies kein rein akademisches Szenario ist, zeigt die Entdeckung einer Google Mitarbeiterin. Ihr fiel auf einem Linienflug auf, dass der verschlüsselte Datenverkehr zu Google Webseiten umgeleitet und dem Browser ein gefälschtes Zertifikat präsentiert wurde. Auf diese Weise konnte ein Anbieter, der etliche amerikanische Fluglinien mit dieser Technologie ausstattet, Anfragen an Videoportale wie YouTube ausbremsen, häufig verwendeten Inhalt zwischenspeichern und dadurch übertragene Daten (inklusive Zugangsdaten) mitlesen. [196] Wie die beiden Beispiele zeigen, ist eine Gefahr in öffentlichen (Hotspots) wie privaten Netzen (z.B. DSL-Router oder Firmennetze) durchaus gegeben.

Mit Hilfe von *DANE* (=*DNS-based Authentication of Named Entities*) kann vom Webseitenbetreiber bestimmt werden welche Zertifizierungsinstanz die Domain Zertifikate ausstellen darf. Alternativ kann auch die Nutzung eines bestimmten Zertifikats erzwungen werden. Dieses Zertifikat kann von einer externen CA aber auch selbst signiert sein. Hierzu wird über *DNS-Einträge* (*TLSA-records*) definiert, welchem Zertifikataussteller oder welchem konkreten Zertifikat vertraut werden darf. *DANE* kennt vier Betriebsmodi:

0. PKIX-TA: Bindung an eine (oder mehrere) bestimmte CA(s)
1. PKIX-EE: Überprüfung der Gültigkeit gemäß bestehender Zertifikatskette
2. DANE-TA: Betreiben einer eigenen CA
3. DANE-EE: Bindung an ein Domain-Zertifikat (selbst oder fremdsigniert)

Abbildung 85: Aufbau eines TLSA Records für DANE, AFNIC [9]

Für den geplanten Schutz des bereits bestehenden *SSL Zertifikats* ist Modus 3, also Beschränkung auf dieses Zertifikat sinnvoll. Der *Selector* 0 definiert die Einbeziehung des gesamten Zertifikats. Der *Matching type* gibt Auskunft über die für die Kontrolle des Zertifikats verwendete *Hashfunktion*. Hierbei steht der verwendete Wert 1 für *SHA-256*.
Der *TLSA-Record* lautet demnach:

_443._tcp.studienportal.eu. IN TLSA 3 0 1
d635e54002fb0095891461f3b28f3c4f16f3496f8ab1d697f6c2754c41b87b9c

Besitzt eine Webseite *DANE* Unterstützung, kann mit einem entsprechenden *Resolver und TLSA Validator* die Authentizität des übertragenen Zertifikats überprüft werden.

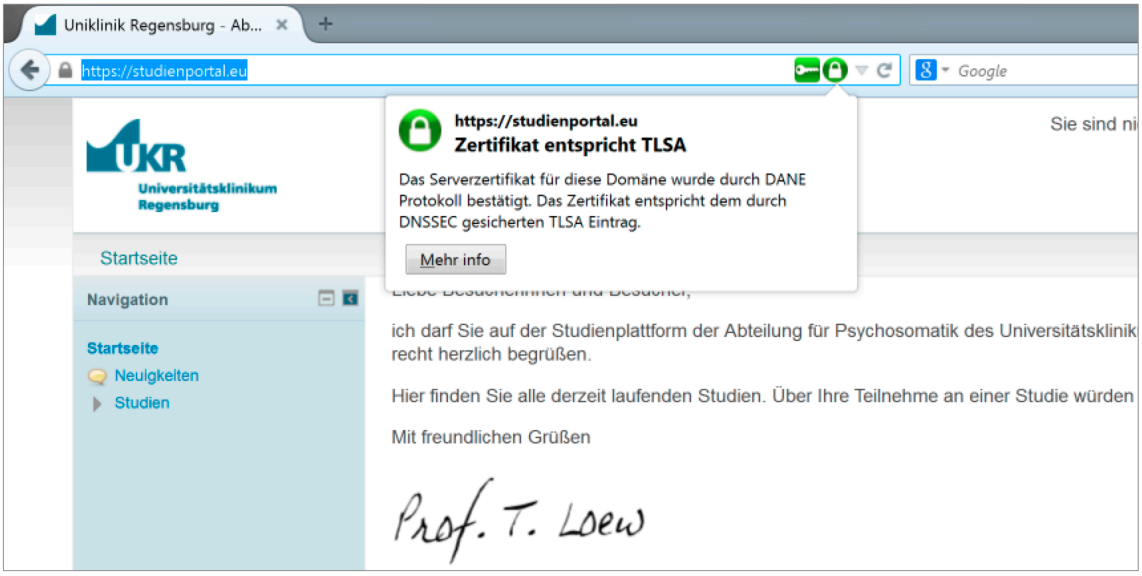


Abbildung 86: Echtheit des SSL-Zertifikats wird durch TLSA/DANE bestätigt

Versucht nun jemand mit einem *Man-in-the-Middle* Angriff wie im vorherigen Beispiel dem Browser ein falsches Zertifikat unterzuschieben kann dies sofort erkannt werden.

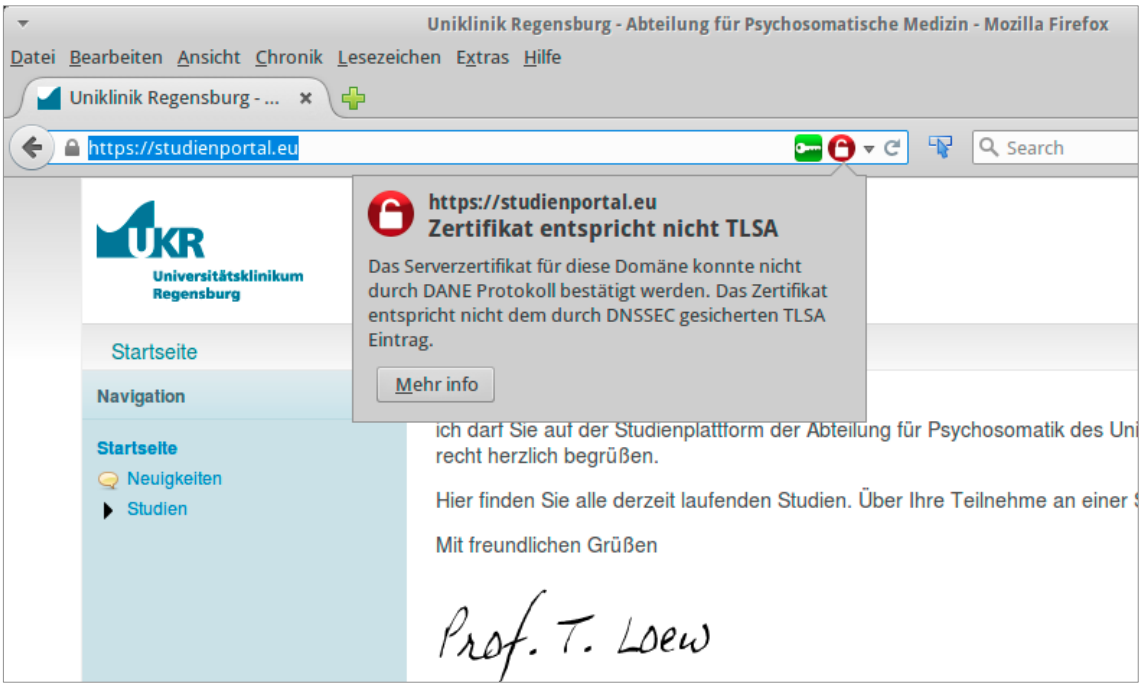


Abbildung 87: Gefälschte Zertifikate fliegen dank TLSA auf

Je nach Integration von *TLSA* auf der Seite des Anwenders kann mit Anfragen an Webseiten mit gefälschten Zertifikaten unterschiedlich umgegangen werden. Der Aufruf von Webseiten kann verhindert oder auch nur mit einer Warnmeldung oder einem Symbol versehen werden.

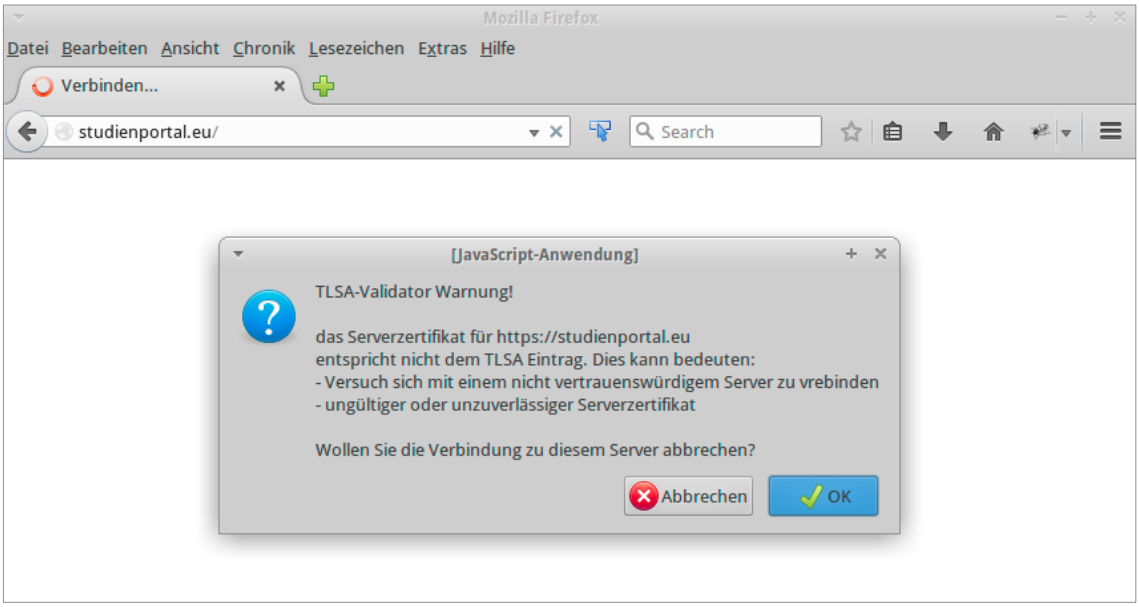


Abbildung 88: Warnmeldung vor manipuliertem Zertifikat

Die Teilnehmer sollten darauf hingewiesen werden, dass das Studienportal die Möglichkeit bereitstellt zu überprüfen, ob die Verbindung zum Studienportal tatsächlich unverfälscht und sicher ist. Dies kann dazu beitragen, dass sich Vorbehalte gegenüber Onlinestudien reduzieren. Die Integration von *DNSSec* und *TLSA* in Mozilla Firefox ist besonders unkompliziert und

auch für Laien sehr gut geeignet. Firefox ist für alle Betriebssysteme verfügbar und hat unter den Browsern die größte Verbreitung. [197] Den Teilnehmern könnte z.B. in einer kurzen Videoanleitung erklärt werden, wie sie das Plugin in den Browser integrieren können.

7.2.7 Intrusion detection & Web-Application Firewall

Um den Webserver vor Angriffen zu schützen, können *Intrusion Detection Systeme* eingesetzt werden. Diese Systeme analysieren üblicherweise Protokolle (Log-Files), die jeder Dienst erzeugt. Ein unter Linux weit verbreitetes System ist fail2ban. Die in Phyton geschriebene Software basiert auf einem individualisierbaren Regelwerk. Mit Filtern können verdächtige Muster gesucht und dadurch Angriffe erkannt werden. Für jeden Dienst kann beschrieben werden, welche konkrete Maßnahme im jeweiligen Fall ergriffen werden soll. Üblicherweise werden IP-Adressen von Angreifern über die Firewall des Betriebssystems (*iptables*) blockiert. [22] Für den Angreifer ist der Server dann nicht mehr erreichbar.
Für das Studienportal wurden einige Filter benutzt, um den Apache Webserver vor Angriffen zu schützen. Meist scannen Angreifer ihre Ziele vor einem Angriff. Mit den gesammelten Informationen können Rückschlüsse auf die eingesetzte Webanwendung gewonnen und diese gezielter angegriffen werden. Beliebt ist zum Beispiel die Suche nach der Datenbankoberfläche phpMyAdmin, da diese direkten Zugang zur Datenbank verspricht. Analysiert man die Log-Dateien des Webservers, finden sich wiederholt Einträge, die beweisen, dass gezielt nach gewissen Webanwendungen gesucht wurde. Man kann davon ausgehen, dass sich hinter IP-Adressen, die (erfolglos) nach unterschiedlichen Verzeichnissen oder Dateien suchen, Angreifer verbergen. Mit Hilfe entsprechender Aktionen können gezielt solche IP-Adressen blockiert und weitere Angriffe verhindert werden. Auch Versuche, Passwörter, sei es für Moodle oder für den Fernwartungsanschluss zu erraten, können erkannt und unterbunden werden. Die Sperrzeit von IP-Adressen ist individuell für jeden Dienst einstellbar. Fail2ban ist ein komplexes System, mit dessen Hilfe sich grobe Einbruchsversuche gut verhindern lassen. Es ist jedoch keine Präventivmaßnahme und kann nur auf bereits protokollierte Ereignisse reagieren. Es benötigt für die Analyse Rechenleistung und Zugriff auf die Festplatte und hat dadurch direkten Einfluss auf die Leistungsfähigkeit des Servers. Große Log-Dateien, die zum Beispiel bei stark frequentierten Webseiten oder bei *DoS* Angriffen entstehen, können die Ressourcen des Servers stark beanspruchen und ihn ggf. überlasten. Für das Studienportal wurden folgende Filter verwendet:

Filter	Funktion
apache-nohome	Suche nach Installationsverzeichnissen erkennen
apache-BadBots	Erkennen von unerwünschten Suchmaschinen/Bots
apache-auth	Erkennen von Passwort Bruteforce Angriffen
apache-noscript	Erkennen von Suche nach Skripten
apache-overflows	Erkennen von Versuchen den Webserver zu überlasten
rfi-attack	Einschleusen von Schadcode via PHP erkennen
sqli-attack	Erkennen von SQL Injection Versuchen
ssh	Erkennen von Einbruchsversuchen über den Fernwartungszugang
ssh-ddos	Erkennen von Versuchen, den Server mit vielen gleichzeitigen Fernwartungsanfragen auszulasten

Tabelle 14: verwendete Filter

Web Application Firewalls (*WAF*) tragen ebenfalls zur Sicherheit der Webanwendung bei. Im Gegensatz zu Intrusion Detection Systemen schützen sie die Webanwendung indem sie die eingehenden Browser-Anfragen nach Auffälligkeiten untersuchen. Für den Apache Webserver existiert die Erweiterung *mod security*, die eingehende Anfragen an den Webserver genauer unter die Lupe nimmt. [17] Systeme wie *mod_security* sind besonders hilfreich, da sie potentiell schädliche Anfragen wie *XSS* Angriffe erkennen und unterbinden können, bevor sie die Webanwendung erreichen. [26] *WAF* Systeme haben demnach präventiven Charakter. Damit *WAFs* ihr volles Potential entfalten können, müssen sie möglichst viele Angriffsmuster kennen. Um dies zu gewährleisten, können Datenbanken mit bekannten Mustern angezapft werden. Ähnlich wie bei Virenscannern für heimische PCs gibt es kostenpflichtige und freie Quellen. Sicherheitsunternehmen investieren viel Zeit in das Analysieren von Sicherheitslücken. Ihre Datenbanken sind daher in der Regel umfangreicher und kostenpflichtig. Eine kostenfreie und dennoch umfangreiche Quelle mit Basisregeln stellt das Open-Web-Application-Security-Project (OWASP) zur Verfügung.
Die Grundregeln umfassen potentielle *XSS* Angriffe ebenso wie *Webcrawler* und *SQL Injections*. Im Gegensatz zu kommerziellen Regelsätzen sind sie jedoch nicht tagesaktuell und demnach nicht an aktuelle Angriffsszenarien angepasst. Zum Zeitpunkt der Implementierung waren die Regeln 9 Monate bis 2 Jahre alt. [10] Für den Prototypen des Studienportals wurde primär aus Kostengründen der Regelsatz von OWASP verwendet. Das Produktivsystem sollte jedoch durch die aktuelleren kommerziellen Quellen erweitert werden. Anbieter wie Atomicorp (99$) oder SpiderLabs (495$) bieten Angebote mit Echtzeit-Aktualisierung an und untersuchen unter anderem Dateien während des Uploads auf Viren oder weisen Anfragen bekannter Angreifer direkt ab. Sobald eine neue Gefährdung bekannt wird, aktualisiert sich das Regelwerk automatisch.

7.2.8 Schutz durch Faktor-2-Authentifizierung

Besonders gefährdet gegenüber Angriffen sind Fernwartungsanschlüsse der Server. Da Server für Wartungsaufgaben permanent zugänglich sein müssen, sind sie besonders exponiert und zugleich für Hacker besonders attraktiv, da sie direkten Zugang zu Daten versprechen. Abhängig von der eingesetzten Rechenleistung und Zeit kann jedes noch so sichere Passwort geknackt werden. Daher ist es besonders wichtig sensible Bereiche vor *Bruteforce* Angriffen zu schützen. Linux Server werden über *SSH* (*secure socket shell*) administriert. Über eine verschlüsselte Verbindung kann die textbasierte Verwaltung des Servers erreicht werden. Es gibt viele Möglichkeiten den Zugang zum Server einzuschränken.
Einige Administratoren favorisieren eine VPN -Einwahl zu einem internen Wartungsnetz. Dies ist prinzipiell eine gute Idee, dennoch besteht zum einen weiterhin ein Risiko für Angriffe innerhalb des eigenen Netzes, zum anderen muss für eine verlässliche Einwahlmöglichkeit gesorgt sein. Ist der VPN-Server nicht erreichbar, ist eine Administration der Server nicht mehr möglich. Von einer VPN -Lösung wurde abgesehen um keine redundante VPN- Umgebung schaffen zu müssen. Andere Systemverwalter setzen auf die Authentifizierung via *SSH Key*. Hierbei wird auf dem eigenen Computer nach *public key -Verfahren* ein Schlüsselpaar erzeugt. Dieser öffentliche Schlüssel wird auf dem Server gespeichert. Verbindungen zum Fernwartungssystem werden nur aufgebaut, wenn der Client über den zum öffentlichen Schlüssel passenden privaten Schlüssel verfügt. Der *SSH Dienst* ist dadurch ohne Login nur noch von dem Rechner mit dem passenden *SSH* Key erreichbar. Der Vorteil ist, dass *Bruteforce* Angriffe ins Leere laufen, da kein klassisches Login mit Benutzername und Passwort mehr stattfindet. Problematisch kann dafür ein Defekt oder Diebstahl des autorisierten Rechners sein. Ist der Schlüssel z.B. durch den Crash des Betriebssystems verloren, kann kein ad-

ministrativer Zugriff auf den Server mehr erfolgen. Auch sind Wartungsarbeiten von anderen Rechnern aus nicht ohne weiteres möglich. Es müsste dazu der *SSH Schlüssel* (z.B. via USB Stick, Netzwerk oder Email) auf den anderen Rechner übertragen werden, was wiederum ein Risiko darstellt.
Anstelle der beiden beschriebenen Verfahren wurde daher eine andere Methode für alle verwendeten Server implementiert. Wie beim Pin/Tan Verfahren des Homebankings wird zum klassischen Login ein weiterer Faktor hinzugenommen. Der zweite Faktor besteht aus einem Einwegpasswort (*OTP*), das ähnlich einer Tan-Nummer nur einmal verwendet werden kann. Dieses Einwegpasswort entspricht einer 32 stelligen hexadezimalen Tan. Für 128bit müssten über $3{,}4 \times 10^{38}$ Kombinationen ausprobiert werden um die Zeichenfolge erfolgreich zu erraten. Innerhalb der kurzen Zeitspanne zwischen Eingabe der Login-Daten und dem Einwegpasswort wäre selbst für Supercomputer nicht genügend Zeit die Kombinationen durchzuspielen. Das *One-Time-Password* (*OTP*) wird von einem kleinen USB-Stick auf Knopfdruck erzeugt. Vergleichbar mit einem Hausschlüssel kann er am Schlüsselbund getragen und von unterschiedlichen Personen an unterschiedlichen Computern verwendet werden. Der Vorteil liegt auf der Hand. Selbst wenn die Zugangsdaten des Administrators in falsche Hände geraten, kann ohne den elektronischen Schlüssel kein Zugriff auf die Server erfolgen. Durch den zweiten Faktor kann der Zugang zum Server durch das Wartungspersonal eingeschränkt werden. Obwohl der Administrator die Zugangsdaten besitzt, kann er Tätigkeiten nur dann ausführen, wenn er auch den elektronischen Schlüssel besitzt. Auf diese Art kann ein zeitlich und räumlich unbeschränkter und unkontrollierbarer Zugriff pragmatisch unterbunden werden. Für das Studienportal wurde die Faktor-2-Implementierung mit Hilfe eines Yubikey umgesetzt. Der USB-Stick ist preislich erschwinglich und sehr robust. Wie beim analogen Pendant können mehrere Zweit- oder Ersatzschlüssel je Server verwendet werden. Der Hersteller bietet die Möglichkeit die Echtheit des erzeugten *One-Time-Passwords* über seinen kostenfreien Cloud-Dienst zu überprüfen. Wahlweise kann auch eine lokale Instanz des Authentifizierungsservers im eigenen Rechenzentrum betrieben werden. In Anbetracht der geringeren Komplexität wurde für die Überprüfung des *OTP* auf die Server des Herstellers zurückgegriffen.

Abbildung 89: Faktor-2-Authentifizierung mittels Yubikey

Die Anwendbarkeit des Yubikeys beschränkt sich jedoch nicht nur auf Fernwartung. Auch die Anmeldung privilegierter Benutzer am Studienportal könnte auf diese Weise sicherer gemacht werden. Für einige Content-Management-Systeme (wie z.B. Wordpress, Joomla, Drupal) gibt es bereits Plugins. Dank quelloffener Software und einer guten Programmierschnittstelle können auch eigene Webanwendungen mit dem Yubikey *OTP* ausgestattet werden. Zum Zeitpunkt des Abschlusses der Implementierung (12/2014) war leider für Moodle noch kein Modul verfügbar, das die Nutzung von *OTP* ermöglicht hätte. Für den Produktivbetrieb kann überlegt werden diese Funktion hinzu zu programmieren.

7.2.9 Sicherheit des Mailsystems

Der Emaildienst mag auf den ersten Blick für das geplante Studienportal eine eher untergeordnete Rolle spielen, jedoch soll aufgrund der dennoch großen Relevanz auf die wichtigsten Aspekte des Mailsystems eingegangen werden. Während der Registrierungsprozedur und beim Rücksetzen des Passworts wird eine Email an die Teilnehmer verschickt. Es ist daher wichtig, dass Emails, die vom System generiert werden ihre Empfänger erreichen. Es gibt eine unüberschaubare Vielzahl an Mailanbietern und Mailservern zu denen ein Versand von Emails sichergestellt werden muss. Im Folgenden sollen aus Gründen der Einfachheit nur die häufigsten Vertreter betrachtet werden. Die in Deutschland dominierenden Mailanbieter wurden mittels umfassender Studien (n= 12 Millionen) ermittelt. Die Marken GMX, Web.de und 1und1 stellen gemeinsam etwa die Hälfte aller deutschen Emailpostfächer. Alle drei Marken werden von der United Internet AG betrieben und sind sich daher technisch sehr ähnlich. [136] Unter die Rubrik „Andere“ fallen hauptsächlich Mailserver von Firmen und Universitäten.

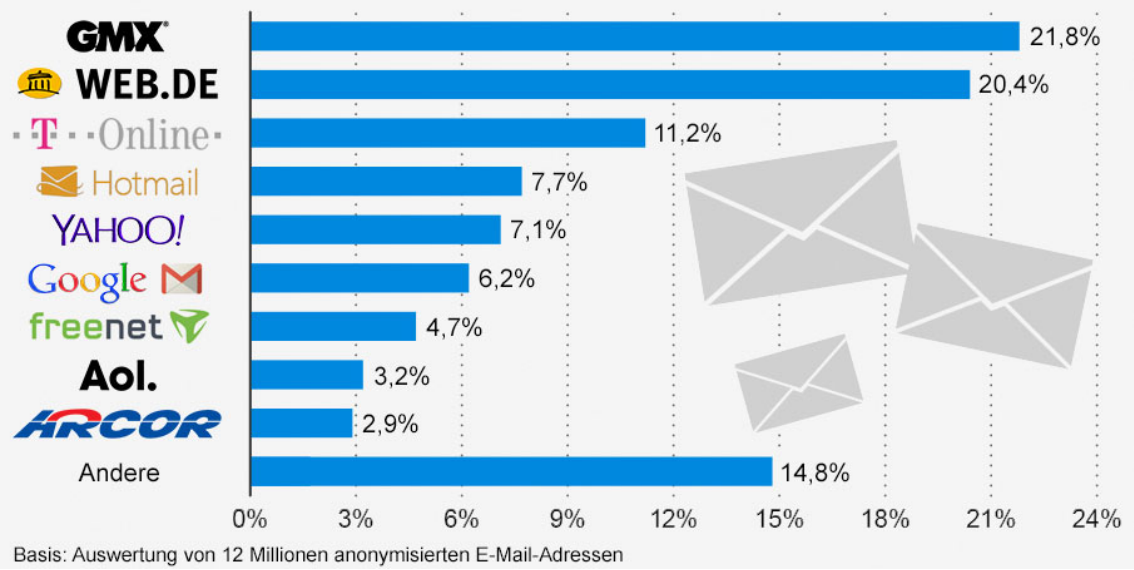


Abbildung 90: Verbreitung deutscher Emailanbieter 2013, statista [136]

Ein Grund sich mit Emailsicherheit bewusst auseinanderzusetzen liegt in der Gefahr begründet, die von manipulierten Emails ausgehen kann. Über einen gefälschten Emailabsender können nicht nur Spam- Emails in Umlauf gebracht, sondern auch Teilnehmer und Studienbetreuer Opfer von *Pishing Angriffen* werden. Es ist daher notwendig, technische Maßnahmen zu ergreifen, die eine sichere und verlässliche Zustellung der elektronischen Post gewährleisten. *Pishing* Angriffe bedienen sich immer raffinierterer Methoden. Um das Vertrauensniveau zu erhöhen, verwenden sie gültige Zertifikate um die gefälschten Webseiten vertrauenswürdig erscheinen zu lassen. Dies ist einfach möglich, da einige Serveranbieter Zertifikate für ihre Kunden bereitwillig ausstellen. Aber auch große renommierte Zertifizierungsstellen wie Symantec, Comodo oder GlobalSign werden für das Ausstellen von Zertifikaten für gefälschte Webseiten genutzt. [198] Mehrfach wurde diese Methode bereits zum Erbeuten von Zugangsdaten für den Online-Bezahldienst Paypal oder von Homebanking verwendet. [199]

7.2.9.1 RDNS und generische Hostnamen

Aus Angst vor Missbrauch bedienen sich die populären Maildienstanbieter unterschiedlicher Mechanismen, um die Flut ungewollter Nachrichten einzudämmen. Unglücklicherweise gibt es keine allgemeingültige Lösung an die sich alle Anbieter halten. Vielmehr veröffentlichen die großen Mailbetreiber ihre eigene Emailpolitik sowie technische Anforderungen, an die man sich anpassen muss, möchte man mit deren Nutzern Post austauschen. Dies hat zur Folge, dass ein gewisser Aufwand betrieben werden muss, um Kompatibilität mit allen Anbietern herzustellen. Um sicherzustellen, dass Emails nicht von beliebigen Mailservern angenommen und weitergeleitet werden, schränken Mailanbieter wie Web.de [200], GMX [201], T-Online [202], AOL [7], Arcor [203], Microsoft Hotmail [204], Yahoo [205] ihre Empfangsbereitschaft durch die Überprüfung von Hostnamen und IP-Adressen deutlich ein. Die genannten Mailprovider schließen die Verwendung *generischer Hostnamen* aus, wie sie von fast allen Internet-Anbietern und vielen günstigen Serveranbietern verwendet werden. Viele Serveranbieter liefern ihre Server mit *generischen Hostnamen* wie *s1234567.onlinehome-server.info* und nicht mit dem tatsächlichen Domainnamen z.B. *server.studienportal.eu* aus. Mit *RDNS* (*reverse DNS*) wird zusätzlich überprüft, ob der Weg von der IP-Adresse zurück zum Hostnamen möglich ist. Es muss also immer eine 1:1 - Verbindung zwischen dem Servernamen und der IP-Adresse bestehen. So muss beispielsweise der Mailserver der Universität Regensburg *mail.uni-regensburg.de* mit der IP-Adresse *194.94.155.125* wiederum von *194.94.155.125* auf *mail.uni-regensburg.de* verweisen, was im verwendeten Beispiel auch der Fall ist. Überprüft ein Mailanbieter *RNDS,* werden hierdurch automatisch Betreiber von Home-Servern (z.B. am heimischen DSL-Anschluss), aber auch viele virtuelle Server oder Besitzer von Webhosting-Tarifen, vom Versand über die eigene Domain ausgeschlossen. In den genannten Fällen fehlen die notwendige dedizierte feste IP-Adresse, der passende Hostname oder *RDNS-Einträge*. Bei einigen Mailanbietern ist es zusätzlich notwendig, sich an einem Partnerprogramm (Feedback-Loop) anzumelden, um die Vertrauenswürdigkeit der eigenen IP-Adresse zu erhöhen. Dies trifft z.B. auf AOL [206] und Hotmail [207] zu. In diesem Fall werden unbekannte IP Adressen zunächst als nicht vertrauenswürdig eingestuft und erst durch Ausfüllen entsprechender Formulare ein Mailversand ermöglicht.

Für den korrekten Mailversand wurde dem Mailserver des Studienportals eine feste IP-Adresse zugewiesen und entsprechende *DNS/RDNS-Einträge* gesetzt.

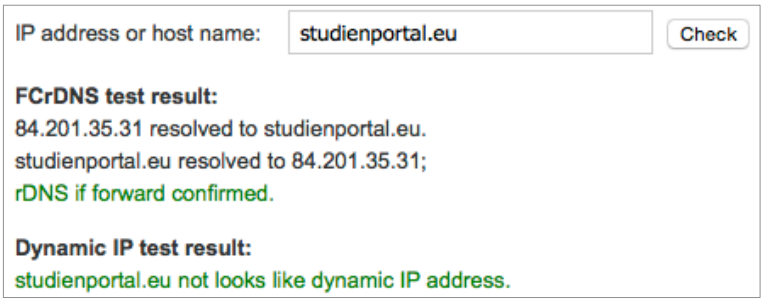


Abbildung 91: Online-Test der RDNS Einstellungen auf www.debouncer.com

7.2.9.2 Absender-Fälschung (SPF)

SPF steht für *sender policy framework*. Mit Hilfe von *SPF* kann auf *DNS-Ebene* eine Liste vertrauenswürdiger Mailserver definiert werden. Der Mailversand kann dadurch auf einzelne

oder mehrere Server verteilt werden. Der potentielle Empfänger der Mail kann anhand der *DNS-Einträge* vergleichen, ob die IP-Adresse bzw. der Hostname des Senders in der Liste erlaubter Server aufgeführt ist. Daraufhin kann er entscheiden, ob Post von diesem Absender angenommen oder abgelehnt wird. So kann verhindert werden, dass Emails von **@studienportal.eu* von fremden Servern aus verschickt werden und evtl. für Spam oder für *Pishing* Angriffe missbraucht werden können. [65]

Hierzu muss lediglich in die bestehende *DNS-Konfiguration* der Domain ein zusätzlicher Eintrag hinzugefügt werden, der die Regel (sender policy) enthält. Der *DNS-Eintrag*

@ IN TXT "v=spf1 mx -all"

besagt, dass nur der oder die bereits in den *DNS-Einstellungen* hinterlegten Webserver (*A*) oder Mailserver (*MX*) Emails für die Domain verschicken dürfen. Alternativ kann der Versand auch auf eine Liste von IP-Adressen innerhalb eines oder mehrerer Netzwerke beschränkt werden. [86] Größere Emailanbieter mit mehreren Servern und unterschiedlichen Netzen bevorzugen die Variante der Beschränkung auf IP-Adressen.

Das folgende Beispiel soll die Funktionsweise von *SPF* verdeutlichen:
Ein Hacker versucht von seinem Server aus eine Email mit dem Absender *info@studienportal.eu* an den privaten Empfänger *hanshans@gmx.de* zu verschicken. Der Mailserver des Empfängers (GMX.de) überprüft nun die *DNS-Daten* der Absenderadresse. Hierbei findet er heraus, dass die IP-Adresse des Absenders von der IP-Adresse der Domain *studienportal.eu* abweicht und nicht auf der Liste erlaubter Mailserver für *studienportal.eu* steht. Die Email wird daher vom Empfänger-Server (GMX) abgelehnt.

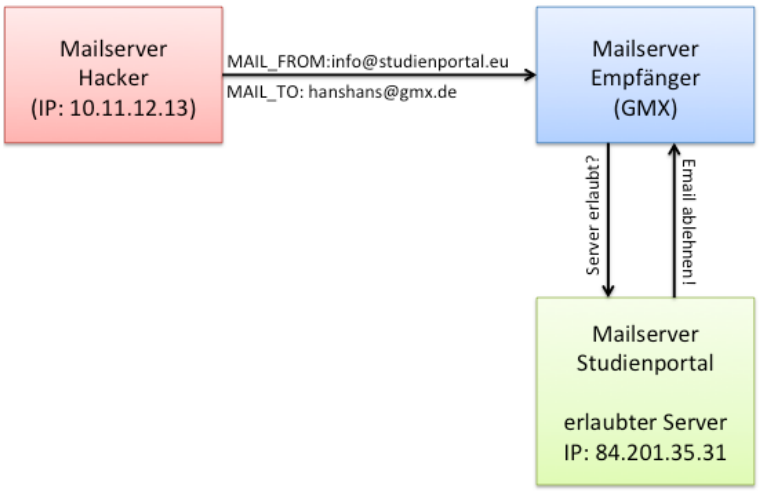


Abbildung 92: Schutz vor Absenderfälschung durch Sender Policy Framework

Obwohl *SPF* eine gute Möglichkeit darstellt die Auswahl legitimer Mailserver zu definieren, findet es nur bei wenigen großen Mailanbietern Anwendung. Es sei jedoch kritisch angemerkt, dass *SPF* nicht die Sicherheit des eigenen Mailservers erhöht, sondern primär den Empfänger vor Fälschungen schützt. *SPF* verlässt sich hierbei auf die Gültigkeit der *DNS-Daten*. Da diese jedoch in der Regel (abgesehen beim Einsatz von *DNSSec*) unverschlüsselt übertragen werden und deren Korrektheit nicht überprüft werden kann, liegt in der Manipulation von *DNS-Informationen* ein möglicher Angriffspunkt. Diese Lücke kann durch den Einsatz von *DANE/TLSA* (☞ Kapitel 7.2.6) geschlossen werden, indem *DNS-Informationen* verschlüsselt

und signiert übertragen werden. Dennoch wird *DANE/TLSA* derzeit von keinem der großen Anbieter eingesetzt. [208] (Stand 01/2015)
Für das Studienportal wurde *SPF* implementiert. Die Einträge für den Mailserver sind auf das Netzwerk des Clusters beschränkt. Alle Mailserver außerhalb dieses Netzes werden als ungültig abgelehnt.

7.2.9.3 Domainkeys (DKIM)

Eine weitere nützliche Technologie zur Vermeidung von Absenderfälschungen ist die *DomainKeys Identified Mail* - kurz *DKIM* oder *Domainkeys* genannt. Hinter dem Namen verbirgt sich die Möglichkeit, jede versendete Email (im Gegensatz zu *PGP* oder *S/MIME*) serverseitig mit Hilfe eines *public-key-Verfahrens* digital zu signieren. Durch diese Maßnahme wird die Herkunft von Emails überprüfbar. Ursprünglich wurde *DKIM* von Yahoo entwickelt um Spam und *Pishing* einzudämmen. Hierzu wird ähnlich wie bei *SPF* die notwendige Zusatzinformation (der öffentliche Schlüssel) als *DNS-Eintrag* hinterlegt. Dieser kann von allen Empfängern überprüft werden. Das Schlüsselpaar wird mit dem *RSA-SHA1* oder *RSA-SHA2* Algorithmus erzeugt und der öffentliche Schlüssel als TXT -Eintrag via *DNS* veröffentlicht. [21] Der *DNS-Eintrag* sieht etwa wie folgt aus:

default._domainkey.studienportal.eu. IN TXT p=MHwwDQYJKoZI....awAwaAJhAOFtq;

Der Mailserver erweitert die Metadaten jeder ausgehenden Email mit der individuellen Signatur, die mit einem geheimen privaten Schlüssel generiert wird. Jeder Empfänger (Mail-Server oder auch ein Mailprogramm) kann diese Signatur mit Hilfe des öffentlichen Schlüssels überprüfen und somit sicherstellen, dass die Email vom angegebenen Server stammt. [21]

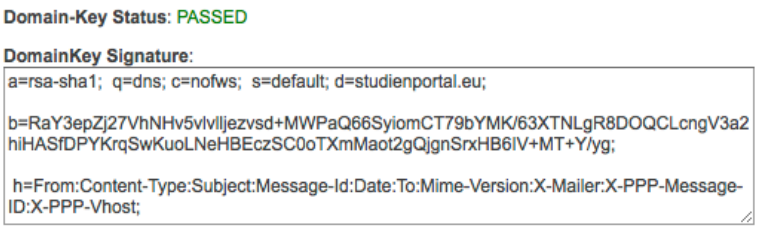


Abbildung 93: Absender-Identifikation anhand einer Signatur im Email-Header

Für die Implementierung von *DomainKeys* sind Eingriffe am Mailserver notwendig, da für jede Email kryptografische Zusatzinformationen erzeugt und Emails bearbeitet werden müssen. Zudem ist analog zu *SPF,* die Modifikation von *DNS-Einträgen* notwendig. Dies ist vermutlich der Grund, warum dieses Verfahren nicht flächendeckend eingesetzt wird.
Mittels manueller Analyse des *Headers* einer Email kann getestet werden, ob ein Anbieter seine Post elektronisch signiert. Alternativ existieren kostenfreie Werkzeuge im Internet mit deren Hilfe auf *DKIM* überprüft werden kann (z.B. *http://mailradar.com/domainkeys/*).

Die in Deutschland gebräuchlichen Emailanbieter wurden auf *DKIM* untersucht. Derzeit wird *DKIM* nur von Yahoo, Googlemail und AOL praktiziert. Die Anbieter der Kampagne Email made in Germany und der .DE-Mail (T-Online, Web.de, GMX, 1und1, Freenet) aber auch MSN Hotmail übermitteln in ihren Emails keine digitale Signatur. (Stand 12/2014).

```
X-Spam-Level:
Received-Spf: pass (lobustho.de: domain of t-online.de designates 194.25.134.84 as permitted sender) client-ip=194.25.134.84;
envelope-from=lobustho@t-online.de; helo=mailout09.t-online.de;
X-Ums: email
X-Spam-Checker-Version: SpamAssassin 3.3.2 (2011-06-06) on lobustho.de
Domainkey-Status: no signature
Return-Path: <lobustho@t-online.de>
Mime-Version: 1.0
X-Toi-Msgid: 46c12c80-f9d9-4a47-9f27-3f08bc0fa3f0
X-Priority: 3
Content-Transfer-Encoding: 8bit
X-Mailer: DTAG LISA 0.1
Message-Id: <1Y4zYj-3gStJg0@fwd09.aul.t-online.de>
X-Id: Z6tsQeZXwhVbvuwKvUaMUK4bNI8Ph4LDMooXHRMxMWVksB6FR73CZ+fKfZZfSVLZGj@t-dialin.net
X-Spam-Status: No, score=-1.9 required=2.0 tests=BAYES_00,TVD_SPACE_RATIO, T_RP_MATCHES_RCVD autolearn=ham
version=3.3.2
Content-Type: text/plain; charset="UTF-8"
Delivered-To: hanshans@lobustho.de
X-Original-To: hanshans@lobustho.de
Received: from mailout09.t-online.de (mailout09.t-online.de [194.25.134.84]) by lobustho.de (Postfix) with ESMTPS id 6FCAC5007AF for
<hanshans@lobustho.de>; Sat, 27 Dec 2014 23:16:07 +0100 (CET)
Received: from fwd09.aul.t-online.de (fwd09.aul.t-online.de [172.20.27.151]) by mailout09.t-online.de (Postfix) with SMTP id
75D2147D890 for <hanshans@lobustho.de>; Sat, 27 Dec 2014 23:10:14 +0100 (CET)
Received: from localhost (Z6tsQeZXwhVbvuwKvUaMUK4bNI8Ph4LDMooXHRMxMWVksB6FR73CZ+fKfZZfSVLZGj@[172.20.102.134]) by
fwd09.aul.t-online.de with esmtp id 1Y4zYj-3gStJg0; Sat, 27 Dec 2014 23:10:05 +0100
Received: from 87.138.86.77:7946 by cmpweb25.aul.t-online.de with HTTP/1.1 (Lisa V3-1-6-0.11371 on API V3-17-12-0)
Domainkeys Test
```

Abbildung 94: Email-Header ohne DKIM (gesendet via T-Online Webmail)

Die meisten Emailanbieter betreiben Spamfilter-Systeme, die mit Schätzregeln die Vertrauenswürdigkeit des Absenders bewerten. Ein positiver *Domainkey* Status bildet eine wichtige Vertrauensgrundlage und verringert das Risiko, dass eine Nachricht versehentlich als Spam klassifiziert wird. Weiterhin reduziert sich für Studienteilnehmer die Gefahr, Opfer von *Pishing* Mails zu werden, da viele Mailsysteme ungültige *Domainkeys* Abfragen als Spam einstufen und aussortieren bzw. gar nicht erst zustellen. Aus diesem Grund wurde das Mailsystem des Studienportals mit *Domainkeys* Unterstützung konfiguriert.

7.3 Namensserver

Die Domain *studienportal.eu* wurde beim Regensburger Registrar Schlund Technologies GmbH registriert. Dieser bietet zwar für seine Domainkunden äußerst zuverlässige Namensserver an, jedoch wie die übrigen großen Domainregistrare keine vollständige *DNSSec* Unterstützung. Es besteht zwar die Möglichkeit den öffentlichen Schlüssel für die Domain zu hinterlegen, das Bereitstellen sowie das Ausliefern der signierten *DNS-Einträge* (*RRSIGs*) übernimmt der Anbieter jedoch nicht. Den öffentlichen Schlüssel reicht Schlund als *DS-Record* an die Namensserver der *.eu Zone* weiter.

Für die Implementierung des fälschungssicheren *DNS-Systems* wurde der weit verbreitete Namensserver BIND verwendet. BIND ist eine sehr schlanke und robuste Serversoftware, die mit wenigen Ressourcen auskommt, gut dokumentiert ist und *DNSSec* beherrscht. Dem Redundanzprinzip folgend können mehrere primäre und sekundäre *DNS-Server* definiert werden. Der Abgleich der *DNS-Einträge* zwischen allen Namensserver-Instanzen kann automatisiert werden. Änderungen müssen nur an einem Server vorgenommen werden. Die übrigen DNS-Server übernehmen die Einstellungen automatisch. [55] Für die Implementierung der Infrastruktur wurden zwei virtuelle Server mit dem BIND (Version 9) im *Primary/Secondary* Modus konfiguriert. Da für jeden Seitenaufruf die Namensauflösung über diese beiden Server

abgewickelt wird und durch die Verschlüsselung der *DNS-Anfragen* der Rechenaufwand steigt, wurden beide virtuelle Server mit hoher CPU Priorität versehen.
Beide Maschinen sind mit 4 CPU Kernen ausgestattet und laufen aus Redundanzgründen auf unterschiedlichen physikalischen Maschinen. Zur Erhöhung der Ausfallsicherheit sollten für den Produktiveinsatz zusätzliche sekundäre *DNS-Server* außerhalb des Rechenzentrums erwogen werden. Während der Entwicklungs- und Testphase hat sich das beschriebene Setup auch unter Lastsimulation gut bewährt.

Der öffentliche *DNSSec* Schlüssel wurde mit 2048bit und dem *NSEC3* Algorithmus verschlüsselt und mit *SHA-1* signiert. (☞ Kapitel 7.2.5) Um ausschließlich das vom Webserver verwendete Zertifikat zu akzeptieren, wurden für die *DANE* Implementierung der Zone für Port 80 (http) und Port 443 (https) zwei *TLSA* Einträge hinzugefügt. (☞ Kapitel 7.2.6) Die Zoneninformationen (inklusive der *RRSIGs*) haben eine Gültigkeit von einer Woche. Danach müssen sie neu signiert werden. Dies bietet einen guten Kompromiss aus Schutz und Komfort. Zur Automatisierung der Ausstellung gültiger *RRSIGs* wurde ein Skript erstellt das alle 7 Tage automatisch aufgerufen wird.

7.4 Hochverfügbarkeitscluster

7.4.1 Hardware

Als Auswahlkriterium für die technische Hardwareausstattung für die Umsetzung des Studienportals wurden die bereits erläuterten wesentlichen Kriterien für einen ausfallsicheren Betrieb herangezogen. (☞ Kapitel 3.2)

Die Wahl fiel auf Hardware von Dell. Die Besonderheit des verwendeten Dell Poweredge C6105 liegt in der kompakten und zugleich soliden Bauweise. [209] In lediglich 2 Höheneinheiten finden 4 separate Server (*Nodes*) Platz. Die Server teilen sich die Stromversorgung, die Festplatten (*Backplane*) sowie die Kühlung. Stromversorgung und Lüfter bieten jeweils eine N+1 Redundanz. Durch das Teilen gemeinsamer Ressourcen können die 4 Server mit einer Energieeffizienz von ca. 92% besonders ökonomisch betrieben werden, was die laufenden Kosten gering hält und die Umwelt schont. [209] Im Testbetrieb konnte der gesamte Cluster mit vollbestückten Festplatten inklusive der Netzwerktechnik mit ca. 550Watt bis 750Watt Leistung betrieben werden. Konventionelle Server mit vergleichbarer Ausstattung und Rechenleistung weisen typische Leistungsaufnahmen von mindestens 250 - 450 Watt je Server auf. Die geringe Bauhöhe erlaubt es den Server günstig in einem Rechenzentrum unterzubringen, da sich die Kosten neben Strom und Datenvolumen auch nach der Anzahl belegter Höheneinheiten berechnen. Jeder Server verfügt über je zwei 6-kernige AMD Opteron 4332 Prozessoren, jeweils 32GB Arbeitsspeicher und 6x Gigabit Netzwerkkarten. Zwei Netzwerksanschlüsse werden für die Außenanbindung (WAN) auf zwei Dell PowerConnect 6224 Switches verteilt. [25] Die übrigen 4 Ports werden für das clusterinterne Netz (Cluster LAN) sowie für den Aufbau einer netzwerkbasierten Speicherlösung reserviert. (☞ Kapitel 7.4.2) Zur Steigerung der Netzwerkleistung und Ausfallsicherheit erfolgt eine Bündelung der Netzwerkkarten via *Link Aggregation* und *Port Trunking*.
Das Gehäuse bietet Festplatteneinschübe für 24x 2,5“ Festplatten, die den 4 Servern zugeordnet sind. Jeder *Node* erhält zwei gespiegelte 2,5“ SAS 10k U/min Festplatten für die Installation des Betriebssystems und vier Festplatten im *Raid 10* Verbund zur Speicherung der Daten. Zwei *Nodes* wurden mit langsameren aber großvolumigeren *Nearline Storage* ausgestattet um dort große Datenvolumina für z.B. Videos oder Backups zu speichern. Die übrigen beiden *Nodes* erhielten schnelle und für den Dauerbetrieb ausgelegte *SSD* Speicher der Enterprise-Klasse für Datenbank- und Webserveranwendungen. Für die Ansteuerung der Festplatten kommen Hardware-Raid-Controller (LSI 9260-8i) mit 512MB Pufferspeicher (*Cache*) und einer Pufferbatterie (*BBU*) zum Einsatz. Die Pufferbatterie ermöglicht den sicheren Einsatz des schnellen Zwischenspeichers des Controllers (*Write-Back Cache*) und erhöht dadurch den Speicherdurchsatz zwischen der Software und dem Festplattenverbund.
Die Stromversorgung des zweiten Netzteils des Servers und der Switches ist an einer separaten Stromschiene mit batteriegepufferter Notstromversorgung angeschlossen um eine redundante Stromversorgung aller Komponenten zu gewährleisten.

7.4.2 Hochverfügbarkeitscluster (HA-Cluster)

Trotz des soliden Hardwareunterbaus mit einer N+1 Redundanz von Stromversorgung, Lüfter, Netzwerk und Festplatten, bleiben als möglicher Risikofaktor das Mainboard des jeweiligen Servers bestehen. Um diesen *Single-point-of-failure* zu eliminieren wurde ein Hochverfügbarkeits-Cluster (*HA-Cluster*) implementiert.
Der Cluster besteht insgesamt aus 4 Servern und zwei Switches. Auf den Servern werden

wiederum virtuelle Server betrieben, die alle notwendigen Serverdienste für den Betrieb bereitstellen. Einer der genannten virtuellen Server ist der eigentliche Webserver, auf dem das Studienportal betrieben wird. Hinzu kommen weitere Server für Verwaltungsaufgaben. Hierzu gehören zwei DNS Server, ein Backupserver, ein Server für die Überwachung der Dienste, sowie ein Mailserver. Eine Kopie des Webservers wird als Entwicklungsserver genutzt. Außerhalb des Rechenzentrums sind zwei weitere physikalische Maschinen für externes Monitoring sowie Benchmarks im Einsatz. Das folgende Schema visualisiert die aktuelle Anordnung des Rechnerverbunds.

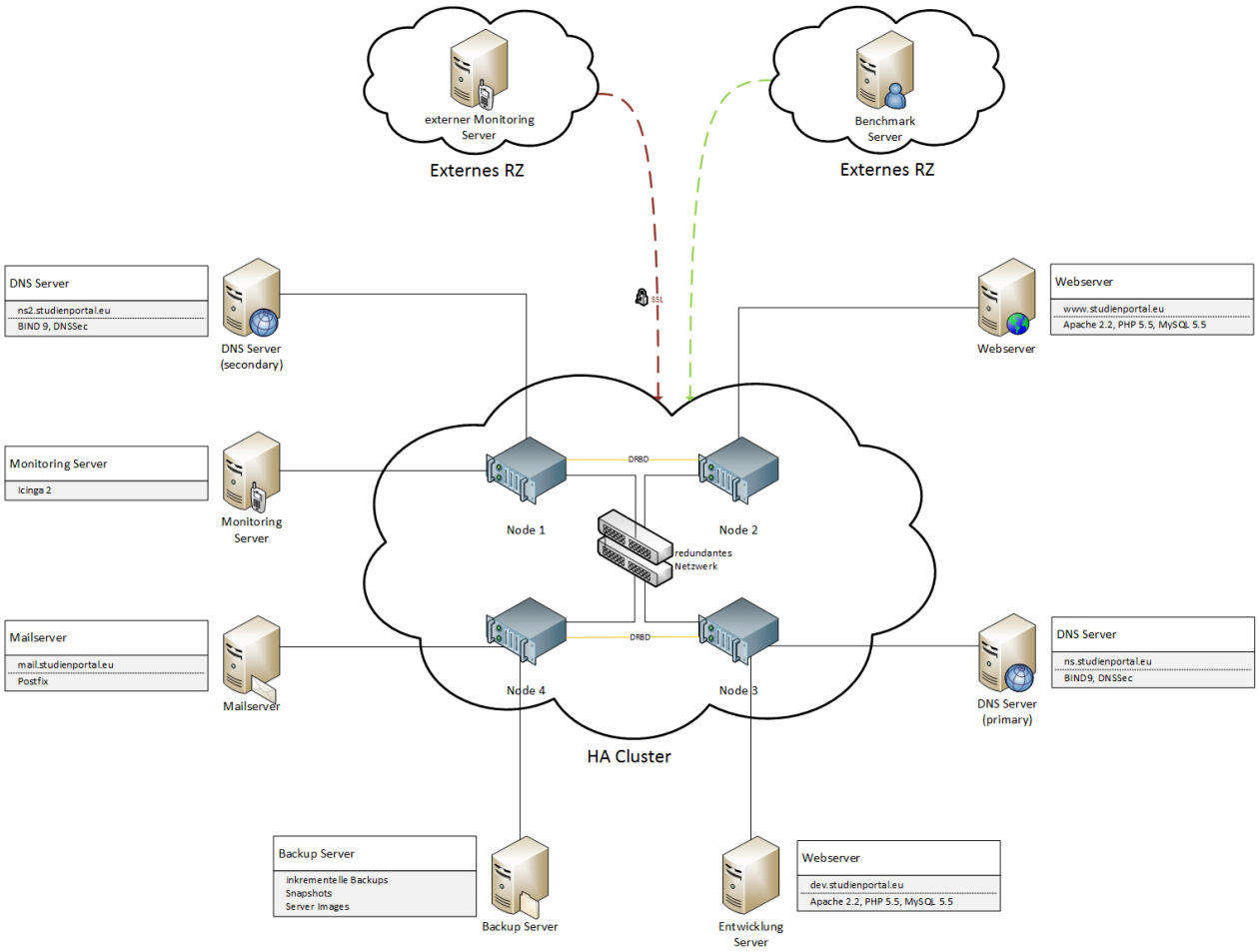


Abbildung 95: Aufbau des Clusters

Konventionelle Cluster-Systeme verwenden einen zentralen redundanten Datenspeicher auf Basis eines *SAN* (*Storage Array Network*) oder ein *NAS* (*Network Attached Storage*). Diese bestehen bei einfacher Redundanz aus zwei Speicherservern. Der Zugriff auf den zentralen Datenspeicher erfolgt über ein Netzwerk und nutzt hierfür schnelle aber teure Spezialhardware (*FibreChanel* oder *Infiniband*). Für die Redundanz der Anwendungsserver müssen zwei weitere Server eingeplant werden. Traditionellerweise besteht ein *HA-Cluster* daher aus mindestens vier Servern, von denen lediglich zwei für die eigentliche Anwendung zur Verfügung stehen. [210] (☞ Kapitel 6.6.1) *SAN* Systeme sind auf hohe Stabilität und Verfügbarkeit ausgelegt. Die meisten Geräte erlauben inzwischen eine Kombination aus unterschiedlichen Festplattentechnologien (*SATA*, *SAS* und *SSD*). Mit Hilfe intelligenter Software wird versucht Daten, je nach deren Zugriffshäufigkeit und Wichtigkeit, möglichst geschickt auf unterschiedliche Festplattentypen zu verteilen. Auf diese Weise wird versucht den Kostenaufwand für

weniger häufig benötigte oder große Daten (wie z.B. Backups) zu reduzieren indem sie auf langsameren aber günstigen Medien vorgehalten werden. Schnelle aber teure und kleine Speichermedien (wie *SSDs*) hingegen werden für häufig benötigte oder priorisierte Daten reserviert. [88] Dennoch sind professionelle *SAN* Installationen in Preisregionen jenseits der 20.000€ Marke angesiedelt. Anwendungsgebiete für *SANs* sind Serverfarmen von Krankenhäusern, Universitäten oder Firmen mit vielen Servern, die auf einen zentralen Pool zurückgreifen oder aber große Cluster mit vielen Nodes.

Eine kostengünstigere und für den geplanten Einsatz leistungsfähigere Alternative stellt *DRBD (Distributed Block Device)* dar. Mit Hilfe von *DRBD* werden die Festplatteninhalte zweier Server über ein Netzwerk gespiegelt. Dabei können die beiden Server zeitgleich für Anwendungsaufgaben verwendet werden. Die Daten werden nicht wie bei *SAN* oder *NAS* von einem zentralen Speicher über das Netzwerk geladen, sondern lokal von den Festplatten der jeweiligen Server. Nur für das Spiegeln der Daten ist die Netzwerkverbindung zwischen den Servern notwendig. Ein spezieller Algorithmus sorgt dafür, dass lediglich geänderte Datenblöcke auf den zweiten Server übertragen werden. [102] Ein *HA-Cluster* mit *DRBD* kommt daher im Vergleich zu klassischen *HA-Clustern* mit lediglich zwei Servern aus und kann mit konventioneller und deutlich günstigerer Netzwerktechnik (z.B. Gigabit Ethernet) realisiert werden. Somit ist es möglich eine N+1 Redundanz mit deutlich geringerem Hardware- und damit auch Kostenaufwand abzubilden. [211]

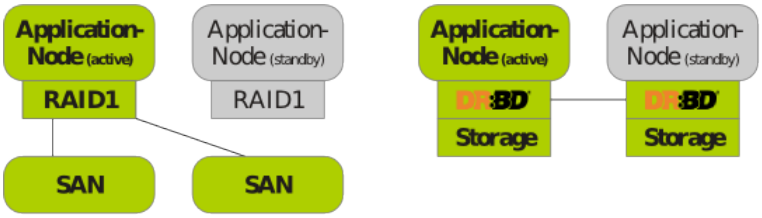


Abbildung 96: Unterschied zwischen SAN und DRBD, Linbit [210]

Für die Umsetzung der N+1 Redundanz wurden je zwei *Nodes* mit identischer Festplattenkonfiguration über ein isoliertes TCP/IP-Netzwerk (1GbE Base-T Cross-Over) verbunden. Über dieses Netzwerk werden die Daten beider Server im Sinne eines *Raid 1* gespiegelt. Alle Daten werden von der *DRBD* Software permanent repliziert und auf beiden Servern vorgehalten. Daten werden im *DRBD* Jargon allgemein als *(Speicher-)Ressourcen* bezeichnet. Eine *Ressource* kann eine gesamte Festplatte aber auch nur ein Teil davon sein.
Entgegen der Standardkonfiguration wird das *DRBD* im *Aktiv-Aktiv* Modus betrieben. (☞ Kapitel 6.6.1) Dadurch kümmert sich *DRBD* lediglich darum, dass zu jeder Zeit auf beiden Servern der gleiche Datenbestand vorliegt. Diese Betriebsart ermöglicht den Zugriff zweier Server auf eine gemeinsame *Ressource*. Im Falle eines Ausfalls kann somit der eine Server die *Ressourcen* des anderen nahtlos übernehmen. Allerdings ist zusätzliche Software erforderlich um zu verhindern, dass die beiden Server gleichzeitig auf ein und dieselbe *Ressource* zugreifen, da dies sonst zu Datenkorruption führen würde. Um diese Situation (*Split-Brain Situation*) zu verhindern, reguliert ein *Ressourcemanager* (Pacemaker) die Zugriffe auf alle *Ressourcen* (Festplattenspeicher, IP-Adressen usw.). Mit Hilfe eines clusterweiten Nachrichtendienstes (Corosync) werden für den Betrieb notwendige Informationen innerhalb des Clusters verteilt. Corosync überprüft zum Beispiel die Funktionstüchtigkeit und Berechtigung aller Cluster-Mitglieder und informiert den *Ressourcemanager* über Änderungen des Betriebszustands von *Nodes* oder ihrer Berechtigung in Bezug auf *Ressourcen*. Durch das Zusammenspiel von Pacemaker und Corosync wurde eine *Aktiv-Passiv* Konfiguration umgesetzt.

Im Regelbetrieb darf nur ein Server der beiden *DRBD Nodes* auf eine bestimmte *Speicherressource* zugreifen. Der andere Server verhält sich in Bezug auf die *Ressource* passiv. Welcher Server die aktive und welcher die passive Rolle einnimmt, wird von der Clustersoftware (Pacemaker und Corosync) gesteuert. Fällt der aktive Server aus, wird dies erkannt und der zuvor passive Server erhält die Zugriffsrechte und wird in den aktiven Zustand versetzt. *Ressourcen* können hierbei durchaus homogen auf zwei Server verteilt sein, so dass sich beide Server für eine *Ressource* aktiv, für eine andere wiederum passiv verhalten. Im Störfall kann der noch verfügbare Server die *Ressourcen* und Aufgaben des ausgefallenen Rechners ohne Ausfallzeit übernehmen. *DRBD* arbeitet hiervon unabhängig im Hintergrund und hält lediglich die *Speicherressourcen* redundant. Im Fehlerfall stehen die *Speicherressourcen* dann nur noch auf dem aktiven Server zur Verfügung. Die darüber liegende Anwendungssoftware (z.B. der Webserver) bekommt hiervon jedoch nichts mit.

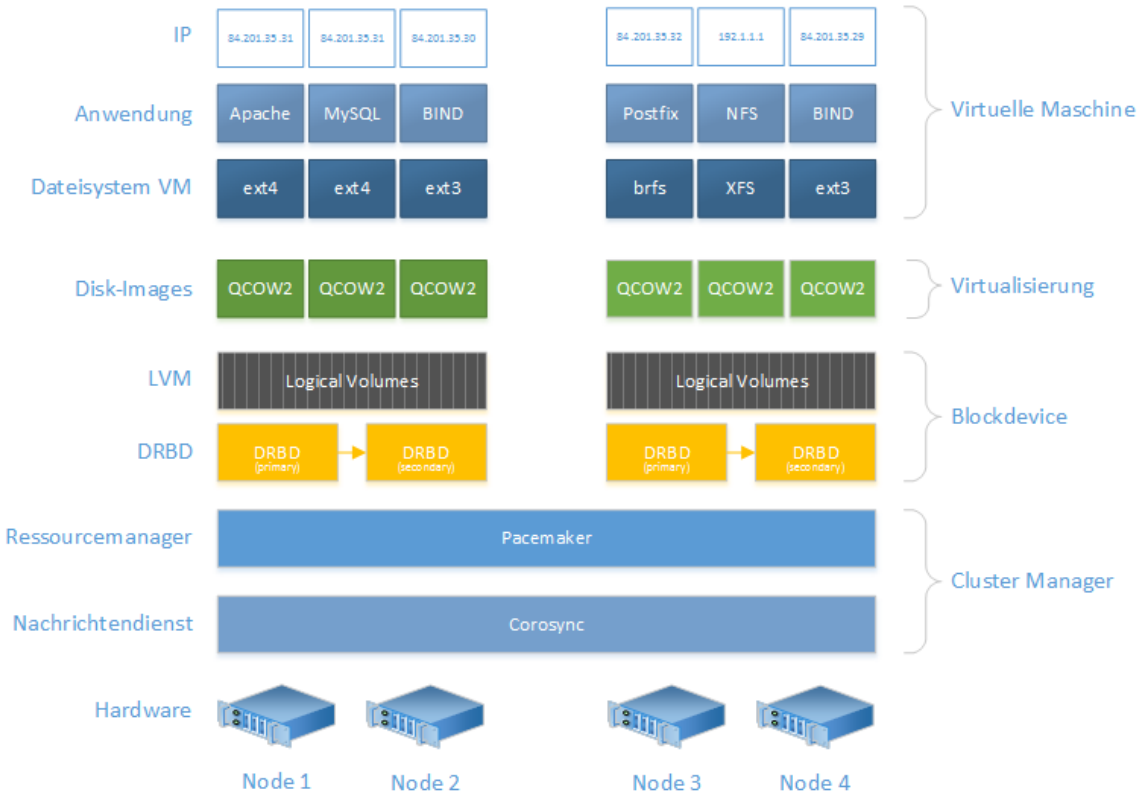


Abbildung 97: Datenverwaltung im eingesetzten Aktiv/Passiv HA-Cluster

Um die Infrastruktur des Studienportals unabhängiger von der verwendeten Hardware zu halten, und um die Rechenleistung der verwendeten Server besser ausnutzen zu können, wurde eine Virtualisierungstechnolgie eingesetzt. Virtualisierung bietet im Vergleich zum Betrieb eines Servers ohne Virtualisierung (= physikalischer Server) die Möglichkeit mehrere (virtuelle) Server auf einem Gerät zu betreiben. Besonders hilfreich ist, dass virtuelle Server unabhängig von der zugrundeliegenden Hardware betrieben werden können. Dem virtuellen Server wird eine virtuelle Hardware vorgegaukelt (virtuelle Festplatten, Prozessoren, Netzwerkkarten oder Peripheriegeräte) mit denen das Betriebssystem interagiert. Über die tatsächlich verwendete Hardware der physikalischen Server wissen die virtuellen Server nichts. Hierdurch können virtuelle Maschinen einfach auf neue Hardware umgezogen oder aufgerüstet werden, sollte die Leistung des physikalischen Servers nicht mehr ausreichen. Der Leistungsverlust durch die Virtualisierung ist dabei vernachlässigbar gering.

KVM steht für *Kernel-based Virtual Machine* und ist eine in aktuellen Linux Kerneln fest integrierte Virtualisierungstechnologie. Im Gegensatz zu den gängigen kommerziellen Virtualisierungslösungen von Citrix, VMWare oder Microsoft sind hier weder Lizenzkosten noch Hardwarebeschränkungen zu berücksichtigen. [123] *KVM* ist im Vergleich zu manch anderer Virtualisierungstechnologie wie Citrix Xen bei höherer Leistung energiesparender und zudem kostenfrei und quelloffen. [68] [54] Aus diesem Grund erfolgte die Installation aller verwendeten Serverdienste mit Hilfe virtueller Maschinen auf *KVM* Basis. Die virtuellen Festplatten wurden im *QCOW2* Format angelegt. Das *QCOW2* Format bietet die Möglichkeit virtuelle Festplatten zu verschlüsseln, zu komprimieren und den aktuellen Betriebszustand mit Hilfe von *Snapshots* einzufrieren. [212] (☞ Kapitel 6.6.2) Mit entsprechender Optimierung sind *QCOW2 Images* vergleichbar schnell wie konventioneller *RAW Images*, bieten aber deutlich mehr Komfort. [53]

Virtuelle Server werden vom *Ressourcemanager* ebenfalls als *Ressource* behandelt. Sollte ein physikalischer Server ausfallen, kann die virtuelle Maschine von einem anderen Server im Cluster übernommen werden. Besonders vorteilhaft ist es, wenn der Reserveserver Teil eines *DRBD* Pärchens ist. In diesem Falle geschieht die Übernahme mit sehr geringer Ausfallzeit, da die (virtuellen) Festplatten des virtuellen Servers nicht auf den Reserveserver kopiert werden müssen. Vom *Ressourcemanager* muss lediglich die Zuständigkeit für die virtuelle Maschine auf den Reserveserver übertragen werden. Dieser startet die virtuellen Server des ausgefallenen Rechners einfach neu. Je nach Größe des virtuellen Servers und der darunterliegenden physikalischen Festplatten dauert dieser Startvorgang zwischen 2 Sekunden und wenigen Minuten. Ein manuelles Eingreifen ist hierbei nicht erforderlich.

Das Herzstück für den Betrieb der Webseite bildet der Webserver. Von einer virtuellen Maschine wird sowohl der Webserver als auch das Datenbankmanagementsystem bereitgestellt. Ein Klon des Produktivsystems wird unter alternativer IP-Adresse als Testumgebung vorgehalten, um dort neue Programmmodule oder das Einspielen von Updates simulieren zu können.

Für die Namensauflösung wurden zwei DNS Server installiert und mit *DNSSec* konfiguriert. (☞ Kapitel 7.2.5) Beide DNS-Server liegen aus Gründen der Ausfallsicherheit auf unterschiedlichen *Nodes*.

Den Versand von Emailbenachrichtigungen übernimmt eine separate Mailserverinstanz.

Für das Vorhalten von Datensicherungen wurde eine virtuelle Maschine mit einem *NFS* Dienst installiert. *NFS* (Network File System) entspricht einem Netzlaufwerk, das für jeden der Server separate Sicherungsordner bereitstellt. Der Backup Server ist über das Clusternetz (3x 1Gbit *LACP Link Aggregation*) mit allen Servern im Cluster verbunden. Er sichert zentral alle anderen (virtuellen) Server im Cluster. Rechnerisch ist durch die 3Gbit Anbindung eine Datenübertragungsrate von 375 MB/sec zwischen zwei Servern möglich. Die im Backup Server verwendeten *SATA-3* Festplatten erreichen im *Raid 10 Verbund* eine Datenrate von etwa 250MB/sec. Somit ist das Netzwerk nicht der Flaschenhals für Backupprozesse. Für den späteren Einsatz empfiehlt es sich, die Backupinstanz ebenfalls redundant abzubilden.
Für die Überwachung des Clusters und aller Serverdienste wurde eine weitere virtuelle Maschine bereitgestellt. Sie sammelt innerhalb des gekapselten Clusternetzes wichtige Informationen über den Betriebs- und Gesundheitszustand aller Komponenten.
Häufig kündigen sich Hardwareprobleme bereits vor einem Ausfall an. Mit geeigneten Frühwarnsystemen (☞ Kapitel 7.5) können diese rechtzeitig erkannt werden. Zum Beispiel gehen

Arbeitsspeicherausfällen eine erhöhte Anzahl nicht korrigierbarer Bit-Fehler voraus und vor Festplattenausfällen häufen sich Lesefehler, die über das Selbstüberwachungssystem der Festplatten (*S.M.A.R.T*) erkannt werden können. [121] [99] Auch *Raid Controller* stellen Vitalparameter wie z.B. Temperatur des Controllers, Füllstand der Pufferbatterie und natürlich den Status des *Raid* zur Verfügung. [137] Werden Hardwarefehler frühzeitig erkannt, können die virtuellen Server ohne Ausfallzeit auf einen anderen *Node* verschoben werden. Dieser, *Live-Migration* genannte, Vorgang kann im laufenden Betrieb stattfinden. Sind alle virtuellen Maschinen auf einen anderen physikalischen Server ausgelagert, können die notwendigen Wartungsarbeiten vorgenommen werden ohne dass die Verfügbarkeit der Serverdienste davon beeinträchtigt wird. Die temporär migrierten virtuellen Server können nach Beendigung der Wartung (z.B. Austausch von Hardware, Update des Betriebssystems usw.) wieder zurückverschoben werden.

7.4.3 Rechenzentrum

Die einzigen durch das bisher beschriebene Konzept nicht direkt beeinflussbaren Größen stellen die Strom- und Netzwerktechnik des Rechenzentrums sowie die Anbindung an das Internet dar. Es ist wichtig bei der Auswahl des Rechenzentrums darauf zu achten, dass auch hier für Redundanz gesorgt ist. Auf entsprechende Zertifikate (wie z.B. das TÜV Zertifikat) sowie vertragliche Zusicherungen von Mindestverfügbarkeiten (*Service-Level-Agreements*) ist daher besonders zu achten. Als Serverstandort wurde ein Rechenzentrum in Frankfurt gewählt. Entscheidend für die Wahl war die direkte Anbindung an einen der weltweit größten Internetknoten, den DE-CiX. Das TÜV- geprüfte *Tier 3* Hochverfügbarkeitsrechenzentrum besitzt alle relevanten Zertifizierungen und eine 24 Stunden Besetzung mit qualifiziertem Personal, das im Ernstfall umgehend Störungen im Rechenzentrum beheben kann. Eigene Reservehardware kann im Rechenzentrum hinterlegt und gegen Entgelt von einem Techniker eingebaut werden. Der Zugang zum Server ist nach telefonischer Anmeldung und nach Identifikation mit Personalausweis und speziellem Lichtbildausweis jederzeit möglich. Neben dem internationalen Standard für Informationssysteme ISO 27001 wird auch der Qualitätsstandard 9001 vom Betreiber eingehalten. Eine 100Gbit Anbindung an das Internet erfolgt über internationale Austauschpunkte der Telekom, den DE-CiX, TINET, AMSiX, Level(3) Communications sowie Arcor, Freenet, Swisscom und viele andere. Die Stromversorgung erfolgt über batterie- und aggregatgepufferten Strom aus erneuerbaren Energiequellen. Neben Zugangsprotokollen für das Personal des Rechenzentrums erfolgt die Überwachung aller Serverschränke mit Video- und Infrarotkameras, Klima- und Brandschutzsystemen. [1]

Durch den beschriebenen Cluster und die Wahl eines professionellen Rechenzentrums kann die angestrebte Verfügbarkeit von 99,99% zur Jahresmitte erreicht werden.

7.5 Monitoring

Eine wichtige Grundlage für das Betreiben von Hochverfügbarkeitslösungen ist eine Überwachung jeder einzelnen Komponente, die zum Ausfall des Gesamtsystems beitragen kann. Das Überwachungssystem muss frühzeitig bei Überschreitung von Grenzwerten Alarm schlagen um einen Ausfall zu verhindern. Ein für diese Aufgabe gut geeignetes Werkzeug ist das Monitoring System Icinga. Es handelt sich hierbei um OpenSource Software die sich als Abspaltung von Nagios (dem derzeitigen Industriestandard) entwickelt hat. [107] Nachdem Nagios für einige Zeit nicht mehr aktiv weiterentwickelt wurde, entschied sich die Netways GmbH, ein Nürnberger Netzwerkspezialist, dazu, eine modernere Version von Nagios auf OpenSource Basis zu schaffen. [213] Der große Vorteil ist die gute deutschsprachige Dokumentation und die aktive Entwicklergemeinde, welche die kostenfreie Software permanent weiterentwickelt. Ein weiteres Argument für Icinga ist die Kompatibilität zu Nagios. Jeder Sensor des Systems wird über ein entsprechendes Plugin integriert. Dank dieser Kompatibilität und der großen Verbreitung von Nagios stehen auch für Icinga unzählige Sensoren zur Verfügung. Icinga verfügt über die Möglichkeit mit weiteren Icinga Instanzen zu kommunizieren. Auf diese Weise kann ein Abgleich zwischen einer im Cluster befindlichen und einer externen Instanz umgesetzt werden. [215] Der Vorteil dieser Lösung besteht neben der zusätzlichen Redundanz darin, dass das Abfragen aller Serverfunktionen innerhalb des lokalen Clusternetzwerks stattfindet und nicht direkt über das Internet übertragen werden muss. Dies spart Bandbreite und ist zudem sicherheitstechnisch von Vorteil. Die externe Instanz kann die Verfügbarkeit aus Sicht der Webseitenbesucher überwachen und die Ergebnisse in einem bestimmten Intervall über eine verschlüsselte Verbindung mit der im Cluster befindlichen Instanz synchronisieren. Eine Benachrichtigung im Fehlerfall kann auf unterschiedlichste Weise erfolgen. Üblich ist jedoch eine Benachrichtigung per SMS, Email oder via Smartphone App. Aus Kostengründen wurde keine SMS-Benachrichtigung integriert. Die Benachrichtigung erfolgt per Email an das Mailkonto eines Emailanbieters, der über einen Push-Dienst verfügt. Push-Benachrichtigungen sind von der Geschwindigkeit mit SMS-Nachrichten vergleichbar. Im Gegensatz zu SMS setzen sie aber eine permanente Internetverbindung des Smartphones voraus.

Über den *Baseboard-Management-Controller* (*BMC*) der Server kann der Gesundheitszustand der Serverhardware abgefragt werden. Mit Hilfe eines Nagios Plugins können diese Informationen an das Monitorsystem übertragen werden. Hierbei werden Temperaturwerte aller Serverkomponenten (CPU, RAM, Mainboard), der Stromverbrauch jedes Servers sowie die Drehzahl der Gehäuselüfter erfasst. Hardwareprobleme wie z.B. ein Lüfterausfall oder ein Temperaturanstieg können auf diese Weise zuverlässig erkannt werden. Als weitere Hardwareelemente werden die Festplatten sowie das *Raid System* überwacht. Das eingesetzte Plugin basiert auf dem *check_lsi_raid* Plugin, das von der Thomas Krenn AG entwickelt wurde. [110] Das Plugin wurde jedoch um die Überwachung der Festplattentemperatur, dem gehäuften Auftreten von Lesefehlern und S.M.A.R.T- Informationen erweitert. Eine Temperaturerhöhung über 40° Celsius, ein gehäuftes Auftreten von *Predictive Failure Counts* oder *S.M.A.R.T* Fehlern erzeugt als Hinweis für einen drohenden Festplattenausfall eine entsprechende Warnmeldung und fordert zum Tausch der betroffenen Festplatte auf. Eine defekte Pufferbatterie, Festplatte oder ein Problem im *Raid* führt zu einer kritischen Fehlermeldung. Weiterhin werden für alle Server die Auslastung von Festplatte, CPU und Arbeitsspeicher, die Prozesslast sowie der Netzwerkdurchsatz überwacht. Für die speziellen Serverdienste wurden geeignete Plugins gewählt, die den Betrieb des Datenbanksystems (*check_mysql_health.sh*), des Apache Webservers (*check_apache2*) oder des DNS-Servers (*check_bind.sh*) überwachen können. [73] [3] [2] Darüber hinaus wird die Erreichbarkeit des Studienportals über einen

simulierten Aufruf der Webseite, der Ablauf des *SSL Zertifikats* sowie die *DNSSec* Funktionalität überprüft und im Fehlerfall ein Alarm erzeugt. Um eine Trend-Entwicklung über einen längeren Zeitraum verfolgen zu können, werden alle Messwerte in einer Datenbank gespeichert und mit Hilfe von *PNP4Nagios* grafisch aufbereitet. [61]

7.6 Backup

Um Backups auf Dateiebene zu speichern, wurde ein separater Linux Server verwendet. Mit Hilfe von *rsync*, einem unter Linux und Unix Systemen weit verbreiteten Synchronisationsprogramm, werden die Dateien aller Server zentral auf den Backup-Server übertragen. *rsync* ist ein mächtiges Werkzeug, das Dateien zwischen Quelle und Ziel abgleichen kann. Hierzu bringt das Programm einen sehr effizienten Algorithmus (*Delta Kodierung*) mit, der unter anderem mit Hilfe einer *MD5* Hashfunktion, eine veränderte Datei (*Deltas*) aufspürt und nur diese überträgt. Dadurch wird sowohl das über das Netzwerk übertragene Datenvolumen als auch die Backupzeit deutlich reduziert. Dies ermöglicht es, Datensicherungen in kürzeren Zeitabständen durchzuführen, ohne den laufenden Betrieb zu beeinträchtigen. Für die Datenübertragung wurde hierbei nicht das unverschlüsselte *rsh* Protokoll, sondern das verschlüsselte *SSH* Protokoll gewählt. Auf dem zu sichernden Server erzeugt der Datenabgleich wegen der Überprüfung auf geänderte Dateien und wegen der Verschlüsselung zwar eine höhere Prozessorbelastung, diese kann aber aufgrund der großzügigen Hardwareausstattung der Server vernachlässigt werden. *Rsync* ist zwar äußerst umfangreich und robust, verfügt aber nicht über eine grafische Oberfläche oder eine Zeitsteuerung mit der man fortlaufende Datensicherungen erstellen könnte. Um inkrementelle Backups zu erstellen, musste daher ein kleines Programm geschrieben werden, das für jeden Server zu definierten Zeiten ausgeführt wird. Das Programm bezieht nicht alle Ordner in die Datensicherung mit ein. Da alle verwendeten Server Debian und Linux verwenden, kann das Skript unverändert und für alle Server eingesetzt werden. Von der Datensicherung ausgeklammert werden temporäre Dateien sowie Pseudoverzeichnisse des Betriebssystems *(/proc, /sys, /mnt*). Das Skript sichert die Daten auf dem Backup-Server in einem Ordner, der Datum und Uhrzeit sowie den Servernamen enthält. Als Basis für das inkrementelle Sichern wird das letzte Backup des Tages ausgewählt. Bei mehrmaligem Ausführen am gleichen Tag wird das jüngste Backup aktualisiert. Das letzte Tages-Backup wird komprimiert und für 7 Tage vorgehalten. Für die zeitgesteuerte Ausführung des Programms sorgt ein *Cron-Job*, der für jeden Server individualisiert ist. Um zu verhindern, dass während der Datensicherung für die Hauptaufgabe des Servers Rechenzeit verloren geht, werden alle automatisierten Backup-Prozesse mit niedriger Priorität ausgeführt.

Da Änderungen am DNS System nicht so häufig zu erwarten sind und die Server im Fehlerfalle schnell wieder rekonstruiert werden müssen, erfolgt die Sicherung der beiden Namensserver täglich auf Festplattenebene. Um den Regelbetrieb nicht zu beeinträchtigen, finden die Sicherungen für beide Server nachts und jeweils um eine Stunde zeitversetzt statt. Während das Festplattenabbild erstellt wird, kann es zum Verlust einiger Datenpakete kommen. Der zeitliche Abstand beider Sicherungen soll sicherstellen, dass trotz Datensicherung immer ein Namensserver erreichbar ist. Da die DNS Server mit geringer Festplattenkapazität (10 GB) ausgestattet sind, gelingt die Wiederherstellung und Wiederaufnahme des Regelbetriebs in weniger als fünf Minuten.

Das Monitoring-System protokolliert wertvolle Daten über Stromverbrauch, Gesundheitszustand, Systemauslastung sowie Fehler der Hardware. Ein Datenverlust muss daher vermieden werden. Die Messdaten ändern sich sekündlich während das Betriebssystem und die instal-

lierte Software eher unverändert bleiben. Es wird daher stündlich eine Datensicherung auf Dateiebene durchgeführt um alle Sensordaten zu sichern. Zusätzlich wird einmal pro Woche ein Festplattenabbild erstellt. Im Fehlerfall muss zunächst der Server aus dem Festplattenabbild wiederhergestellt werden. Anschließend können die gesicherten Sensordaten rückgesichert werden. Der Datenverlust beträgt maximal eine Stunde, der zeitliche Aufwand für die Wiederherstellung des Regelbetriebs zwischen vier und fünf Stunden. Da das Monitoring keinen direkten Einfluss auf die Verfügbarkeit des Studienportals hat, ist diese Zeitspanne akzeptabel.

Der Webserver als Herzstück des Studienportals benötigt ein differenzierteres Backup-Konzept. Zum einen muss im Fehlerfall die Rückkehr zum Regelbetrieb in möglichst kurzer Zeit möglich sein. Zum anderen darf die Datensicherung den laufenden Betrieb nicht behindern. Besonders unter Belastungsbedingungen kann dies problematisch sein. Da der Webserver neben dem Quellcode des Studienportals auch das *Datenbankmanagementsystem* samt Studiendaten enthält wurde hier eine zusätzliche Backup-Ebene eingefügt. Für das Datenbankmanagement System *MySQL* existiert das Programm *mysqldump,* mit dessen Hilfe Sicherungen der Datenbank erzeugt werden können. [216] Die wertvollen Studiendaten können auf diese Weise einfach aus dem Datenbanksystem extrahiert und gesichert werden. Hierzu steht dem Webserver eine separate Festplatte zur Verfügung. Diese basiert wie auch die Hauptfestplatte auf *SSD* Technologie und ist in der Lage viele Eingabe/Ausgabeoperationen durchzuführen und Daten besonders schnell speichern zu können. Diese Eigenschaft kommt der Datenbanksicherung zugute. Auf eine Komprimierung wurde zu Gunsten der Rechenleistung verzichtet. Um die Zeit für die Sicherung weiter zu verkürzen, werden *Dumps* nicht von allen Tabellen erstellt, sondern nur von denjenigen, welche relevante Teilnehmerdaten enthalten. In einem Zeitintervall von 30 Minuten werden sie auf die zweite Festplatte in einer *.sql* Datei gesichert, aus der sie sich im Ernstfall wieder rekonstruieren lassen. Die Datenbankexporte werden mit der aktuellen Uhrzeit versehen und einen Tag lang vorgehalten. Ein tägliches Festplattenabbild gewährleistet eine Historie über eine Woche. In der Testumgebung gelang die Sicherung unter Lastsimulation mit 200 gleichzeitigen Zugriffen auf das Studienportal in 10 Sekunden, ohne den laufenden Betrieb merklich zu beeinträchtigen. Für den Produktivbetrieb sollte jedoch der Einsatz eines Datenbank-Clusters erwogen werden. Hierzu müssen mindestens zwei *MySQL* Instanzen betrieben werden. Ein Server übernimmt die Rolle des *Masters*, der andere die Rolle des *Slaves*. *MySQL* kümmert sich darum, dass der Hauptserver die Daten in Echtzeit auf den *Slave* repliziert. Auf ihm kann die gesamte Datenbank zwecks Datensicherungen blockiert werden (*lock tables*). So ist es möglich eine Sicherung durchzuführen, ohne dass der Hauptserver davon beeinträchtig wird.

Analog zur Sicherung der anderen Server wird die Serverinstallation durch ein nächtliches Festplattenabbild gesichert. Dieses wird zu einem Zeitpunkt ausgeführt, zu dem mit wenigen Webseitenbesuchern (4:00 Uhr) zu rechnen ist. Eine Wiederherstellung des Servers ist dadurch in weniger als 10 Minuten möglich. Um Zugriffsstatistiken und Konfigurationsdateien der Serverdienste zu sichern, wird das bereits beschriebene Datei-Backup stündlich aufgerufen. Diese Strategie erlaubt ein möglichst engmaschiges Sichern von Daten des Studienportals auf unterschiedlichen Ebenen. Ein Wiederherstellen der Verfügbarkeit des Servers ist in kurzer Zeit möglich. Die Daten können auch auf einen Server geklont, auf einem weiteren Server wieder rekonstruiert und zur Not im laufenden Betrieb wieder rückgespielt werden.

Vor geplanten Wartungsarbeiten wie z.B. dem Einspielen von Software, Updates oder Änderungen von globalen Einstellungen werden *Snapshots* erzeugt. Diese protokollieren Änderungen, die am Festplattenimage vorgenommen werden ähnlich dem *rsync* Algorithmus. *Snap-*

shots fassen zudem den gesamten Inhalt des Arbeitsspeichers. [123] Mit Hilfe eines *Snapshots* kann demnach der aktuelle Betriebszustand des Servers eingefroren werden. Nach erfolgreich abgeschlossener Wartung kann der *Snapshot* wieder entfernt oder auch zur Sicherheit aufgehoben werden. Der Vorgang dauert je nach Größe der Festplatte, des Arbeitsspeichers und der Anzahl während des Vorgangs veränderter Dateien zwischen einer und fünf Minuten. Während des *Snapshots* reagiert der Server verlangsamt, ist jedoch weiterhin erreichbar. Daher sollten Wartungsarbeiten zu Zeitpunkten stattfinden, in denen die Server nicht oder nur wenig genutzt werden.

7.7 Benchmarks

7.7.1 DNS Server

Die Geschwindigkeit mit der DNS-Abfragen abgearbeitet werden können, wurde mit dem Softwarewerkzeug DNSPerf ermittelt. Hinter der Entwicklung steht Nominum Inc, ein großer Netzwerk- und Telekommunikationsanbieter. [93] Das Unternehmen, das vom Internationalen Internetkonsortium mit der Entwicklung des *DNS Servers* Bind 9 beauftragt wurde, bietet auch hochverfügbare *DNS* Infrastrukturen an. Das Benchmark Programm wurde mit dem Ziel entwickelt, die eigenen Dienste zu messen und mit anderen Anbietern wie Google oder CloudFlare vergleichen zu können. [217] Auf der Webseite des Projekts findet man neben dem Quellcode auch eine aktuelle Datenbank mit Statistik zur Verfügbarkeit und Performance aller großen DNS Anbieter. [217]

Das Messsystem wurde mit *DNSPerf 2.0.0* und *Bind 9.9.0* aus dem Quellcode auf einem Debian 7 Server kompiliert und installiert. *DNSPerf* misst die Anzahl der Anfragen, deren Laufzeit sowie eine errechnete Anzahl Transaktionen pro Sekunde. Insbesondere die Latenz ist ein entscheidender Parameter, da die Geschwindigkeit jedes Seitenaufrufs von der Geschwindigkeit der Namensauflösung abhängig ist. Um eine Aussage über die Leistungsfähigkeit der *DNS* Infrastruktur des Studienportals zu treffen, wurde sie mit den *DNS* Servern von Google verglichen. Die Tests erfolgten im Netz der Universität Regensburg und Würzburg sowie von einem regulären DSL-Anschluss aus (VDSL-100). Es wurden jeweils 10.000 Anfragen an die beiden *DNS* Server gestellt und deren Reaktion gemessen. Die Latenz des Namensservers des Studienportals liegt deutlich unter denen der Referenz. Auch zu den anderen großen *DNS* Dienstleistern ist die eigene Namensserver-Infrastruktur durchaus konkurrenzfähig. Der Statistik nach rangiert der Studienportal *DNS* Dienst mit 22,9ms knapp hinter den schnellsten beiden *DNS* Anbietern. [91] Im Vergleichstest bearbeitete der schnellste der vier Google *DNS* Server (*ns1.google.com*) etwa 1.737 Anfragen pro Sekunde mit einer Laufzeit von 56ms. Das Studienportal kann in der gleichen Zeit etwa 4.322 Anfragen mit einer Latenz von lediglich 23ms bearbeiten.

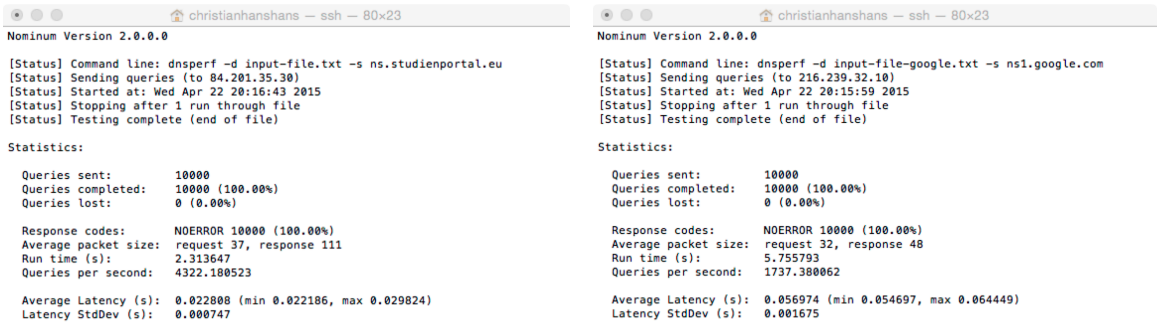


Abbildung 98: dnsperf Performance-Test des Studienportals im Vergleich zum Google DNS-Dienst

Dieser orientierende Test lässt sicher noch keine Aussage über Bedingungen unter hoher Last zu. Es scheint jedoch, dass die *DNS* Infrastruktur für den geplanten Einsatz ausreichend dimensioniert ist. Vor dem Einsatz im Produktivbetrieb sollte dennoch ein weiteres Benchmark unter simulierten Lastbedingungen durchgeführt werden.

7.7.2 Webserver

Um eine Aussage über die Leistungsfähigkeit des Prototypen treffen zu können, wurde das Studienportal einem Stresstest unterzogen. Das Benchmark misst die Zeit für den ersten Verbindungsaufbau (*ctime*), die Zeit, die der Webserver für die interne Abarbeitung der Anfragen benötigt (*wait*), die Dauer für die Übertragung der Webseite inklusive der Verarbeitungszeit (*dtime*), sowie die Zeitspanne, die für das Laden der Webseite aus Anwendersicht benötigt wird (*ttime*). [218]

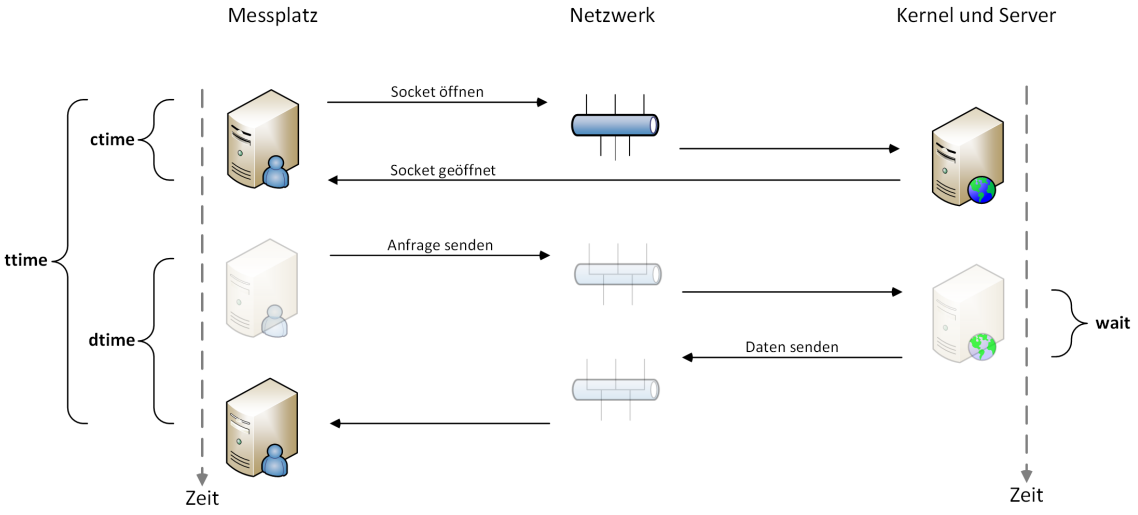


Abbildung 99: Schema des Messprinzips der Webserver-Tests

Als Messpunkt wurde ein vom Studienportal unabhängiger Linux Server (4-Core Intel Xeon CPU, 4GB Ram, 1 Gbit LAN) gewählt. Mit Hilfe des Programms *ApacheBench* wurden 400 gleichzeitige *TLS1.2* Verbindungen zum Testkandidaten aufgebaut und die Ergebnisse der Verbindungszeiten in einer Textdatei gespeichert. Insgesamt wurde der Test 4.000 mal wiederholt. Das wesentliche Leistungskriterium stellt die maximale Anzahl an Anfragen pro Sekunde dar, die vom Webserver verarbeitet werden können. Mit Hilfe der *requests per second* wird die Leistungsfähigkeit des Gesamtsystems (Webserver, Datenbanksystem und Netzwerk) beurteilt.

Schenkt man den Statistiken auf *internetlivestats.com* Glauben, werden pro Sekunde etwa 48.500 Suchanfragen an Google gestellt, 9.000 Tweets bei Twitter veröffentlicht und 99.800 YouTube Videos angeklickt. [219] Mit dem Messsystem konnten für Facebook 582 req/sec, YouTube 869 req/sec, Twitter 684 req/sec und Google 976 req/sec gemessen werden. Als größter messbarer Wert (und damit als Grenzen des Messsystems) wurden daher 1.000 *requests/seconds* angenommen. Die Differenz zwischen Messung und statistisch zu erwartender Leistungsfähigkeit der gemessenen Referenzsysteme hat mehrere Ursachen. Es gilt zu bedenken, dass die genannten Portale Lastverteilungsverfahren und Clustertechniken mit regionaler Zuständigkeit nutzen. Anfragen werden in der Regel nicht von einem zentralen Server, sondern von einem für die Region zuständigen Satellitensystem verarbeitet. [75] [220] [50] Mit dem Messsystem lässt sich demnach nie die Gesamtleistung genannter Plattformen messen. Zudem ist ein einzelner Rechner durch die Anbindung an das Internet sowie durch die Soft/Hardware (Netzwerkkarte, Betriebssystem und Datenbus, CPU) limitiert. Für die zu erwartende Leistungsfähigkeit des Studienportals ist das Messsystem jedoch ausreichend dimensioniert. Die Messfehler durch Limitierungen des Messsystems sind wenig relevant, da nicht ein einzelner absoluter Zahlenwert, sondern vielmehr die Leistungsfähigkeit im Vergleich zu etablierten Systemen von Interesse ist. Die primäre Fragestellung lautet: „Wie gut schlägt sich das Studienportal im Vergleich zu bestehenden Alternativen."

Der Aufruf der Messung erfolgte mit folgenden Parametern:

 # ab -c 400 -n 4000 -g result.txt -f TLSv12 https://<TESTKANDIDAT>

Das Resultat des Tests sind Messwerte, die protokolliert und im Zeitverlauf mit Hilfe von *GnuPlot* dargestellt wurden.

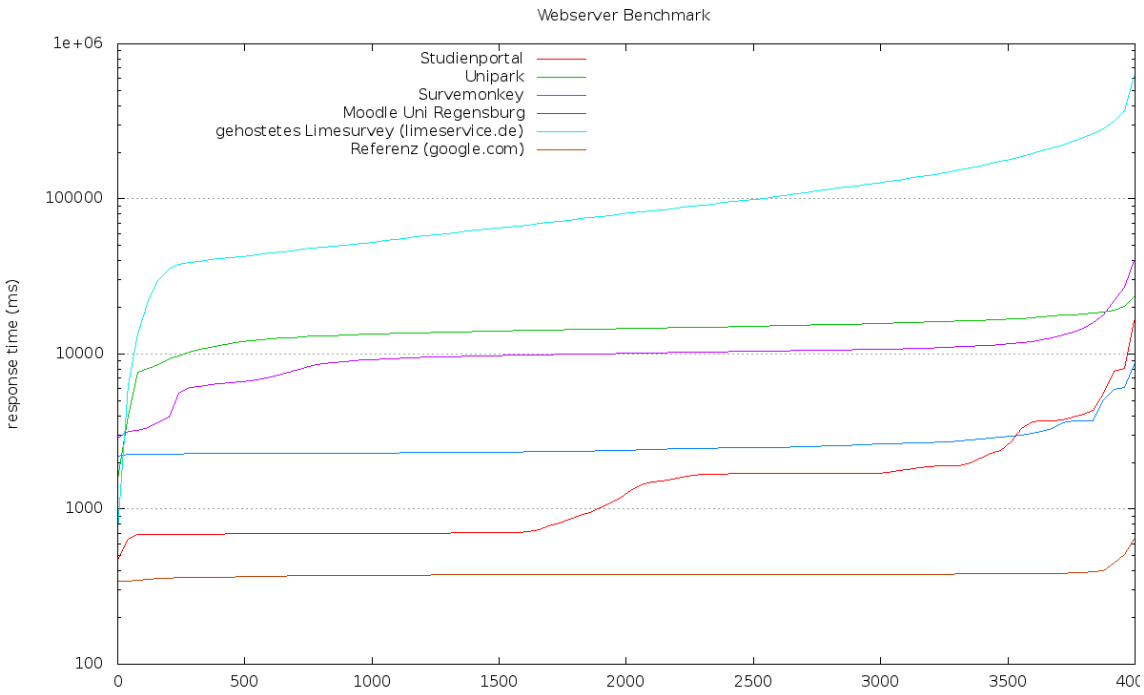


Abbildung 100: Benchmark mit 400 gleichzeitigen Verbindungen, Stand 25.04.2015

Im Diagramm wurde aufgrund der großen Differenz zwischen den einzelnen Antwortzeiten eine logarithmische y-Achse gewählt. Die x-Achse ist linear und umfasst die gestellten 4.000

Anfragen. Um möglichst realistische Vergleichswerte zu erhalten, wurden alle Tests nicht in der Hauptnutzungszeit durchgeführt, sondern nachts um 4:00 Uhr. In dieser Zeit ist eine möglichst geringe Verfälschung der Messung durch reguläre Nutzung der Dienste zu erwarten. Für die Messungen wurde von den untersuchten Fragebogen-Systemen von Unipark, Surveymonkey und LimeSurvey jeweils die URL eines im Netz frei verfügbaren Fragebogens verwendet um so den Seitenaufruf aus Nutzersicht nachzuahmen.
Idealerweise reagiert ein Server auch unter Last mit einer konstanten Antwortzeit unter einer Sekunde. Mit Ausnahme der Referenz (google.com) gelingt dies jedoch keinem Kandidaten. Surveymonkey zeigt ebenfalls eine stabile Kurve bei den Antwortzeiten. Das kommerzielle Fragebogensystem ist im Rechenzentrum der Amazon Cloud untergebracht, was die stabile Leistung aber auch die längeren Verbindungszeiten erklärt. Deutlicher Verlierer ist ein Anbieter, der sich auf LimeSurvey Hosting spezialisiert hat. Bereits nach wenigen Verbindungen steigt die Wartezeit auf über 10 Sekunden. Der kommerzielle Dienst ist bei mehr als zehn gleichzeitigen Verbindungen kaum zu gebrauchen.
Zum Vergleich der Moodle Installation des Studienportals mit einer anderen Moodle-Instanz wurde die Plattform der Universität Regensburg herangezogen. (Stand 04/2015)

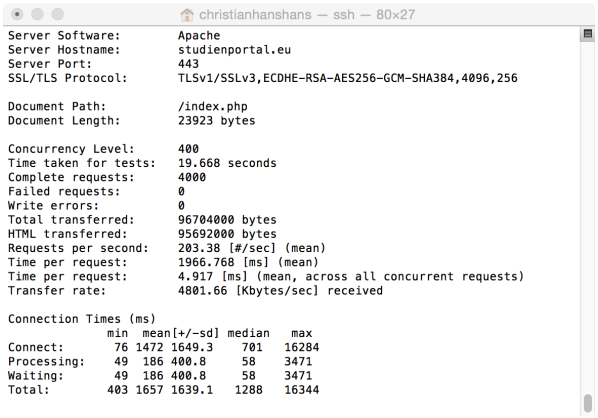


Abbildung 101: Performancetest Studienportal

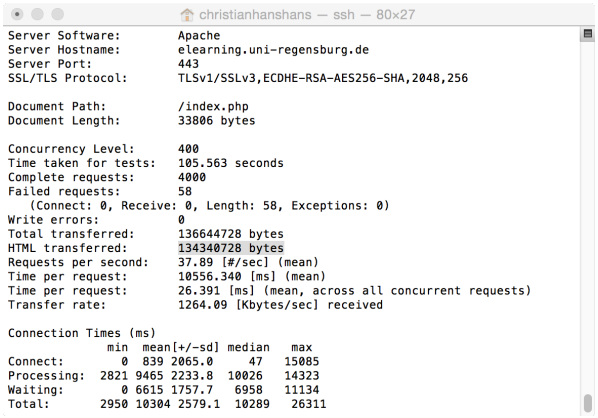


Abbildung 102: Performancetest Moodle Referenz

Auffällig ist beim direkten Vergleich zur Vergleichsinstallation die längere Zeitspanne für den Verbindungsaufbau (*mean connect*) zum Studienportal (1472ms v.s. 830ms). Sie lässt sich durch die Verwendung der SHA-384 statt SHA-128 Prüfsumme, Übertragung und Überprüfung des 4096bit statt 2048bit *SSL Zertifikats*, sowie durch den Mehraufwand für die verschlüsselte Übertragung der *DNS-Informationen* (*DNSSec*/*DANE*) erklären.
Der Größenunterschied der beim *Handshake* übertragenen Daten zwischen einem 4096bit und 2048bit Zertifikat liegt bei etwa 14%.

 # openssl s_client -connect studienportal.eu:443 -tls1
Server public key is 4096 bit
TLSv1/SSLv3, Cipher is DHE-RSA-AES256-SHA
*SSL handshake has read **4603 bytes***

 # openssl s_client -connect studienportal.eu:443 -tls1
Server public key is 2048 bit
TLSv1/SSLv3, Cipher is DHE-RSA-AES256-SHA
*SSL handshake has read **3953 bytes***

Betrachtet man nur die Reaktionszeit des Studienportals jeweils mit und ohne Einsatz von Verschlüsselung wird dies erkennbar. Mit steigender Anzahl an Verbindungen fällt der Aufwand für die Verschlüsselung zunehmend ins Gewicht und stellt einen leistungslimitierenden Faktor dar.

Abbildung 103: Antwortzeit in Anhängigkeit von Verschlüsselung und der Anzahl an Verbindungen

Dennoch kann das Studienportal auch mit *PerfectForwardSecrecy*, *DNSSec* und starken Schlüsseln im Schnitt 200 Anfragen pro Sekunde bearbeiten. In über 80% aller Fälle gelingt das Ausliefern der Webseite unter der Last von 400 gleichzeitigen Anfragen in unter 2 Sekunden. Eine Zusammenfassung der Leistungsdaten der Vergleichssysteme:

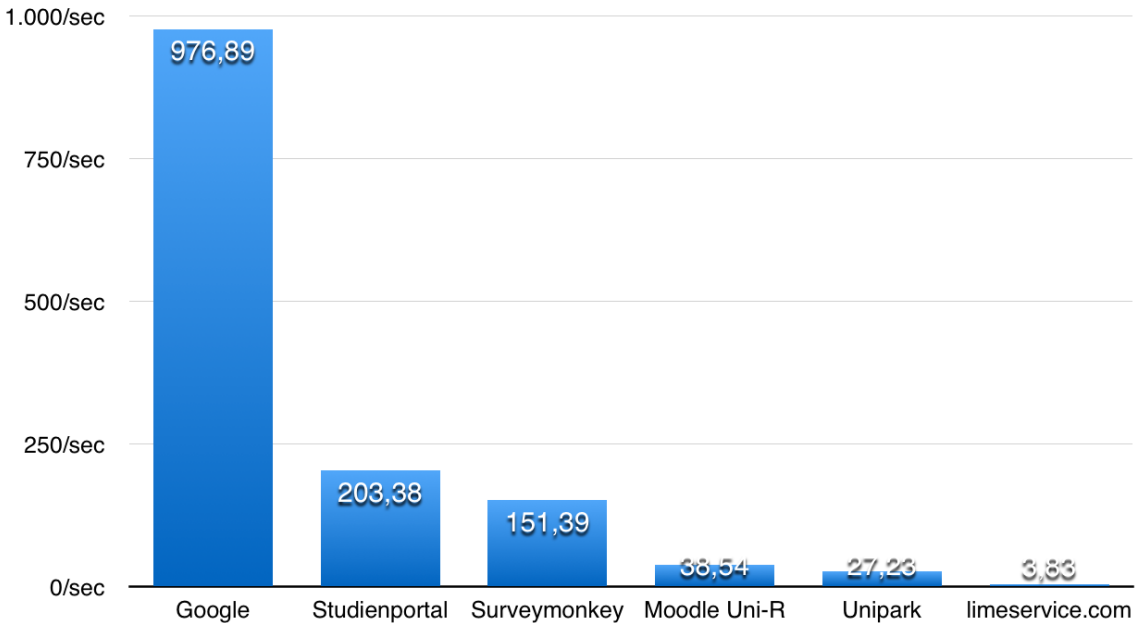


Abbildung 104: Performance der verglichenen Systeme in Anfragen pro Sekunde, Stand 25.04.2015

8 Diskussion

Das Problem bei der Durchführung von klinischen Studien im Bereich der psychosomatischen Medizin besteht nicht nur in der Rekrutierung entsprechender Studienpopulationen. Sicherlich stellt dies für die meisten Studien die größte Herausforderung dar. Jedoch konnte im Rahmen dieser Arbeit gezeigt werden, dass über rein organisatorische Probleme hinaus viele weitere wichtige Kriterien bei der Planung und Durchführung von internetbasierten Studien zu berücksichtigen sind. Diese beziehen sich zum Großteil auf nicht-medizinische Bereiche und sind daher für medizinisches und wissenschaftliches Personal nicht immer auf den ersten Blick zu erfassen. Dies beinhaltet Wissen und Erfahrung aus dem Kerngebiet des Marketings, erstreckt sich über juristische Probleme und gesetzliche Rahmenbedingungen bis hin zu komplexen technischen Details der Informatik. [133] Aus diesem Grund ist es nicht verwunderlich, dass bei der Vielzahl der zur Verfügung stehenden Werkzeuge, die nicht primär für den Einsatz in der Medizin konzipiert wurden, strukturelle oder konzeptionelle Mängel zu finden sind. Vor jedem Einsatz eines Softwarewerkzeugs im medizinischen Umfeld gilt es zu prüfen, ob es den gesetzlichen Rahmenbedingungen bezüglich des Datenschutzes, dem „Arztgeheimnis“ und den ökonomischen Rahmenbedingungen standhält. [145] [20] [28] In diesem Zusammenhang spielt auch die Konformität zu den *Leitlinien zur guten klinischen Praxis* (*GCP*) eine wichtige Rolle. Diese beschreiben zwar kein geltendes Recht, stellen jedoch den Standard wissenschaftlichen Arbeitens dar und werden nicht zuletzt deshalb von vielen Drittmittelgebern vorausgesetzt. In den „harmonisierten ICH Leitlinien für EU, Japan und die USA“ (*ICH-GCP*) finden sowohl Forderungen an organisatorische wie auch technisch valide Testverfahren, Qualitätszertifikate und Qualitätskontrollen Eingang. [221] Alleine die unsichere Rechtslage beim Einsatz von Dienstanbietern aus dem nicht-europäischen Ausland, die einen Großteil der kostengünstigen Fragebogensoftware ausmacht, sollte zur Entscheidung führen, Forschungsdaten nur unter strenger Kontrolle innerhalb Deutschlands zu erheben und zu speichern. [67] Für den willkürlichen Zugriff amerikanischer Behörden auf personenbezogene Daten reicht es bislang aus, wenn ein Unternehmen in den USA tätig ist oder eine Niederlassung dort besitzt. [16] [117] Auch wenn bisher mit Hilfe des „Safe Harbour Abkommens“ viele amerikanische Firmen, die ein „angemessenes Schutzniveau“ versprachen, der Europäischen Datenschutzrichtlinie (Richtlinie 95/46/EG) konform Daten aus Europa nach USA transferieren durften, so widerspricht dies dennoch in einigen Punkten dem Bundesdatenschutzgesetz. [222] [223] Spätestens mit der Entscheidung des europäischen Gerichtshofs vom Oktober 2015 wurde die Einschätzung der USA als sicherer Hafen für europäische Daten revidiert. Die Safe Harbour Legitimation verliert somit ihre Gültigkeit, so dass Software von Anbietern aus USA, Serverstandort USA oder mit Filiale in USA nicht mehr eingesetzt werden sollten. [95] [67] Dies trifft besonders auf Surveymonkey und Google zu.

Eine Hürde beim Betreiben von Datenbanken auf eigenen Infrastrukturen stellen die damit verbundenen finanziellen und personellen Kosten dar. Werden die technischen Maßnahmen an interne oder externe Dienstleister übergeben, müssen die Integrität der Daten und die Anonymität der Studienteilnehmer in besonderem Maß gewährleistet sein. Steht das benötigte Material (z.B. Server oder Rechenzentrum) nicht zur Verfügung, kann dies auch von externen Dienstleistern bezogen werden. Dies kann auch aus wirtschaftlichen Gründen sinnvoll sein. In jedem Fall muss eine besondere vertragliche Ausgestaltung stattfinden. [114] Die Investition in die eigene Infrastruktur, sei es Hardware oder Software, lohnt sich jedoch auf verschiedene Weise. Unter einem forschungsethischen Blickwinkel betrachtet, kann durch Transparenz zusätzliches Vertrauen und Akzeptanz und dadurch eine höhere Studienbeteiligung erreicht werden. Entscheidend ist, dass der Teilnehmer zu jeder Zeit verlässlich um den Speicherort und die Verwendung seiner Daten weiß, und diese auf Wunsch auch jederzeit löschen kann.

[34] In diesem Zusammenhang sollte auch nach Regeln der *Good Clinical Practise* eine leicht verständliche Einwilligungserklärung der Studienteilnehmer eingeholt werden.
Unter dem ökonomischen Gesichtspunkt kann eine gut konzipierte zentrale Plattform von mehreren Forschungsgruppen oder sogar Fakultäten genutzt werden. Einfache Bedienbarkeit und technisch ausreichende Dimensionierung vorausgesetzt, können Forschungsgruppen der Medizin (insbesondere der Psychosomatik) und der Psychologie Synergieeffekte nutzen, indem auf gemeinsam verwendbare Fragebögen oder gar Studienpopulationen zurückgegriffen wird. Der personelle Aufwand für die Erstellung einzelner Befragungsprojekte kann hierdurch signifikant reduziert werden. Wird die Plattform professionell verwaltet (interne oder externe Fachkraft), steigt potentiell die Sicherheit und Qualität der Datenhaltung im Vergleich zu arbeitsgruppeninternen Insellösungen. Wie im Rahmen dieser Arbeit bewiesen wurde, kann auch ohne den Einsatz lizenz- und kostenpflichtiger Software eine geeignete Studienplattform mit Hilfe von OpenSource Komponenten aufgebaut werden. Hierbei entsprechen die Kosten für die notwendige Technik in etwa den Lizenzkosten kommerzieller Software, für die übliche Weise zusätzlich Hardwarekosten anfallen. [46] Durch den Verzicht auf lizenzpflichtige Software kann ein nicht zu vernachlässigender Teil des Budgets gespart und z.B. in Personal investiert werden. In diesem Fall stünde das Personal für die Anwender als Anlaufstelle bei technischen Problemen zur Verfügung und wäre in der Lage im Bedarfsfall auch individuelle Anpassungen vorzunehmen. Dies wiederum erhöht die Flexibilität und den Komfort für die Anwender.

Bei der Planung und Umsetzung des Prototypen stellte sich heraus, dass nicht nur der Funktionsumfang der Webanwendung sondern auch die technische Infrastruktur auf der die Plattform betrieben wird, besonderer Aufmerksamkeit bedarf. Es wurde eine Infrastruktur nach aktuellem technischen Stand der Technik entworfen und umgesetzt, die zukunftsorientiert auf veränderte Situationen angepasst werden kann. Ein konkretes Beispiel hierfür ist die sich ständig ändernde Bedrohungssituation durch Sicherheitslücken in Softwarekomponenten oder die inflationäre Entwicklung von Verschlüsselungstechnologien. Dies spiegelt sich auch in den durchgeführten Sicherheitsanalysen wieder. Hierbei wurden bei nahezu allen untersuchten Angeboten kleinere oder größere leicht vermeidbare Mängel gefunden. Im Fall von Cloud-Diensten oder proprietärer Software besteht für den Anwender keinerlei Möglichkeit die Behebung der Mängel zu beeinflussen. Während der Entstehung dieser Arbeit wurden diverse Sicherheitslücken bekannt, die von einigen Betreibern bzw. von kommerziellen Diensten bis zum heutigen Tag nicht behoben wurden. [224] [225] [226] [227] [228] [229] [230] (☞ Kapitel 4.12) Umso wichtiger scheint die Möglichkeit selbst Hand anlegen zu können, was die Entscheidung für den konsequenten Einsatz von OpenSource auf eigener Hardware weiter legitimiert. Allerdings fordert dies zugleich regelmäßige Kontrolle und Eigeninitiative als Betreiber einer solchen Installation und nicht zuletzt die Bereitstellung qualifizierten Personals. Allerdings ist jeder Betreiber einer IT-Installation (unabhängig davon ob in Eigenregie betrieben oder ausgelagert) dazu verpflichtet die Erfassung personenbezogener Daten regelmäßig zu überprüfen, wodurch sich der zusätzliche Mehraufwand durch das Betreiben eigener Technik auf die tatsächlich durchgeführten Wartungsarbeiten reduziert. [145] Der Erfahrung seit Inbetriebnahme des Studienportals nach liegt das Verhältnis von Zeit für Kontrollaufgaben und Zeit für tatsächliche Wartung bei je 50%.
Durch die Anpassung der Grundinstallation der zugrunde liegenden Moodle Installation auf einem Linux Webserver war bereits ein Teil der Anforderungen an die geplante Studienplattform erfüllt. Die Fähigkeit auch mit großen Benutzerzahlen sicher umgehen zu können beweisen die Moodle Installationen vieler Universitäten (darunter auch die Universität Regensburg) täglich. [38] Die Benutzerverwaltung und das Rechtemanagement übernimmt das Lernmanagement System. Für die Gestaltung der Fragebögen befand sich leider kein Modul

im Funktionsumfang des quelloffenen Lernmanagement Systems welches den *GCP* Anforderungen standgehalten hätte. Durch die Ausrichtung von Moodle als Lernplattform steht der Personenbezug zwischen Dozent zu Student bzw. dessen Aktivitäten im Vordergrund. Eine anonymisierte Speicherung von Daten ist daher primär nicht vorgesehen. Es existieren zwar Erweiterungen, die es erlauben Fragebogenelemente zu gestalten, jedoch genügen diese nicht den gestellten Anforderungen der pseudonymisierten Speicherung von Daten. [221] Aus diesem Grund musste eine Erweiterung in Moodle implementiert werden, die sich zudem möglichst nah an dem analogen Pendant anlehnt. Sowohl die Erzeugung der Pseudonyme als auch die Speicherung der pseudonymisierten Datensätze erfolgt derzeit auf dem Server des Studienportals. Zwar ist das Pseudonym kryptografisch abgesichert, nichts desto trotz wäre eine *informationelle Gewaltenteilung* im Sinne von Trennung der Daten von Pseudonymisierung wie bei konventionellen Studien wünschenswert. [28] [20] Entsprechende technische Verfahren wären durchaus denkbar, wenn auch deutlich aufwändiger, im Vergleich zur derzeitigen Umsetzung.

Eine weitere Schwierigkeit bestand in der Integration des Schulungs-Videos für die zu erlernende Entspannungstechnik. Da Videos immer in einem bestimmten Format vorliegen, jedoch nicht alle Endgeräte zu jedem Format kompatibel sind, ergibt sich hieraus ein praktisches Problem. Das in Moodle bereits integrierte Plugin zum Einbetten von YouTube Videos wäre technisch betrachtet die einfachste Lösung, da das Videoportal in der Lage ist Videos im jeweils passenden Format an alle Endgeräte auszuliefern. Allerdings scheiden die von Moodle unterstützten, in USA ansässigen, Videoportale aus Datenschutzgründen aus. Werden Videos über Moodle ohne die Verlinkung auf ein externes Videoportal eingebunden, können diese nur in einem einzigen Format hinterlegt werden. Aufgrund dieser Einschränkung musste eine technische Lösung gefunden werden, die Computer, Smartphones oder Tablets mit dem Video im geeigneten Format versorgt. Es wurde daher ein Programm entwickelt das zunächst den Gerätetyp und dessen Fähigkeiten bestimmt und darauf hin das passende Format abspielt. Die Integration des Videos in das Studienportals erfordert eine gewisse manuelle Vorarbeit. Zunächst muss das Originalvideo in drei gängige Videoformate umgewandelt und auf den Server übertragen werden. Für die Videoumwandlung muss ein externes Videoprogramm eingesetzt werden. Glücklicherweise gibt es eine Vielzahl kostenfreier Konvertierungssoftware, die aus dem Ursprungsvideo drei unterschiedliche Dateien erzeugen. Die drei Videodateien (*.mp4, *.ogv, *.webm) müssen auf den Server übertragen und in einem gemeinsamen Ordner gespeichert werden. Über einen Link (vergleichbar der Integration von YouTube Videos) kann nun auf den Speicherort des Videos verwiesen werden. Der Link enthält zwei Angaben, die es ermöglichen, unterschiedliche Videos im selben Ordner oder auch mehrere unterschiedliche Videoordner zu verwenden. Auf diese Weise ist das Einbinden weiterer Videos möglich. Die manuellen Teilschritte (Videokonvertierung, Video Upload, Video Verlinken) können eine technische Hürde und potentielle Fehlerquelle darstellen. Eine komfortablere technische Lösung in Form eines speziellen Moodle Plugins wäre daher durchaus wünschenswert. Hierzu müsste der Server jedoch über eine entsprechende Konvertierungssoftware verfügen, die die Ursprungsdatei beim Hochladen automatisch konvertiert, speichert und die Verlinkung vornimmt. Unter Berücksichtigung der Tatsache, dass das Video für die SURE-Studie nur einmal umgewandelt werden muss und anschließend für alle Teilstudien verwendet werden kann, schien der hierfür notwendige Programmieraufwand unverhältnismäßig.

Um den Aufwand des manuellen Rekrutierens und Zuordnens von Teilnehmern zu den jeweiligen Studiengruppen zu reduzieren, wurde vom Konzept der *Landing-Pages* Gebrauch gemacht. Durch diese individuellen Startseiten ist es möglich mit Werkzeugen des Online-Marketings wie z.B. durch Suchmaschinenoptimierung, individuelle Gestaltung und Integra-

tion sozialer Netzwerke potentielle Teilnehmer über Internetkanäle zu gewinnen. Diese können sich über die eigenständige Webseite, die idealerweise auch über eine eigene Internetadresse verfügt (z.B. sure-bundeswehr.de) selbstständig in die Studie eintragen. Für die Verknüpfung zwischen der Studiendatenbank und der *Landing-Page* sorgt das Registrierungsformular. Im Rahmen der Arbeit wurde eine Möglichkeit vorgestellt, wie an jede beliebige Internetseite das Registrierungsformular ohne Programmierkenntnisse angefügt werden kann, jedoch ist das optische Erscheinungsbild durchaus verbesserungswürdig. Homepagebaukästen für den heimischen Computer, Content-Management-Systeme oder *Landing-Page Generatoren* gibt es in vielfältiger Ausführung kostenpflichtig und kostenfrei im Internet. Das generische Formular kann zwar optisch noch weiter ausgebaut werden, jedoch dürften *Landing-Pages* mit individuell gestaltetem und an das Design angepasste Registrierungsformular eine deutlich höhere Akzeptanz bei potentiellen Studienteilnehmern und dadurch eine höhere Konversionsrate erzielen. Dieses Phänomen lässt sich beim Einsatz von *Landing-Pages* vielfach beobachten. [8] Dieser Vorteil wird nicht nur von Unternehmen, sondern auch zunehmend von universitären Einrichtungen ausgenutzt. [159] [32] Allerdings erfordert die individuelle Gestaltung einer solchen Startseite grundlegende Kenntnisse in den Programmiersprachen HTML, CSS und PHP. Für den geübten Webdesigner sollte die Anbindung einer *Landing-Page* mit individuellem Registrierungsformular jedoch kein Problem darstellen.

Das Studienportal wurde mit der Vorgabe konzipiert leicht erweiterbar und fehlertolerant zu sein. Aus diesem Grund wurde besonderer Wert auf die Konzeption gelegt. Das Konzept berücksichtigt neben der Auswahl adäquater Hardware auch Hochverfügbarkeitstechniken. Mit Hilfe eines Verbunds aus mehreren Servern können Ausfallrisiken deutlich verringert und dadurch eine größtmögliche Verfügbarkeit und Leistungsfähigkeit der Studienplattform gewährleistet werden. Da neben einer hohen Verfügbarkeit auch ein Höchstmaß an Sicherheit eine fundamentale Rolle spielt, wurden in die Planung aktuellste Sicherheitstechnologien einbezogen, auch wenn diese derzeit noch nicht die volle Unterstützung auf der Anwenderseite mitbringen. Diese erstrecken sich von Techniken zur Vermeidung von Absenderfälschung und Spam für den Mailtransfer (*SPF*, *Graylisting*, *DKIM*) über vollständige Verschlüsselung des Transportwegs (*DNSSec*, *DANE/TLSA*, *Perfect Forward Secrecy*) bis hin zu Schutzmaßnahmen vor Angriffen auf das Studienportal (*XSS*, *SQL Injects*, *DoS, Intrusion detectsion system*). Ebenso wurden Überwachung und Früherkennung möglicher Hardware und Softwareprobleme durch die Installation eines Monitoring Systems berücksichtigt.
Trotzdem konnten alle geplanten und gemäß dem Konzept umgesetzten Konzepte lediglich unter Laborbedingungen getestet werden, da bisher keine Studie über das Studienportal abgewickelt wurde. Unter simulierten Belastungssituationen jenseits der realistisch zu erwartenden Nutzung konnte die Leistungsfähigkeit, Stabilität und Sicherheit aller Komponenten zwar bestätigt werden, jedoch existieren bislang keine verwendbaren Erfahrungswerte.

9 Ausblick

Auch wenn das Studienportal in seiner derzeitigen Form optisch und technisch weit fortgeschritten scheint, so bleiben dennoch einige technische Verbesserungsmöglichkeiten und Ergänzungswünsche, die sich bei jeder Software typischer Weise im Laufe des Entwicklungsprozesses oder während der Nutzung ergeben.

Hierzu zählt zum Beispiel der Wunsch einer komfortableren Gestaltung von *Landing-Pages*. Zwar können mit Hilfe externer Softwarewerkzeuge (Homepagebaukästen) oder sogar aus Microsoft Powerpoint Webseiten generiert werden, diese verfügen dann jedoch noch nicht über ein individuelles Registrierungsformular, das mit der Studienplattform verknüpft werden könnte. Wird ein individuelles Registrierungsformular gewünscht, ist derzeit die manuelle Integration von PHP-Code und Javascript notwendig. Es wäre jedoch vorstellbar Design-Vorlagen (Templates) für Content Management Systeme (zB.: Wordpress) zu erstellen und eine Erweiterung (Plugin) zu entwickeln, die ein mit dem Studienportal verknüpftes Registrierungsformular in die Webseite integriert. Sowohl Wordpress als auch Templates und Module lassen sich einfach über eine Weboberfläche installieren. Auf diese Weise könnten neue *Landing-Pages* von wissenschaftlichen Mitarbeitern bequem per Weboberfläche aufgesetzt werden. Als Nebeneffekt der Verwendung, eines auf Webblogs ausgelegten *CMS,* ist die gute Anbindung an soziale Netze und eine bessere Reichweite im Internet.

Eine weitere wünschenswerte Verbesserung wäre die einfachere Integration von Videos innerhalb der Studien. So könnte die bereits entwickelte Browserweiche mit Hilfe eines eigenen Moodle Moduls beim Hochladen, Konvertieren und Abspeichern der Videos behilflich sein. Hierzu muss der Server über ausreichende Rechenleistung und Videokonvertierungssoftware verfügen, um die Umwandlung in die gängigen Videoformate durchzuführen. Analog zum Erstellen eines Fragebogens kann der Benutzer über ein Datei-Upload Verfahren das Video von seiner Festplatte auswählen. Nach abgeschlossenem Upload könnte das Plugin automatisch das Umrechnen und Einbetten in die Studie übernehmen. Dies wäre eine deutliche Komfortsteigerung. Ein alternativer Weg wäre der Einsatz eines eigenständigen Medienservers. Viele Universitäten besitzen bereits fakultätsübergreifend eine solche Infrastruktur. Der Vorteil besteht darin, dass für das Studienportal weder Festplatten noch Bandbreitenressourcen für das Abrufen der Videos verloren gehen. Für die gängigen kommerziellen Streaming Server gibt es passende Moodle Erweiterungen von den Herstellern, über die eine Einbindung des Medienservers in das Studienportal erfolgen kann. Die Umwandlung und Bereitstellung erfolgt dann durch den Medienserver. Ein weiterer Vorteil besteht in der Auslagerung von Wartungs- und Datensicherungstätigkeiten. Jedoch gilt auch hierbei zunächst zu prüfen, ob die Anbindung an ein universitäres Mediensystem mit den Datenschutzanforderungen in Einklang zu bringen ist. Sollte die Fakultät über eine entsprechende Plattform nicht verfügen, oder aus rechtlichen Gründen eine gemeinsame Nutzung nicht in Frage kommen, wäre auch das Betreiben einer eigenen Videoplattform denkbar. Hierbei wurden an unterschiedlichen Hochschulen bereits Erfahrungen gesammelt, auf die man zurückgreifen kann.

Dank des *Responsive Design* ist die Darstellung des Studienportals und seiner Inhalte (Navigation und Fragebögen) bereits auf die Verwendung mit mobilen Endgeräten ausgelegt.
Um jedoch noch besser auf die speziellen Bedürfnisse von Smartphones und Tablets eingehen zu können, wäre die Erstellung einer App sinnvoll. Eine wesentliche Einschränkung bei mobiler Nutzung stellt die begrenzte Bandbreite sowie ein limitiertes Datenvolumen dar. Eine Studienportal App könnte große und wenig veränderliche Daten (Dokumente und insbesondere Videos) offline verfügbar machen. Nutzer könnten dann auch mit eingeschränkter Internet-

verbindung weiterhin mit den Studienunterlagen arbeiten. Fragebogendaten könnten bei schlechter Verbindung lokal auf dem Gerät zwischengespeichert und bei guter Verbindung übertragen werden. Durch ein spontaneres Ansprechen einer App im Vergleich zu einer konventionellen Webseite sowie beschriebener Mehrwerte kann die Anwenderzufriedenheit und Adhärenz weiter erhöht werden.

Um für weitere sozialwissenschaftliche Forschung auf eine bereits bestehende Studienpopulation zurückgreifen zu können, sollte das Studienportal um eine Panel-Funktion erweitert werden. Panels bestehen generell aus registrierten Benutzern, die für weitere Datenerhebung zur Verfügung stehen. Studienteilnehmer sollten während oder nach Beendigung einer Studie gefragt werden, ob sie bereit wären an weiteren Studien oder wissenschaftlichen Forschungsprojekten teilzunehmen. Durch die Einverständniserklärung kann für weitere Studien eine Studienpopulation allein durch Rückgriff auf die Datenbank zusammengestellt werden. Allerdings setzt dies auch detaillierte Angaben von personenbezogenen Informationen im Teilnehmerprofil voraus. Möchte man alle nicht-verheirateten, männlichen Nicht-Akademiker mit einem technischen Beruf im Alter zwischen 25 und 35 Jahren selektieren, sind im Vergleich zu den für die SURE-Studie benötigten Daten deutlich mehr personenbezogene Informationen von Nöten. Diese zusätzlichen Daten müssen vom Teilnehmer (nach dessen Einverständnis) abgefragt und gespeichert werden. Auf diese Weise kann für künftige Studien mit einer deutlichen Kosten-/Zeitminderung, Flexibilität, schnellen Umsetzbarkeit und einer höheren Repräsentativität gerechnet werden. Mit wachsender Anzahl an Studien steigt das Potential der Plattform. [130] Wie dies technisch umgesetzt werden kann, zeigt das *Hamburger registration and organization online tool* (hroot) auf eindrückliche Weise. Die einfache und benutzerfreundliche Oberfläche ermöglicht auf einfache Art das Selektieren von Teilnehmern, die Kommunikation mit potentiellen Teilnehmern und die zeitliche Planung von Onlinebefragungen. [12] Nachdem das Projekt als OpenSource Software kostenfrei zur Verfügung steht, wäre eine Anbindung an die Datenbank oder gar eine Integration in das Studienportal denkbar, insbesondere da beide Systeme sich in der grafischen Gestaltung sehr ähneln.

In der aktuellen Version des Studienportals erfolgt die Erzeugung der Pseudonyme und die Speicherung der Studiendaten am selben Ort (auf dem Server des Studienportals). Die Zuordnung von Fragebogen zu Teilnehmer ist durch ein entsprechendes Schutzkonzept ausgeschlossen. Keine der im Studienportal angemeldeten Rollen kann personenbezogene Fragebogeninformationen einsehen. Für Manager, Studienersteller, Studienbetreuer und Studienassistenten handelt es sich demnach um anonyme Daten. Nur das System kann mit Hilfe eines Algorithmus den Bezug von Fragebogen zu Nutzer wieder herstellen. Hierbei erfolgt sowohl die Generierung des Pseudonyms als auch die Speicherung der pseudonymisierten Daten auf dem Studienportal. In Analogie zu konventionellen Studien, bei denen eine externe Instanz die Pseudonymisierung vornimmt, kann die „informationelle Gewaltenteilung“ auch in elektronischer Form erfolgen. [28] Dies ist insofern sinnvoll, als dass durch den Zugriff auf den Quellcode (durch Administrator oder Hacker) der Algorithmus offen liegt. Mit Hilfe des Algorithmus, der Zugriffsmöglichkeit auf die Datenbank und entsprechender Rechenleistung ist eine Entpseudonymisierung denkbar. Um dies zu verhindern kann die Erzeugung von Pseudonymen und Daten an getrennten Orten stattfinden. So kann zum Beispiel ein zweiter Server (*Token Server*) die Pseudonymisierung übernehmen. Um ein Pseudonym abzufragen, müsste sich das Studienportal zunächst elektronisch beim *Token Server* ausweisen. Anschließend kann das Studienportal seine Anfrage stellen und einen Datensatz übertragen. Aus diesem Datensatz erzeugt der *Token Server* das Pseudonym, das er dem Studienportal zur Verfügung stellt. Auf diese Weise ist sichergestellt, dass die Daten selbst bei Zugriff auf die Datenbank anonym bleiben. Der *Token Server* beherbergt keine sensiblen

Daten, sondern lediglich eine Rechenvorschrift, nach der abhängig von der übermittelten Information, Zahlen und Buchstabenkombinationen erzeugt werden.

Nachdem das Studienportal bislang nicht für eine reale Studie genutzt wurde, sollte der Einsatz zunächst an einer Studie mit einer überschaubaren Studienpopulation erfolgen. Diese Test-Studie sollte die gleichen Anforderungen wie die im größeren Maßstab geplante SURE-Studie stellen. Daher sollte für diese Studie zunächst eine individuelle *Landing-Page* entworfen und die Studie über alle verfügbaren Kanäle beworben werden. Über die Rücklaufquote und die Bedeutung von Social Media und viralem Marketing lassen sich wertvolle Informationen und Erfahrungen gewinnen. Besonders wichtig ist in dieser Phase der enge Kontakt zu den Anwendern (Studienpersonal wie Teilnehmer) und deren Feedback. Die Nutzer sollten aufgefordert werden Fehlermeldungen oder Verbesserungswünsche zu äußern. Diese müssen, sofern sinnvoll und technisch machbar, zeitnah umgesetzt werden. Von diesem Lerneffekt profitiert die Qualität und Praxistauglichkeit des Systems, es schafft Vertrauen bei den Studienbetreuern und ermöglicht die Einarbeitung bei einer noch überschaubaren Arbeitsbelastung. Auf der technischen Seite ermöglichen die gemessenen Daten eines aktiv genutzten Systems eine viel bessere Prognose als die bislang durchgeführten Simulationen. Durch die erhobene Serverstatistik kann viel über die bevorzugt genutzten Endgeräte und deren technische Ausstattung gelernt werden. Daraufhin kann speziell für diese Endgeräte eine Optimierung des Portals erfolgen. Die Nutzungsverteilung im Tagesprofil ermöglicht das beste Zeitintervall für Wartungs- oder Datensicherungsaufgaben herauszufinden. Über die erzeugte Datenübertragung lassen sich benötigte Volumenkontingente und Bandbreite besser abschätzen. Schlussendlich dient dieser Lernprozess der Vorbereitung auf großangelegte Studien wie SURE mit mehreren tausend Nutzern und vielen gleichzeitigen Zugriffen. Im laufenden Betrieb sind grundlegende Änderungen oder Probleme schwieriger zu handhaben. Daher sollte der reibungslose Betrieb gewährleistet sein, bevor das Portal in den Regelbetrieb übergeht.

Ein wichtiger Punkt ist die Transparenz im Umgang mit den erhobenen Daten und deren Schutz. Die bei der Umsetzung des Studienportals getroffenen Sicherheitsmaßnahmen liegen weit über dem Maß der gängigen Praxis. Dies sollte als Argument genutzt werden um das Vertrauen zu fördern und verunsicherte oder misstrauische Teilnehmer zu überzeugen. Hierfür empfiehlt es sich ein kurzes Video (z.B. auf der Startseite) zu platzieren, welches das Portal, den Zweck der Datenerhebung und die sicherheitstechnischen Maßnahmen zum Schutz der Privatsphäre kurz und verständlich präsentiert. Kurze Animationssequenzen und Schaubilder können dabei helfen die groben technischen Prinzipien verständlich zu machen, ohne lange Studienbeschreibungen oder Hinweise lesen zu müssen. Darüber hinaus sollte den Teilnehmern erklärt werden, wie sie sich vor Missbrauch schützen und das Studienportal sicher nutzen können. Ein Beispiel hierfür wäre die Empfehlung eine aktuelle Version des Firefox Browsers zu nutzen und die Installation des DNSSec Validator Plugins durchzuführen, um somit die Echtheit des Studienportals sowie die unverfälschte und sichere Datenübertragung überprüfbar zu machen.

10 Zusammenfassung

Die vorgestellte Arbeit beschreibt die Konzeption einer elektronischen Studienplattform zur Durchführung umfangreicher internetbasierter Studien in der psychosomatischen Medizin. Als Anwendungsfall dient die SURE-Studie, in der die Schlüsselmethode, eine schneller zu erlernende Alternative zu etablierten Entspannungstechniken, untersucht werden soll. Zu diesem Zweck sollen mehrere tausend Teilnehmer aus Berufsgruppen mit hoher berufsbedingter, psychischer Belastung ausgewählt werden. Studienteilnehmer sollen zunächst über eine Onlineplattform die neue Entspannungstechnik per Videoanleitung erlernen. Anschließend soll über einen Beobachtungszeitraum von etwa einem Jahr der präventive Charakter der Entspannungsmethode zur Vermeidung stressbedingter Folgeerkrankungen untersucht werden.
Im Wesentlichen setzt sich die Arbeit mit organisatorischen, rechtlichen, technischen und ökonomischen Problemen auseinander, die es zu lösen gilt, wenn sozialwissenschaftliche Studien auf elektronischem Weg durchgeführt werden. Analysiert wurden Hürden bei der Teilnehmerrekrutierung, der elektronischen Bereitstellung der Studienunterlagen sowie Erhebung der Daten unter Einhaltung gesetzlicher Rahmenbedingungen und nach Maßgaben der *Good Clinical Practice*. Es wurden häufig verwendete kostenfreie und kommerzielle Softwarewerkzeuge unter dem Aspekt des benötigten Funktionsumfangs, des Datenschutzes, von Ergonomie und nicht zuletzt ökonomischen Einschränkungen untersucht. Hierbei wurden auch in einschlägiger Literatur erwähnte Panels genauer betrachtet. Nachdem sich die großen, kommerziellen Panel-Lösungen aus Komplexitäts- und Kostengründen und die kostengünstigeren Cloud-basierten Varianten aus fehlender Skalierbarkeit, unzureichendem Funktionsumfang oder aus Datenschutzgründen für den geplanten Einsatz als untauglich erwiesen haben, erfolgte die Planung einer eigenen Onlineplattform. In der Konzeption wurden die Schwachstellen bestehender Systeme berücksichtigt und eine Infrastruktur gemäß den Empfehlungen des Bundesamtes für Sicherheit in der Informationstechnik (BSI) und den Anforderungen an das Bundesdatenschutzgesetz (BSDG) geplant. Die vielfältigen, sich hieraus ergebenden Teilprobleme wurden im Detail beleuchtet und technische Lösungsstrategien entwickelt. Für die praktische Umsetzung eines Prototypen der Onlineplattform wurde als Grundlage hierzu das quelloffene Lernmanagementsystem (Moodle) herangezogen und um die benötigten Funktionen erweitert. Durch die entwickelten Erweiterungen ist die Plattform in der Lage, Teilnehmer über individualisierte Startseiten (*Landing-Pages*) automatisch in Studiengruppen zu erfassen, zu pseudonymisieren, Studienunterlagen und Medien für konventionelle und mobile Endgeräte bereitzuhalten und Onlinebefragungen durchzuführen. Studien und Fragebögen können über eine Weboberfläche aus Vorlagen oder neu erstellt und verwaltet werden. Ein Exportieren der gewonnenen Daten erlaubt die Auswertung mit gewohnter Software.
Als technischer Unterbau der Studienplattform wurde im Sinne einer *Best Practice* eine hochverfügbare technische Infrastruktur (HA-Cluster) aufgebaut, die in Punkto Sicherheit dem neusten Stand der Technik entspricht. Das Gesamtkonzept sieht vor, dass auch in Zukunft teilnehmerstarke Studien mit etlichen tausend Teilnehmern und ressourcenintensiven Anwendungen (z.B. dem Abspielen von Videoclips) möglich sind. Um die Tauglichkeit der Studienplattform für den geplanten Einsatz (SURE) zu überprüfen wurden Bedrohungsanalysen gemäß IT-Grundschutz, sowie Tests zur Messung der Leistungsfähigkeit des Systems (Benchmarks) durchgeführt. Die Testergebnisse zeigten, dass die erstellte Plattform auch mit einer Vielzahl gleichzeitiger Nutzer zurechtkommt und den meisten kommerziellen Angeboten insbesondere in Punkto Sicherheit überlegen ist. Letztlich bleibt das entworfene System die Praxiserfahrungen noch schuldig. Vor dem Einsatz in einer großangelegten Studie wie SURE, sollte die Plattform zunächst im Rahmen einer Teststudie in kleinerem Umfang erprobt und ihre Alltagstauglichkeit evaluiert, Wünsche von Anwendern aufgenommen, sowie umgesetzt und mögliche Fehler eradiziert werden. Nach erfolgreichem Test steht einer Um-

setzung des Prototypen in ein Produktivsystem, gegebenenfalls mit Unterstützung von Fördermitteln, nichts mehr im Wege.

11 Anhang

11.1 Abbildungsverzeichnis

Abbildung 1: Studienunterlagen der SURE-Studie 2010 7
Abbildung 2: TOP Modell zur Maßnahmenhierarchie übertragen auf IT Systeme 10
Abbildung 3: Beispiel einer Risikomatrix zur Datensicherheit 12
Abbildung 4: aktuelle Durchführung sozialwissenschaftlicher Studien 14
Abbildung 5: WiSo Panel der Universität Freiburg verwendet Unipark unverschlüsselt 16
Abbildung 6: Studienportal der Universität Mannheim, Basel, Ulm 16
Abbildung 7: Surveymonkey, ein Onlinefragebogenwerkzeug mit Serverstandort USA 17
Abbildung 8: Cookies zur Analyse des Benutzerverhaltens 18
Abbildung 9: Erstellung eines Fragebogens mit SosSci Survey 19
Abbildung 10: Panelfunktion mit Kontrolle des Rücklaufs auf Email-Rekrutierung 19
Abbildung 11: EvaSys, ein häufiges Werkzeug für universitätsinterne Umfragen 21
Abbildung 12: Fragebogenerstellung mit dem kostenlosen LimeSurvey 23
Abbildung 13: Onlinefragebögen mit Google Forms 24
Abbildung 14: Sphinx Desktop Erstellung und Auswertung von Umfragen 25
Abbildung 15: Universität Heidelberg verwendet die Eigenentwicklung WebLab 26
Abbildung 16: WiSo Panel - eine Eigenentwicklung der Universität Freiburg 27
Abbildung 17: Einladung zur Teilnahme an einer Studie des WiSo Panels mit Unipark 27
Abbildung 18: hroot ein Programm zur Rekrutierung und Verwaltung von Teilnehmern 28
Abbildung 19: Vereinfachte UseCases des Studienportals im Überblick 36
Abbildung 20: Schema des Registrierungsprozesses 37
Abbildung 21: Schema des Abrufs von Studienunterlagen 38
Abbildung 22: Schema Abruf und Auswertung von Studiendaten 38
Abbildung 23: Funktionsweise der Informationsverbreitung über Social Media 40
Abbildung 24: Metaanalyse Studienkosten 47
Abbildung 25: Schema verschlüsselter Datenübertragung 52
Abbildung 26: vereinfachter Handshakeprozess 53
Abbildung 27: Prozessorlast und Rechenzeit verschiedener Schlüsseltauschverfahren 55
Abbildung 28: Prinzip der Pseudonymisierung 60
Abbildung 29: Beispiel eines Session Cookie von Yahoo.com 63
Abbildung 30: Hierarchie der Einflussfaktoren zur Datensicherheit 65
Abbildung 31: Mirroring Prinzip von Nodes entspricht Raid 1 66
Abbildung 32: Shared All links Regelbetrieb, rechts Ausfall von Knoten A 67
Abbildung 33: Cluster-Dateisystem 68
Abbildung 34: Funktionsprinzip externen Monitorings (Webserver, Mailserver) 76
Abbildung 35: Schematischer Aufbau eines LAMP Servers 80
Abbildung 36: Startseite des Studienportals 82
Abbildung 37: Optimierung für die Nutzung von mobilen Endgeräten 83
Abbildung 38: Anmeldemaske für registrierte Benutzer 84
Abbildung 39: Einfache Anpassung der Terminologie 84
Abbildung 40: Fragebögen können als Aktivität hinzugefügt werden 85
Abbildung 41: Fragebogenbeschreibung 86
Abbildung 42: Zeitsteuerung der Fragebogenbearbeitung 86
Abbildung 43: Abgabeeinstellungen der Fragebögen 86
Abbildung 44: Fragebogen aus Vorlage erstellen 87
Abbildung 45: Freigabe eines Fragebogens als Vorlage für den Fragebogen-Pool 87
Abbildung 46: Verhalten bei eingereichten Fragebögen 88

Abbildung 47: Frage und Antwortmöglichkeiten eingeben 89
Abbildung 48: Fragebogen aus Teilnehmersicht. 90
Abbildung 49: Pausieren der Bearbeitung 90
Abbildung 50: Anzeigen eigener Fragebögen im Verlauf 91
Abbildung 51: Datenbank Export der Studiendaten 91
Abbildung 52: Fragebogendaten in Microsoft Excel 92
Abbildung 53: Prävalenz von Betriebssystemen weltweit 93
Abbildung 54: Funktionsprinzip der Browserweiche 94
Abbildung 55: Flussdiagramm mit Algorithmus der Browserweiche 96
Abbildung 56: Übersicht des Videoverzeichnisses 97
Abbildung 57: Einbinden eines Videos in Moodle über die Browserweiche 97
Abbildung 58: Variablen für die Integration des Registrierungsformulars 98
Abbildung 59: Verknüpfung von Landing-Page und Studie 99
Abbildung 60: allgemeines Registrierungsformular v.s. individuelles Registrierungsfor .. 100
Abbildung 61: Manuelles Anlegen einer neuen Studie 101
Abbildung 62: Beispiel der Vergabe von Berechtigungen für die Rolle des Studienassis.. 102
Abbildung 63: Bericht der Berechtigungen des Fragebogenmoduls 102
Abbildung 64: mögliche Angriffspunkte für Webanwendungen 104
Abbildung 65: Passwörter müssen ein Groß- und Kleinbuchstaben und ein Sonderzeich . 105
Abbildung 66: Der Benutzer wird nach 5 Fehleingaben für 30 Minuten gesperrt 106
Abbildung 67: Notwendige Moodle HTTP Sicherheitseinstellungen 106
Abbildung 68: Cookie wird nur über https übertragen 107
Abbildung 69: Zugriff auf Cookies durch JavaScript verbieten 107
Abbildung 70: Bekannte Sicherheitslücken in Moodle 2.5 108
Abbildung 71: Anmeldenamen sollten nicht im Browser gespeichert werden 108
Abbildung 72: Speichern des Passworts im Browser unterbinden 109
Abbildung 73: HTTP Header des Webservers vor Anpassung 111
Abbildung 74: Anpassung der Apache Sicherheitseinstellungen 112
Abbildung 75: HTTP Header nach der Anpassung 112
Abbildung 76: Erzwingen verschlüsselter Verbindungen beim Apache Webserver 114
Abbildung 77: Abfrage von HSTS Header via Linux oder Mac Terminal 115
Abbildung 78: HSTS fähige Browser Stand 12/2014 115
Abbildung 79: Studienportal.eu ist fest im Quellcode von Google Chrome integriert 116
Abbildung 80: Ablauf einer DNS-Anfrage 117
Abbildung 81: Man-in-the-Middle Angriffe via DNS-Manipulation 117
Abbildung 82: Browser Plugin überprüft DNSSec und TLSA 118
Abbildung 83: Überprüfung und grafische Darstellung der DNS-Kette mit dnsviz.net 119
Abbildung 84: Erbeuten der Zugangsdaten durch Man-in-the-Middle Angriff 120
Abbildung 85: Aufbau eines TLSA Records für DANE 121
Abbildung 86: Echtheit wird des SSL-Zertifikats wird durch TLSA/DANE bestätigt 121
Abbildung 87: Gefälschte Zertifikate fliegen dank TLSA auf 122
Abbildung 88: Warnmeldung vor manipuliertem Zertifikat 122
Abbildung 89: Faktor-2-Authentifizierung mittels Yubikey 125
Abbildung 90: Verbreitung deutscher Emailanbieter 2013 126
Abbildung 91: Online-Test der RDNS Einstellungen auf www.debouncer.com 127
Abbildung 92: Schutz vor Absenderfälschung durch Sender Policy Framework 128
Abbildung 93: Absender-Identifikation anhand einer Signatur im Email-Header 129
Abbildung 94: Email-Header ohne DKIM (gesendet via T-Online Webmail) 130
Abbildung 95: Aufbau des Clusters 133
Abbildung 96: Unterschied zwischen SAN und DRBD 134

Abbildung 97: Datenverwaltung im eingesetzten Aktiv/Passiv HA-Cluster 135
Abbildung 98: dnsperf Performance-Test des Studienportals im Vergleich zum Google .. 142
Abbildung 99: Schema des Messprinzips der Webserver-Tests 142
Abbildung 100: Benchmark mit 400 gleichzeitigen Verbindungen 143
Abbildung 101: Performancetest Studienportal..... 190
Abbildung 102: Performancetest Moodle Referenz 144
Abbildung 103: Antwortzeit in Anhängigkeit von Verschlüsselung..... 145
Abbildung 104: Performance der verglichenen Systeme in Anfragen pro Sekunde 146

11.2 Tabellenverzeichnis

Tabelle 1: Sicherheitsanalyse aktuell verwendeter Fragebogensoftware, Stand 12/2014..... 30
Tabelle 2: Eingesetzte Verschlüsselungskomponenten 55
Tabelle 3: Verfügbarkeit von IT-Diensten und deren Einsatzgebiete 69
Tabelle 4: Kosten laut Hersteller Microsoft, Oracle, MySQL, PostgreSQL 72
Tabelle 5: Übersicht Datensicherungsmöglichkeiten 75
Tabelle 6: Übersichtstabelle Browserkompatibilität und Browserverbreitung 94
Tabelle 7: Wiedergabe von Videos über das Studienportal 94
Tabelle 8: Browserkompatibilität HTML5.....95
Tabelle 9: XSS Analyse mit XSS Heuristik mit ExploitMe..... 110
Tabelle 10: Ergebnis des Sicherheitsscans nach der Grundinstallation 110
Tabelle 11: Apache Webserver Headereinstellungen..... 111
Tabelle 12: Ergebnis des Sicherheitsscans nach Sicherheitsoptimierung 112
Tabelle 13: Sicherheitsanalyse des Studienportals analog der Marktanalyse 116
Tabelle 14: verwendete Filter..... 123

12 Quellenangabe

1. Accelerated IT Service GmbH. 2014. Beschreibung des Rechenzentrums. Retrieved from https://www.accelerated.de/pdf/FRA4-Accelerated.pdf
2. Mike Adolphs. 2009. Nagios Plugin: check_bind.sh. check_bind.sh - Nagios Exchange. Retrieved April 4, 2015 from http://exchange.nagios.org/directory/Plugins/Network-Protocols/DNS/check_bind-2Esh/details
3. Mike Adolphs. 2009. Nagios Plugin: check_apache2. check_apache2.sh - Nagios Exchange. Retrieved April 4, 2015 from http://exchange.nagios.org/directory/Plugins/Web-Servers/Apache/check_apache2-2Esh/details
4. AJ Lewis, Joe Thornber, Patrick Caulfield, Alasdair Kergon, and Jochen Radmacher. LVM HOWTO. Retrieved October 17, 2013 from http://www.tldp.org/HOWTO/LVM-HOWTO/index.html
5. Bruce Allen. 2004. Monitoring Hard Disks with Smart. Linux Journal 2004, 117: 9. Retrieved March 22, 2014 from http://delivery.acm.org/10.1145/960000/959345/6983.html?ip=132.187.246.37&id=959345&acc=ACTIVE%20SERVICE&key=2BA2C432AB83DA15%2EDACA52EEFEA65449%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&CFID=519074104&CFTOKEN=57004854&__acm__=1434240492_d4caa1a755aaf2e86280547aff393ba6
6. Kazumaro Aoki, Jens Franke, Thorsten Kleinjung, Arjen K. Lenstra, and Dag Arne Osvik. 2007. A kilobit special number field sieve factorization. In Advances in Cryptology–ASIACRYPT 2007. Springer, 1–12. Retrieved March 17, 2014 from http://link.springer.com/chapter/10.1007/978-3-540-76900-2_1
7. AOL Inc. AOL Postmaster Technical and Policy Requirements for Sending Email to AOL. AOL Postmaster | Postmaster / Technical and Policy Requirements for Sending Email to AOL. Retrieved December 23, 2014 from http://postmaster.aol.com/Postmaster.Tech.php
8. Tim Ash, Maura Ginty, and Rich Page. 2012. Landing page optimization: the definitive guide to testing and tuning for conversions. John Wiley & Sons.
9. Sandoche Balakrichenan, Stephane Bortzmeyer, and Mohsen Souissi. 2013. A Step-by-Step guide for implementing DANE with a Proof of Concept. Retrieved from https://www.ietf.org/mail-archive/web/dane/current/pdfk2DbQF0Oxs.pdf
10. Ryan Barnett, Brian Rectanus, Ivan Ristić, and Breno Silva. 2010. OWASP ModSecurity Core Rule Set (CRS). SpiderLabs/ModSecurity. Retrieved January 18, 2015 from https://github.com/SpiderLabs/ModSecurity
11. Daniel Bleichenbacher. 1998. Chosen ciphertext attacks against protocols based on the RSA encryption standard PKCS #1. In Advances in Cryptology — CRYPTO '98, Hugo Krawczyk (ed.). Springer, 1–12. Retrieved March 19, 2014 from http://link.springer.com/chapter/10.1007/BFb0055716
12. Olaf Bock, Andreas Nicklisch, and Ingmar Baetge. 2012. Hamburg registration and organization online tool. Hamburg registration and organization online tool, 01. Retrieved December 6, 2014 from http://www.wiso.uni-hamburg.de/fileadmin/einrichtungen/forschungslabor/WorkingPaper_01_BockNicklischBaetge.pdf
13. Martin Bruder, Anja S. Göritz, Ulf-Dietrich Reips, and Ramon K. Gebhard. 2014. Ein national gefördertes Onlinelabor als Infrastruktur für die psychologische Forschung. Psychologische Rundschau 65, 2: 75–85. http://doi.org/10.1026/0033-3042/a000198
14. Prof Dr Manfred Bruhn, Dipl-Kffr Daniela B. Schäfer, Dipl-Kfm Jürgen Schwarz, and

Mareike Lauber. 2011. Facebook, Twitter, YouTube und Co. – Erwartungen der Nutzer an Social-Media-Plattformen. Marketing Review St. Gallen 28, 5: 36–42. http://doi.org/10.1007/s11621-011-0061-x
15. Johannes Buchmann. 2007. Einführung in die Kryptographie. Springer, Darmstadt.
16. Jens Budszus, Oliver Berthold, Alexander Filip, Sven Polenz, Thomas Probst, and Maren Thiermann. 2014. Orientierungshilfe – Cloud Computing. Arbeitskreise Technik und Medien der Konferenz der Datenschutzbeauftragten des Bundes und der Länder sowie der Arbeitsgruppe Internationaler Datenverkehr. Retrieved May 9, 2014 from https://www.datenschutz-bayern.de/technik/orient/oh_cloud.pdf
17. Bundesamt für Sicherheit in der Informationstechnik (ed.). 2006. Sicherheit von Webanwendungen Maßnahmenkatalog und Best Practices. Retrieved from https://www.bsi.bund.de/cae/servlet/contentblob/476464/publicationFile/30642/WebSec_pdf.pdf
18. Bundesamt für Sicherheit in der Informationstechnik. 2015. Kryptographische Verfahren: Empfehlungen und Schlüssellängen. Retrieved from https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/TechnischeRichtlinien/TR02102/BSI-TR-02102-2_pdf.pdf?__blob=publicationFile
19. Bundesministerium der Justiz und für Verbraucherschutz. 2015. Telemediengesetz TMG. Retrieved from http://www.gesetze-im-internet.de/tmg/
20. Jalalle Chahboune. IT-Unterstützung vs. Datenschutz im Bereich klinischer Studien. 77, 2. Retrieved from http://www.evimed.com/wp-content/uploads/2015/03/PI7701_1147_Chahboune_IT-Unterstuetzung_sd.pdf
21. Dave Crocker, T. Hansen, and M. Kucherawy. 2011. RFC 6376 DomainKeys Identified Mail (DKIM) Signatures–Update. Internet Engineering Task Force (IETF). Retrieved December 23, 2014 from http://tools.ietf.org/html/rfc5672.html
22. Cyril Jaquier. 2013. Fail2ban Manual. Retrieved January 18, 2015 from http://www.fail2ban.org/wiki/index.php?title=MANUAL_0_8&oldid=4887
23. Casey Deccio. DNSViz | A DNS visualization tool. Retrieved July 4, 2015 from http://dnsviz.net/
24. Roland van Rijswijk- Deij. 2012. Deploying DNSSEC - Validation on recursive caching name servers. Retrieved from https://www.surf.nl/binaries/content/assets/surf/en/knowledgebase/2012/rapport_Deploying_DNSSEC_v20.pdf
25. Dell Deutschland. PowerConnect 6224 – Detailinformationen. Retrieved October 4, 2015 from http://www1.euro.dell.com/de/de/premier/networking/pwcnt_6224/pd.aspx?refid=pwcnt_6224&s=premier
26. Maximilian Dermann, Mirko Dziadzka, Boris Hemkemeier, et al. 2008. Best Practices: Use of Web Application Firewalls - OWASP. Retrieved January 7, 2015 from https://www.owasp.org/index.php/Category:OWASP_Best_Practices:_Use_of_Web_Application_Firewalls
27. Alexis Deveria. Browser Support for HSTS. Retrieved December 28, 2014 from http://caniuse.com/#search=hsts
28. Christian Dierks, Alexander Roßnagel, Gerrit Hornung, and Silke Jandt. 2009. Rechtsgutachten zum Datenschutz in der medizinischen Forschung.
29. T. Dierks and E. Rescorla. 2006. The Transport Layer Security (TLS) Protocol. Network Working Group. Retrieved December 20, 2015 from https://www.ietf.org/rfc/rfc4346.txt
30. Benjamin Doerr, Mahmoud Fouz, and Tobias Friedrich. 2011. Social Networks Spread Rumors in Sublogarithmic Time. Proceedings of the Forty-third Annual ACM Sympo-

sium on Theory of Computing, ACM, 21–30. http://doi.org/10.1145/1993636.1993640
31. Benjamin Doerr, Mahmoud Fouz, and Friedrich, Tobias. 2012. Social Networks Spread Rumors in Sublogarithmic Time. Retrieved March 14, 2014 from http://scidok.sulb.uni-saarland.de/volltexte/2012/4903/
32. Frank Robert Drechsel. 2012. Landing Page für Studieninteressierte. Retrieved from http://tu-dresden.de/studium/angebot/studienmoeglichkeiten/newsletter/bewerberbefragung
33. Zakir Durumeric, Mathias Payer, Vern Paxson, et al. 2014. The Matter of Heartbleed. ACM Press, 475–488. http://doi.org/10.1145/2663716.2663755
34. Waldemar Dzeyk. 2001. Ethische Dimensionen der Online-Forschung. Kölner Psychologische Studien 6, 1. Retrieved October 6, 2015 from http://kups.ub.uni-koeln.de/volltexte/2008/2424/
35. Claudia Eckert. 2013. IT-Sicherheit: Konzepte - Verfahren - Protokolle. Oldenbourg Verlag.
36. Birgit van Eimeren. „Always on“ – Smartphone, Tablet & Co. als neue Taktgeber im Netz. Retrieved from http://www.ard-zdf-onlinestudie.de/fileadmin/Onlinestudie/PDF/Eimeren.pdf
37. Birgit van Eimeren and Beate Frees. Ergebnisse der ARD/ZDF-Onlinestudie 2014. Retrieved from http://www.ard-zdf-onlinestudie.de/fileadmin/Onlinestudie_2014/PDF/0708-2014_Eimeren_Frees.pdf
38. Tanya Elias. 2010. Universal instructional design principles for Moodle. The International Review of Research in Open and Distance Learning 11, 2: 110–124. Retrieved October 19, 2013 from http://www.irrodl.org/index.php/irrodl/article/viewArticle/869
39. Sabine Feierabend, Ulrike Karg, and Thomas Rathgeb. 2013. Basisstudie zum Medienumgang 12- bis 19-Jähriger in Deutschland. Retrieved from http://www.mpfs.de/fileadmin/JIM-pdf13/JIMStudie2013.pdf
40. G. Flatten, U. Gast, A. Hofmann, et al. S3 Leitlinie posttraumatische Belasttunsungsstörungung ICD 10: F 43.1. Retrieved from http://www.awmf.org/uploads/tx_szleitlinien/051-010l_S3_Posttraumatische_Belastungsstoerung_2012-03.pdf
41. forsa Politik- und Sozialforschung GmbH. Data Monitor 2015 – Wofür die Deutschen ihre persönlichen Daten preisgeben. Retrieved from http://www.sas.com/content/dam/SAS/bp_de/doc/studie/ba-st-forsa-der-umgang-mit-daten-2343928.pdf
42. Gary Chen. 2012. Open Source Cloud System Software. Retrieved September 22, 2013 from http://www.linuxfoundation.org/publications/linux-foundation
43. Gesellschaft für strategische Kommunikation mbH. Datenschutzvorfälle | Projekt Datenschutz. Retrieved October 10, 2015 from http://www.projekt-datenschutz.de/datenschutzvorfaelle
44. Google Inc. Search Engine Optimization Starter Guide. Retrieved from http://www.google.co.jp/intl/en/webmasters/docs/search-engine-optimization-starter-guide.pdf OBJ OBJ
45. Anja S. Göritz, Martin Bruder, and Ulf-Dietrich Reips. 2014. Ein national gefördertes Online- labor als Infrastruktur für die Forschung: Ergebnisse einer Meinungs- und Bedarfserhebung. RatSWD. Retrieved October 8, 2015 from http://www.ratswd.de/dl/RatSWD_WP_241.pdf
46. Jörg Gutsche. 2006. Ökonomische Analyse offener Software. Retrieved October 6, 2015 from https://ub-madoc.bib.uni-mannheim.de/1169
47. Major Hayden. MySQLTuner. MySQLTuner. Retrieved March 28, 2015 from http://mysqltuner.com/

48. Volker Hockmann and Heinz-Dieter Knöll. 2008. Profikurs Sicherheit von Web-Servern. Vieweg + Teubner, Wiesbaden. Retrieved from http://www.springerprofessional.de/978-3-8348-9471-7---profikurs-sicherheit-von-web-servern/1797784.html
49. Jeff Hodges, Collin Jackson, and Adam Barth. 2012. Http strict transport security (hsts). URL: http://tools. ietf. org/html/draft-ietf-websec-strict-transport-sec-04. Retrieved December 28, 2014 from http://www.hjp.at/doc/rfc/rfc6797.html
50. Todd Hoff. 2009. Scaling Twitter: Making Twitter 10000 Percent Faster - High Scalability -. Retrieved April 25, 2015 from http://highscalability.com/blog/2009/6/27/scaling-twitter-making-twitter-10000-percent-faster.html
51. Holger Hennig. 2013. High-Availability Clustering.
52. Ralph Holz, Yaron Sheffer, and Peter Saint-Andre. 2014. Recommendations for Secure Use of TLS and DTLS. Retrieved March 17, 2014 from http://tools.ietf.org/html/draft-sheffer-tls-bcp-02
53. Khoa Huynh. 2013. Exploiting The Latest KVM Features For Optimized Virtualized Enterprise Storage Performance. Exploiting The Latest KVM Features For Optimized Virtualized Enterprise Storage Performance. Retrieved from http://events.linuxfoundation.org/sites/events/files/slides/CloudOpen2013_Khoa_Huynh_v3.pdf
54. Khoa Huynh, Andrew Theurer, and Stefan Hajnoczi. 2013. KVM_Virtualized_IO_Performance_Paper_v2.pdf. Retrieved from ftp://public.dhe.ibm.com/linux/pdfs/KVM_Virtualized_IO_Performance_Paper_v2.pdf
55. Internet Systems Consortium, Inc. 2014. BIND 9 Administrator Reference Manual. Retrieved from https://www.isc.org/wp-content/uploads/2014/01/B99ARM.pdf
56. Ivan Ristić. 2006. The Transport Layer Security (TLS) Protocol Version 1.1. Retrieved September 9, 2013 from http://tools.ietf.org/html/rfc4346/
57. Ivan Ristić. 2014. SSL/TLS Deployment Best Practices. Retrieved October 12, 2014 from https://www.ssllabs.com/downloads/SSL_TLS_Deployment_Best_Practices_1.2.pdf
58. Clemente Izurieta and James Bieman. 2006. The evolution of FreeBSD and Linux. Proceedings of the 2006 ACM/IEEE international symposium on Empirical software engineering, ACM, 204–211. http://doi.org/10.1145/1159733.1159765
59. Wilhelm Janke and Günter Debus. 1978. Die Eigenschaftswörterliste: EWL; eine mehrdimensionale Methode zur Beschreibung von Aspekten des Befindens. Verlag für Psychologie Hogrefe.
60. Anaïs Le Jeannic, Céline Quelen, Corinne Alberti, Isabelle Durand-Zaleski, and $author firstName $author.lastName. 2014. Comparison of two data collection processes in clinical studies: electronic and paper case report forms. BMC Medical Research Methodology 14, 1: 7. http://doi.org/10.1186/1471-2288-14-7
61. Joerg Linge. 2014. PNP4Nagios Documentation. Retrieved April 4, 2015 from https://docs.pnp4nagios.org/
62. Jonathan Corbet, Greg Kroah-Hartman, and Amanda McPherson. 01.01. Linux Kernel Development: How Fast it is Going, Who is Doing It, What They are Doing, and Who is Sponsoring It. Retrieved September 22, 2013 from http://www.linuxfoundation.org/publications/linux-foundation
63. K. W. Kallus. 1995. Der Erholungs-Belastungs-Fragebogen. Handanweisung.
64. Emilia Käsper. 2012. Fast elliptic curve cryptography in OpenSSL. In Financial Cryptography and Data Security. Springer, 27–39. Retrieved March 18, 2014 from http://link.springer.com/chapter/10.1007/978-3-642-29889-9_4
65. S. Kitterman. 2014. Sender Policy Framework (SPF) for Authorizing Use of Domains

in Email, Version 1. Retrieved December 23, 2014 from https://tools.ietf.org/html/rfc7208
66. Thorsten Kleinjung, Joppe W. Bos, Arjen K. Lenstra, et al. 2012. A heterogeneous computing environment to solve the 768-bit RSA challenge. Cluster Computing 15, 1: 53–68. http://doi.org/10.1007/s10586-010-0149-0
67. Yonathan Klijnsma, Zheng Hu, Yun, Lennart Haagsma, van Maarten, Dantzig, and Barry Weymes. 2014. CryptoPHP Whitepaper. Retrieved from https://foxitsecurity.files.wordpress.com/2014/11/cryptophp-whitepaper-foxsrt-v4.pdf
68. Jukka Kommeri, Tapio Niemi, and Olli Helin. 2012. Energy efficiency of server virtualization. ENERGY 2012, The Second International Conference on Smart Grids, Green Communications and IT Energy-aware Technologies, 90–95. Retrieved October 30, 2014 from http://www.thinkmind.org/index.php?view=article&articleid=energy_2012_4_30_40149
69. Thomas Kranig. 2014. Datenschutzprüfung bei Mailservern bayerischer Unternehmen. Retrieved from http://www.lda.bayern.de/lda/datenschutzaufsicht/p_archiv/2014/pm012.html
70. Sebastian Kraska and Michael Stolze. Datenschutz und Cloud-Computing: Entschließung der 82. Konferenz der Datenschutz-Aufsichtsbehörden. Retrieved October 4, 2015 from https://www.iitr.de/veroeffentlichungen-des-instituts-fuer-it-recht/274-datenschutz-und-cloud-computing-entschliessung-der-82-konferenz-der-datenschutz-aufsichtsbehoerden.html
71. Philipp Kutz. 2010. SURE-Studie. Ein neues Entspannungsverfahren zum Abbau von Stress im Rettungsdienst. Retrieved October 28, 2013 from http://epub.uni-regensburg.de/14111
72. James Glanz Jeff Larson and Andrew W. Lehren. 2014. Spy Agencies Tap Data Streaming From Phone Apps. The New York Times. Retrieved March 18, 2014 from http://www.nytimes.com/2014/01/28/world/spy-agencies-scour-phone-apps-for-personal-data.html
73. Gerhard Laußer. 2012. check_mysql_health. check_mysql_health - Nagios Exchange. Retrieved April 4, 2015 from http://exchange.nagios.org/directory/Plugins/Databases/MySQL/check_mysql_health/details
74. Arjen K. Lenstra and Eric R. Verheul. 2001. Selecting cryptographic key sizes. Journal of cryptology 14, 4: 255–293. http://doi.org/10.1007/s00145-001-0009-4
75. Reuven M. Lerner. 2007. Open-Source Databases, Part III: Choosing a Database. Linux Journal, 158: 42–45. Retrieved from http://www.linuxjournal.com/article/9649
76. Libby Clark and Brian Warner. 2013. 20 years of Top500.org Supercomputer Data Links Linux With Advances in Computing Performance. Retrieved September 22, 2013 from http://www.linuxfoundation.org/publications/linux-foundation
77. Oliver Liebel. 2011. Linux Hochverfügbarkeit. Galileo Press, Bonn.
78. Gary Ling. 2014. Unveiling scalable HTTP load balancing across cloud regions. Google Cloud Platform Blog. Retrieved June 19, 2015 from http://googlecloudplatform.blogspot.de/2014/06/unveiling-scalable-http-load-balancing-across-cloud-regions.html
79. Thomas W Lipp. 1994. Die grosse Welt der Grafikformate: Grafikprogrammierung unter Windows und Windows NT. Synergy-Verl., München.
80. Thomas Loew and Christian Hanshans. 2013. Sure ideal for crisis intervention: a somatic universal regulative exercise improves the mood and reduces psychic tension. PSYCHOTHERAPY AND PSYCHOSOMATICS. Retrieved from

https://www.researchgate.net/publication/279033304_Sure_ideal_for_crisis_interventio n_a_somatic_universal_regulative_exercise_improves_the_mood_and_reduces_psychic _tension
81. Graham Lowe, Patrick Winters, and Michael L. Marcus. 2007. The great dns wall of china. MS, New York University. Accessed December 21. Retrieved January 18, 2015 from http://www.cs.nyu.edu/~pcw216/work/nds/final.pdf
82. Ciobanu Luminita and Ciobanu Nicoleta. 2012. E-learning Security Vulnerabilities. E-learning Security Vulnerabilities Procedia - Social and Behavioral Sciences, 46. http://doi.org/10.1016/j.sbspro.2011.04.171
83. Andreas Maercker. 2003. Posttraumatische Stress Skala-10 (PTSS-10). Retrieved from http://www.psychologie.uzh.ch/fachrichtungen/psypath/ForschungTools/Fragebogen/pt ss10testbeschreib.pdf
84. Evan Marcus and Hal Stern. 2000. Blueprints for high availability: designing resilient distributed systems. Wiley, New York.
85. Michael Maston. 1999. Managing Windows with WMI. Managing Windows with WMI. Retrieved March 22, 2014 from http://msdn.microsoft.com/de-de/library/ms811533.aspx
86. Julian Mehnle. SPF Record Syntax. Retrieved December 23, 2014 from http://www.openspf.org/SPF_Record_Syntax
87. Joseph Menn. 2013. Exclusive: Secret contract tied NSA and security industry pioneer. Reuters. Retrieved March 18, 2014 from http://www.reuters.com/article/2013/12/21/us-usa-security-rsa-idUSBRE9BJ1C220131221
88. Mike McNamara. NetApp - Skalierbarkeit der Enterprise-Klasse mit Data ONTAP 8 Cluster-Mode. Retrieved June 21, 2015 from http://www.netapp.com/de/communities/tech-ontap/tot-data-ontap-8-cluster-mode-1209-de.aspx
89. Darko Miletić. 2011. Moodle Security. Packt Publishing Ltd. Retrieved March 21, 2014 from https://www.packtpub.com/hardware-and-creative/moodle-security
90. Bodo Möller, Thai Duong, and Krzysztof Kotowicz. 2014. This POODLE Bites: Exploiting The SSL 3.0 Fallback. Retrieved November 10, 2014 from https://www.openssl.org/~bodo/ssl-poodle.pdf
91. M. Wielsch, J. Prahm, and H.-G. Eßer. 1999. Linux intern Technik, Administration und Programmierung. Data Becker GmbH & Co. KG.
92. Jakob Nielsen. 2000. Erfolg des Einfachen: Jakob Nielsen's Web Design. Markt + Technik-Verl., München.
93. Nominum, Inc. 2012. DNSPerf - DNS Performance Tool Manual. Retrieved from ftp://ftp.nominum.com/pub/nominum/dnsperf/1.0.1.0/dnsperf-1.0.1.0-Info-20071228.pdf
94. Nominum, Inc. DNS Performance - Compare the speed of enterprise and commercial DNS services. DNS Performance - Compare the speed of enterprise and commercial DNS services. Retrieved April 22, 2015 from http://www.dnsperf.com/
95. OpenSSL Project. SSL cipher display and cipher list tool. Retrieved July 22, 2015 from https://www.openssl.org/docs/apps/ciphers.html
96. Serkan Özkan. Moodle : List of security vulnerabilities. Retrieved January 19, 2015 from http://www.cvedetails.com/vulnerability-list/vendor_id-2105/product_id-3590/Moodle-Moodle.html
97. Europäisches Parlament. 2015. Gerichtshof der Europäischen Union Pressemitteilung Nr. 117/15. Amtsblatt Nr. L 281, 23: 31–50. Retrieved October 6, 2015 from http://curia.europa.eu/jcms/upload/docs/application/pdf/2015-10/cp150117de.pdf
98. Ivan Pavlović, Tomaz Kern, and Damijan Miklavcic. 2009. Comparison of paper-based

and electronic data collection process in clinical trials: costs simulation study. Contemporary clinical trials 30, 4: 300–316. http://doi.org/10.1016/j.cct.2009.03.008
99. Eduardo Pinheiro, Wolf-Dietrich Weber, and Luiz André Barroso. 2007. Failure Trends in a Large Disk Drive Population. FAST, 17–23. Retrieved March 22, 2014 from http://static.usenix.org/event/fast07/tech/full_papers/pinheiro/pinheiro_html/
100. Andrey Popov. Prohibiting RC4 Cipher Suites. Retrieved July 22, 2015 from https://tools.ietf.org/html/rfc7465
101. Jarno Rajahalme, Shane Amante, Sheng Jiang, and Brian Carpenter. 2011. IPv6 flow label specification. RFC, 6265. Retrieved March 20, 2014 from http://tools.ietf.org/html/rfc6437.html
102. Philipp Reisner and Lars Ellenberg. 2005. Replicated storage with shared disk semantics. Proceedings of the 12th International Linux System Technology Conference (Linux-Kongress), Germany, 111–119. Retrieved June 21, 2015 from http://drbd.linbit.com/fileadmin/drbd/publications/drbd8.pdf
103. Ronald L. Rivest. 1992. The MD5 Message-Digest Algorithm. MIT Laboratory for Computer Science. Retrieved March 19, 2014 from http://www.ietf.org/rfc/rfc1321.txt
104. Ronald L. Rivest, Adi Shamir, and Len Adleman. 1978. A method for obtaining digital signatures and public-key cryptosystems. Communications of the ACM 21, 2: 120–126. Retrieved September 10, 2013 from http://dl.acm.org/citation.cfm?id=359342
105. Yu Sasaki and Kazumaro Aoki. 2009. Finding preimages in full MD5 faster than exhaustive search. In Advances in Cryptology-EUROCRYPT 2009. Springer, 134–152. Retrieved March 20, 2014 from http://link.springer.com/chapter/10.1007/978-3-642-01001-9_8
106. D.J. Saul. The 2014 Facebook Demographic Report. Retrieved March 14, 2014 from http://istrategylabs.com/2014/01/3-million-teens-leave-facebook-in-3-years-the-2014-facebook-demographic-report/
107. Tobias Scherbaum. 2009. Praxisbuch Nagios. O'Reilly Media, Inc.
108. Sven Schleier and Thomas Schreiber. 2012. Aktuelle Verbreitung von HTTP Strict Transport Security (HSTS). Retrieved from https://www.securenet.de/fileadmin/papers/HTTP_Strict_Transport_Security_HSTS_Verbreitung.pdf
109. Georg Schönberger. check_lsi_raid.sh. git.thomas-krenn.com Git - check_lsi_raid.git/blob - README. Retrieved April 4, 2015 from http://git.thomas-krenn.com/check_lsi_raid.git
110. Matthias Schubert. 2007. Datenbanken: Theorie, Entwurf und Programmierung relationaler Datenbanken. B. G. Teubner Verlag / GWV Fachverlage GmbH, Wiesbaden, Wiesbaden. Retrieved September 23, 2013 from http://dx.doi.org/10.1007/978-3-8351-9108-2
111. Dan Sinclair. XSS Me Open Source Firefox Plugins for Penetration Testing. Retrieved from http://www.layerone.org/wp-content/uploads/2011/01/Dan-Sinclair-L1-2008-Exploit-Me1.pdf
112. William Stallings. 1998. SNMP, SNMPV2, Snmpv3, and RMON 1 and 2. Addison-Wesley, Boston, MA, USA. Retrieved from http://www.gbv.de/dms/ilmenau/toc/572061714.PDF
113. Thorsten Kleinjung, Kazumaro Aoki, Jens Franke, et al. 2010. Factorization of a 768-bit RSA modulus. 2010/006. Retrieved September 10, 2013 from http://eprint.iacr.org/2010/006.pdf
114. Ernst Tiemeyer and Thomas Feil (eds.). 2006. Handbuch IT-Management: Konzepte, Methoden, Lösungen und Arbeitshilfen für die Praxis. Hanser, München.
115. Zafer Unal and Asli Unal. 2011. Evaluating and comparing the usability of web-based

course management systems. Journal of Information Technology Education 10: 19–38. Retrieved October 19, 2013 from http://informingscience.org/jite/documents/Vol10/JITEv10p019-038Unal904.pdf
116. NIST US Department of Commerce. Cryptographic Standards Statement. Retrieved March 18, 2014 from http://www.nist.gov/director/cybersecuritystatement-091013.cfm
117. U.S. Government Printing Office. Patriot Act. Patriot Act. Retrieved October 6, 2015 from http://www.gpo.gov/fdsys/pkg/PLAW-107publ56/html/PLAW-107publ56.htm
118. Philippe Vialle and Greg Wroblewski. Security Best Practices to Protect Internet Facing Web Servers. Security Best Practices to Protect Internet Facing Web Servers - TechNet Articles TechNet Wiki. Retrieved September 23, 2013 from http://social.technet.microsoft.com/wiki/contents/articles/13974.security-best-practices-to-protect-internet-facing-web-servers.aspx
119. Vincent Bernat. SSL/TLS & Perfect Forward Secrecy Benchmark. SSL/TLS & Perfect Forward Secrecy Benchmark. Retrieved September 9, 2013 from http://vincent.bernat.im/en/blog/2011-ssl-perfect-forward-secrecy.html
120. Kashi Venkatesh Vishwanath and Nachiappan Nagappan. 2010. Characterizing cloud computing hardware reliability. Proceedings of the 1st ACM symposium on Cloud computing, 193–204. http://doi.org/http://dx.doi.org/10.1145/1807128.1807161
121. Matthäus Wander and Torben Weis. 2013. Measuring Occurrence of DNSSEC Validation. Springer, Universität Düsseldorf. Retrieved December 29, 2014 from http://dnssec.vs.uni-due.de
122. Robert Warnke and Thomas Ritzau. 2010. Qemu-kvm & libvirt. Books on Demand, Norderstedt. Retrieved from http://quemu-buch.de
123. Mike West and Dan Veditz. 2015. W3C Content Security Policy. Retrieved January 8, 2015 from https://w3c.github.io/webappsec/specs/content-security-policy/
124. Lutz Wienhold. 2005. Qualität des Handelns der Fachkräfte für Arbeitssicherheit. Fb 1046: 186.
125. WiSo Panel. 2015. [WiSoPanel] Einladung zu einer Studie.eml.
126. David Wood. 1999. Programming Internet Email. O'Reilly Media, Inc.
127. De Santis Workshop on the Theory and Application of Cryptographic Techniques and EUROCRYPT '94. 1995. Advances in cryptology, EUROCRYPT '94: Workshop on the Theory and Application of Cryptographic Techniques, Perugia Italy, May 9-12, 1994 : proceedings. Springer-Verlag.
128. www.vulnerability-lab.com (ed.). XSS Dokumentation, Analyse & Techniken. Retrieved from http://www.vulnerability-lab.com/resources/documents/198.pdf
129. Dan York. 2012. Challenges and opportunities in deploying dnssec. Securing and Trusting Internet Names (SATIN). Retrieved December 17, 2014 from http://www.internetsociety.org/deploy360/wp-content/uploads/2012/03/SATIN2012-DNSSEC-Challenges-v12-FINAL.pdf
130. Konrad Zerr. 2001. Online-Marktforschung Theoretische Grundlagen und praktische Erfahrungen. In Online-Marktforschung, Dr Axel Theobald, Marcus Dreyer and Thomas Starsetzki (eds.). Gabler Verlag, 7–26. Retrieved October 7, 2015 from http://link.springer.com/chapter/10.1007/978-3-322-99429-5_2
131. 2007. stern.de-Studie: Web 2.0 - gern genutzt, aber was ist das? - Digital. stern.de. Retrieved September 25, 2013 from http://www.stern.de/digital/online/sternde-studie-web-20-gern-genutzt-aber-was-ist-das-591483.html
132. 2010. OpenSSL Ciphersuite Downgrade Attack. OpenSSL Ciphersuite Downgrade Attack. Retrieved March 19, 2014 from http://www.openssl.org/news/secadv_20101202.txt
133. 2011. Empfehlungen zu Forschungsinfrastrukturen in den Geistes- und Sozialwissen-

schaften. Retrieved from http://www.wissenschaftsrat.de/download/archiv/10465-11.pdf
134. 2012. Security Vulnerabilities in the open source Moodle eLearning System. Colloquium for Information Systems Security Education, CISSE. Retrieved October 20, 2013 from http://www.cisse.info/archives/category/27-2012-paper-sessions
135. 2013. Die Publicare E-Mail-Studie 2013 für Deutschland. HIVE. Retrieved December 23, 2014 from http://hive.publicare.de/de/e-mail-studie-2013/
136. 2013. LSI StorCLI Reference Manual. Retrieved from http://www.lsi.com/downloads/Public/RAID%20Controllers/RAID%20Controllers%20Common%20Files/StorCLI_RefMan_revf.pdf
137. 2013. Web Server Survey | Netcraft. Web Server Survey | Netcraft. Retrieved September 23, 2013 from http://news.netcraft.com/archives/2013/05/03/may-2013-web-server-survey.html
138. 2013. Click to Chat : un nouvel outil d'e-relation client. Blog CCM Benchmark Institut. Retrieved September 25, 2013 from http://www.ccmbenchmark.com/institut/blog/click-to-chat-relation-client/
139. 2014. OpenSSH. OpenSSH. Retrieved March 22, 2014 from http://www.openssh.com/
140. Internetnutzung in Deutschland - Statista-Dossier. Internetnutzung in Deutschland - Statista-Dossier. Retrieved October 9, 2015 from http://de.statista.com/statistik/studie/id/22540/dokument/internetnutzung-in-deutschland-statista-dossier/
141. LTE Verfügbarkeit – Karte & Test zum aktuellen Ausbau. Retrieved October 9, 2015 from http://www.lte-anbieter.info/verfuegbarkeit/lte-verfuegbarkeit-testen.php
142. Ausstattungsgrad - Personal Computer in deutschen Haushalten bis 2014 | Statistik. Statista. Retrieved October 9, 2015 from http://de.statista.com/statistik/daten/studie/160925/umfrage/ausstattungsgrad-mit-personal-computer-in-deutschen-haushalten/
143. heise online Timeline – NSA-Skandal. Retrieved October 10, 2015 from http://www.heise.de/extras/timeline/#vars!date=2014-01-01_13:23:00!
144. Fragebogen: Bedarf an einem national geförderten Onlinelabor. Retrieved October 8, 2015 from http://www.unipark.de/uc/a7aa/ospe.php?SES=7dbb3a7c9cd45cbe90469a5ca1485fdb&syid=501791&sid=501792&act=start
145. BDSG - Bundesdatenschutzgesetz. Retrieved December 30, 2014 from http://www.gesetze-im-internet.de/bdsg_1990/BJNR029550990.html#BJNR029550990BJNG000102301
146. Forschung erleben - Forschung erleben. Retrieved October 10, 2015 from http://www.forschung-erleben.uni-mannheim.de/index.php?q=aktuelleforschung/online_studien
147. SurveyMonkey Pläne und Preise. Retrieved June 21, 2015 from https://de.surveymonkey.com/pricing/?ut_source=header
148. Angebot Electric Paper.pdf.
149. Security_Paper_EvaSys_de_6.0.pdf.
150. Wisopanel.de - Reverse IP Lookup - DomainTools. Retrieved January 4, 2015 from http://reverseip.domaintools.com/search/?q=wisopanel.net
151. Studienportal des Psychologischen Instituts der Universität Heidelberg. Retrieved January 4, 2015 from https://studienportal.psychologie.uni-heidelberg.de/
152. Sicherheitsinformation. Retrieved September 9, 2015 from https://www.1blu.de/sicherheit/
153. Sicherheitsvorfall. Retrieved September 9, 2015 from

https://forum.ovh.de/showthread.php?12015-Sicherheitsvorfall&p=71423#post71423
154. Security Issue – Hetzner DokuWiki. Retrieved September 9, 2015 from http://wiki.hetzner.de/index.php/Security_Issue
155. “Datenklau”: GMX und Web.de sperren betroffene Mailkonten | iX. Retrieved September 9, 2015 from http://www.heise.de/ix/meldung/Datenklau-GMX-und-Web-de-sperren-betroffene-Mailkonten-2165338.html
156. 1&1-Server im Visier von Hackern - com! professional. Retrieved September 9, 2015 from http://www.com-magazin.de/news/sicherheit/1-1-server-im-visier-hackern-114848.html
157. SSL Server Rating Guide. Retrieved April 12, 2014 from https://www.ssllabs.com/downloads/SSL_Server_Rating_Guide.pdf
158. DNSSEC Analyzer. Versisign Labs DNS Debugger. Retrieved January 4, 2015 from http://dnssec-debugger.verisignlabs.com/
159. Case Study: How We Improved Landing Page Conversions by 79.3%. ConversionXL. Retrieved March 14, 2014 from http://conversionxl.com/case-study-how-we-improved-landing-page-conversion/
160. The rise of the networked enterprise: Web 2.0 finds its payday | McKinsey & Company. Retrieved September 25, 2013 from http://www.mckinsey.com/insights/high_tech_telecoms_internet/the_rise_of_the_networked_enterprise_web_20_finds_its_payday
161. SSL/TLS & Perfect Forward Secrecy | Vincent Bernat. Retrieved September 9, 2013 from http://vincent.bernat.im/en/blog/2011-ssl-perfect-forward-secrecy.html
162. SSL/TLS Strong Encryption: An Introduction - Apache HTTP Server. Retrieved March 18, 2014 from http://httpd.apache.org/docs/current/ssl/ssl_intro.html
163. SSL: Intercepted today, decrypted tomorrow | Netcraft. Retrieved March 17, 2014 from http://news.netcraft.com/archives/2013/06/25/ssl-intercepted-today-decrypted-tomorrow.html
164. Heartbleed trifft auch Onlinebanking. Retrieved December 7, 2014 from http://www.wiwo.de/technologie/digitale-welt/ssl-sicherheitsluecke-heartbleed-trifft-auch-onlinebanking/9754080.html
165. Heartbleed / OpenSSL / CVE-2014-0160. Retrieved December 7, 2014 from https://www.dfn-cert.de/aktuell/OpenSSL-Heartbleed-Schwachstelle-CVE-2014-0160.html
166. An Excerpted Document From the N.S.A. From May 2010. Retrieved March 18, 2014 from https://www.documentcloud.org/documents/1009660-nsa.html
167. Synology Network Attached Storage - How to choose between different RAID... Retrieved September 21, 2013 from http://www.synology.com/support/tutorials_show.php?lang=us&q_id=512#t4
168. Ethernet-to-the-Factory 1.2 Design and Implementation Guide - Implementation of High Availability [Design Zone for Manufacturing]. Cisco. Retrieved September 21, 2013 from http://www.cisco.com/en/US/docs/solutions/Verticals/EttF/ch6_EttF.html
169. Betriebssysteme - Marktanteile weltweit bis März 2015 | Statistik. Retrieved June 9, 2015 from http://de.statista.com/statistik/daten/studie/157902/umfrage/marktanteil-der-genutzten-betriebssysteme-weltweit-seit-2009/
170. Oracle Database Products. Retrieved September 23, 2013 from https://shop.oracle.com/pls/ostore/product?p1=OracleDatabase&p2=&p3=&p4=
171. SQL Server 2012 Produktlizenzierung Übersicht. Retrieved September 23, 2013 from http://www.microsoft.com/de-de/licensing/produktlizenzierung/sql-server-2012/lizenzierung.aspx#112

172. MySQL Customers by Industry. Retrieved September 23, 2013 from http://www.mysql.com/customers/industry/?id=
173. SQL Server 2012 Hardware- und Softwareanforderungen. Retrieved September 23, 2013 from http://msdn.microsoft.com/de-de/library/ms143506.aspx
174. Oracle Database Preinstallation Requirements. Retrieved September 23, 2013 from http://docs.oracle.com/cd/B28359_01/install.111/b32006/reqs.htm#i1005703
175. MySQL Referenzhandbuch. MySQLReferenzhandbuch. Retrieved September 23, 2013 from http://dev.mysql.com/doc/
176. PostgreSQL 9.3 Documentation. PostgreSQL: Documentation: 9.3: PostgreSQL 9.3.0 Documentation. Retrieved September 23, 2013 from http://www.postgresql.org/docs/9.3/static/index.html
177. Volumeschattenkopie (Übersicht). Volumenschattenkopie. Retrieved October 17, 2013 from http://technet.microsoft.com/de-de/library/cc784351(WS.10).aspx
178. Moodle.org: Registered sites. Retrieved October 18, 2013 from https://moodle.org/sites/index.php?country=DE
179. Debian -- Debian-Gesellschaftsvertrag. Retrieved March 28, 2015 from https://www.debian.org/social_contract
180. Installation von Moodle – MoodleDocs. Retrieved March 28, 2015 from https://docs.moodle.org/28/de/Installation_von_Moodle#Server_aufsetzen
181. Debian -- Details of package php5 in wheezy. Retrieved March 28, 2015 from https://packages.debian.org/de/wheezy/php5
182. MySQL :: MySQL 5.5 Reference Manual :: 14 The InnoDB Storage Engine. Retrieved March 28, 2015 from http://dev.mysql.com/doc/refman/5.5/en/innodb-storage-engine.html
183. MySQL Engines: MyISAM vs. InnoDB | Tag1 Consulting. Retrieved March 28, 2015 from http://tag1consulting.com/MySQL_Engines_MyISAM_vs_InnoDB
184. LAMP (Softwarepaket) – Wikipedia. Retrieved June 21, 2015 from https://de.wikipedia.org/wiki/LAMP_(Softwarepaket)
185. phpESP - php Easy Survey Package. SourceForge. Retrieved March 24, 2014 from http://sourceforge.net/projects/phpesp/
186. Betriebssysteme - Marktanteile weltweit bis September 2013 | Statistik. Retrieved November 5, 2013 from http://de.statista.com/statistik/daten/studie/157902/umfrage/marktanteil-der-genutzten-betriebssysteme-weltweit-seit-2009/
187. Encoding · Docs · Flowplayer. Retrieved November 4, 2013 from http://flowplayer.org/docs/encoding.html
188. XSSed.com. Retrieved January 19, 2015 from http://xssed.com/articleslist
189. transport_security_state_static.json. Contents of /trunk/src/net/http/transport_security_state_static.json. Retrieved December 28, 2014 from http://src.chromium.org/viewvc/chrome/trunk/src/net/http/transport_security_state_static.json
190. Dnssec-Trigger. Dnssec-Trigger. Retrieved December 30, 2014 from http://www.nlnetlabs.nl/projects/dnssec-trigger/
191. DNSSEC/TLSA Validator. Retrieved December 30, 2014 from https://www.dnssec-validator.cz/
192. Lars Eggert – DNSSEC Deployment Trends. Retrieved December 30, 2014 from https://eggert.org/meter/dnssec
193. Google Online Security Blog: Maintaining digital certificate security. Retrieved December 31, 2014 from http://googleonlinesecurity.blogspot.de/2014/07/maintaining-

digital-certificate-security.html
194. Key Internet operator VeriSign hit by hackers | Reuters. Retrieved December 31, 2014 from http://www.reuters.com/article/2012/02/02/us-hacking-verisign-idUSTRE8110Z820120202
195. Gefälschte SSL-Zertifikate auf Reiseflughöhe | heise Security. Retrieved January 17, 2015 from http://www.heise.de/security/meldung/Gefaelschte-SSL-Zertifikate-auf-Reiseflughoehe-2512310.html
196. Browser - Marktanteile in Deutschland bis Januar 2014. Statista. Retrieved March 25, 2014 from http://de.statista.com/statistik/daten/studie/13007/umfrage/marktanteile-der-browser-bei-der-internetnutzung-in-deutschland-seit-2009/
197. Certificate authorities issue SSL certificates to fraudsters | Netcraft. Retrieved October 17, 2015 from http://news.netcraft.com/archives/2015/10/12/certificate-authorities-issue-hundreds-of-deceptive-ssl-certificates-to-fraudsters.html
198. Paypal-Phisher missbrauchen kostenlose SSL-Zertifikate von Cloudflare | heise Security. Retrieved October 17, 2015 from http://www.heise.de/security/meldung/Paypal-Phisher-missbrauchen-kostenlose-SSL-Zertifikate-von-Cloudflare-2514585.html
199. Postmaster WEB.DE. Retrieved December 23, 2014 from http://postmaster.web.de/en/email-policy/
200. Postmaster MailSecurity (GMX, Postmaster Mailsecurity). Retrieved December 23, 2014 from http://postmaster.gmx.de/de/e-mail-policy/
201. T-Online Postmaster. Retrieved December 23, 2014 from http://postmaster.t-online.de/-/id_17044590/index
202. ARCOR POSTMASTER INFO. Retrieved December 23, 2014 from http://postmaster.arcor-online.net/
203. Hotmail Grundsätze, Verfahren und Richtlinien zum Versand von Emails. Retrieved December 23, 2014 from http://mail.live.com/mail/policies.aspx
204. Yahoo Postmaster - Best practices for senders. Retrieved December 23, 2014 from https://help.yahoo.com/kb/postmaster/practices-senders-sln3435.html?impressions=true
205. AOL Postmaster | Postmaster / IP Reputation, the Whitelist, and Inbox Delivery at AOL. Retrieved December 23, 2014 from http://postmaster.aol.com/Postmaster.Reputation.php
206. MSN Partner-Programm zum Melden von Junk-Mail. Retrieved December 23, 2014 from https://postmaster.live.com/snds/JMRP.aspx
207. Mail Server Security Test for DNSSEC and DANE/TLSA - tlsa.info. Retrieved December 23, 2014 from https://www.tlsa.info/
208. PowerEdge C6105 Rack-Server mit AMD Prozessor und 2 HE. Dell. Retrieved March 21, 2015 from http://www.dell.com/de/unternehmen/p/poweredge-c6105/pd
209. Mirrored SAN vs. DRBD | LINBIT Blogs. Retrieved December 12, 2013 from http://blogs.linbit.com/p/347/san-vs-drbd/
210. LINBIT Competitive Enterprise Storage Solution based on DRBD. Retrieved December 9, 2013 from http://www.linbit.com/en/company/news/116-competitive-enterprise-storage-with-drbd
211. Features/Snapshots - QEMU. Features/Snapshots - QEMU. Retrieved March 27, 2015 from http://wiki.qemu.org/Features/Snapshots
212. NETWAYS GmbH Icinga. Retrieved April 4, 2015 from https://www.netways.de/produkte/icinga/
213. Icinga Dokumentation - verteilte Überwachung. Retrieved December 20, 2015 from http://docs.icinga.org/latest/de/distributed.html
214. mysqldump — A Database Backup Program. Retrieved April 3, 2015 from https://dev.mysql.com/doc/refman/5.1/en/mysqldump.html

215. DNSPerf - DNS Measurement Tools. Retrieved April 22, 2015 from http://nominum.com/measurement-tools/
216. DNSPerf Network Map. Retrieved April 22, 2015 from http://www.dnsperf.com/network
217. ab - Apache HTTP server benchmarking tool - Apache HTTP Server Version 2.2. Retrieved April 26, 2015 from http://httpd.apache.org/docs/2.2/programs/ab.html
218. 1 Second - Internet Live Stats. Retrieved April 25, 2015 from http://www.internetlivestats.com/one-second/
219. Large-scale graph partitioning with Apache Giraph | Facebook Code. Large-scale graph partitioning with Apache Giraph | Engineering Blog | Facebook Code | Facebook. Retrieved June 19, 2015 from https://code.facebook.com/posts/274771932683700/large-scale-graph-partitioning-with-apache-giraph/
220. ICH-GCP. Retrieved from http://ichgcp.net/de/
221. Internetauftritt der Bundesbeauftragten für den Datenschutz und die Informationsfreiheit - International - Safe Harbor. Retrieved October 6, 2015 from http://www.bfdi.bund.de/DE/Europa_International/International/Artikel/SafeHarbor.html
222. Safe Harbor - List. Retrieved October 6, 2015 from https://safeharbor.export.gov/list.aspx
223. On the Security of RC4 in TLS. Retrieved October 6, 2015 from http://www.isg.rhul.ac.uk/tls/
224. Details on the “Crime” Attack. Retrieved October 6, 2015 from https://www.nccgroup.trust/us/about-us/newsroom-and-events/blog/2012/september/details-on-the-crime-attack/
225. Security Labs: Is BEAST Still a Threat? | Qualys Community. Retrieved October 6, 2015 from https://community.qualys.com/blogs/securitylabs/2013/09/10/is-beast-still-a-threat
226. Heartbleed / OpenSSL / CVE-2014-0160. Retrieved December 7, 2014 from https://www.dfn-cert.de/aktuell/OpenSSL-Heartbleed-Schwachstelle-CVE-2014-0160.html
227. POODLE / SSL 3.0 / CVE-2014-3566. Retrieved December 7, 2014 from https://www.dfn-cert.de/aktuell/POODLE-SSL3-0-Schwachstelle-CVE-2014-3566.html
228. Security Labs: SSL Pulse: 49% Vulnerable to CVE... | Qualys Community. Retrieved January 4, 2015 from https://community.qualys.com/blogs/securitylabs/2014/06/13/ssl-pulse-49-vulnerable-to-cve-2014-0224-14-exploitable
229. Weak Diffie-Hellman and the Logjam Attack. Retrieved October 6, 2015 from https://weakdh.org/

13 Lebenslauf

Christian Matthias Bernhard Hanshans

Kitzinger Str. 1a
97320 Sulzfeld
hanshans@lobustho.de
Geburtsdatum: 31.10.1979
Geburtsort: Würzburg
Familienstand: ledig

Ausbildung

1986 – 1990	Besuch der St. Hedwig Grundschule, Kitzingen
1990 – 2000	Besuch des Armin-Knab-Gymnasiums, Kitzingen
2000	Abitur am Armin-Knab-Gymnasium Leistungskurse: Biologie, Englisch Grundkurse: Mathematik, Religion
2000 – 2001	Zivildienst bei Kolping-Mainfranken gGmbH, Würzburg
2003	Studium der Medizininformatik Hochschulverbund Oldenburg-Ostfriesland-Wilhelmshaven Schwerpunkt: Biomedical Engineering und Medizininformatik. Studium Sicherheitsingenieur Bundesanstalt für Arbeitsschutz und Arbeitsmedizin
2006	Auslandsstipendium an der Oregon-State University
Mai 2007	Diplomarbeit „Charakterisierung von Polymerfilmen für neurobiologische Retinaimplantate“ Universität Oldenburg – Institut für Physik/Energie u. Halbleiterforschung (Prof. Dr. Jürgen Parisi)
Aug. 2007	Fachkunde mit Abschlussarbeit Sicherheitsingenieur Firma Knauf Gips KG Explosionsschutz für Braunkohlenstaub-Befeuerung
Aug. 2007	Abschluss mit mündlicher Diplomprüfung Medizininformatik
Sept. 2007	Abschluss mit mündlicher Prüfung Sicherheitsingenieur
Okt. 2007 – Jan. 2008	Praktikum Pflegedienst Horn´sche Spitalstiftung Alten- und Pflegeheim Dettelbach
Mai 2008 – Sept. 2009	Leitung des Kompetenzzentrum eLearning in der Medizin Bayern

	Medizinische Fakultät der Universität Würzburg
Okt. 2009 – Nov. 2015	Studium Humanmedizin an der Universität Regensburg
Nov. 2015	Abschluss mit mündlicher Prüfung Humanmedizin

Berufspraxis

2000 – 2001	Zivildienst bei Kolping-Mainfranken gGmbH
2001 – 2004	Tätigkeit als EDV-Dozent (Word, Exel, Netzwerktechnik, Betriebssysteme, Internet&Security, Finanzmathematik) für Kolping-Mainfranken gGmbH, Kolping-Bildungswerk e.V., Kolping Trainingszentrum Miltenberg GmbH, Arbeitsamt Aschaffenburg/Würzburg
seit 2004	Lobustho - Ingenieurbüro für neue Medien Beratung, Konzeption und Entwicklung von Internettechnologien und webbasierten Softwarelösungen
2006	Viewplus Technologies Entwicklung eines neuen Webstandards (XForms) zur barrierefreien sBereitstellung von Formularen und Vektorgrafiken als Alternative zu PDF Dokumenten und Pixelgrafiken. Implementierung in die Braille Drucker sowie das Tactile-Audio System IVEO
2006 – 2007	Translations.com Übersetzung/Anpassung englisch-sprachiger Software/Handbücher für den deutschen Markt
2007	Knauf Gips KG Tätigkeit als Sicherheitsingenieur
2008-2009	Universität Würzburg Institutsleitung, Tätigkeit in Forschung und Lehre
seit 2009	Universitätsklinik Würzburg Klinik für Innere Medizin II - Infektiologie Wissenschaftlicher Mitarbeiter (Teilzeit) Koordination des Projekts „vhb Infektiologie“ Entwicklung eines fallbasierten Trainingssystems für Studierende der Humanmedizin

Eidesstattliche Erklärung

Ich erkläre hiermit, dass ich die hier vorgelegte Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer die angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe de Quelle gekennzeichnet. Insbesondere habe ich nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder andere Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für die Arbeit erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

_____ _____

Ort, Datum Christian Hanshans