Universität Regensburg
Fakultät für Wirtschaftswissenschaften
Lehrstuhl für Wirtschaftsinformatik I - Informationssysteme

Sustainable Identity and Access Management

Dissertation

Zur Erlangung des akademischen Grades “Doktor der Wirtschaftswissenschaft (Dr. rer. pol.)” an der Universität Regensburg gemäß der Promotionsordnung vom 23.07.2014, eingereicht an der Fakultät für Wirtschaftswissenschaften

vorgelegt von
Matthias Hummer, M.Sc.

Berichterstatter:
Prof. Dr. Günther Pernul
Prof. Dr. Doğan Kesdoğan

Tag der Disputation: 28.05.2019

To my parents, Irmgard and Peter,
and my girlfriend Theresa

Abstract

For today's enterprises, information technology (IT) evolved into a key success factor affecting nearly all areas of value chains. As a consequence, identity and access management (IAM) is established for centralized and structured management of digital identities together with their access to internal assets. During this effort, a centralized management platform is created, which serves as middle-ware among available software systems and human resource applications, thereby creating a unified view and enabling business-oriented management. This enables the implementation of an according level of IT-security, business process automation and the alignment to external compliance requirements. However, as IT-infrastructures evolve over time, thereby leading to continuous changes and varying demands, these developments need to be addressed within IAM in a constant manner. As IAM is designed as a cross-cutting topic between business and IT , business requirements such as restructurings need to be realized likewise. Additionally, more and more legal requirements are set in place by external authorities which affect the way digital information are to be managed. Bringing together requirements of these different stakeholders in a comprehensive way imposes high complexity for enterprises, thereby leading to high administrational effort. This leads to a situation where enterprises are in need to constantly evaluate and adapt their implemented IAM strategy and execution. Thus the dissertation at hand is devoted to provide means of aligning IAM to a more sustainable way of operation. Within information systems research, sustainability comprises the ability to meet the needs of today without hindering future developments. To achieve this, the two concepts IAM measurement and IAM policies are leveraged. Firstly, IAM measurement enables enterprises to achieve detailed information concerning the state of an IAM infrastructure. Secondly, this effort is fostered to shift IAM to a more dynamic way of operation and provide suitable recommendations concerning how to adjust different aspects of IAM in a long-term manner. During the research process, the presented approaches have been evaluated within real-world scenarios to outline their relevance and demonstrate practical applicability.

Acknowledgement

Within the last years, I have been able to work with various really clever and inspiring colleagues, who strongly supported me writing this dissertation. To begin with, I would really like to say thank you very much to all of you.

First of all, I would like to thank Prof. Dr. Günther Pernul for his help and guidance throughout my doctoral studies. He embodies the rare ability to give all necessary freedoms to his research students, enabling them to devote themselves to interesting research topics while simultaneously providing valuable hints, directions and feedback when needed. From my personal experience, I can say that he has a very fine intuition concerning how to express critics in a fruitful and purposeful way, which at some points provided me the push I needed. Likewise, I would like to thank my second supervisor Prof. Dr. Doğan Kesdoğan who provided valuable and constructive feedback.

Special thanks goes to Dr. Michael Kunz with whom I began the journey of our PhD studies. While we shared office from the beginning, we supported each other in our free time as well as during working hours. Your open-mindedness and discussions strongly supported me in this dissertation. I'm glad that our journey will continue within future. Another big part of this dissertation can be attributed to Dr. Ludwig Fuchs. Your incredible knowledge and ideas helped me since the beginning. Thank you for the last years. Not to forget, during my time as PhD student I was working with Nexis GmbH. I would like to thank you guys for the discussions and for giving me time to finish this thesis. Hereby I would like to highlight Sebastian Groll. Your positive mind really supported me during the last time of this dissertation.

Additionally, I would like to thank my colleagues from the Department of Information Systems, who went through the process with me. Thank you Michael, Sabri, Hannes, Stefan and Christian, best wishes for the future. Of particular note is Dr. Michael Netter who taught me his understanding of research and strongly influenced the way I perceive research. In this context, I need to mention Florian Menges. From my time as a M.Sc. student, you were always available for discussions or poker nights - thank you very much, the next Laphroaig is on me.

Mist importantly, I would like to thank my familiy - you have always been by my side and constantly believed in me more than I did. Last but not least, I need to mention my girlfriend Theresa - I know you had to really show patience within the last years. Thank you for being the way you are.

Contents

List of Tables

List of Figures

Part I

Overview of the Thesis

1 Introduction

With the start of the digital revolution [Sch17], information technology (IT) eventually became an essential element of organizations. Initially, computer systems have been operated in an isolated way, with data having been transmitted using physical mediums such as the floppy disc. Managing such IT-systems represented a mainly technical task, as computers were geographically static, not continuously connected to other IT-systems and thereby not substantially mission-critical for enterprises yet [Sim73]. Rising computational power, the development of enterprise networks and the internet changed the way information technology has been addressed. This allowed decentralized data storage and distributed processing by different users. Although providing various advantages and opportunities, enterprises have been confronted with the task of managing access to these resources in order to retain a sufficient level of IT-security.

While medium to large-sized companies started to tackle this topic either by implementing custom applications to handle recurring tasks or by investing an increasing amount of manual effort, vendors started to implement standardized products to enable structured identity and access management (IAM). This was necessary due to the development of the so called identity chaos [FP10]. As applications were designed without taking the administration of identities and access as primary factor into account, these are now considered identity silos, where each application managed user accounts independently. As a result, enterprises found themselves in a situation where it was extremely hard to identify who had access to which resources together with recognizing ownerships of accounts spread among various applications. Providing access was executed in a semi-structured manner, thereby leading to an inestimable proliferation of independent identities across the IT-infrastructure [FP10]. Additionally, employees accumulated access privileges over time, as no structured means were established to withdraw access privilege assignments e.g. while moving an employee to a new organizational unit. As a result, IAM was designed to implement a structured identity lifecycle. This includes the creation of a digital identity in case of a new employee joining the enterprise, mover processes to reassign employees to another department or function and the deletion of the digital identity in case of dropout. Yet, early IAM implementations focused mainly on technical demands, which did not recognize the need of collaborating with business users, who embody necessary knowledge to meet suitable decisions.

1.1 IAM Architecture

As enterprises employ potentially thousands of people, it is nearly impossible for IT-staff to know the exact field of activity of each person, together with knowing which applications he needs to access in which way. Due to the IT-centered administration effort, amongst other things, different security incidents occurred such as insider misuse, which resulted in the establishment of regulatory compliance requirements like the Sarbanes-Oxley Act (SOX) [Uni02]. Additionally, further compliance requirements have been set in place within different scopes. For the banking sector, the "minimum requirements for risk management" (MaRisk) [Fin12] and Basel III [Bas11] have been established, while within europe the "europe's general data protection regulation" (GDPR) [VB17] was set in place, together with complementary general guidelines such as ISO 27001 [Wat13]. This led to additional demands for IAM like the implementation of ownerships where each managed entity (e.g. access privileges or user accounts) must be managed by a designated employee. Additionally, the periodic re-certification of access privilege assignments to address over-privileged users and comprehensive external audits of access became mandatory requirements. This enforced the development of IAM as a data governance platform on the one hand and as a business-centered and company-wide management factor on the other hand. To achieve this goal, IAM architectures utilize different building blocks to integrate technology and business as depicted in Figure 1.

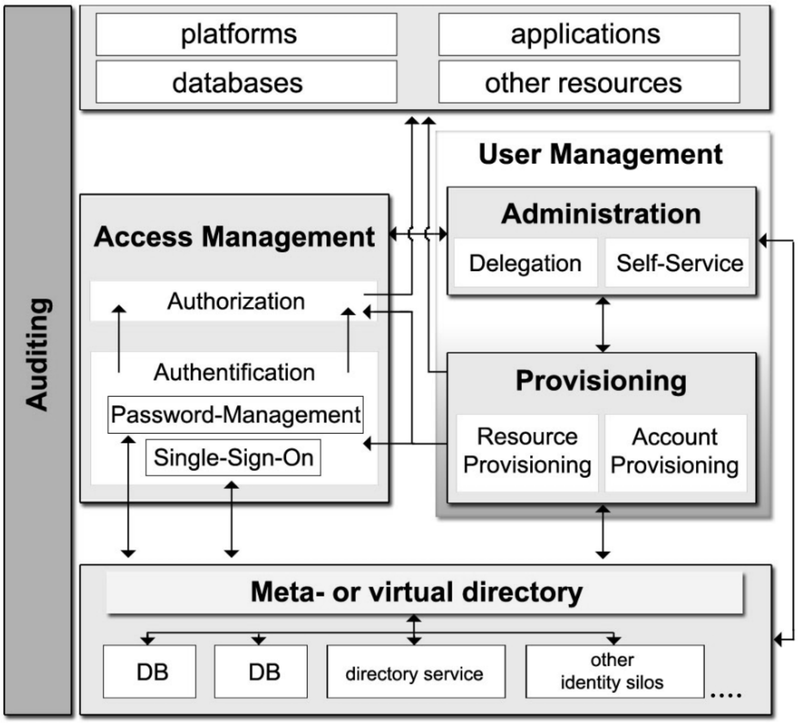


Figure 1: IAM building blocks [FP07]

The main components of IAM infrastructures include *access management*, *user management*, *provisioning*, *auditing* and *data storage*. Meta- or virtual directories enable a centralized view over distributed identity data and thereby enable integrated management of these data. Access management includes the management of the implemented access control model. During this task, technical access privileges are abstracted in order to support business-friendly management and maintenance. Additionally, building blocks such as single-sign-on and credential management provide means of lowering manual effort for the IT staff as well as for business departments, thereby clearly reducing costs. For example, a recent study showed that password reset costs may reach about one million dollars [MC18] within large enterprises. User management is devoted to implement according processes to manage the identity lifecycle and provide a user-friendly interface to enable user self-service. These management efforts need to enforced within connected resources such as applications or cloud services. For this task, automated provisioning is used, which transmits and translates changes of identity data to other applications. On top of this foundation, additional concepts like IAM governance, privileged user management or identity federations can be implemented.

Due to its central role and various functionalities within corporations, IAM represents a key-success factor when it comes to managing a distributed, proprietary and heterogeneous IT infrastructure. This enables enterprises to shift from monolithic application management to a service-oriented IT-environment by integrating applications, custom-developed applications and historically required software systems. Thereby the foundation is created to implement a required level of IT-security for protecting internal assets from external attacks and insider misuse.

1.2 IAM Maturity

Keeping operations, strategy, concepts and technologies constantly aligned to these internal and external demands represents a challenging and complex task. Taking into consideration that each assigned access privilege may impose a crucial risk (as erroneous assignments may for example lead to unauthorized flow of information or data corruption), management must be conducted in a careful way, which requires company-wide collaboration of domain experts. This becomes even more important as the increasing trend of pushing IT to a high level of dynamics, flexibility and integration results in enterprises being forced to constantly evaluate, rethink and adapt IAM processes and technology. Recent trends like lean value chains, tight cooperation with suppliers, multi-tenancy cloud access management, the integration of customer identities or industry 4.0 amplify this development, thereby pushing companies to continuously adapt their IAM strategy to meet occurring demands. In order to handle this complexity, enterprises continue to evolve the maturity level of their IAM as depicted in Figure 2. Windley et al. [Win05] differentiate between the following IAM maturity levels:

- *Ad hoc*: IAM infrastructure is designed ad hoc with hardly any structured processes available.
- *Focused*: structured processes have been implemented for a limited number of activities.
- *Standardized*: Enterprise-wide policies and processes are established together with a suitable IAM infrastructure.
- *Integrated*: Mature processes and policies are put into place with a high degree of automation.

The given classification focuses on how well current requirements may be managed, yet does not take into consideration to which degree the currently established IAM is capable of being adapted to varying future environments in a *sustainable* way. Applied to IAM, sustainability may be defined as "meeting the needs of the present without comprising the ability of future organizational development to meet their needs" [Kee88]. Taking into consideration how fast IAM is evolving together with the relevance for today's enterprises, this dissertation is devoted to close this gap by providing means for sustainable IAM. This is achieved by integrating operational and strategical IAM, in order to identify mutual dependencies and thereby provide suitable measurement and management methodologies.

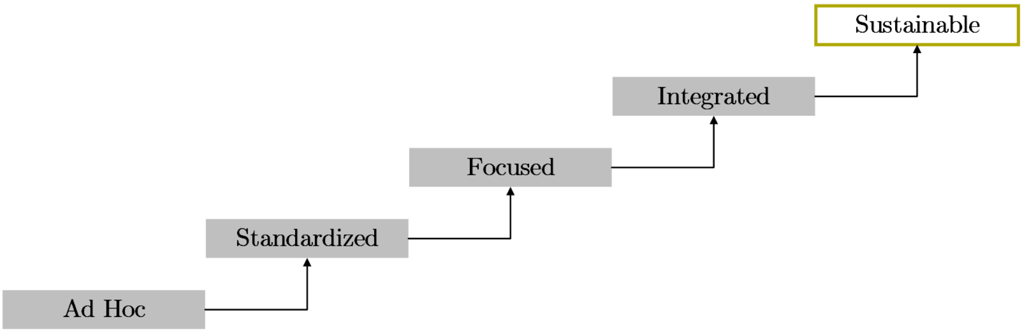


Figure 2: Extended IAM maturity levels based on [Win05]

1.3 IAM Strategy and Operation

In general, strategy is devoted to setting goals together with actions to achieve these goals, while taking limited resources and environmental factors into consideration [Fre15]. Within IAM, these goals are a set of defined business goals. Based on research and practice, these include risk management, cost reduction, compliance, IT governance, data and process quality and business facilitation [Win05, MBPS10, Osm13]. Limited resources may be understood as a given amount of human resources (working staff) together with limited economical investment options of an enterprise. Environmental factors concern different fields of relation regarding organizational as well as technical concerns. Different stakeholders such as management, users or external authorities [WYSS09] express different demands concerning what is to be achieved by IAM. IT-environments of enterprises comprise both, internally used software systems as well as external services and products. Additionally, organizational structures may change over time, e.g. due to carve-ins, carve-outs or due to restructurings. Strategical IAM needs to take these different factors into consideration in order to define a suitable high-level line of approach which needs to be constantly evaluated and potentially refined in order to continuously meet organizational and technical demands.

While strategy is devoted to long-term goals, "operations is the activity of managing the resources and processes that produce and deliver goods and services" [SL02]. In other terms, it is focused around the "input-transformation-output" model of operations [SL02]. Within IAM, "input" references different interactions in an organizational or technical manner. From an organizational point of view, employee actions (e.g. joining the enterprise, changing department or accessing IT-systems) need to be addressed, while technical events occur during the interaction with IT-systems (e.g. reconciliations [WYSS09] or the connection of additional applications). The "transformation" is based on strategical IAM decisions of the enterprise. Depending on the defined technologies (e.g. single-sign-on), paradigms (e.g. compliance) and process models (e.g. joiner / mover processes), certain actions are executed. The "output" itself is aligned to different IAM business goals that are are commonly strived for, such as the implementation of compliance requirements or facilitation of business operations. Consequently, strategical and operational IAM strongly influence each other. On the one hand, operations implements strategical decisions, yet the applicability is strongly dependent on the overall architecture together with the availability of resources. Strategical IAM, on the other hand, needs to take the current implementation state into consideration and needs to align the long-term roadmap accordingly.

Consequently, sustainable IAM as further evolved maturity level requires an integrated approach regarding these two dimensions. Within the context of this dissertation, sustainable IAM aims at the integration of recommendation mechanisms into IAM processes and components in order to improve user and authorization management on a strategical as well as an operational level. The recommendation mechanisms aim at supporting contemporary decision-making and ease the adaption to varying environments while retaining the flexibility to meet future IT and business-related demands. For this purpose, strategical information and operational data are utilized and evaluated as a foundation to adjust enterprises' IAM. While continuous adaption enables IAM to stay aligned to requirements, a limited number of resources needs to be taken into consideration. In order to enable the implementation within a feasible framework, an additional pillar of this dissertation is dynamic IAM. On the one hand, this includes the integration of dynamic concepts into strategical management. On the other hand, shifting operational implementations to a more dynamic foundation allows IAM to semi-automatically adapt to changes. Thereby the focus lies on adjusting and extending existing concepts to enterprises' needs, which enables the infrastructure to respond to sustainable IAM recommendations. A conceptual overview is depicted at Figure 3.

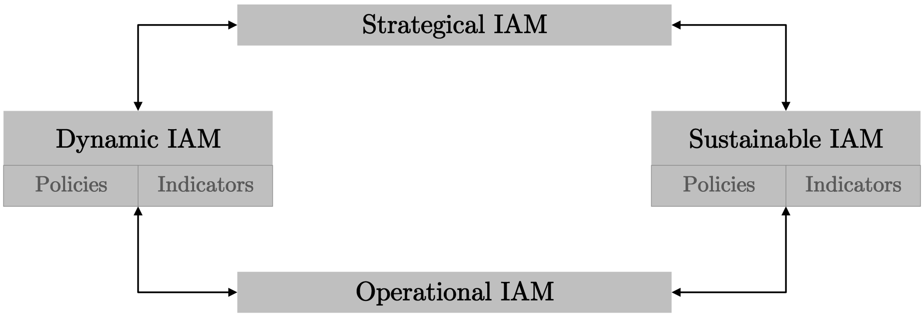


Figure 3: Conceptual overview of the dissertation

In order to achieve sustainable IAM, the dissertation at hand leverages two concepts. Based on the concept "you can't control what you can't measure" [Gil88], IAM indicators are used to improve the general measurement and evaluation of current IAM implementation and strategy. These indicators are designed to be flexibly aligned to enterprises' needs and therefore represent potentially complex coherencies as understandable values. Secondly, IAM recommendation functionality is provided based on these IAM indicators together with policies as an additional decision foundation. Such policies may either be strictly enforced (e.g. in form of compliance policies) or be implicitly implemented, thereby providing guidelines concerning the commonly implemented mode of operation (e.g. by depicting standard access privilege assignments recommended to users). Consequently, the dissertation at hand is devoted to establish a structured connection between the management of IAM strategy together with the execution of operational IAM.

2 Research Questions

In recent years, enterprises gained a thorough understanding concerning what goals are to be achieved by utilizing IAM. Yet the topic's complexity encumbers constant effort which may lead to its potential laying fallow. Consequently, sustainable long-term development represents a task, which needs to be reflected throughout planning and executing of IAM in order to reduce this complexity and allow a long-term oriented and dynamic execution. During our research, we identified different unanswered research questions targeting operational and strategical IAM. This dissertation essentially aims at answering the following research question:

RQ: *How can operational and strategical IAM be integrated to achieve sustainable IAM and how can this development be supported by dynamic IAM?*

The dissertation is built upon the need for an integrated approach of the given dimensions strategy and operation. While continuous feedback and adjustments are necessary to keep IAM aligned to various requirements, this must not constantly provide additional effort for enterprises. This might otherwise lead to a situation where the given approach may not be addressed in a suitable manner due to a prioritization of short-term issues, leaving out long-term goals. By adjusting existing means of dynamic IAM to enterprises' demands, these are leveraged to enable more effective adaption to changing requirements. The research question is divided into five separate sub-questions. As a starting point, the current state of IAM within practice needs to be derived in order to establish a solid knowledge foundation. After this effort, the topics sustainable and dynamic IAM are approached, starting from operational IAM and consequently advancing to a comprehensive approach also including of strategical IAM.

RQ 1 - Strategical IAM: *What strategical IAM goals are currently pursued by enterprises and how are these handled?*

Within the past, enterprise IAM developed from decentralized managing internal employees and IT-resources to a fully integrated governance and business-oriented platform, which manages internal systems, external (cloud) systems, customers and supply chains. This trend resulted not only in new and extended use cases for IAM application but also in a need for constant improvement of performance and functionalities. As a consequence, enterprises are in need to prioritize different aspects of their IAM to fulfill contemporary requirements and thereby facilitate business demands. Several research publications have been devoted to present different applicable technologies (e.g. [Win05, AA11], process models [Osm13] or economical investment factors [Roy13, MBPS10]) to support this trend. While implementing an according strategy needs to take a lot of different IAM goals and environmental factors into consideration, research currently lacks information concerning IAM goal relevancy in which way these are currently addressed. By developing a structured set of possible IAM goals and IAM indicators and consequently matching these with data from medium to large sized companies, requirements for IAM strategy are to be derived.

RQ 2 - Operational IAM (Dynamics): *Which data are available within operational IAM and how can these be used to make IAM more dynamic?*

Operational management of digital identities and their access represents a labor-intensive and error-prone task when executed based on a static foundation. This includes topics like the static assignment of access privileges or a static set of activities which need to be conducted as a result of organizational events. IAM administrators need to consider various internal and external requirements during process executions, access model management or the assignment of access privileges. While various IAM processes are executed on a daily basis within medium to large-sized enterprises, each participating decider needs to take all of these requirements into consideration, which represents a hardly feasible task. As a consequence, critical access privilege assignments accumulate over time, resulting in an increased management effort over time and a reduced IT-security level [FP07]. This topic has largely been acknowledged within research, as commonly IAM is viewed as being built on the three pillars technology, processes and policies [FP07, Roy13]. Within this context, policies represent the foundation to enable compliant operation regarding internal and external requirements, thereby acting as a countermeasure to potentially erroneous operations. Yet this concept is not leveraged to its full potential, as policies are often only implemented in a rudimentary way and hardly updated after being put in place. By answering the given research question, the goal is to provide a feasible foundation of policies which may be managed in a semi-automated way and therefore be utilized to increase dynamics of IAM together with decreasing administrational efforts.

RQ 3 - Dynamic IAM: *How can enterprises be enabled to adjust IAM in a manner to provide a dynamic infrastructure for strategical and operational management?*

Enterprise IAM utilizes different concepts and process models to manage the identity lifecycle, which are defined during strategical planning and later implemented during IAM operations. Commonly, each organizational IAM event requires the execution of a structured process, e.g. obtaining approvals to assign additional access privileges or changing workplace, which requires recurring and time-consuming business input. Within access management, role-based access control (RBAC [SCFY96]) developed into the de-facto standard for managing access of digital identities within enterprises. This concept introduces roles as intermediates, which enables shifting access management away from technical resources to the business level. Yet the static approach showed various drawbacks like a proliferation of access privileges due to the complexity involved within the past. In order to shift to a dynamic operation, IAM needs to be approached in a strategical and in an operational level by emphasizing topics like data quality, homogenization and automation. Consequently, a process model is to be introduced which enables enterprises to shift to dynamic IAM together with providing means of integrating this concept into IAM executions.

RQ 4 - Operational IAM (Sustainability): *How can operational IAM be improved by utilizing contextual data?*

IAM data are created, updated and deleted on a daily basis to reflect changes which occur within enterprises. These changes stem from different sources, e.g. changes within the human resources system, administration within the IAM system or modifications made within connected applications. Although the complexity and amount of data grows over time, there are hardly any means available to establish measurements of quality. Within large-sized enterprises, commonly up to millions of access privilege assignments are assigned to thousands of employees, granting them access to various applications. While traditional data quality concepts are already applied to IAM data (e.g. to improve attribute quality), the evaluation of business-induced data requires additional data analysis techniques. Taking assigned access privilege assignments as an example, the evaluation of access model quality needs to take additional dimensions into consideration like the context of an employee such as his department or the location and time he activates the access privileges. By answering this research question, these contextual data can be used in order to measure different dimensions of IAM. The results enable the creation of recommendations to leverage ongoing IAM tasks.

RQ 5 - Sustainable IAM: *How can enterprises use available contextual data to continuously evaluate and steer IAM strategy and operations?*

In general, the definition of an IAM strategy comprises the definition of organizational requirements, derivation of environmental factors, discussion with stakeholders, identification of current gaps and the creation based on best practices [WYSS09]. The implementation is executed by using different IAM concepts and technologies, for example by setting appropriate processes in place and installing necessary technologies. Due to the complexity of IAM, thoroughly understanding the impact of such decisions to the overall system is a challenging task. Take the connection of an additional application of a company to the IAM system as an example. While this enables improved centralized management, already implemented IAM processes may not be completely suitable for the administration of this application, e.g. due to poor attribute quality or missing knowledge. In order to prevent decreasing attribute quality or rising manual effort (e.g. due to an increased number of help-desk calls), this requires contemporary IAM feedback and adjustments. Additionally, due to the volatility of today's enterprises and their IT infrastructures, IAM needs to be aligned continuously. This constantly extends the application of IAM, however this also makes it hard to evaluate the mutual impacts. By answering this research question, we aim at providing a tool-based methodology which is designed to provide a flexible and customizable decision support methodology, which takes future predictions into consideration.

3 Research Methodology

Research of information systems (IS) sources historically out of the management and business community and therefore utilizes mainly two research paradigms [VK15, ÖBF^{+}11]. Firstly, *behavioral science* is rooted in natural science research and focuses on the development and the evaluation of theories, which explain or predict organizational behavior [HMPR04]. Consequently, it is built on the analysis of social or technical data based on the design, implementation and management of information systems. During this process, information about the interaction among people, technology and organizations is derived for research and practitioners [HMPR04]. Secondly, *design science* is rooted in engineering and focuses on the analysis, design, creation, management and evaluation of innovative artifacts which support organizational and human demands [HMPR04]. Within information system research, phenomena are artificial and can therefore be created and studied. Although these two paradigms use different approaches, they act as complementary forms of research and influence each other, which may be described as a research cycle, as outlined in Figure 4. Theories and knowledge of behavioral science form the theory and knowledge basis, while design science is focused around the creation of artifacts, which may be utilized to extend knowledge.

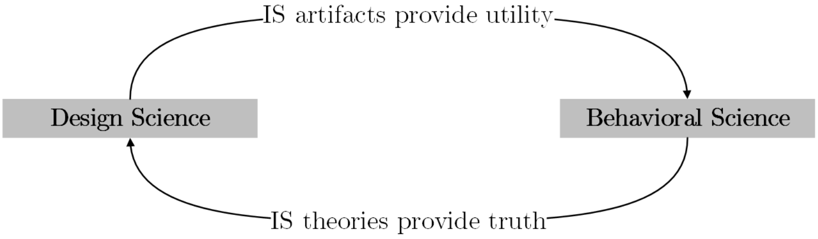


Figure 4: Complementary research cycle [HC10]

The dissertation at hand is focused around the context of the German *Wirtschaftsinformatik* [ÖBF^{+}11], which emphasizes the use of the design science paradigm in contrast to the Anglo-American information systems research community, which focuses mainly on behavioral science. This decision is rooted in the fact that enterprise IAM is a very practical-oriented research topic, which allows the creation and evaluation of artifacts within real-world scenarios, thereby emphasizing applicability.

Both research paradigms aim at overcoming existing boundaries and contribute to the existing knowledge base [HMPR04]. Design science creates value by solving existing problems and creates novel and innovative artifacts or enhances existing artifacts for this purpose [Ake04]. However, the construction of artifacts needs to differ from routine design, which is described as solving problems by applying already existing concepts and knowledge, therefore not leading to novel knowledge [HMPR04]. Design science requires innovativeness and creativity to contribute to the existing knowledge base. In order to achieve relevance and rigor, this dissertation is built on the research framework based on [HMPR04] as depicted in Figure 5. This framework consists of the building blocks *environment* and *knowledge base* to execute *information systems research*. *Environment* includes people, organizations and technology, which interact to achieve a certain goal [HMPR04]. In this context, the environment is associated with enterprise IAM, together with its impact on organizations, users and information technology. The *knowledge base* incorporates theories, concepts and methodologies rooted in published research which utilized according research paradigms and is used within the process of developing novel artifacts [HMPR04]. This dissertation is built around the knowledge foundations IAM, data analysis methodologies and information system management theories. *Information systems research* consists of two available phases, 'Develop / Build' and 'Justify / Evaluate'. During the develop and build phase, an artifact is designed based on environmental requirements, while the justify and evaluate phase addresses the analysis and justification of the artifact.

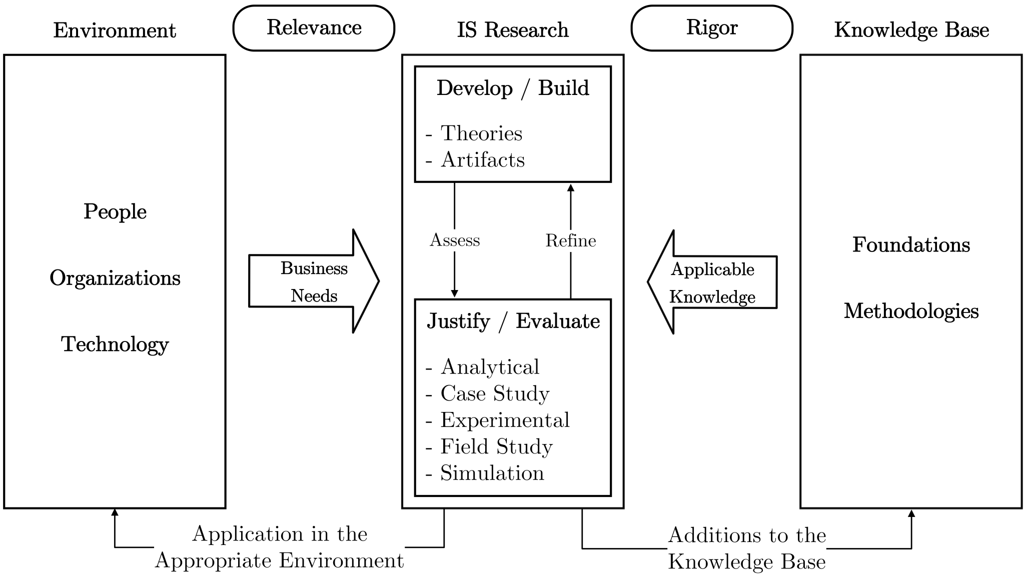


Figure 5: Information systems research framework [HMPR04]

The presented research framework provides a foundation to ensure relevance and rigor within information systems research. Different publications elaborated this framework by introducing research guidelines, which define individual activities that need to be gone through in order to achieve valid results [PTRC07, HMPR04, MS95]. Within this dissertation, the seven guidelines of Hevner et al. have been applied due to their wide acknowledgement within information systems research [HMPR04]:

- Guideline 1: Design as an Artifact

  The output of design science research is a purposeful artifact which meets business needs. In this context an artifact can be a construct, a model, a method or an instantiation. Therefore the artifact is equally interdependent and coequal regarding people, organizations, social context and business needs.

To address this guideline, an artifact is constructed and described in detail within each of the provided research articles. Firstly, the design is based on people who interact with IAM. Within the context of this dissertation, this includes technical IT-staff and business users. Secondly, the organizational view is taken into consideration by focusing on how to integrate the presented artifacts into available enterprises, with each artifact addressing a certain business need.

- Guideline 2: Problem Relevance

  The overall goal of design-science research is to apply the designed artifact in order to solve a relevant and important business problem. Business problems and opportunities commonly focus on increasing revenue or decreasing costs through the design of appropriate business processes, which are supported by information systems.

  Consequently, this work focuses on different problems within IAM. These are commonly caused by the complexity and amount of required effort to manage digital identities within an enterprise. If not addressed properly, different assets like IT-security, economical investments or the reputation of an enterprise can be affected negatively. Within each research article a comprehensive description of the addressed problem is given.

- Guideline 3: Design Evaluation

  The designed artifact must be well evaluated concerning utility, quality and efficiency. It is only complete and effective when it satisfies the needs and constraints of the respective problem. For this purpose suitable evaluation methods have to be selected and rigorously demonstrated. These methods may be observational, analytical, experimental, testing and descriptive.

  As IAM is implemented within a wide variety of enterprises, the designed artifacts were mainly evaluated using the testing method. Thus it was analyzed by embedding it into a real-world IAM infrastructure together with productive data. This enabled to identify potentially necessary design adjustments and proved its additional value.

- Guideline 4: Research Contributions

  Research needs to clearly lay out in which way it contributes to the knowledge base in a novel, general and significant way. There are three possible types of contributions within design science research. Firstly, the contribution may be the artifact itself. Secondly, a foundation may be provided such as models, formalisms or ontologies. Thirdly, methodologies includes the development and use of evaluation methods.

Within this dissertation, mainly artifacts have been constructed. These are designed as a software system for a specific scope within IAM. Additionally, a description is given concerning how this application can be integrated into existing IAM infrastructures. Research article 1 represents an exception as it provides a methodology in surveying the state of IAM within practice.

- Guideline 5: Research Rigor

  Rigor describes in which way research is conducted. In opposition to behavioral-science research's rigor, which is based on appropriate data collections and analysis methods, the rigor on design-science research focuses on the construction and evaluation of the design artifact. Hence rigor must be applied regarding the applicability and generalizability of the artifact.

  Within this dissertation, each artifact is described thoroughly on a conceptual as well as a practical usage level. Each publication provides a research classification and how the presented work can be included into the research foundation. Additionally, each artifact was designed based on available research methodologies, therefore fostering available knowledge.

- Guideline 6: Design as a Search Process

  The design of the artifact is not a singular task but in fact it is an ongoing process in order to find the best solution which also fits the laws of the aspired business environment. This includes taking means, ends and laws into consideration. Means represent available resources and actions to build a solution. Laws act as uncontrollable constraints, while ends depict the goals and constraints of the solution.

  During the conduction of research, the develop / build and justify / evaluate cycle was used. After creating a version of an artifact, it was tested within a real-world scenario and consequently improved. This circle was applied until an according quality level of output was achieved.

- Guideline 7: Communication of Research

  In design-science research the prepared work should not only be presented to the technology-oriented but also to the business-oriented public in order to be most valuable and efficient. The created theories need to include sufficiently detailed technical descriptions and the process concerning how the artifact has been created. This enables reproductionality of results.

  As a consequence, all research results have been published within according conferences and journals focusing researchers as well as practitioners. The creation of each artifact was described in detail, including the reasons why it was created and the process of how it was created.

4 Results

4.1 Overview of the Research Contributions

The research questions which are presented in Section 2 have been addressed by publishing results in well-reputed conference proceedings and journals, with each focusing on IT-security or information systems. Table 1 provides an overview of the five given research papers. Note that the number does not correspond to their chronological order but to corresponding research questions. During the time of writing, the papers 1 - 4 have already been accepted, while article 5 is currently undergoing a review process.

Paper 1 is built around the identification of the current strategical IAM goals which are pursued by enterprises, thereby providing the knowledge foundation to adjust strategical IAM research. Paper 2 and 3 are devoted to the topic of dynamic IAM. This research area has firstly been approached regarding operational IAM with the establishment of a foundation of dynamic policies in publication 2. Consequently, article 3 integrates operational and strategical IAM to include this paradigm within both layers of IAM. The second pillar of this dissertation, sustainable IAM, has been addressed within the publications 4 and 5. Again, within a first step attention has been laid onto operational IAM. Paper 4 represents a substantially extended version of research article 2 in providing novel and relevant research results essential to this dissertation. It introduces the concept of operational IAM indicators, which is fundamental for this work together with implicit policies, which are used to derive recommendations. Additionally, these concepts are used as a foundation for sustainable IAM, that induced the inclusion as self-sufficient publication. Publication 5 connects strategical and operational aspects of sustainable IAM by integrating these layers.

No.	Paper Title	Submitted to	Publication
1	Measuring Identity and Access Management Performance - An Expert Survey on Possible Performance Indicators	International Conference on Information Security Systems Security and Privacy (ICISSP)	Conference
2	Advanced Identity and Access Policy Management using Contextual Data	International Conference on Availability, Reliability and Security (ARES)	Conference
3	Introducing Dynamic Identity and Access Management in Organizations	International Conference on Information Systems Security (ICISS)	Conference
4	Adaptive Identity and Access Management - Contextual Data based Policies	EURASIP Journal on Information Security (JIS)	Journal
5	Analyzing Context Data for Sustainable Identity and Access Management		Journal

Table 1: Overview of research papers within this dissertation

Each publication is designed as a self-contained article. Figure 6 shows how each work contributes to the dissertation at hand. Dynamic IAM research is aligned to the adaption and establishment of dynamic concepts to enterprises' demands, thereby affecting these two IAM layers. Sustainable IAM on the other hand consumes contextual data of strategical and organizational IAM, processes these in an sufficient way (please refer to the respective publications for a detailed description) and provides recommendations to adjust IAM based on this data processing. Therefore it is designed as a building block which consumes IAM input and creates according output to support organizations.

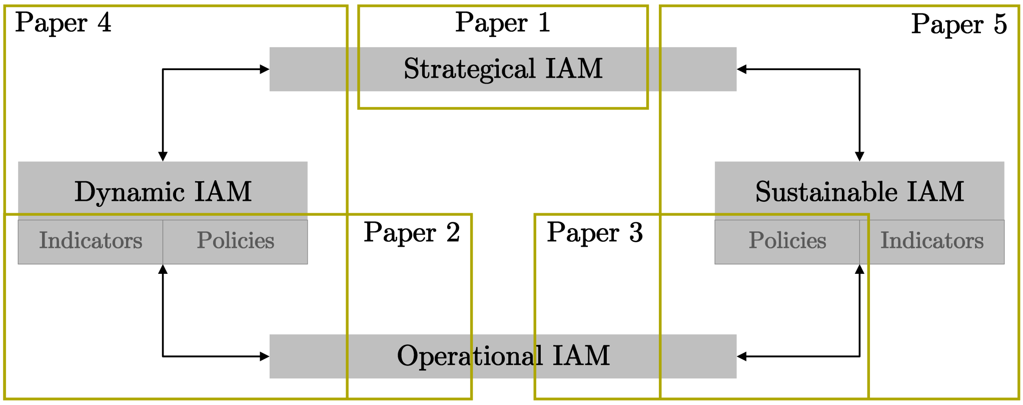


Figure 6: Research papers in correspondence to research areas

4.2 Strategical IAM (Research Paper 1)

A recently conducted study among German enterprises revealed that about 64% of medium to large-sized enterprises already established software-based and centralized IAM, while the bulk of remaining enterprises plan future introduction [MFL$^+$18]. However, the degree of implementation strongly differs concerning the full utilization of IAM technologies such as single-sign-on or mobile device management [MFL$^+$18]. This development shows that enterprises approach this topic individually and prioritize different aspects of IAM. Within this dissertation, the first aim was to identify which IAM goals are currently addressed by enterprises together with measuring how well these goals have already been implemented. Inherently, IAM strategies are custom-tailored to enterprises' needs regarding their respective requirements, current infrastructure and willingness for economical investments. Research results showed that companies struggle with measuring the performance of their currently implemented IAM adequately. Consequently, the future development or improvement of the underlying IAM strategy in place remains an unsolved challenge.

To address this research gap, a structured survey among IAM experts was conducted to evaluate the current status quo. Within a first step, five IAM goals have been derived from research and practice, including risk reduction, data and process quality, compliance requirements, business facilitation and cost reduction. While these IAM goals provide a founded starting point, measuring is difficult due to various aspects which are related to each goal. Following the divide and conquer approach, these goals have been decomposed by applying the goal question paradigm [Bas92]. While originally developed for the field of software engineering, it is currently widely used within research due to its capability to develop qualitatively or quantitatively measurable factors. For each IAM goal, different measurable IAM indicators have been derived using abstraction sheets [AKP02]. These were used to construct and test a structured survey. Each participating enterprise was queried concerning how far respective indicators have been implemented or if an establishment was planned. The results are depicted within Table 2. Based on our scores, data and process quality achieved the highest value with about 83% of participants having already achieved substantial improvement. Surprisingly, the already achieved progress within the topic risk reduction was only at about 45%, yet it is part of mid-term roadmap of about 31% of enterprises.

IAM Goal	IAM Indicator	Improved	Planned
Risk reduction		*25*	
	Number of security incidents due to user and entitlement management	14	6
	Number of security incidents due to critical role and access right combinations	11	5
	Duration until deactivation of employee access rights	25	1
	Duration until emergency deactivation of employee access rights	18	1
Data and process quality		*26*	
	Development of data quality	25	0
	Error rate within access management	18	3
	Error rate within identity and account creation	21	1
Compliance requirements		*26*	
	Reduction of compliance violations	18	6
	Number of successful audits	21	4
	Duration until complete solution of a compliance incidents	11	8
Business facilitation		*22*	
	Reduction of administrational effort	19	3
	Improvement of user satisfaction	13	5
	Duration until employee readiness	19	2
	Duration until access model adjustment	13	4

Table 2: IAM goals and IAM indicator scores based on [HGK$^+$18]

The distribution of participants showed that about 34% of the participants are currently working within strongly regulated sectors (e.g. finance or health care), about 37% work in medium regulated sectors (e.g. manufacturing or service), while the others did not make a statement regarding industry sector. On the whole, out of the five IAM goals, cost reduction had to be discarded due to low relevance. This hints that participating enterprises still perceive IAM as an investment and effort factor, without leveraging its economical possibilities. Additionally, out of 19 identified IAM indicators, 5 were discarded due to low relevance. This includes for example the indicator 'duration until user requests are processed', further underlining the mentioned presumption. Finally, we analyzed the influence of demographic factors on IAM indicators. This evaluation showed e.g. that industry sector and maturity level strongly influence which IAM goals are prioritized. Strongly regulated industries focus on security and compliance, while medium regulated sectors focus on business facilitation and data quality. Furthermore, working positions (e.g. managers or IAM administrators) strongly influence how the overall IAM performance is perceived, as managers in general rate the performance significantly better. Summed up, we identified three major research findings:

- Enterprises struggle in leveraging the full potential of IAM as their focus mainly lies on meeting short-term requirements.
- Requirements are strongly dependent on industry sectors and IAM maturity level.
- Companies currently lack common and thorough understanding of their IAM performance and strategy.

4.3 Operational IAM - Dynamics (Research Paper 2)

Strategical IAM defines in which way IAM is carried out together with managing available human and technical resources. However, the operational level is in charge of actually setting the given principles in place. With an increasing number of employees and access privileges to manage, this imposes an effortful task. For example, identity lifecycle management constantly requires multi-stage approvals for static assignments together with compliance checks e.g. in order to avoid separation of duty conflicts. Thus enterprises aim at shifting IAM to a more dynamic and stringent way of operation. To achieve this, policies are put in place to keep business and IT aligned to internal and external requirements and increase automation. Practical experience, however, showed that such policies are only poorly managed. Instead of a rich set of business-driven policies, only a rudimentary technical set of policies is mostly being implemented. As a result, these are henceforth hardly fostered and updated due to organizational and technical effort. Additionally, these policies operate solely on static identity data such as HR information, identity data or assigned access privileges, strongly limiting their effectiveness.

However, policies are well applicable within different fields of IAM, including compliance, workflows or access management. To identify a structured foundation which may be used to implement dynamic policy management, available IAM data were identified. The result is displayed in Table 3. Column one depicts the system type in which IAM data are managed. Commonly, these include three categories. Firstly, human resource management applications serve as source systems and are used to manage employee master data (e.g. name or function) together with contextual information such as absences or working space. Secondly, target applications are controlled by IAM e.g. to manage access by means of provisioning. Thirdly, the IAM system itself represents the centralized management system. Column two shows which types of data are available, while column three depicts if these data are currently utilized for IAM polices within enterprises. The results show that contextual data are currently hardly fostered accordingly, although being crucial for the definition of dynamic policies.

In order to open up these available data sources, an IAM policy mining extension has been proposed, which utilizes the comprehensive set of available data. The extension

System	Data type	Used
HR system$_{1..n}$	Employee master data	x
	Employee context	
Application$_{1..n}$	Account information	x
	Entitlement information	x
	Account Activity	
IAMS	Identity information	x
	Entitlement/role information	x
	Provisioning information	x

Table 3: Available data types within IAM [HKN$^+$15]

follows the paradigm that existing policies need to be evaluated constantly, while new policies are derived based on current mode of operation. Consider the re-organization of an organization as example. This might lead to a situation where access privileges are used within different contexts and environments (e.g. in different departments), which may automatically be detected and nominated, while available policies need to be outdated as they do not meet the changed demands.

To structure the usage of the given approach, the dynamic policy management process (DPMP) has been introduced, as depicted in Figure 7. This process model provides guidelines concerning how to conduct structured management of policies in the context of enterprise-wide IAM. The first phase consists of the infrastructure setup. During this task, available data sources are connected, normalized and prepared for constant data synchronization. After this step, the process enters the cyclic policy mining activity. During each iteration, data are extracted and collected throughout connected applications. These data are correlated, thereby deriving static and contextual relationships, which serve as foundation of the actual policy mining. As these relations change over time, the mining component is able to detect outlier connections or potential policy adaption requirements. Within the last step, these policies need to be evaluated by IAM experts. This is necessary to reduce the number of potential false-positives and thereby prevent potential hindering of IAM execution. Additionally, recommendations regarding operational management are created, e.g. by detecting uncommon access privilege assignments due to uncommon context.

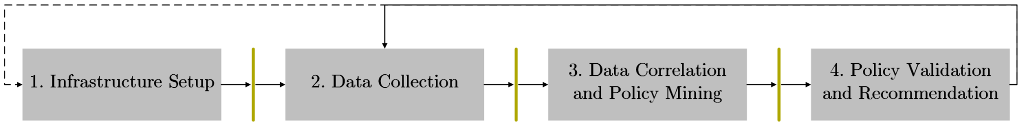

1. Infrastructure Setup
2. Data Collection
3. Data Correlation and Policy Mining
4. Policy Validation and Recommendation

Figure 7: Dynamic Policy Management Process [HKN$^+$15]

Finally, the approach has been evaluated within a real-world scenario in cooperation with a world-wide manufacturing company. Thereby it was shown that using policies, different shortcomings of IAM operations may be identified. Taking activations of access privileges as example, the enterprises was able to identify the focus areas of common usage, while different outlier activations showed erroneous assignments. The presented research article lead to three major research findings:

- Current implementations of IAM policies do not support steering of operational IAM in the desired way, which leads to high manual administrational effort.
- Available context data are not utilized accordingly for IAM purposes.
- Dynamic policy management may improve compliant and stringent IAM execution.

4.4 Dynamic IAM (Research Paper 3)

While the identification and implementation of dynamic IAM policies enable enterprises to clearly improve operational IAM, this topic needs to be addressed in a comprehensive way. To successfully implement dynamic IAM, strategic and operational management need to be shifted in taking this paradigm into consideration during planning and executing of access management and identity lifecycle management. Using access management as a starting point, this has been built on a static basis, with RBAC being by far the most dominant principle within practice. This led to different shortcomings, including the so called role explosion [EK10], duplication and proliferation of multiple access objects due to missing insight and high manual effort. As a result, enterprises aim at shifting to a more dynamic approach, such as attribute based access control (ABAC) [HKS$^+$14]. This concept is based on using attributes of a subject (e.g. an employee), object (e.g. a permission or a technical resource) together with contextual data as foundation to meet access decisions. Thereby employees and their access privileges are adjusted in an automated way when undergoing different changes (e.g. joining / leaving an enterprise or switching department). Yet identity lifecycle management needs to be aligned to achieve sufficient data quality, as erroneous or ambiguous data lead to different drawbacks like unexpected access privilege assignments. Despite providing obvious advantages over static IAM, this approach requires organizational and technical arrangements in order to achieve a suitable data foundation and an according set of access rules.

To address these organizational and technical requirements, two main building blocks have been identified in literature, namely *policy management* and *attribute management*. The goal of this publication is to provide enterprises with a structured migration guide which uses a static IAM state as starting position and the establishment of dynamic IAM structures as its goal. For this task, a process model was proposed which includes the three phases preparation, implementation and maintenance. During the preparation phase, the data foundation and a common understanding of the topic is established. The actual implementation phase is devoted to achieve suitable data quality and organizational efforts to accomplish requirements for policy development and implementation. During the maintenance phase, implemented policies and data quality are continuously optimized and measured.

The first building block *policy management* includes the definition of organizational guidelines and the agreement of a suitable language, the development, simulation and

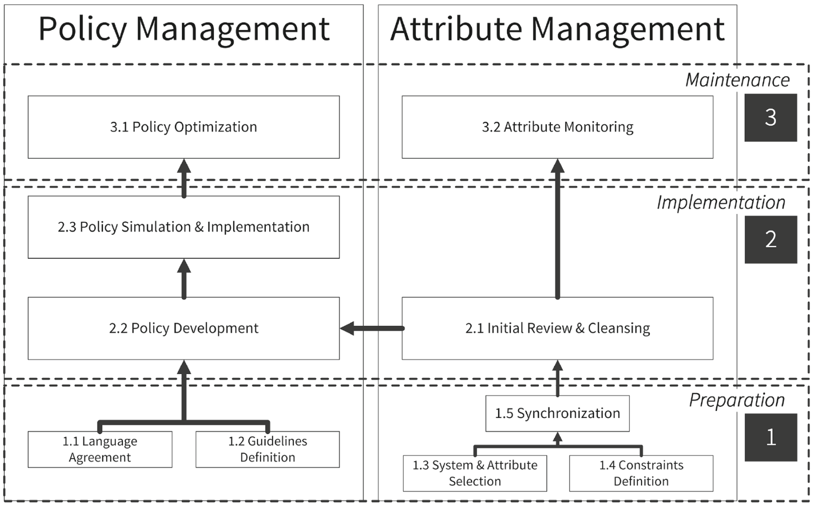


Figure 8: Dynamic IAM migration process model [KFHP15]

implementation of policies, with finally constant optimization. These activities are mainly fostered by research results of dynamic policy management presented at research article 2. In order to provide a solid foundation for these policies, the second building block *attribute management* is devoted to achieve a suitable data quality. Hereby the main focus is set on cleansing available attributes together with providing according monitoring processes to achieve high quality in a long-term manner.

The presented approach has been evaluated within two use-cases, with each utilizing a real data set of an enterprise. While data quality of the first use-case needed to be improved substantially, use-case two already provided very well maintained data. Despite conducting data cleansing, enterprise one has only been able to cover 32% of available access right assignments using dynamic policies, while enterprise two achieved a value of 45.9%. This underlines the initial thesis of the publication, which states that by integrating dynamic IAM aspects like adequate organizational and quality management, the overall IAM performance can be improved substantially. A more homogenized and effective way of operation consequently provides value to enterprises in reducing administrational effort and reducing complexity. Summed up, three major research findings regarding dynamic IAM have been identified:

- Dynamic IAM access policies may be derived from identity data.
- The implementation of according data quality processes represents a fundamental requirement for shifting to dynamic IAM.
- Introducing dynamic IAM serves as incubator to improve different aspects like IT-security and administrational effort.

4.5 Operational IAM - Sustainability (Research Paper 4)

In order to foster the establishment of dynamic IAM, research articles 2 and 3 are devoted to provide a dynamic policy-based foundation together with strategical and operational adaptions to extend and improve IAM capabilities. However IAM needs to be continuously aligned to rapidly developing requirements of information technology, organizations and users. Thus enterprises constantly need to be in control of their implemented IAM performance. Operational management of the identity lifecycle and access management tackles various aspects of organizations, which in turn leads to a great number of technical and social interactions. This makes it hard to comprehensively monitor IAM efficiency. Yet in order to fully leverage this endeavour, constant IAM evaluation and easy understandable information for decision makers are necessary. While technically enforced policies strongly lower manual expenditures, the large amount of daily executed IAM processes still represents a huge success factor, as decision makers are often not in hold of all necessary information which are required for thorough decision-making. Additionally, wrong decisions are commonly not detected contemporary, which may impose various security risks. Consequently, the research article 4 is focused on IAM measurement and the derivation of compliant operational IAM behaviour.

In order to improve IAM decision-making, the concept of explicit and implicit policy derivation has been introduced. Explicit policies are precisely and clearly expressed or readily observable and are technically enforced by the underlying IAM system (e.g. by a script, code, etc.). Implicit policies on the other hand are commonly implied but not directly expressed and are generally realized by a set of stringent decisions. An overview of observed policies is depicted in Figure 9. Take the assignment of access privileges as an example. While potentially every employee to access privilege assignment is possible, only certain combinations are put in place by IAM administrators, e.g. only financial work staff is assigned to accounting entitlements. Implicit policies provide these additional insights for decision makers (e.g. by hinting discrepancies during approval workflows). Thereby they act as contemporary measure to explicit policies, by further narrowing IAM execution without hindering actual execution (as e.g. in some cases the assignment of uncommon access privileges might be necessary due to business requirements). Thus by using implicit policies, operational IAM is extended in shifting from considering only atomic and static operations to taking the comprehensive status quo into consideration, thereby enabling a more sustainable mode of operation.

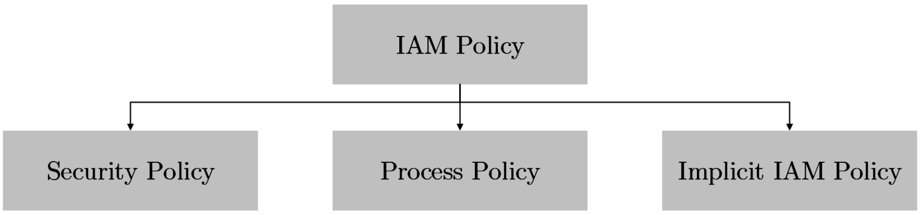


Figure 9: Overview of IAM policies [HKN^{+}16]

Implicit policies need to be of an automated and dynamic nature in order to provide up to date and relevant recommendations, as erroneous information might impose a negative impact on the overall business (e.g. by declining a request to assign an access privilege due to false recommendations). In order to overcome this shortcoming, IAM key performance indicators (KPIs) are proposed. These enable the identification of human-understandable and processable data out of static identity information and contextual data which can be included within common operational IAM workflows. Keeping to the example above, each possible permission assignment can be flagged with a criticality index, which may be automatically computed based on different algorithms. Consequently, implicit policies are burnished with a set of indicators, which act as threshold concerning their validity and applicability. As soon as the threshold exceeds a certain value (e.g. when the area of responsibility of a department is extended which leads to the need of additional permissions), a policy may automatically be outdated. Thereby manual administrational effort is lowered for operative policy management. In total, the given research article addresses three major research findings:

- The introduction of implicit IAM policies to support operational IAM decision-making as it enables to utilization of contextual data.
- IAM KPIs are used to evaluate the mode of operation and implemented polices within IAM.
- During the real-world evaluation, KPIs were used to identify the criticality level of assigned access privileges which serves as foundation for policy evaluations.

4.6 Sustainable IAM (Research Paper 5)

Research paper 5 further leverages the concept of IAM indicators together with providing recommendations for strategical IAM. In general, the definition of an IAM strategy and an according roadmap relies on qualitative information like organizational goals defined by management stakeholders together with past experiences and best practices. These strategical decisions strongly influence the way IAM is carried out in a certain enterprise. While environments (business area, work staff, IT-environment) constantly change, these decisions are hardly re-evaluated after having been put in place in an appropriate manner. This article is devoted to closing this gap by introducing a tool-based analysis methodology for strategic IAM decisions, which utilizes historical and predictive meta data that are commonly produced as a result of strategical IAM decisions. This methodology consists of four steps, namely data collection, parameter extraction, indicator calculation and indicator simulation.

As data foundation, a generic approach to data collection (which may be defined as IAM measurement values based on [YS15]) has been proposed. Firstly, process data are diversified regarding their five different perspectives, namely the functional, behavioural, organizational, informational and operational perspectives [RM05]. Secondly, logical IAM data (data regarding managed entities like employees, user accounts or access privileges) are segmented based on different characteristics. By combining these logical and process parameters, enterprises are enabled to define custom-tailored IAM indicators to measure the impact of decision-making. For these indicators correlation algorithms are exerted for recognizing potentially hidden relationships. Taking the strategical reorganization of an enterprise as example, this might affect logical data (e.g. through frequent changes of department) together with process data (e.g. in increased execution times due to increased workloads). Consequently this enables the identification of potentially required strategical or operational adjustments (e.g. to assess if the required workload is addressable by the current working staff).

In order to provide IAM administrators additional insight for sustainable IAM decision-making, simulation techniques are utilized. As the development of IAM indicators may be depicted as a graph (as exemplary depicted in Figure 10), simulation techniques are applicable to predict the development trend of certain indicators. By combining simulation results with certain thresholds, enterprises are entitled to determine actions before e.g. working staff is no more capable of handling all necessary activities, which might result in further negative outcomes like decrease of data quality.

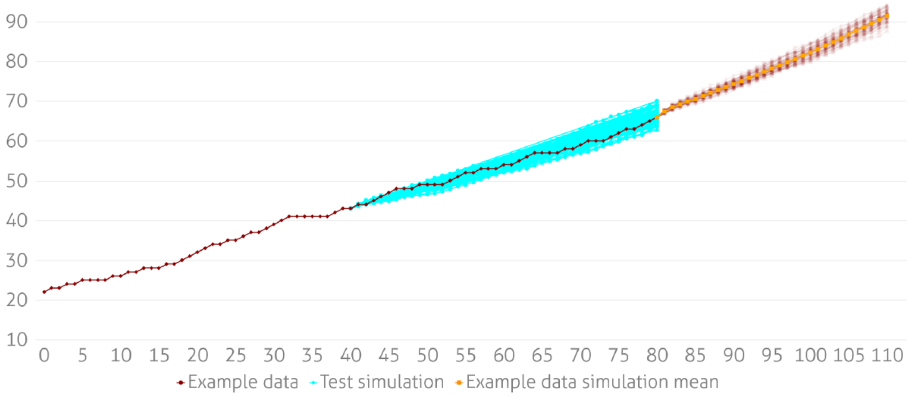


Figure 10: IAM indicator simulation based on research article 5

Again, the presented approach has been evaluated within a real-world scenario. Together with a internationally operating company, two use-cases have been deployed. Firstly, the effect of the enterprises course of growth has been analyzed with about 225 employee joins per month. Despite the analysis showed that this trend is manageable by current staff, missing correlation of the two different IAM indicators 'employee joins per day' and 'number of access privileges assigned to joined employees' revealed substantial shortcomings within the implementation of joiner processes. Secondly, the evaluation of risk management indicators revealed a strong correlation of the growth of critical access privilege assignments and employee mover processes. This insight induced the discussion to implement strategical means of improving IT-security by decreasing the number of critical assignments, e.g. by establishing periodic access re-certifications. Within this research paper on sustainable strategical IAM decision-making, two research insights have been identified:

- IAM strategy management is not addressed in a way which is required to fulfill complex and varying IAM requirements, thereby possibly imposing additional workload and decreased IT-security implications.

- Monitoring and visualizing IAM indicators provides well-usable insight for improving IAM decision-making.

5 Conclusion

During the last years, information technology has evolved in a rapid manner and eventually became an essential part of everyday life. This trend is commonly described as ubiquitous computing, which induces constant interaction with digital technologies during any time, location and in any arbitrary format. In this regard, consumers are affected in the same way as enterprises. Within the near future, the number of managed digital identities is expected to fundamentally increase, taking for example customers or IoT devices into consideration. Additionally, the degree of integration is further expanded to address topics like lean value chains, standardized external cloud services and multi-tenancy software systems. This induces extending existing IAM system to external users devices and platforms and further amplifies its central role within enterprises. By design, ubiquitous computing involves users accessing IT-resources within different contexts, detached from boundaries of enterprises. As a consequence, identity-related services need be further developed, e.g. to integrate identities from potentially different sources, such as social or government IDs. Therefore specialized identity lifecycle management along with automated and secure access management are required. Taking the current state within practice into consideration, IAM needs to evolve in order be capable of meeting these demands. Thus the motivation for this dissertation sourced in the alignment of IAM to as well current requirements as observable future developments.

Summed up, the dissertation at hand is focused on the initially in Section 2 described research question *"How can operational and strategical IAM be integrated to achieve sustainable IAM and how can this development be supported by dynamic IAM?"*. Thus the given research question is centred around two contemporary IAM concepts. Firstly, research of dynamic IAM aims to introduce and extend existing dynamic concepts, which enable a long-term decrease of administrational effort to implement a more homogenized and automated mode of IAM operation. Secondly, sustainable IAM is devoted to adjusting IAM by meeting today's demands without hindering future needs. The focus in this regards is to provide founded measurement of IAM implementations and generate according recommendations to establish potentially necessary adaptions. These concepts were implemented based on IAM indicators and the extended application of policies.

The overall research question was derived into five sub-questions. Each aspect is formulated as a self-sufficient research question and was addressed within one scientific publication. Within the treatment of research question one, the status-quo of enterprise IAM was derived by conducting a structured survey among IAM experts of different industry sectors. Answering research question 1 showed that this topic is still considered as an investment and support factor to a large extent, while potentials are not fully leveraged. The results of the survey showed that the topic cost-reduction is hardly fostered at the moment, what underlines this theory. This insight underlines the need for future research of enterprise IAM, as companies are in need of additional guidelines and methodologies to fully benefit from IAM.

Research questions 2 and 3 are devoted to establish dynamic IAM in order to decrease administrational effort and increase the level of IT-security. Research paper 2 is devoted to operational IAM policy management. To approach this topic in a structured way, a dynamic policy management process was proposed. This utilizes currently unused contextual data for the definition of policies to improve automation and support daily IAM operation. During a real-world evaluation, different policies where derived in order to help the corresponding enterprise in improving different aspects of IAM operations. An example for this effort is the correlation of permission activations together with contextual data. This allowed the company to derive profiles of access privilege activations which occur during standard operation together with the definition of outliers which need further investigation. The concept of dynamic IAM was pursued within research article 3, which introduced a process model to establish dynamic IAM within enterprises. This model is based on the two pillars policy management and attribute management. In other words, the overall data and process quality needs to be increased to provide a profound foundation for operational IAM services as well as for strategic IAM maintenance (e.g. by adjusting available process models). Within our evaluation with two companies, we showed that by utilizing the given concept, the amount of manually managed access privilege assignments may be decreased substantially and as a consequence the overall management effort may be decreased.

Fostering these results, research questions 4 and 5 are devoted to the concept of sustainable IAM. Research article 4 introduced the concept of IAM KPIs to enable founded measurement of current operational IAM implementations. As a foundation, contextual data identified within publication 2 were utilized. Based on these results, implicit policies can be implemented in order to identify standard and deviant activities and support domain experts to continuously meet sustainable decisions. These are used to identify standard and deviant activities and therefore support domain experts to continuously meet sustainable decisions. During the real-world evaluation, we were able to determine the potential criticality of access privilege assignments, which can be used to support future approval workflows. In utilizing this concept, enterprises are enabled to provide additional information to deciders for the execution of IAM tasks, thereby enabling a more coherent mode of operation. Finally, publication 5 is focused on further utilizing IAM indicators. The article introduces a more generic use of IAM indicators to support strategical as well as operational IAM. In order to extend this measurement concept, simulation techniques are utilized to estimate how the current mode of operation is going to affect future IAM. Take the constant hiring of employees as an example, which represents a fundamental part of well-operating enterprises. The correlation of indicators enables the assessment how such a trend may influence IAM and if actions have to be taken to handle the expected workload. Consequently, enterprises are enabled to define custom-tailored indicators and achieve founded estimations about their long-term development, thereby providing IAM architects a solid basis for decisions.

During the conduction of research, different research challenges were identified together with possible shortcomings of the presented results. Firstly, the given work focuses on the demands of enterprise IAM together with commonly implemented and available infrastructures. Consequently, individual challenges of IAM (e.g. multi-tenancy cloud-based IAM) are not addressed specifically, concerning how the given concepts can be integrated. Yet the presented theories are designed as general approaches to meet the demands of enterprise IAM and thus may be adapted to individual topics in future. Secondly, the comprehensive use of IAM recommendations within daily operations still requires additional research. This resulted due to the fact that the respective evaluations have been conducted in cooperation with IAM administrators who inherently incorporate broad knowledge of IAM. This leaves different challenges concerning how to provide these information to users without profound IAM expertise, e.g. by facilitating suitable visualization techniques. Finally, the definition and implementation of IAM indicators requires effort of enterprises and IAM experts in order to achieve valuable results. To reduce this complexity, standardized metrics and forms of user interactions are necessary. These need to be integrated in terms of technical and social layers of IAM, which represents an interesting future research challenge. To support this process, future enterprise IAM functionality needs to be extended regarding its monitoring, measurement and visualization capabilities. Doing so, it is going to allow for more agility in order to tackle new technical as well as organizational challenges and improve its overall business-value.

Part II

Research Papers

1 Measuring Identity and Access Management Performance - An Expert Survey on Possible Performance Indicators

Current status:	Published
Conference:	International Conference on Information Systems Security and Privacy (ICISSP)
Date of acceptance:	18 October 2017
Full citation:	Matthias Hummer, Sebastian Groll, Michael Kunz, Ludwig Fuchs, Günther Pernul. Measuring Identity and Access Management Performance - An Expert Survey on Possible Performance Indicators *Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP), 2018, Pages 233-240, SciTePress.*

Conference Description: The International Conference on Information Systems Security and Privacy aims at creating a meeting point for researchers and practitioners that address security and privacy challenges that concern information systems, especially in organizations, including not only technological issues but also social issues.

Measuring Identity and Access Management Performance - An Expert Survey on Possible Performance Indicators

Matthias Hummer[1,2], Sebastian Groll[2],Michael Kunz[1,2], Ludwig Fuchs[2] and Günther Pernul[1]
[1]*Department of Information Systems, University of Regensburg, Regensburg, Germany*
[2]*Nexis GmbH, Regensburg, Germany*
{f_author, s_author}@ur.de, {f_author, s_author}@nexis-secure.com

Keywords: Identity and Access Management, Performance Indicators, Survey

Abstract: Currently existing digital challenges such as securing access, proof of compliance with regulations and improvement of business performance are urging companies to implement structured Identity and Access Management (IAM). Over the past decades, the introduction of IAM represented a critical task for companies trying to get their complex IT infrastructures comprising hundreds of systems, thousands of accounts and millions of access right assignments under control. However, once introduced, the identification of potential IAM malfunctions remains an unsolved challenge. Within this paper, we want to provide a first step into the direction of sustainable IAM maintenance, by introducing indicators that are able to capture the efficiency of a rolled-out IAM. We firstly derive IAM performance indicators via a structured scientific approach and later evaluate their relevance by surveying IAM experts.

1 INTRODUCTION

Identity and Access Management (IAM) has become one of the core topics to tackle insider misuse of access, complying with regulations and achieving transparent management of digital identities and entitlements in enterprises. Complexity and the so called identity explosion (Fuchs and Pernul, 2008) forces companies to tackle the problem of users' correct access to systems, a crucial task in terms of security and efficiency (Hovav and Berger, 2009). While traditionally system-specific administration of accounts and permissions was conducted, nowadays companies centralize their user management to provision, control and analyze their digital identities throughout all connected systems. However, newly arising technologies like highly volatile cloud infrastructures or industry 4.0 require even more sophisticated IAM solutions and demand a steady increase in performance of an organization's IAM. Up to now, the measurement of IAM performance is an issue only attracting little attention and remains an unsolved problem. While there are many approaches (Windley, 2005; Royer, 2007; Fuchs et al., 2009; Kunz et al., 2015) that offer guidelines on how to adopt IAM in good practice, only little notion has been dedicated to judge whether an existing IAM is able to cope with current requirements from business, technology or regulation. Up to now research offers partial approaches to estimate the quality of certain IAM capabilities, however, there is only little support in rating the overall performance of a specific instance of IAM. Consequently, companies struggle with knowing the general maturity of their IAM leading to possibly flawed decisions on future IAM investments or risking an insecure IAM infrastructure not capable of meeting today's increasing demands. In order to address these issues, our paper's main contribution is to suggest but more importantly to verify performance indicators for IAM.

The remainder of the paper is structured as follows. Section 2 presents related work concerning which approaches exist for measuring the performance of specific IAM capabilities. Within Section 3, we outline our overall methodology, before Section 4 introduces the construction of 19 performance indicators by applying the Goal-Question-Metric (GQM) paradigm (Basili et al., 1994). In order to evaluate these for relevance, we conducted a survey with 32 participants specialized within the field of IAM, leading to a generalizable expert opinion on our indicator candidates, which is described in Section 5. These results together with other interesting findings of the expert's answers are discussed in Section 6. Finally Section 7 shows limitations of our approach and concludes with future work.

2 RELATED WORK

Existing IAM research mainly focuses on specific technical or organizational features of IAM infrastructures and does not cover performance indicators of IAM in general. Literature in research and practice (Witty, 2003; Bresz et al., 2007; Hermans, 2008; Dell, 2011; Harvard, 2014; Fisher, 2016) underlines that risk reduction, IT cost reduction, compliance requirements, data and process quality and business facilitation are the main drivers for modern IAM in organizations. These drivers can act as starting point for the development of performance indicators for long-term IAM maintenance. For instance, Royer et al. mention the importance of assessing and evaluating IAM systems within several publications (Royer, 2007; Royer and Meints, 2008; Royer, 2013). They transfer the concept of balanced score cards to IAM thus presenting a generic methodology for estimating an IAM system's performance. Following similar goals as ours, they mostly focus on financial aspects to evaluate the value of IAM systems.We argue that the overall performance of IAM as a cross-cutting enterprise functionality must be taken into consideration.

(Höllrigl et al., 2008) define several evaluation dimensions to compare architectures for access control in federated environments. In (Schell et al., 2009) they use these dimensions as a basis to derive metrics for an IAM system's performance evaluation. However, they mostly focus on architecture and consider performance as a quantifiable measure defined by how long various systems' activities are taking. Performance in our terms is having a broader perspective than their focus on an IAM systems' capability of timely processing decisions. Staite et al. (Staite and Bahsoon, 2012) perform a systematic literature review as well as an architectural trade-of, analysis method (Kazman et al., 1999) to derive requirements and metrics for authentication and user profiles in Identity Management architectures. These metrics, however, focus on the technical implementation of an IAM architecture. Peterson et al. (Peterson, 2006) provide indicators to measure and manage the risk within IAM systems. They show some valuable metrics that can assist in judging whether the execution time of requests and the delivery of access rights are in acceptable condition. Their approach is focusing only on the fields of risk reduction and process quality improvement and leaves out other necessary categories.

An overall perspective and judging from a top-level goal that IAM centers around has not yet been addressed. Furthermore, the approaches do not verify their indicators via conducting a survey, thus not validating their suggestions in practice.

3 DEVELOPMENT OF IAM PERFORMANCE INDICATORS

3.1 Overall Methodology

Having outlined the research field that this publication contributes to, in the following Section our methodology for conducting this research is briefly described. To the best of our knowledge, there exists no comparable approach for developing general performance indicators for IAM in such a focused and structured way. The overall process follows the five steps depicted in Figure 1 and ist based on the GQM paradigm which is widely respected for its capability to develop qualitatively or quantitatively measurable factors derived from overall goals. Initially developed for the field of Software Engineering, its basic assumption is that measurement must fulfill three goals (as described below). Transferred to IAM, measurement must be:

- Centered around an overall strategy (i.e. various goals)
- Holistic (i.e. considering all involved organizational and technical entities such as both, processes and resources)
- Interpretable within the IAM context

Generally speaking, GQM is tackling the problem of metric development via a divide and conquer process. According to (Assmann et al., 2002) it comprises three layers, namely a goal layer, a question layer and a metric layer. Each layer deals with a specific question as (Basili et al., 1994) indicates:

1. **Goal**: which goals are to be achieved by the measurements?
2. **Question**: which questions can define these goals more precisely?
3. **Metric**: which metrics can answer these questions?

For executing the GQM paradigm within IAM, we followed the presented methodology which suggests a generic six-step approach that can be followed by answering all three questions above. For a detailed description of this process please refer to the initial publication. Note, that in this publication, we treat the terms metric and IAM indicator synonymously, as the metrics that are identified via the GQM can be considered as performance indicating measurements.

3.2 IAM Goals

Following the GQM paradigm, in a first step goals for IAM have to be formulated as mission statements.

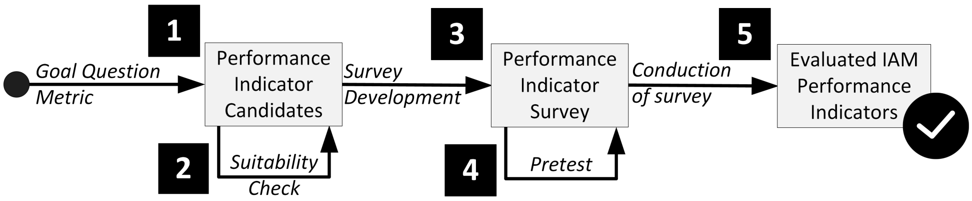

1
Goal Question Metric
Performance Indicator Candidates
2
Suitability Check
Survey Development
3
Performance Indicator Survey
4
Pretest
Conduction of survey
5
Evaluated IAM Performance Indicators

Figure 1: Methodology for Suggesting IAM Indicators

Goals are structured hierarchically, which raises the requirement that the GQM goals for IAM have to be on the same hierarchical level in terms of granularity. The overall goal of IAM, as (Bresz et al., 2007) summarizes, is to 'initiate, capture, record and manage the user identities and related access permissions to the organization's proprietary information'. Building on existing research and practice (cf. Section 2) this generic goal can be decomposed into the following sub goals which are further used in the remainder.

Risk reduction: IAM focuses on preventing resources from unauthorized access. To reduce risks arising from an unstructured management of identity data, IAM provides several means, e.g. giving an overview over the data and allowing inspection and correction of wrong access privileges (Meier et al., 2013).

IT cost reduction: Literature (Bresz et al., 2007; Witty, 2003) and practice (Gartner, 2009; Software Engineering of America, 2015) indicate, that IT-Helpdesk costs are mostly connected with password problems of end users. IAM proposes new technical measures to tackle these problems. On the one hand technologies such as SSO are enterprise-wide and application-wide deployable, and on the other hand user-friendly portals for self-service can be delivered.

Improvement of process and data quality: With an established IAM, companies are better supported in maintaining and improving their data quality. Without a centralized IAM system multiple error sources occur while connecting and integrating data from company-wide systems (Windley, 2005; Bertino and Takahashi, 2011).

Regulatory compliance: With the ongoing trend of digitalization, national and international regulatories are imposing the need of auditing and managing access within a companies' applications upon enterprises. While there are general regulations such as the Sarbanes Oxley Act (United States Congress, 2002) or the soon effective EU General Data Protection Regulation (Council of the European Union, 2016), more and more industry-specific regulations such as the HIPAA (United States Congress, 1996) or Basel III (Basel Committee on Banking Supervision, 2011) are challenging organizations into presenting audited and well-proven access infrastructures.

Business facilitation: Lastly, another important sub-goal of IAM is, similar to all IT-related activities, allowing a smoother and non-disrupting business experience. While in traditional scenarios, users have to order access rights in various forms, centralized IAM provides a standardized and understandable request process for identities, user accounts or access rights (Windley, 2005).

3.3 Abstraction Sheets, Deriving of Questions and Development of Performance Indicators

Following the applied methodology for the conduction of the GQM, in a next step, we created so called abstraction sheets (Assmann et al., 2002; Basili et al., 1994) for each of the stated goals. These serve as a decomposition of the proposed sub-goals into several parts which can be transferred later into items that can be questioned within our survey. Abstraction sheets are composed of two main elements. Firstly, intention, quality aspect, subject und perspective are summarizing in short what the main components of the goal are (e.g. *compliance with regulations via IAM as observed by managers*). Secondly, for this first aspects quality issues (e.g. *number of violations of compliance rules*) and environmental factors (e.g. *automated reporting*) are raised. These quality issues are further mapped onto our IAM performance indicators.

While this list of questions is not designed to be exhaustive, we argue that these are the main compliance issues that can be tackled via structured IAM and suffice for describing the overall objective within IAM. These questions were developed with respect to

(Assmann et al., 2002)'s eight points for meaningful survey questions.

Following these principles for each sub-goal, we arrived at a set of questions, each indicating a possible performance indicator that in return can be assigned to a goal. Note, that the indicators might correlate with various goals, however we assign them to the goal that initially raised the respective question.

4 STUDY OF PERFORMANCE INDICATORS IN IAM

4.1 Development of the survey

Following the presented methodology we conducted a survey among IAM experts in order to validate and evaluate the presented IAM indicators. For this purpose we created an online questionnaire which is structured as follows:

Firstly we inquired demographic features (e.g. project status or company size). Secondly the participant was asked for IAM goals relevant to his company. As a result only questions concerning the selected goals were presented, whereas each question references an IAM performance indicator. Thirdly for each indicator we raised two questions. The first collects if the company did already achieve an improvement through IAM regarding the indicator. If no improvement was achieved up to that point, the second question relates to if there is an improvement planned. Before conducting the actual survey we started a pretest to validate our questions regarding suitability, interpretability, problems during procession, question order, possible technical problems and temporal requirements.

For this initial evaluation we inquired three IAM experts and lead a short interview afterwards. While no major issues in length, order, structure and time were criticized, phrasing of the questions had to be improved for interpretability. Furthermore, another major adaption to the sample population was needed: As one of the three pre-testers was an IAM consultant, he stressed that answering the questions was hard as he had various projects in mind and could not guarantee replying consistently without getting confused due to the number of his different clients. In order to avoid data distortion we reason that IAM consultants involved in several projects should be suspended from the sample and respected this in the conduction of the survey as the description of our sample shows (cf. Section 4.2.1). Having developed the questionnaire, an evaluation method for assessing the validity of an indicator for IAM is needed before conducting the study, in order to consistently judge the suitability of IAM indicators. Figure 2 shows the process we applied to each of the candidates in order to evaluate its applicability in practice: Summing up the process, we have two main criteria for the validation of an IAM indicator:

1. Relevancy of an IAM goal
2. Statement of participants of improvement of IAM through the indicator

The relevancy of the IAM goal (meaning that the IAM goal was answered as relevant to the participant's company) is a filter criterion that justifies whether the sample is large enough to have significance. As pointed out, only indicator questions concerning selected IAM goals were raised. The second criteria is split into two sub-criteria.

Firstly, if at least 50% of the respondents (16 out of 32) consider an indicator as already IAM-improving, we argue that it is relevant to companies. Secondly, if the aggregated amount of the participants' IAM improvements of the past combined with planned enhancements through the indicator shows more than 75% (24 out of 32) response rate, we argue, that the combined percentage is high enough to indicate relevance of the investigated indicator.

Additionally, we introduce a category of results where an indicator is likely to exist, but where the quantity of answer is not significantly high enough (aggregated amount of already improved or prospective planned answers between 50% and 75%). We reason, that if a company has already achieved betterment of their IAM through an indicator, this should be weighted stronger than if they only expect future results.

4.2 Results

In alignment with the evaluation process introduced in Section 4.1 (cf. Figure 2), four of five IAM goals received a sufficient score of acknowledgement (cf. Figure 2). IT cost reduction was answered as relevant only by 11 participants (~34%) which means in our terms that we cannot make valid statements on this topic, thereby excluding the corresponding indicator candidates. However, we will include the indicators discarded in the discussion in order to reason upon possible causes for their little impact.

4.2.1 Participants and Demography

We only invited potential participants with dedicated background in IAM. As already mentioned in Section 3.1, they have to judge a single company's IAM, therefore preventing the confusion of multiple

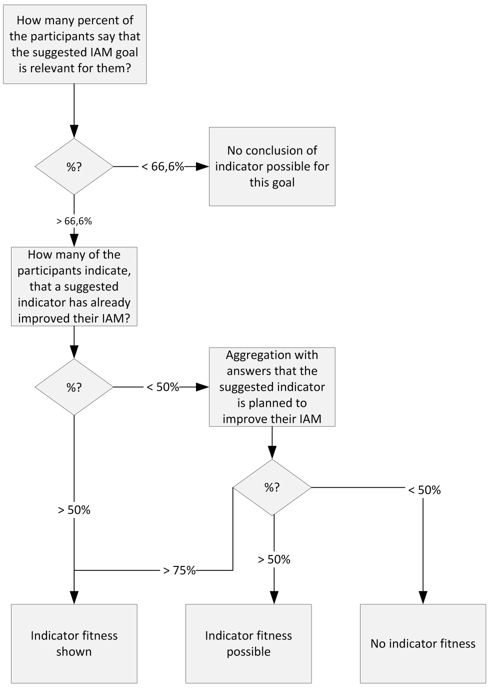

How many percent of the participants say that the suggested IAM goal is relevant for them?
%?
< 66,6%
No conclusion of indicator possible for this goal
> 66,6%
How many of the participants indicate, that a suggested indicator has already improved their IAM?
%?
< 50%
Aggregation with answers that the suggested indicator is planned to improve their IAM
%?
< 50%
> 50%
> 75%
> 50%
Indicator fitness shown
Indicator fitness possible
No indicator fitness

Figure 2: Evaluation Process

projects. Second, they have to be employed in operational or strategic IAM. Based on these requirements we sent out 73 invitations. 32 out of these 73 people fully conducted the survey (~41%). All participants are located within the DACH region (Germany, Austria and Switzerland). The average processing time of the questionnaire survey was 11 minutes and 30 seconds, the median was 8 minutes and 18 seconds which matched our expectations.

For our demographic evaluation we analyzed our respondents in regard of their company's size, industry branch, IAM project progress and their job position within IAM. It showed that the majority work for companies with more than 5.000 employees, which matches the fact that IAM is currently a topic mostly relevant for larger enterprises. The distribution among industrial sectors is displayed in Table 1. Eight of our participants did not make a statement regarding industry, thus their answers can not be applied in the discussion. However, the given sectors can be divided into strongly regulated sectors (rows number 1 and 2) and average regulated sectors (rows number 3 - 8). This fragmentation will later find application within our discussion.

Within the questionnaire we asked the participants how far IAM within their company is processed using a scale from 0% to 100% with steps of 10%. Thus our interviewees can be divided in 'early project state' (from 0% to 30%), 'medium project state' (from 40% to 70%) and 'advanced project state' (from 80% to 100%). We determined that eight participants are located in an early state, while 12 respectively are either at a medium state or have already advanced their IAM. We further analyzed if there is a connection between the project state and other demographic features like company size or industry sector regulation. We could not identify any significant connection concerning these characteristics. This leads to the assumption that there seems to be a strong individual dependency concerning how IAM is focused by enterprises.

#	Industry Sector	Participants
1	Finance / Insurance	11
2	Pharmacy / Medicine	1
3	Automotive / Supplier	5
4	Metal industry	3
5	Service	2
6	Food	1
7	Software / Hardware	1
8	Others	8

Table 1: Participant's Industry Sectors

4.2.2 IAM Indicator Evaluation

Generally speaking, the goals risk reduction, improvement of data and process quality and compliance requirements achieved highest scores of relevancy in the answers of the participants with 25, 26 and 26 positive answers respectively (out of 32). Naturally, these goals are correlating with the basic functionalities of a company-wide IAM which typically are implemented first. The protection of internal assets (risk reduction) represents a major goal of companies utilizing IAM to strengthen their IT security thus avoiding possible malicious activities by allowing unwanted access. Similarly, data and process quality improvements lead to smoother and less erroneous user management workflows therefore presenting a highly valuable goal for companies burdened by challenges in correlating data from various systems. Compliance achievements are an unwanted but necessary issue in order to allow a company to meet their industry's specific requirements.

The goal of business facilitation achieved a high score as well, indicating that issues such as user satisfaction and less disruption of business are a much preferred topic that IAM can deliver as well. Such topics are typically approached once basic IAM functionality is up and running. On the other hand, IT cost reduction caused by IAM is a goal which is less in favor as

IAM Goal	IAM Indicator	Improved	Planned
Risk reduction		*25*	
	Number of security incidents due to user and entitlement management	14	6
	Number of security incidents due to critical role and access right combinations	11	5
	Duration until deactivation of employee access rights	24	1
	Duration until emergency deactivation of employee access rights	18	1
IT cost reduction		*11*	
	Costs for entitlement and access management	4	0
	Support costs for user management	10	1
	Costs for data storage of user data	7	3
Data and process quality		*26*	
	Development of data quality	25	0
	Error rate within access management	18	3
	Error rate within identity and account creation	21	1
Compliance requirements		*26*	
	Reduction of compliance violations	18	6
	Number of successful audits	21	4
	Duration until complete solution of a compliance incidents	11	8
Business facilitation		*22*	
	Reduction of administrational effort	19	3
	Improvement of user satisfaction	13	5
	Duration until employee readiness	19	2
	Duration until access model adjustment	13	4
	Number of failed authentication requests	8	2
	Processing duration of user requests	13	2

Table 2: IAM Goals and IAM Indicator Scores

our received responses demonstrate. This could reason in the fact, that initially, IAM increases IT costs substantially (Cser, 2017), whereas a later general decrease through e.g. automation is not yet perceived by our survey participants.

Table 2 displays our proposed IAM goals together with their indicators. The ’improved’ column specifies how many participants already achieved an improvement concerning the indicator while the ’planned’ column expresses how many participants plan an improvement of the corresponding indicator in the future. For each IAM goal we present the number of participants which described the goal as relevant for their company. In accordance with the evaluation process (cf. Figure 2), we define the suggested indicator candidates for the IAM goal IT cost reduction as not suitable as it is only relevant to ~34% of our participants. Thus, 16 possible indicators remain out of the four suitable IAM goals for further check of relevancy. Please note that this does not necessarily mean that the indicators raised for IT cost reduction do not exist. However on the sample data, we cannot make a significant statement about the validity of the candidates. The criteria presented in Figure 2 was reached by nine IAM indicators.

The next category consists of indicators which show potential fitness. These do not fully match our requirements yet do have a score high enough to possibly represent indicator fitness within certain scenarios or environments. In the field of risk reduction ’duration until emergency deactivation of employee’ and ’security incidents due to critical access right combinations’ fulfill the categories’ demands. Correlations with the industrial background of participants’ replies show, that these indicators might mainly be relevant within the focus of finance and insurance companies thus not being highly valuable for all of our participants. Furthermore ’duration until complete solution of compliance incidents’ and ’user satisfaction’ fall into this category. This suggests that only some companies focus on topics which do not directly effect the core IAM or correspond to topics which only occur with a certain probability.

Finally, five candidates did not show indicator fitness. Apart from the ones within the goal cost reduction, the two other indicators correspond to the goal of business facilitation. The first indicator is ’number of failed authentication attempts’ with only ~31% (score of 10). The last discarded candidate is the ’duration until user requests are processed’. This also represents an interesting finding as from our perspective this candidate is connected with the indicator 'user satisfaction'. This could ground in enterprises focusing rather on other issues for increasing user satisfaction such as integrated portal usage.

5 DISCUSSION

In the following we present a discussion of the derived results. Five of our presented IAM indicators did not show indicator fitness. Firstly, IT cost reduction does not seem to be within short and mid term focus of IAM in general which results in three discontinued indicators. Secondly, the number of failed authentication requests did not show any indicator fitness. This can be explained as this topic can be handled very well on a technical level by IAM systems thus it does not receive a lot of attention among IAM experts. Thirdly the processing time for user requests has hardly been improved or planned to be improved in future. This is remarkable as it represents a main point of contact of users to IAM and thereby we expected a positive response to this indicator. We tried to find a correlation concerning other demographic features yet none produced significant results. Thus we conclude that this topic is already handled very well by today's companies as the negative impact might strongly impact the overall performance of the company itself.

Influence of project status on IAM indicators: The presented indicators show different performance regarding the IAM project state. As expected not all indicators can be developed on an equal speed. For example the number of successful audits strongly increases at beginning IAM projects. So ~75% of our participants within an early project state, ~75% within a medium project state and ~90% within an advanced project state have increased this indicator. However indicators like user satisfaction increase at a later project state. Only 25% of our participants within an early project state could improve this issue while ~45% within a medium project state achieved an improvement and ~63% within an advanced project state increased this indicator. This firstly shows that companies in general begin with issues concerning core IAM indicators during their project and secondly that the presented indicators can further be split up according the project state in order to optimally support organizations.

Influence of industry sector on IAM indicators: Regarding the presented industry sector partition (strongly regulated and average regulated) we could determine further differences. Average regulated companies perceive IAM as a support function and thereby focus on indicators which facilitate effort and business. For example 75% of our participants in average regulated industry sectors achieved an improvement in the duration it takes to adjust the access model in place while this was only achieved by ~33% of participants within strongly regulated industry sectors.
On the other hand 75% participants in strongly regulated industry could reduce the number of security incidents due to role and access right combinations while this was achieved by only ~30% of participants with average regulations. According to these observations the presented indicator catalog could further be elaborated regarding industry sectors and currently available legal requirements.

6 CONCLUSION AND FUTURE WORK

Having discussed and presented our findings, we want to briefly outline limitations and summarize our contribution before providing a short outlook for future work. In general we perceive three limitations of the conducted survey. Firstly, a response rate higher than the 41% of our invited candidates would have backed our results even more. However, for qualitative research such as our study, we argue that our received responses are high enough. Additionally, a potential selection bias (e.g. over-representing project managers) might exist, but cannot be suppressed due to the fact that to the best of our knowledge no research exists on the general distribution of job profiles for IAM. Lastly, all of our participants are all located in the German-speaking countries, however, due to the international background of most of our participants' companies operating worldwide, our findings can be transferred to other nations with similar preconditions as well.
In a nutshell, within this paper we were able to demonstrate relevancy and existence of several IAM indicators that help analyzing an existing IAM. By evaluating their relevance in practice, we were able to provide researchers and practitioners with valuable results towards how IAM performance can be expressed either in a quantifiable or qualitative manner. Utilizing our results, the first step towards a holistic IAM measurement framework has been taken. Consequently we are planning on establishing such a framework with our suggested and approved indicators as baseline. By doing so, we aim at delivering a tool for sustainable IAM measurement and maintenance.

REFERENCES

Assmann, D., Kalmar, R., and Punter, T. (2002). *Messen und Bewerten Von Webapplikationen Mit der Goal/Question/Metric Methode: Handbuch*. IESE-Report / Fraunhofer Einrichtung Experimentelles Software Engineering. Fraunhofer-IESE.

Basel Committee on Banking Supervision (2011). Basel III - A global regulatory framework for more resilient banks and banking systems.

Basili, V. R., Caldiera, G., and Rombach, H. D. (1994). Experience factory. *Encyclopedia of software engineering*, pages 470–476.

Bertino, E. and Takahashi, K. (2011). *Identity Management: Concepts, Technologies, and Systems*. Artech House.

Bresz, F., Renshaw, T., Rozek, J., and White, T. (2007). Identity and Access Management. Technical report, Ernst and Young.

Council of the European Union (2016). EU General Data Protection Regulation.

Cser, A. (2017). Use Commercial IAM Solutions To Achieve More Than 100 Percent ROI Over Manual Processes. Technical report, Forrester.

Dell (2011). Identity and Access Management. Technical report, Dell Inc.

Fisher, P. (2016). Identity and Access Management in the Digital Age. Technical report, CXP Group Company.

Fuchs, L. and Pernul, G. (2008). HyDRo–Hybrid Development of Roles. *Information Systems Security*, pages 287–302.

Fuchs, L., Pernul, G., and Broser, C. (2009). Different Approaches to in-house Identity Management. In *Proc of the 4th International Conference on Availability, Reliability and Security (ARES 2009)*. IEEE Computer Society, Fukuoka, Japan.

Gartner (2009). MarketScope for Enterprise Single Sign-On. Technical report, Gartner.

Harvard, U. (2014). Identity and Access Management - Program Plan.

Hermans, J. (2008). European Identity & Access Management Survey. Technical report, KPMG.

Höllrigl, T., Schell, F., Suelmann, S., and Hartenstein, H. (2008). Towards systematic engineering of Service-Oriented access control in federated environments. In *Congress on Services Part II, 2008. SERVICES-2. IEEE*. IEEE.

Hovav, A. and Berger, R. (2009). Tutorial: Identity Management Systems and Secured Access Control. *Communications of the Association for Information Systems*, 25(1):42.

Kazman, R., Barbacci, M., Klein, M., Carrière, S. J., and Woods, S. G. (1999). Experience with performing architecture tradeoff analysis. In *Proceedings of the 21st international conference on Software engineering*. ACM.

Kunz, M., Fuchs, L., Hummer, M., and Pernul, G. (2015). Introducing dynamic identity and access management in organizations. In *International Conference on Information Systems Security*. Springer.

Meier, S., Fuchs, L., and Pernul, G. (2013). Managing the Access Grid - A Process View to Minimize Insider Misuse Risks. In *11th International Conference on Wirtschaftsinformatik (WI2013)*. University Leipzig.

Peterson, G. (2006). Introduction to identity management risk metrics. *IEEE Security & Privacy*, 4(4):88–91.

Royer, D. (2007). Enterprise identity management-what's in it for organisations?. In *FIDIS*.

Royer, D. (2013). *Enterprise Identity Management: Towards an Investment Decision Support Approach*. Springer Science & Business Media.

Royer, D. and Meints, M. (2008). Planung und Bewertung von Enterprise Identity Managementsystemen. *Datenschutz und Datensicherheit-DuD*, 32(3):189–193.

Schell, F., Dinger, J., and Hartenstein, H. (2009). Performance evaluation of identity and access management systems in federated environments. In *Infoscale*. Springer.

Software Engineering of America (2015). Reduce IBM i Help Desk Costs with Self Service Password Reset. Technical report, Software Engineering of America.

Staite, C. and Bahsoon, R. (2012). Evaluating identity management architectures. In *Proceedings of the 3rd international ACM SIGSOFT symposium on Architecting Critical Systems*. ACM.

United States Congress (1996). Health Insurance Portability and Accountability Act.

United States Congress (2002). Sarbanes-oxley act of 2002, pl 107-204, 116 stat 745. Codified in Sections 11, 15, 18, 28, and 29 USC.

Windley, P. J. (2005). *Digital Identity: Unmasking identity management architecture (IMA)*. " O'Reilly Media, Inc.".

Witty, R. J. (2003). Five Business Drivers of Identity and Access Management. Technical report, Gartner.

2 Advanced Identity and Access Policy Management using Contextual Data

Current status:	Published
Conference:	International Conference on Availability, Reliability and Security (ARES)
Core-Rank[1]:	B
Date of acceptance:	18 May 2015
Full citation:	Matthias Hummer, Michael Kunz, Michael Netter, Ludwig Fuchs, Günther Pernul. Advanced Identity and Access Policy Management using Contextual Data *Proceedings of the 10th International Conference on Availability, Reliability and Security (ARES), 2015, Pages 40-49, IEEE.*

Conference Description: ARES aims at a full and detailed discussion of the research issues of security as an integrative concept that covers amongst others availability, safety, confidentiality, integrity, maintainability and security in the different fields of applications.

Advanced Identity and Access Policy Management using Contextual Data

Matthias Hummer[1], Michael Kunz[1], Michael Netter[2], Ludwig Fuchs[2] and Günther Pernul[1]
[1] *Department of Information Systems*
University of Regensburg
Regensburg, Germany
{*firstname.lastname*}*@wiwi.uni-regensburg.de*
[2] *Nexis GmbH*
Regensburg, Germany
{*firstname.lastname*}*@nexis-secure.de*

***Abstract*—Due to compliance and IT security requirements, company-wide Identity and Access Management within organizations has gained significant importance in research and practice over the last years. Companies aim at standardizing user management policies in order to reduce administrative overhead and strengthen IT security. Despite of its relevance, hardly any supportive means for the automated detection and refinement as well as management of policies are available. As a result, policies outdate over time, leading to security vulnerabilities and inefficiencies. Existing research mainly focuses on policy detection without providing the required guidance for policy management. This paper closes the existing gap by proposing a Dynamic Policy Management Process which structures the activities required for policy management in Identity and Access Management environments. In contrast to current approaches it fosters the consideration of contextual user management data for policy detection and refinement and offers result visualization techniques that foster human understanding. In order to underline its applicability, this paper provides a naturalistic evaluation based on real-life data from a large industrial company.**

***Keywords*-Identity Management, Policy Management, Policy Mining, Access Control, RBAC**

I. Introduction

The efficient administration of employees' and their digital identities' access to sensitive applications and data is one of the biggest security challenges for today's organizations [1]. Typically, large organizations manage millions of user access privileges across thousands of IT resources. Due to ineffective and application-specific user management, employees accumulate excessive access rights over time. At the same time, organizational guidelines and policies can hardly be enforced in a decentralized environment. As a result, organizations implement a company-wide Identity and Access Management (IAM) system for the centralized management of digital identities [2]. They enable organizations to implement standardized user lifecylce processes, reduce security vulnerabilities and comply with existing national and international regulations like the Sarbanes-Oxley Act [3] or Basel III [4].

In general, modern IAM systems are built on three pillars: processes, technologies, and policies [5]. Core identity lifecycle processes like user (de)provisioning or access privilege management are implemented using available automation technologies. Existing products offer a variety of fuctionalities like identity directories for data storage, provisioning engines for user management, or workflow capabilities. Both, processes and technologies are controlled by a set of company-specific policies. These policies control technological aspects like data synchronization or data storage. At the same time they are responsible for process-related aspects like access privilege management, provisioning processes, and security management within the IAM.

While available systems offer a variety of technologies and functionalities for implementing user management processes, polices have only received little attention among researchers and practitioners. Policy management commonly still needs to be carried out manually by IT administrators with hardly any means for structured policy definition or ongoing policy management being available. As a result, only a small number of basic policies are defined and implemented in practice. These policies are commonly extracted from partly documented internal regulations and requirements and remain unchanged during system operation. As a result, they outdate over time, leading to security vulnerabilities, essentially reducing the advantages of a centralized user management. Consecutively, it is mandatory that policies evolve over time in order to reflect organizational and technological changes within a company.

In order to overcome the existing limitations, this paper introduces the Dynamic Policy Management Process (DPMP). It provides a structured approach for policy management for Identity and Access Management by applying automation technologies that detect new and potentially relevant policies as well as outdated policies. In contrast to existing approaches, it integrates the analysis of user management data as well as contextual data. The process model has been designed based on previous academic work as well as on experience gathered during our participation in several industry projects. In order to underline its applicability we extended an existing IAM tool proposed in [6] with DPMP functionality. The tool itself provides standard connectors for widely used application systems. This allowed us to facilitate available functionality and further evaluate the DPMP within a real-life use case of a large industrial company (see Section V). The remainder of the paper is structured as follows. In Section II, an overview of related work is presented before Section III gives a conceptual overview of current Identity and Access Management systems and introduces our proposed improvement. Section IV introduces the DPMP while the practical use case based on naturalistic data from a large industrial company is presented in Section V. Section VI provides a summary and outlook for future work.

II. Related Work

In todays medium to large-sized companies centralized Identity and Access Management systems facilitating Role-based Access Control (RBAC) are the de-facto standard for organizing user access to resources. Their goal is to reduce the burden of user management overhead by offering a central point for user lifecycle and access privilege handling. By doing so, they allow companies to implement standardized processes and guidelines for user management.

A large amount of research considering technological components of IAM systems and their implementation (e.g. [5], [7]), as well as their underlying access control models has been carried out [8]. However, while the importance of IAM policies in general [5] and of organizational policies in particular [9] has been acknowledged, hardly any work specifically considers the challenge of policy detection and management in a large and complex environments.

In the field of policy management, researchers have proposed a variety of top-down and bottom-up policy detection approaches. Examples for discovering security policies top-down by extracting information for policy definition from existing business processes are [10], [11], [12]. Wolter et al. [10], for instance, use Business Process Models to formulate a set of security policies using the eXtensible Access Control Markup Language. Similarly, [11] convert results from Business Process Execution Language-based processes into an RBAC state [13]. Bhatti [14] specifically focus on the detection of security policies, such as separation of duty (SOD) policies. However, SOD policies only represent a small portion of the policies required in IAM systems. Bailey et al. [15] introduce a self-adaptive framework that monitors authorizations made by role- or attribute-based systems, analyzes user behavior, and adapts the target systems accordingly. However, like other approaches, they focus on the detection of security policies rather than providing a guided process for comprehensive policy management in company-wide user management.

Besides the top-down approaches, several researchers have proposed bottom-up policy mining techniques [16], [17], [18]. In [18], for instance, security policies are derived from firewall and network information. Besides general policy mining approaches, the research community recently focused on mining attribute policies for Attribute-based Access Control [19], [20] in order to ease the migration from traditional access control models such as RBAC [21], [16]. While being valuable as a technological solution, these approaches do, amongst others, not consider business semantics or context information required in the context of Identity and Access Management to validate the correctness of suggested policies.

Summing up, available bottom-up and top-down approaches mainly focus on policy detection and do not provide the structured guidance organizations require (1) to implement policy discovery and recommendation mechanisms, and (2) ongoing policy maintenance in IAM environments. They do not consider the integration of available context data, decide upon the value of certain information for policy detection, or show how to transfer detected policies into daily operation. We argue that a comprehensive process model is required for structuring policy management in company-wide Identity and Access Management. Due to the complexity of IAM systems, missing support for human decision makers reduces applicability in practical scenarios, essentially limiting the benefit of centralized user management.

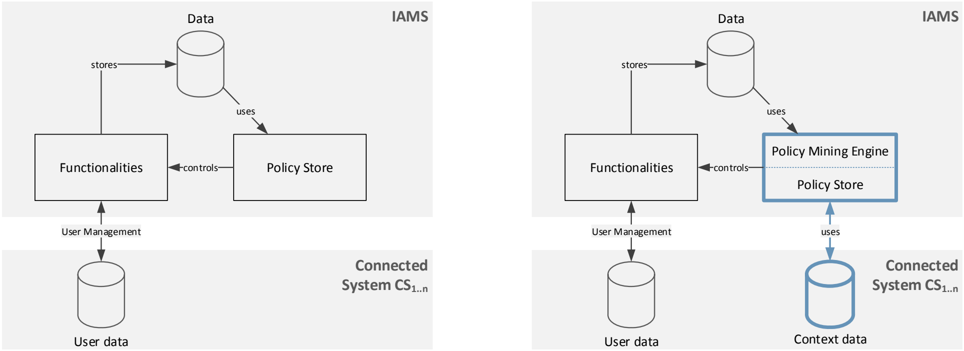
III. Conceptual Overview

In the following, at first an overview of IAM systems and their main components is provided. On this basis we propose the extension of current infrastructures using a policy mining engine for improved policy detection and recommendation. Section IV consecutively introduces the Dynamic Policy Management Process facilitating the capabilities of the newly introduced policy mining engine throughout its structured approach for policy handling.

A. Identity and Access Management Components

Typical Identity and Access Management systems (IAMS) consist of three fundamental components (Figure 1(a)): IAM **data** stored in the infrastructure, tool-supported **functionalities** for executing and automating user management tasks, and **policies** structuring the management of the overall IAMS itself [22].

1) Data: The required data within the IAMS is commonly periodically loaded from connected applications. Those can be enterprise applications having a dedicated user administration (such as the Microsoft Active Directory or SAP systems). They, however, also can represent resources hosted by partner companies (i.e. using identity federations) or cloud-based resources. Table I gives a general overview of existing and used data types. Typically, one or several personnel data systems (HR system) provide employee data such as an employee's name, departmental assignment, and further attributes like his or her cost center or location. At the same time other applications provide user account information such as account identifiers and entitlement information



(a) Status quo

(b) Proposed improvement

Figure 1. Interplay of IAMS components and connected systems

like access privileges and related attributes (e.g. criticality level, owner or description).

The data coming from the various sources is linked and stored in a central database, creating new data types for a global view on identities (e.g. combining an employee's master data with his or her application-specific user accounts) and entitlements (such as business roles that group access privileges from connected applications). Both, the connector technology as well as the data handling mechanisms rely on policies, e.g. for structuring the frequency of data synchronization or data correlation mechanisms.

2) Functionalities: Identity and Access Management functionalities implement the logic required to operate the system and provide automated services. This includes modules for user management, access management, data handling and synchronization, or user provisioning [5], [7]. User management is concerned with managing the identity lifecycle whereas access management provides functionality to authenticate and authorize users. Data handling and synchronization deal with integrating information from applications and exchanging data in a consistent manner. Finally, user provisioning is concerned with the allocation and revocation of user accounts and access privileges or business roles. All of these functionalities require the existence of policies guiding their mode of operation.

The last column of Table I underlines that current IAM systems commonly operate on the basis of information on the subject (like employee data), the object (like access privileges and applications), and the assignments between both. Thus, they are only able to process a limited static view without considering extended contextual information like an employee's activities within certain applications. In fact, most applications generate a huge amount of (audit) data such as information about a requesting entity, the affected resources, the location of access, the time, and the decision of whether the request was granted or denied. Additionally, data like an employees contract status stemming from an HR system might further support policy management. We argue that considering these extended data types allows for the improved detection of access management policies.

3) Policies: Both, data and functionalities of IAMS rely on policies for guiding their mode of operation. On a theoretical level [23] and [9] provide an overview of various policy types and their distinct sectors of applicability. [23] introduces three types of policies, namely Authorization policies, Obligation policies and Delegation policies. Similarly, [9] categorize policies into Process policies, IAM policies and Security policies.

The focus of authorization policies is to manage access to an object [23]. This type of policy regulates access to resources within a company and aims at increasing the security of company information and access to sensitive resources. For example, a depiction of the rule that only managers can view top-secret files into an RBAC role or ABAC rule falls into this category. Delegation policies are a specific set of Authorization policies that allow a subject to transfer the decision making tasks to other subjects.

Obligation policies can be divided into process policies and IAM policies. IAM policies are responsible for the design and governance of the functionality of an IAMS whereas process policies refer to rules that describe how core business processes within organization are executed. Examples for IAM policies are the organizations guidelines on access privilege re-certifications or provisioning policies that are used to automatically grant access to a set of resources when new employees join the company. Process policies, on the contrary, describe which permissions typically are activated together or sequentially in order to execute complete process activities.

Despite its importance, our experience from industry projects shows that policy management and maintenance is only rudimentary realized in practical scenarios. Policies implemented during the setup phase of the Identity and Access Management system outdate over time as no

Table I
DATA GENERATED WITHIN CURRENT IAMS

System	Data type	Examples	Used
HR system$_{1..n}$	Employee master data	Name, personnel number	x
	Employee context	Login state of an employee regarding different applications, vacation	
Application$_{1..n}$	Account information	Account identifiers, Account attributes (e.g. system accounts, privileged accounts)	x
	Entitlement information	Entitlement identifiers, Entitlement attributes (e.g. critical entitlements)	x
	Account Activity	Permission activations, activation sequences, type of permission usage, requested resources	
IAMS	Identity information	Accounts, corresponding systems, entitlements, roles	x
	Entitlement/role information	Corresponding systems, attributes	x
	Provisioning information	Requesting entities, affected resources, approving authorities, decisions	x

technological means or organizational guidance are available for verifying them periodically or detecting newly required policies. Defined policies are rather coarse-grained and simple. This can be partly attributed to the lack of available (contextual) data to identify complex polices and the human policy engineer's lack of understanding of how applications are used by employees. Additionally, scripting languages are often used to store policies. Hence, only technically experienced personnel is able to create them and refine existing ones due to the lack of appropriate user interfaces.

B. Proposed Policy Management Extension

In order to overcome the identified shortcomings, Figure 1(b) depicts our proposed improvement: Firstly, we suggest the facilitation of currently unused contextual data for policy management. Secondly, we extend policy management capabilities of IAMS with a policy mining engine that is able to consider this contextual data during the automated detection and refinement of policies according to a structured process model (presented in Section IV).

1) Context Data: According to Dey, ”context is any information that can be used to characterize the situation of an entity” [24]. In today's IAMS, almost exclusively identity and entitlement attributes are used as context data for policy decisions. Following [25], we differentiate between the following five types of additional context elements available in applications in the remainder.

- **Activity:** Frequency and count of privilege activations as well as the amount of application data accessed.
- **Individuality:** Attributes about employees, user accounts, or access privileges data commonly available within applications (e.g. criticality level or other attributes).
- **Relations:** Activity of similar or related employees, whereas similarity can be based on employee attributes or access privilege usage patterns.
- **Location:** The employee's location from which an activity originated. Technically, IP addresses (internal, external, VPN) are often used in this respect.
- **Time:** The date and time when a permission activation occurred, e.g. within common office hours or at night.

2) Improved Policy Management: To extend the policy functionality of today's IAMS, we introduce a new policy mining engine which gathers, processes and stores contextual data (as defined in Section III-A1) in order to discover and recommend new policies. Additionally, after monitoring and validating employees' activities for a sufficient period of time, it is able to recommend the refinement of existing policies. As an example, access patterns of users across applications can be monitored and policies for resource access can consecutively be refined based on actual usage statistics, usage times, or the criticality of access privileges.

IV. DYNAMIC POLICY MANAGEMENT PROCESS

To implement our vision of an improved policy management in IAM systems in complex IAM environments, a structured process model is mandatory in order to ensure applicability. In the following Section we thus introduce the Dynamic Policy Management Process supporting organizations during their policy management activities (see Figure 2). It consists of four phases that structure the activities required for policy management.

At first the infrastructural setup of the policy management component within the IAMS takes place (phase 1). Input data sources are identified and policy mining mechanisms are parametrized accordingly. Consecutively, the collection of input data is carried out (phase 2). This comprises activities like data loading, data normalization and data linking required as input data might vary regarding its currency, accuracy or provided attribute dimensions. During phase 3 the data correlation and policy mining takes place in order to differentiate between normal and outlier behavior patterns hinting at potential policy definitions and policy violations. Throughout the last step (phase 4), the results are validated and presented to human policy engineers facilitating their organizational expertise in order to model well-designed policies.

Note that phases 2-4 of the DPMP are commonly executed in a cyclic manner while the first phase must be reentered in case the system landscape changes or other strategic changes require adjustment.

The main characteristics of the DPMP are:

- Minimizing efforts to define an initial set of policies.
- Providing tool support to enable IAM engineers to execute policy modelling and refinement.

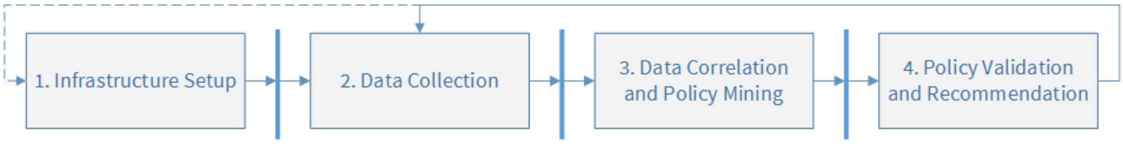

1. Infrastructure Setup
2. Data Collection
3. Data Correlation and Policy Mining
4. Policy Validation and Recommendation

Figure 2. Proposed Policy Optimization Process Model

- Integrating both, actual authorization usage data and business knowledge.
- Improving IT security through continuous refinement of policies based on actual employee behavior.

A. *Infrastructure Setup*

Phase 1 of the DPMP is concerned with the overall preconfiguration of the infrastructure, identifying and setting up data sources, and configuring system behavior regarding policy detection and policy recommendation.

1) Data Identification and Connection: Prior to the actual policy mining, available sources for contextual data need to be identified. Typical data sources are applications connected to the IAMS which store contextual data in log files. Human experts (e.g. the system administrators and policy engineers) need to decide which contextual information from a particular application should be facilitated based on the expected business value, e.g. the potential workload reduction for user management by defining new provisioning policies. For the purpose of improving the provisioning processes, for instance, the number of permission activations, the time, or location (e.g. in-house, through VPN, the originating country, etc.) for each application might be of relevance.

Note that this step heavily depends on the accessibility of data and their potentially temporal availability. While data from centralized applications like SAP ERP systems might be easily accessible, contextual information collection from distributed environments (like file servers in a globally operating organization) might be cumbersome.

After the identification of available contextual information, the data connection settings need to be adjusted. The goal is an automated data synchronization based on existing connectors as well as additional application connectors (e.g. in case required contextual information stems from a system not yet connected to the IAMS). Setting up the data synchronization also includes the mapping of data from applications to the entities stored in the IAMS. Contextual data like user account activity, for instance, needs to be related to the respective user accounts and employees.

2) Policy Mining Settings: After successful data selection and import the respective data analysis configuration needs to take place. This includes the weighting of input data for automated data analysis as well as settings regarding the systems policy recommendation behavior. Regarding the input data weighing, human engineers could e.g. decide to give more weight to data values that are constantly updated, maintained, and revised and thus have a high accuracy during the consecutive algorithmic analysis.

Additionally, the methods of policy recommendation can be parametrized according to a given organizational scenario. Similar to approaches used for the cleansing of static access privilege assignments presented in [26], [27], the DPMP requires human expert interaction after the detection of potentially reasonable policies. In case the system suggests an unreasonable large number of new policies potentially including a high rate of false-positives (detected policy suggestions which are discarded after human review), it would add an additional burden rather than create value for an organization. As a result, the systems anomaly detection techniques need to be parametrized in order to only suggest policy definitions for selected behavior patterns.

These setting commonly require the initial analysis of input data over a reasonable period of time. Imagine the correlation of access privileges usage with employees location data. In case the investigated privilege is only used by employees from a specific location during the period of investigation, the DPMP might recommend the definition of a provisioning policy that only assigns employees from this location to the according access privilege. If the period of investigation has been set too short, employees from other locations might also request the usage of this access privilege, essentially requiring the adaption of the defined policy.

B. *Data Collection*

After successful setup of the policy management system, the data collection phase takes place. During this step, the input data is loaded, normalized, and linked according to the previously defined settings. The goal is a periodic and fully automated data loading process shifting from a manual administration to an automatic machine-based execution. As a result, the latest input data are available for the automated analysis at any point in time for policy management without the need for human interaction.

In a first step the raw data from the relevant applications is imported and normalized. Systems which create a constant data stream require a continuous import, converting, and storing of data while other applications might only support a full data export (e.g. using the CSV format). Furthermore, data storage types might vary among applications, requiring data normalization. Examples are an ERP application providing usage data aggregated per single day and the amount of data accessed by clients in megabytes while a file service application delivers a steady stream of data and the amount of data sent to clients in bytes. During a last data collection activity relationships among data elements stemming from different points of time are set up. For each employee, access privileges or business role a change history is generated. This e.g. allows for the detection of activity patterns fostering the identification of user provisioning policies.

C. Data Correlation and Policy Mining

During the data correlation phase the automated policy mining takes place. The goal is to generate recommendations for relevant policies which have not been implemented up to now. At the same time already established policies are validated for adjustment. In this paper it is not our goal to provide a comprehensive list of pattern detection techniques but rather aim at showing that those techniques can be applied to support policy management efforts in general. For evaluation purposes we implemented a set of analysis techniques (see Section V).

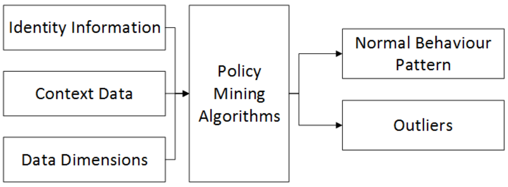


Figure 3. Input and output of policy mining algorithms

The DPMP facilitates existing data mining technologies (e.g. clustering [28] or neural networks [29]) on the basis of existing identity information, contextual data and the various data dimensions defined during the initial setup phase (see 3). Patterns of normal and outlier behavior are automatically extracted for investigated subjects. The subject may either be a single entity or a group of entities which can be uniquely identified by a set of attributes within the context of a policy. Such an entity can be an employee, a user account within an application, or a role bundling access privileges from different applications. Such data are augmented by their contextual data generated, for instance, when an entity is involved in any kind of activity.

Using data mining techniques allows for a multi-dimensional analysis facilitating sets of relevant attributes of subjects (e.g. employees, user accounts, or entitlements) and objects (e.g. amount, frequency, or criticality of data accessed). The overall goal is to identify clusters of subjects that share contextual data patterns which might in turn lead to the definition of IAM policies and the detection of outliers violating the policy.

Imagine an organization that aims at ensuring the principle of least privilege [30] in order to minimize insider misuse by overprivileged employees. Employees only are allowed to have the minimum set of access privileges required by their daily work. The DPMP in this respect continuously monitors existing user provisioning policies by identifying outdated access privilege assignments based on users behavior. The example in Figure 4 depicts the analysis of a privilege providing access to billing data within the company based on employees location (New York) and department (finance). The current provisioning policy might be refined after automated usage pattern detection identified that only employees which are assigned to the job function Clerk actually use the respective access privileges (independent of their assigned location) while Secretaries within the finance department in New York do not activate the access privilege at all.

Examples for the detection of anomalies (in contrast to standard usage patterns) might include entitlements for accessing financial data being activated from a VPN connection (while an according policy forbids this access) or access privileges which are used to manipulate an extraordinary amount of data.

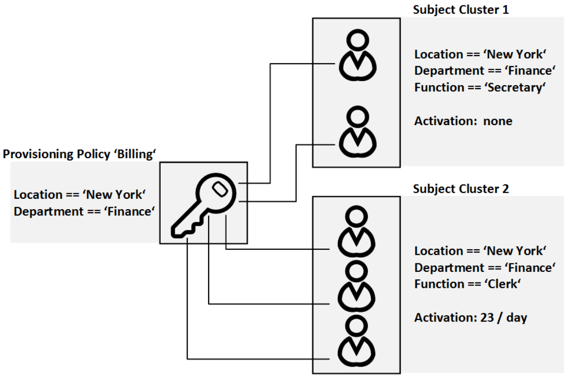


Figure 4. Example of access privilege activation analysis

D. Policy Validation and Recommendation

After successful data correlation and policy mining, a set of potentially relevant policies (e.g. provisioning policies changing the current access control state) has been identified. As IAM systems and connected applications manage a huge amount of data, a high number of potentially relevant policies might be detected by each DPMP iteration. These policy candidates need to be validated by the policy management system before being communicated in an appropriate manner to human engineers for refinement. Policy validation thus can observe the underlying rule for every detected policy over a certain period in time before it is recommended to a human policy engineer. In case a policy suggestion is based on usage activity patterns, for instance, these patterns can be validated over a period of one month. In case the pattern changes during the investigation period, the policy suggestion itself can be revoked.

Policy mining is limited to generating a set of policy suggestions based on clusters of similar subjects and their behavior based on contextual data. As a result, the focus during of the last phase of the DPMP lies on the presentation of results in an intuitive and human-understandable way in order to enable the identity engineer to easily derive appropriate actions. Visualizations can be based on techniques like charts or data tables. From our practical experience it is essential to include the visualization of the reasons why a certain policy suggestion has been created. In case ambiguous or mutually exclusive rules have been identified, this information has to be included in the result presentation as well. A human policy engineer might, for instance, be informed that accepting one policy suggestion might lead to the violation of another already implemented policy. He then might be able to decide whether the old policy is outdated while the new policy suggestion should be activated.

Again, it is not the goal of this paper to provide a comprehensive list of potential visualizations or rule definition scenarios but rather underline the importance of a dedicated result refinement phase including human interaction as a cornerstone of ongoing policy management in IAM.

In this Section we proposed the Dynamic Policy Management Process which enables organizations to gather a deeper understanding of its IAM, the (contextual) data, and the quality of currently implemented policies as well as potential policy suggestions. Based on company-specific settings, the DPMP is able to import the necessary input data, identify patterns of standard subject behavior, and support human policy engineers during policy definition and refinement.

V. Evaluation

In this Section we evaluate the applicability of the DPMP in a real-world scenario. The evaluation is based on data stemming from the SAP ERP system and the IAM system of a globally-operating manufacturing company with more than 12.000 internal and 4.000 external employees. A total number of 8.021 active user accounts, 3.925 single roles, 762 composite roles, and 1.180.962 access privilege assignments from the SAP system were initially imported and anonymized. For the following evaluation, the period under observation comprised five weeks during which daily re-imports took place.

Increasing audit requirements force the company to improve IAM policy management. Up to now only rudimentary provisioning and access re-certification policies have been defined due to missing tool support and knowledge about the underlying data. As a result, a policy detection project has been initiated. Its main goals are:

1) The consideration of contextual data from the SAP ERP system for policy generation
2) The semi-automated detection of new and potentially relevant provisioning and re-certification policies as well as the identification of loosely defined and hence insecure existing policies
3) Providing appropriate visualizations of detected policies to support human policy engineers

While requirement (1) corresponds to phase 1 and 2 of the DPMP, (2) relates its data correlation and policy mining phase (phase 3). Requirement (3) deals with the presentation of discovered policies according to phase 4 of the DPMP. Even though we executed numerous policy detection activities, we focus on two specific examples for evaluation purposes in the remainder. Firstly, the analysis of access privilege activations has been compared to the static distribution generated by the current provisioning policy in the IAM system (corresponding to phase 2 of the DPMP, see Section V-C). Secondly, detected access privilege activation frequencies were visualized in relation to the amount of data objects modified (i.e. data within the SAP system) for investigation by a human policy engineer (corresponding to phase 3 of the DPMP, see Section V-D).

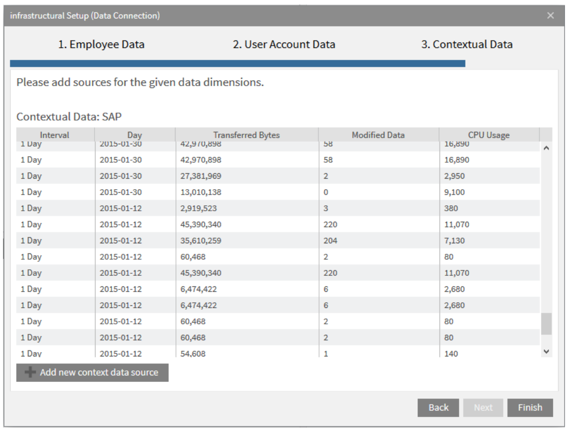
Note that a comparative evaluation of our prototype-based approach with manually executing policy detection and recommendation cannot be executed. This is due to the inapplicability of a manual examination of the several hundred-thousands of access privilege assignments and the available large amount of contextual information.

A. Infrastructure Setup

To address requirement (1) of the project, at first, context data available in the SAP ERP system was analyzed (step 1 of phase 1). Using the classification technique for context data from Section III-B1, the following information on user behavior was extracted and mapped: number and frequency of read and write permission usage and amount of transferred data (activity) for each account (individuality) per day (time) and the corresponding IP address (location). Subsequently, policy mining parametrization was conducted (step 2 of phase 1). Initially, the set of prototypical implemented algorithms (including data classification mechanisms and statistical distribution analysis) were applied using a default configuration. On this basis, distinctive properties of the imported dataset became apparent. For instance, due to SAP system limitations, user behavior can only be extracted on a daily basis. Thus, algorithms need to be configured to identify permission usage irregularities per day (e.g. suspicious permission activations on weekends) but not within the course of a single day (such as off-time activities). Furthermore, data types were weighted in cooperation with a human system expert, emphasizing the importance of data types such as IP address and employee status information during the following analyis.

B. Data Collection

After successful configuration and parametrization the data collection took place (phase 2 of the DPMP). We implemented a software wizard to ease the import of raw data types onto the internal data structures of the extended IAM tool (see Figure 5) in an automated manner. Additionally, data from the SAP system was mapped onto existing user management data from the IAM system. SAP user account activities, for instance, were related to the respective employees' identities coming from the IAM system. As a result, a total number of including 6.214.422 records from 36 days containing contextual information as well as user management data from the SAP system were gathered and mapped using our daily data import functionality.



Please add sources for the given data dimensions.

Contextual Data: SAP

Interval	Day	Transferred Bytes	Modified Data	CPU Usage
1 Day	2015-01-30	42,970,898	58	16,890
1 Day	2015-01-30	42,970,898	58	16,890
1 Day	2015-01-30	27,381,969	2	2,950
1 Day	2015-01-30	13,010,138	0	9,100
1 Day	2015-01-12	2,919,523	3	380
1 Day	2015-01-12	45,390,340	220	11,070
1 Day	2015-01-12	35,610,259	204	7,130
1 Day	2015-01-12	60,468	2	80
1 Day	2015-01-12	45,390,340	220	11,070
1 Day	2015-01-12	6,474,422	6	2,680
1 Day	2015-01-12	6,474,422	6	2,680
1 Day	2015-01-12	60,468	2	80
1 Day	2015-01-12	60,468	2	80
1 Day	2015-01-12	54,608	1	140

Figure 5. Infrastructure setup wizard

C. Data Correlation and Policy Mining

During the third phase of DPMP the actual policy mining was conducted in order to address project requirement (2). Using our implemented policy mining algorithms we were able to detect standard usage patterns potentially leading to the definition of new policies as well as the refinement of currently implemented policies.

Table II
STATIC DISTRIBUTION AND ACTUAL USE OF ACCESS PRIVILEGE P_1

Department	Distribution of static assignment	Activation frequency
D_1	21.15%	**0.04%**
D_2	24.39%	0.00%
D_3	45.08%	**99.96%**
D_4	4.82%	0.00%
D_5	0.72%	0.00%
D_6	1.54%	0.00%
D_7	2.3%	0.00%

Concerning the first exemplary case, we computed the distribution of static assignments of access privileges among the top level departments of the company and compared these to their actual activation information. Table II shows the distribution of access privilege P_1 across top level departments of the company and its actual activation in these departments. As can be seen, nearly half of the employees that are assigned to P_1 are working in the department D_3. The access privilege is almost exclusively used (99.96%) in this department while only a very small number of activations (0.04%) stems from department D_1. This indicates that access privilege P_1 might only be required for tasks conducted in department D_3. Thus, a refinement of the existing provisioning policy that additionally requires employees to work in department D_3 in order to obtain this access privilege is recommended. This restructuring might lead to a reduction of the number of overprivileged employees, thereby strengthening IT security.

In summary, out of the company's total 3.925 single roles defined in the SAP system, we identified 382 (i.e. 9.7%) which – though being assigned to employees in a particular department – were hardly activated (activation frequency for the respective department is below 1%). In an ongoing effort, these results are discussed with the company's policy engineer in order to improve existing provisioning policies and refine existing SAP role definitions leading to access privilege revocation.

D. Policy Validation and Recommendation

In order to address project requirement (3), previously detected standard usage patterns need to be validated and visualized for further human refinement. Concerning the second evaluation example, we conducted a daily analysis of employees according to their access privilege activation frequencies together with the amount of data objects read and modified.

Figure 6 depicts a screenshot from our extended IAM tool which uses a bubble chart visualization in order to display detected usage patterns. The x-axis corresponds to the amount of data that has been "modified" by an employee's access privilege activations on a single day, while the y-axis denotes the amount of data being "read". During phase 1 of the DPMP thresholds were defined for highlighting power users, i.e. employees which either read or modify large amounts of data within the SAP system. In the given example, an employee has been marked as a power user (orange colored highlighting) if he either read more than 10.000 MB or modified more than 200.000 data sets per day. Bubbles in the lower left area of Figure 6 correspond to average system users while highlighted bubbles in the other areas correspond to power users. In the given example, 63 power users were identified for the interval of our investigation. A human policy engineer could use this information for defining a new re-certification policy that demands a periodic assessment of all power users' access privileges. In contrast to standard SAP users whose

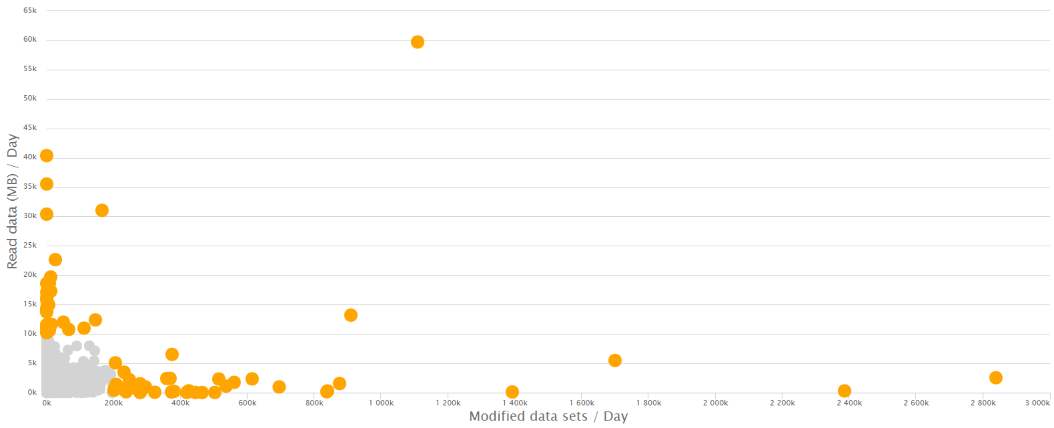


Figure 6. Detection of SAP power users

access privileged are re-certified once a year, power users might be re-certified more frequently in order to reflect their criticality value.

In summary, the naturalistic evaluation based on data from an SAP ERP system presented in this Section of the paper underlined the applicability of the DPMP for structured policy management in practice. Based on the prototypical extension of an existing IAM tool, we were able to import previously unused contextual data, identify clusters of standard as well as outlier usage behavior and visualize the gathered results. Within the company, the results increased management attention by providing in-depth insight into the current access control state and its guiding policies. At our partner's side, efforts for evaluating the application of the DPMP in a periodic manner (daily operation), the extended analysis of further applications, and the adaption of existing IAM policies are currently made.

VI. Conclusion

Over the last decades company-wide Identity and Access Management systems have become a key element for controlling users access to resources in medium to large-sized enterprises. They offer means for a centralized enforcement of standardized user management processes and policies. Despite their importance, the management of IAM policies still needs to be executed manually. While current research concentrates on mechanisms for policy detection, the complexity of user management in large environments rather requires a structured and applicable process for policy management. Human policy engineers need to be supported with guidance and automation during the detection, implementation, and refinement of IAM policies.

In order to improve the current situation we presented the Dynamic Policy Management Process which structures the activities during policy management into four phases. It facilitates a mining engine which generates policy recommendations based on contextual data of employees and further presents gathered results to human policy engineers. In order to underline the practical relevance and applicability of our contribution we conducted a practical case study within a large industrial company and its ERP system managing several thousands of users and more than one million access privileges.

For future work, we plan to extend the DPMP in order to improve the representation and management of policy recommendations. Practical experience shows that a high amount of potentially conflicting recommendations increases manual efforts of human role engineers and requires an in-depth understanding of the underlying data. In the future we hence aim at providing an analysis of policy interdependencies in order to overcome this limitation. We additionally aim at extending our prototype implementation and evaluate the DPMP throughout further practical use cases considering contextual data from de-centralized applications.

Acknowledgements

The research leading to these results was partly supported by the "Bavarian State Ministry of Education, Science and the Arts" as part of the FORSEC research association.

References

[1] A. Hovav and R. Berger, "Tutorial: identity management systems and secured access control," *Communications of the Association for Information Systems*, vol. 25, no. 1, p. 42, 2009.

[2] A. Cleven and R. Winter, "Regulatory compliance in information systems research literature analysis and research agenda," in *Enterprise, Business-Process and Information Systems Modeling*, ser. Lecture Notes in Business Information Processing, T. Halpin, J. Krogstie, S. Nurcan, E. Proper, R. Schmidt, P. Soffer, and R. Ukor, Eds. Springer Berlin Heidelberg, 2009, vol. 29, pp. 174–186.

[3] SOX, "Sarbanes-oxley act of 2002, pl 107-204, 116 stat 745," July 2002.

[4] Basel Committee on Banking Supervision, "Basel III - A global regulatory framework for more resilient banks and banking systems," 2011.

[5] L. Fuchs and G. Pernul, "Supporting compliant and secure user handling - a structured approach for in-house identity management," in *The Second International Conference on Availability, Reliability and Security, 2007: ARES 2007*. Los Alamitos, Calif.: IEEE Computer Society, 2007, pp. 374–384.

[6] L. Fuchs, M. Kunz, and G. Pernul, "Role model optimization for secure role-based identity management," in *European Conference on Information Systems (ECIS)*, Juni 2014.

[7] D. Royer, "Enterprise identity management–what's in it for organisations," *Proceedings of the IFIP/FIDIS summer school on The future of identity in the information society*, pp. 403–416, 2008.

[8] L. Fuchs, G. Pernul, and R. Sandhu, "Roles in information security–a survey and classification of the research area," *Computers & Security*, vol. 30, no. 8, pp. 748–769, 2011.

[9] L. Fuchs and G. Pernul, "Minimizing insider misuse through secure identity management," *Security and Communication Networks*, vol. 5, no. 8, pp. 847–862, 2012.

[10] C. Wolter, A. Schaad, and C. Meinel, "Deriving XACML policies from business process models," in *Web Information Systems Engineering–WISE 2007 Workshops*. Springer, 2007, pp. 142–153.

[11] J. Mendling, M. Strembeck, G. Stermsek, and G. Neumann, "An approach to extract rbac models from BPel4Ws processes," in *Enabling Technologies: Infrastructure for Collaborative Enterprises, 2004. WET ICE 2004. 13th IEEE International Workshops on*. IEEE, 2004, pp. 81–86.

[12] A. Baumgrass, S. Schefer-Wenzl, and M. Strembeck, "Deriving process-related rbac models from process execution histories," in *Computer Software and Applications Conference Workshops (COMPSACW), 2012 IEEE 36th Annual*. IEEE, 2012, pp. 421–426.

[13] R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman, "Role-based access control models," *Computer*, vol. 29, no. 2, pp. 38–47, 1996.

[14] R. Bhatti, E. Bertino, and A. Ghafoor, "X-federate: a policy engineering framework for federated access management," *Software Engineering, IEEE Transactions on*, vol. 32, no. 5, pp. 330–346, May 2006.

[15] C. Bailey, D. W. Chadwick, and R. de Lemos, "Self-Adaptive Authorization Framework for Policy Based RBAC/ABAC Models," *Dependable, Autonomic and Secure Computing, IEEE International Symposium on*, vol. 0, pp. 37–44, 2011.

[16] Z. Xu and S. D. Stoller, "Mining attribute-based access control policies from logs," in *Data and Applications Security and Privacy XXVIII*. Springer, 2014, pp. 276–291.

[17] "Deriving current state rbac models from event logs."

[18] H. Safaa, C. Frédéric, C.-B. Nora, A. Vijay, and M. Stéphane, "Policy mining: a bottom-up approach toward a model based firewall management," in *9th International Conference on Information Systems Security*, A. Bagchi and I. Ray, Eds. Springer Berlin Heidelberg, Dec. 2013, pp. 133–147.

[19] J. Lopez, R. Oppliger, and G. Pernul, "Authentication and Authorization Infrastructures (AAIs): A Comparative Survey," *Comput. Secur.*, vol. 23, no. 7, pp. 578–590, Oct. 2004.

[20] V. C. Hu, D. Ferraiolo, R. Kuhn, A. Schnitzer, K. Sandlin, R. Miller, and K. Scarfone, "Guide to Attribute Based Access Control (ABAC) Definition and Considerations," *NIST Special Publication*, vol. 800, p. 162, 2014.

[21] Z. Xu and S. D. Stoller, "Mining attribute-based access control policies from rbac policies," in *Emerging Technologies for a Smarter World (CEWIT), 2013 10th International Conference and Expo on*. IEEE, 2013, pp. 1–6.

[22] J. Pato and O. C. Center, "Identity management: Setting context," *Hewlett-Packard, Cambridge, MA*, 2003.

[23] M. Strembeck, *Engineering of Dynamic Policy-Based Systems: A Policy Engineering of Dynamic Policy-Based Systems: Language Based Approach*, Habilitation Thesis, WU-Wien, 2008.

[24] A. K. Dey, "Understanding and using context," *Personal Ubiquitous Comput.*, vol. 5, no. 1, pp. 4–7, Jan. 2001.

[25] A. Zimmermann, A. Lorenz, and R. Oppermann, "An operational definition of context," in *Proceedings of the 6th International and Interdisciplinary Conference on Modeling and Using Context*, ser. CONTEXT'07. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 558–571.

[26] L. Fuchs, C. Broser, and G. Pernul, "Different approaches to in-house identity management-justification of an assumption," in *Availability, Reliability and Security, 2009. ARES'09. International Conference on*. IEEE, 2009, pp. 122–129.

[27] A. Colantonio, R. Di Pietro, A. Ocello, and N. V. Verde, "A new role mining framework to elicit business roles and to mitigate enterprise risk," *Decision Support Systems*, vol. 50, no. 4, pp. 715–731, 2011.

[28] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 14. Oakland, CA, USA., 1967, pp. 281–297.

[29] T. Kohonen, "An introduction to neural computing," *Neural networks*, vol. 1, no. 1, pp. 3–16, 1988.

[30] J. H. Saltzer and M. D. Schroeder, "The protection of information in computer systems," *Proceedings of the IEEE*, vol. 63, no. 9, pp. 1278–1308, 1975.

3 Introducing Dynamic Identity and Access Management in Organizations

Current status:	Published
Conference:	International Conference on Information Systems Security (ICISS)
Core-Rank[2]:	B
Date of acceptance:	19 September 2015
Full citation:	Michael Kunz, Ludwig Fuchs, Matthias Hummer, Günther Pernul. Introducing Dynamic Identity and Access Management in Organizations *Proceedings of the 11th International Conference on Information Systems Security, 2015, Pages 139-158, Springer*.

Conference Description: ICISS is a core conference for the international IT security research community. It is typically hosted in India (2015: Kolkata) and annually provides research insights on information systems and their security. ICISS proceedings are published in the Springer *Lecture Notes in Computer Science*.

Introducing Dynamic Identity and Access Management in Organizations

Michael Kunz[1], Ludwig Fuchs[2], Matthias Hummer [2], and Günther Pernul[1]

[1]Department of Information Systems
University of Regensburg, Regensburg, Germany
michael.kunz@ur.de,guenther.pernul@ur.de
[2]Nexis GmbH, Regensburg, Germany
ludwig.fuchs@nexis-secure.com,matthias.hummer@nexis-secure.com

Abstract. Efficient and secure management of access to resources is a crucial challenge in today's corporate IT environments. During the last years, introducing company-wide Identity and Access Management (IAM) infrastructures building on the Role-based Access Control (RBAC) paradigm has become the de facto standard for granting and revoking access to resources. Due to its static nature, the management of role-based IAM structures, however, leads to increased administrative efforts and is not able to model dynamic business structures. As a result, introducing dynamic attribute-based access privilege provisioning and revocation is currently seen as the next maturity level of IAM. Nevertheless, up to now no structured process for incorporating Attribute-based Access Control (ABAC) policies into static IAM has been proposed. This paper closes the existing research gap by introducing a novel migration guide for extending static IAM systems with dynamic ABAC policies. By means of conducting structured and tool-supported attribute and policy management activities, the migration guide supports organizations to distribute privilege assignments in an application-independent and flexible manner. In order to show its feasibility, we provide a naturalistic evaluation based on two real-world industry use cases.

Keywords: Identity and Access Management, IAM, ABAC, Policies

1 Motivation

The effective and secure management of employees' access to sensitive applications and data is one of the biggest security challenges for today's organizations [19]. A variety of national and international regulations or certifications like Basel III [3], the Sarbanes-Oxley-Act of 2002 [45], or the ISO 27000 family [23] together with internal guidelines force enterprises to audit and control actions within their systems. At the same time developments like the application of cloud-based services in corporate environments further underline the need for secure user management.

As a result, centralized Identity and Access Management (IAM) relying on the Role-based Access Control (RBAC) [43] paradigm became the core element for increasing user management efficiency and reduce related IT security risks over the last years. However, due to its static nature, the application of RBAC leads to a considerable amount of administrative overhead. Growing numbers of outdated roles stemming from organizational changes together with the need of manually administrating user role assignments as well as role permission assignments result in complex and outdated RBAC structures. Even disregarding the fact that it takes an average of 18 months for its initial implementation, RBAC consumes an average of 2,410,000$ for a firm of 10,000 employees [34]. As a result, researchers and practitioners recently started to point out the need for dynamic access privilege management IAM infrastructures ([42, 27, 14]).

Using Attribute-based Access Control (ABAC) policies [20] for dynamically granting and revoking access based on employees' and privileges' attributes (from hereinafter referred to as dynamic Identity and Access Management (dIAM)) is seen as the next maturity level of company-wide IAM. The ABAC paradigm in general is based on the presumption that using a subject's, object's, and their shared context's attributes an authorization decision can be made. ABAC research traditionally focused on aspects like expressing ABAC rules (e.g. using XACML as standardized language) while only little attention has been paid to its adoption in company-wide IAM environments. This adaptation requires the definition of a potentially high number of policies within the central IAM system, the enforcement of policy decisions within the legacy applications depending on their underlying access control models, as well as the continuous policy maintenance. In order to complete these tasks, companies require a guided approach which is able to manage organizational project complexity as well es the technical heterogeneity of involved applications and protocols. To the best of our knowledge, no such structured approach has been provided up to now.

In this paper we are closing the existing research gap by firstly investigating the main building blocks required for dIAM infrastructures (Section 3). Secondly we propose a migration guide for implementing dIAM which serves as a project guideline dividing the necessary steps into manageable activities (Section 4). We thirdly evaluate our work within two real world use cases in the insurance and research industry. Besides the theoretical structuring of activities we identified the need for automation and thus additionally provided a prototypical software implementation for executing single activities of our migration guide. In order to achieve this we extended an existing IAM-tool proposed in [10] with attribute management and policy generation functionality. This allowed us to facilitate available functionality (e.g. data import or data visualization) and further evaluate our migration guide within real-life projects (see Section 5).

2 Related Work

Traditionally, Identity and Access Management in organizations has been associated with storing user data, maintaining user accounts, and controlling users' access to applications [11]. In today's medium to large-sized companies a centralized management of users following the RBAC paradigm has become the de facto standard approach for handling the challenges imposed by a steadily growing number of digital identities as well as access privileges. Recent surveys underline this growing importance of roles in information security in general and in IAM environments in particular [13]. However, over time and without proper controls such as de-provisioning processes, the number of roles is steadily growing, contradicting the benefits of administrative cost reduction [9]. In order to keep role systems up to date, methodologies and metrics for the ongoing optimization of role-based IAM infrastructures are required [10, 26]. Nonetheless, the static concept of roles in general lacks the ability to adopt to company changes and struggles with situational adaptivity [42]. Both requirements, however, are main challenges of modern IAM infrastructures.

As a result, companies aim at enhancing their existing IAM systems with dynamic ABAC policies in order to increase provisioning capabilities, strategically reduce administrative tasks, and keep IAM infrastructures manageable [21]. While standard ABAC protocols like the eXtensible Access Control Markup Language (XACML) [33] have been around since 2003, Priebe at al. [36] and Yuan et al. [52] were the first to formally define ABAC as an access control model. However, their focus was on formalizing the model and did not consider an application-independent IAM scenario. Jin et al. suggest an attribute-based architecture for IAM focusing on attribute correlation and attribute importance in different IAM-related domains [25]. Their work, however, does not aim at supporting organizations during the set up of a dIAM system. Recently, Hu et al. [20] were amongst the first to provide generalized definitions and best practices while also giving recommendations on deploying ABAC in cross-application settings. They, however, neither provide the structured guidance nor an overview on how to adopt ABAC in an organization-wide IAM system.

Up to now, to the best of our knowledge, no approach constituting the single building blocks of ABAC-based company-wide IAM and aligning them into a structured process model exists. We close this gap in the remainder by firstly gathering the aforementioned building blocks on the basis of a thorough research review (Section 3). Secondly, we structure them in the form of a migration guide which can be employed by organizations that aim at extending their static identity- or role-based IAM towards the integration of ABAC policies (Section 4).

3 Building Blocks of Dynamic Identity and Access Management

In the following we present the core elements of dIAM systems derived from ABAC literature (e.g. building on the findings of [20]) as well as literature from related areas, such as data and information quality management or policy management. Even though most works do not consider their application for company-wide IAM in particular, researchers in general already identified attribute management as well as policy management as the two main aspects of any ABAC implementation. Attribute management [20, 6, 35, 16, 8, 37, 52] in general deals with requirements related to the attributes used within ABAC policies, ranging from the aggregation of attributes up to their ongoing maintenance. Policy management [20, 4, 15, 24, 30, 22, 37] deals with the development and continuous improvement of access policies.

3.1 Policy Management

While policies and their life-cycle in general have been studied in various research areas (e.g. [7]), researchers recently stated the need for a structured approach for policy management in IAM. Building on the generic policy life-cycle model proposed by Buecker et al. ([7], see Figure 1) we outline relevant aspects of policy management in IAM in the following.

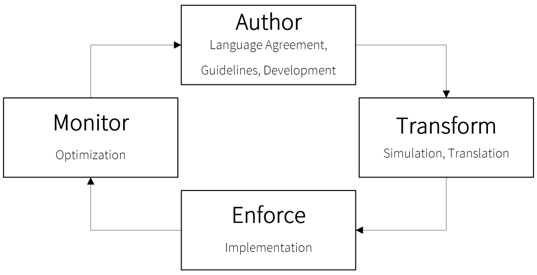


Fig. 1. Policy management based on [7] including corresponding dIAM aspects

Language Agreement
The first challenge prior to defining policies is the agreement upon a common expression language providing the syntax for depicting the semantics of policies interpreted by an IAM infrastructure. Looking at the research area, language requirements have been investigated [44] and comparisons of the suitability of policy languages (e.g. [17]) such as XACML [33] or EPAL [1] have been provided. Other authors like Strembeck [48] rather suggest generating a customized policy language tailored to the specific needs of a certain scenario. Within the area of IAM, however, a standardized approach seems more promising due to the high number of different applications and stakeholders involved.

Guidelines
Besides a common policy language, the establishment of policy guidelines plays an important role during the development as well as maintenance of dIAM systems. Policy guidelines are representing general rules on how policies are to be developed within a specific context. Note that in complex scenarios contradicting policies could potentially be defined. As a result, the establishment of design

guidelines is mandatory in order to avoid semantically correct but inefficiently modeled and contradicting policies. Beckerle and Martucci [4] were the first to formally define security and manageability goals for policies. They exclusively examined general goals for security and authorization rules. However, their results also can be applied in the context of IAM. Examples include the following goals provided in [4]:

- *Rule sets have to grant authorized access*
- *Redundant rules need to be removed.*
- *Contradicting rules need to be removed*
- *Concise rule sets are better than large rule sets*

By means of such exemplary guidelines organizations can increase policy homogeneity and ease policy maintenance.

Development

Policy Development deals with the actual creation of policies. Choosing an appropriate policy development methodology within a given scenario (i.e. an IAM project) is crucial for project success. Available methodologies can be divided into policy engineering and policy mining approaches (see Figure 2). Policy engineering deals with the top-down extraction of policies from business processes or workflows [2, 5], optionally based on security policy templates as shown in [41]. Authors agree that the policy notation used during policy development [47] and the provided tool-support [46] are critical success factors for policy engineering. Policy mining, in contrast, applies data mining technologies for extracting policies from Natural Language Policies [49, 29], currently assigned access privileges [50], or access logs [51, 22]. While providing an increased level of automation, policy mining lacks the integration of business know-how and struggles with low-quality attribute values - above all in the context of company-wide IAM involving numerous stakeholders and policies. Research results from related areas [11] underline that in such scenarios a hybrid approach building on both, an increased level of automation as well as the integration of expert knowledge, is the most promising method for policy modeling.

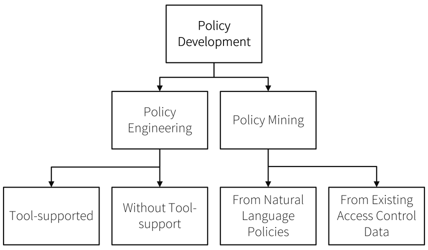


Fig. 2. Policy development methods

Simulation, Translation and Implementation
In company-wide IAM systems a potentially large number of ABAC policies affecting thousands of access privilege assignments might be required. As a result, a tool-supported simulation for anticipating the consequences of newly introduced policies becomes a central step during the setup of a policy base. Simulation tools can support the integration of policy owner feedback prior to policy activation as well as depict the future state of access within systems managed by an IAM infrastructure (e.g. using visual investigations as proposed in [31]). After simulation the policies need to be mapped onto the access control models of the legacy applications connected to an IAM. Those applications commonly are based on static access control models (e.g. SAP based on static roles or the Microsoft Active Directory (AD) based on groups). As a result, the IAM system in place has to carry out the required translations, i.e. the provisioning of dynamically calculated access privileges using static access control concepts (e.g. SAP roles).

Optimization
Once simulated and implemented, policies require the continuous monitoring of their correctness and validity by applying automated analytical methods. Note that due to the high number of expected policies a manual analysis is not feasible in the context of IAM. Lu et al., for instance, provide an approach for discovering inconsistencies and errors within policies at design-time [28]. Recently, Hummer et. al [22] proposed an approach that allows for a structured optimization of policies without interfering with a running IAM system. They apply anomaly detection methods in order to highlight deviations of normal policy patterns and visually present them to human policy engineers.

3.2 Attribute Management

Besides policy-related activities, attributes and their management form the foundation of any ABAC implementation. Attribute management is of great importance for company-wide IAM Despite its importance for company-wide IAM where employees are managed based upon master data attributes and access privileges are handled using attributes. However, attribute management in IAM has not attracted researchers' attention to a great extent up to now.

System & Attribute Selection
The initial selection and definition of application systems as well as related attributes managed within the ABAC policies [20] is the foundation for structured attribute management for dIAM. Note that in case an organization already has a deployed IAM system, basic attribute selection already took place during the initial system setup. Nevertheless, a re-investigation and potential extension of attribute sets commonly needs to be executed. Several master data attributes stored within a personnel management system might, for instance, be unused up to now but needed during later policy definition (e.g. an employee's job position or cost center).

Constraints & Data Types

After selecting required attributes, a definition of their data types, values and constraints needs to be carried out. Data types commonly range from boolean to single-valued and multi-valued attributes [6]. Researchers recently analyzed the effects of policy evaluation performance and highlighted its relation to the used attributes and attribute values [32]. Regarding attribute constraints, Bijon et al., for instance, introduce constraints on attribute assignments and values [6]. As further examples, Jin et al. provide a methodology for the classification of attributes according to their criticality and importance for access [25], while there also exists an overview of data and systems that are typically involved in an IAM environment [22].

Data Integration

As aforementioned, company-wide IAM commonly handles large amounts of data stemming from numerous applications, databases, or directory services. Organizations already operating an IAM hence need to review and extend existing integration processes to reflect the needs of future dynamic ABAC policies. IAM systems in general differentiate between source and target systems whereas a source system for certain attributes can act as target system for other attributes at the same time. An example could be an HR system providing master data of employees while at the same time receiving employees' email addresses from a mail application. Note that the definition of master sources for attributes has implications on attribute ownership. It is e.g. likely that human resources representatives are responsible for reviewing and validating attributes stemming from the personnel system.

Cleansing & Quality Controls

Policies created on the basis of erroneous attribute values essentially lead to security vulnerabilities, compliance violations, and administrative overhead. As a result, a structured review and cleansing of incorporated attribute values is a mandatory building block of dIAM prior to policy development. For an overview of potential data quality problems, cf. [39]. Hummer et al. recently argued that for optimizing policies, a centralized view on available and utilizable attributes spanning all involved systems is necessary in order to detect data errors and inconsistencies [22]. Data cleansing additionally builds on available attribute quality controls (e.g. rules for valid attribute values). Such quality controls, e.g., support the automated monitoring of attribute value changes and the advent of new attribute values and attribute types. We suggest to apply measures and metrics (for an overview cf. [18]) as well as best practices [40] from the field of data and information quality management to address these challenges.

4 Migration Guide

After describing the building blocks of an ABAC-based IAM, this Section of the paper introduces our tool-supported migration guide supporting a step-by-step

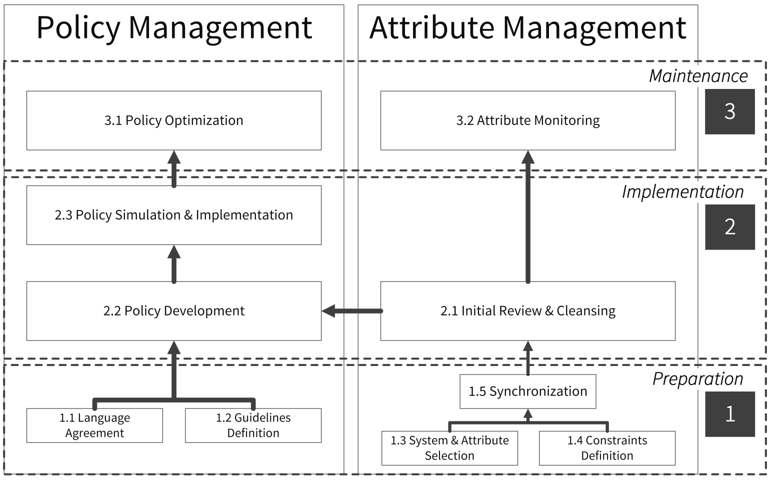


Fig. 3. Process model for migrating towards dIAM

migration from an existing static towards a dynamic IAM solution. It consists of three phases, namely a preparatory phase, an implementation phase and a maintenance phase (see Figure 3). The goal of the preparatory phase is to achieve a common understanding of policies and provide an attribute base used during later process phases. The subsequent implementation phase covers the cleansing of attributes and actual development of policies while the maintenance phase provides measures for continuous monitoring and improvement of the policy system. Note that due to space restrictions we cannot provide a detailed presentation of all involved sub-activities but rather aim at giving an overview of required tasks. In order to increase automation, we implemented a prototypical software for supporting the execution of attribute and policy management activities (Phase 2 of our migration guide). It is able to exchange data with an existing IAM system supporting the respective ABAC implementation process.

4.1 Preparation Phase

Due to the complexity and heterogeneity of static IAM environments, several preparatory activities have to be completed before ABAC policies can be defined. Relevant systems, attributes, responsibilities, and guidelines have to be reviewed and defined in order to foster a common understanding on a technical as well as organizational level among involved stakeholders.

Attribute Management
During system and attribute selection source systems for attribute data (e.g. personnel management systems) need to be investigated for attributes required dur-

ing policy definition (Activity 1.3). Additional sources like IAM systems themselves or other applications providing information about user accounts or access privileges (e.g. ownership, criticality) might be identified. Note that organizations having basic attribute synchronization processes in place commonly have not dealt with the facilitation of extended attributes for complex access control decisions. By investigating system documentation or conducting expert interviews they hence need to review and extend the currently used attribute types in order to reflect ABAC requirements.

At the same time, data types need to be defined and constraint definitions for the attributes need to be established (Activity 1.4, cf. Section 3, *Constraints & Data Types*). This, amongst others, includes the definition of data types, master data sources, data ownerships, valid attribute values, or attribute ranges, i.e. intervals (if the data type is a numeric type) of validity. This way, erroneous attribute values can be identified during the subsequent cleansing activities.

After successfully completing the system and attribute selection and definition of constraints and data types, the attribute synchronization (Activity 1.5) takes place. Attribute values are imported into the IAM during this phase. At the same time conflicts like different encodings or granularity issues (e.g. *address* vs. *street* and *zip code*) can be detected.

Policy Management

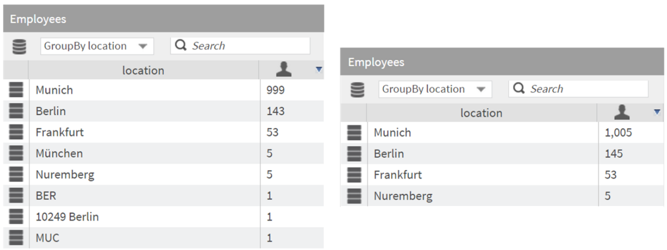
Regarding policy management, a general language agreement (Activity 1.1) for policy expression as well as the definition of policy guidelines need to be established prior to policy creation. Most of the currently available IAM implementations, for instance, are able to foster XACML as standardized policy language. Additionally, a shared understanding among project stakeholders on an organizational level needs to be established in the form of a company-wide glossary with definitions for important terminology. Available policy types like grant or denial policies should, for instance, be described. Furthermore, guidelines for policies (Activity 1.2, cf. Section 3, *Policy Guidelines*) can act as sources on how the human policy engineers are requested to model policies. Imagine a scenario in which only grant policies are allowed. Policy engineers should hence not have the option to design denial policies throughout a tool-supported policy creation process at all. Additionally, guidelines for the strategic maintenance of policies (Phase 3 of our migration guide) need to be defined. By introducing policy and attribute ownerships and requiring a periodic certification process, companies can essentially increase long-term policy quality.

4.2 Implementation Phase

After the preparatory activities have been completed, organizations enter the implementation phase (Phase 2) of our migration guide, i.e. the initial development and setup of a dIAM based on ABAC policies. Concerning attribute management, a systematic initial review and cleansing (Activity 2.1) of attribute data is required before the initial creation of policies as well as their subsequent simulation and implementation (Activities 2.2 and 2.3) can be carried out (cf. Section 3, *Policy Development; Simulation, Translation and Implementation*).

Attribute Management

Medium and large-sized organizations commonly struggle with data quality issues regarding their digital identities and access privileges. As a result, a dedicated cleansing process for improving attribute data quality is a crucial success factor for implementing dIAM. Following an initial assessment of attribute data (e.g. the identification of empty or invalid attribute values) the manual or automated cleansing of attributes needs to take place. We argue that a tool-based detection and cleansing process fosters user adoption by reducing the overall project complexity. Automated error identification can, for instance, be carried out by means of data mining or data quality metrics. Data mining, for instance, can be applied to detect outliers and unusual attribute values (see [12]). Based on predefined quality metrics (e.g. general rules like the currency [18] of an attribute value or a list of valid location attribute values) it leads to an overall higher quality of defined policies. Figure 4 (left side) gives a simple attribute cleansing example by grouping current location attribute values from a personnel system within our prototype after the attribute synchronization took place. Existing data errors such as typos, different language codings, or misspellings can be identified easily. The right side of Figure 4 displays the attribute values after cleansing by human experts in collaboration with attribute owners.



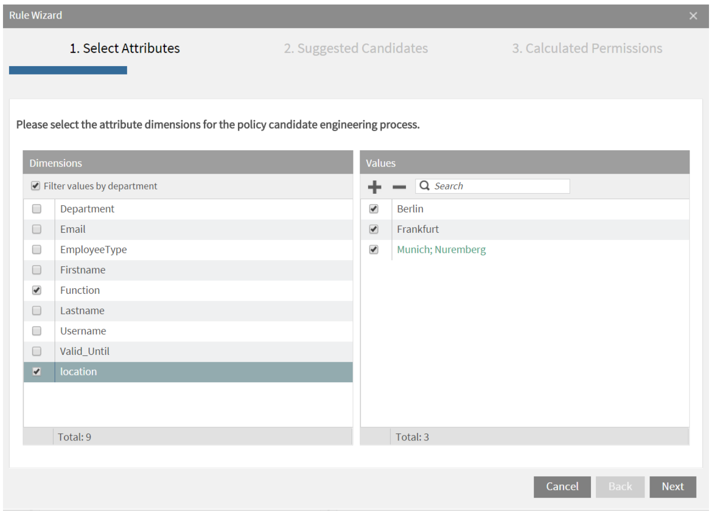
location	👤
Munich	999
Berlin	143
Frankfurt	53
München	5
Nuremberg	5
BER	1
10249 Berlin	1
MUC	1

location	👤
Munich	1,005
Berlin	145
Frankfurt	53
Nuremberg	5

Fig. 4. Before and after manual cleansing by grouping of attribute *location* and its various occurrences

Policy Management

As aforementioned, a potentially high number of policies bundling a wide range of access privileges or responsibilities are managed in corporate IAM environments. As a result, a manual policy generation by human policy engineers is not feasible. Organizations thus aim at employing automation techniques for creating policies and reviewing them in a hybrid manner (e.g. by experts who provide business knowledge and semantics, see Section 3). As one example of a potential role development approach in large IAM environments, we thus implemented policy mining algorithms that are able to automatically generate candidates for grant policies based on given attribute information. In order to support human review processes we additionally developed a simple representation of policies using a wizard-based graphical interface within our prototype (see Figure 5). Using this approach, a human policy engineer can select combinations of available attributes (left side of Figure 5, e.g. *function* and *location*) and optionally merge semantically or syntactically equivalent attribute values (right side of Figure 5, bundling the attribute values *Munich* and *Nuremberg*). In a second review step, suggested policy candidates are then displayed to the policy engineer. Continuing our example above, access is granted on the basis of the combination of employees' *location* and *function*. As a result, three policies for each function attribute value are generated, e.g. one policy for *sales representatives* in *Berlin*, *Frankfurt*, and *Munich/Nuremberg* each. During review, a human policy engineer can alter or remove unneeded policies (e.g. in case no *sales representatives* are located in *Frankfurt*). During a third step our prototype calculates the access rights shared by policy members based on customizable data mining algorithms. This way, a policy engineer could, for instance, enforce that only access rights that are not yet included in other policies are considered during the access privilege calculation or that critical access privileges are in general excluded from policy generation.



Please select the attribute dimensions for the policy candidate engineering process.

Dimensions

☑ Filter values by department

□ Department
□ Email
□ EmployeeType
□ Firstname
☑ Function
□ Lastname
□ Username
□ Valid_Until
☑ location

Total: 9

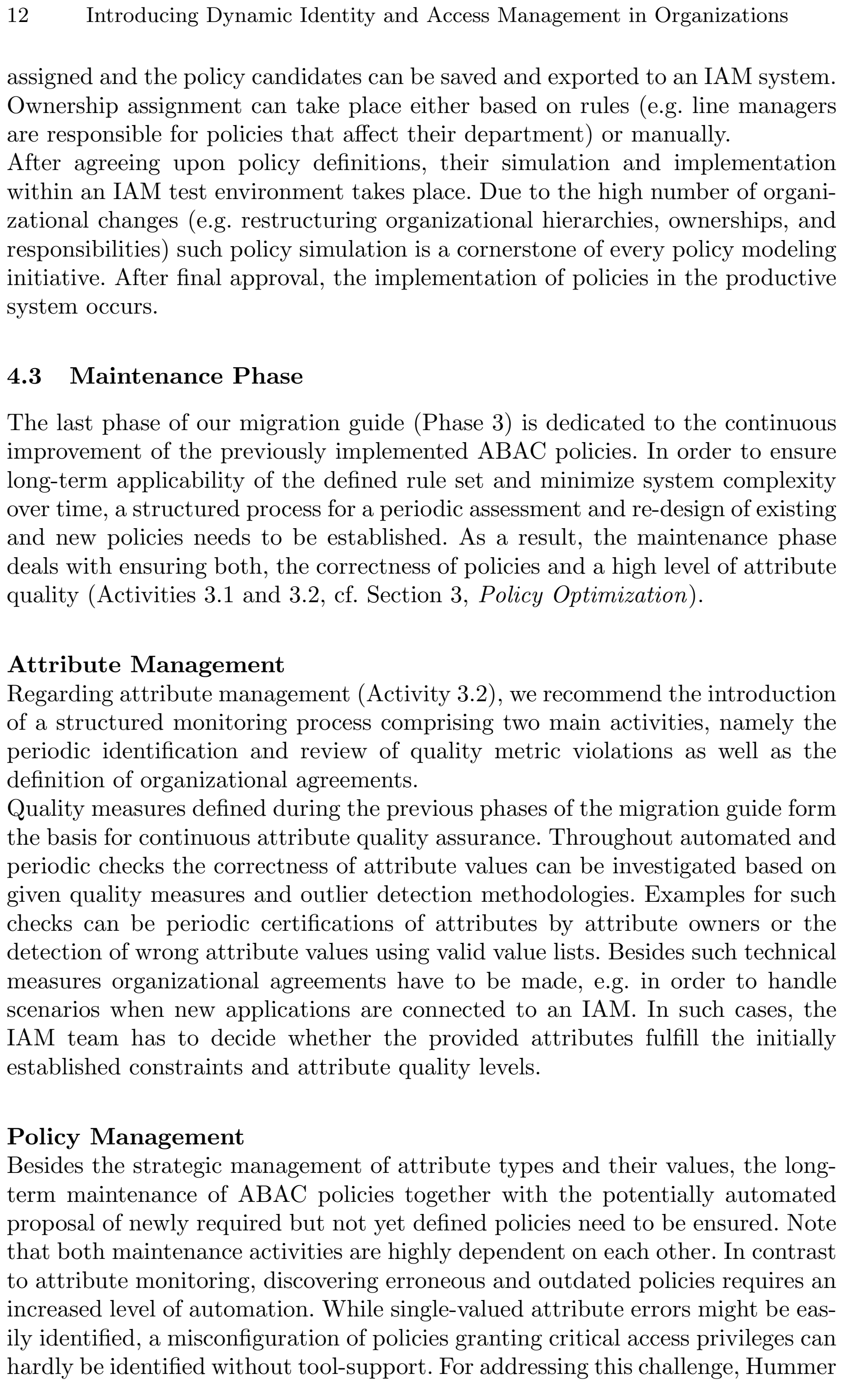
Values

☑ Berlin
☑ Frankfurt
☑ Munich; Nuremberg

Total: 3

Fig. 5. Automated tool-based policy mining and review

Completing the third step of our policy development wizard, policy owners are



assigned and the policy candidates can be saved and exported to an IAM system. Ownership assignment can take place either based on rules (e.g. line managers are responsible for policies that affect their department) or manually.

After agreeing upon policy definitions, their simulation and implementation within an IAM test environment takes place. Due to the high number of organizational changes (e.g. restructuring organizational hierarchies, ownerships, and responsibilities) such policy simulation is a cornerstone of every policy modeling initiative. After final approval, the implementation of policies in the productive system occurs.

4.3 Maintenance Phase

The last phase of our migration guide (Phase 3) is dedicated to the continuous improvement of the previously implemented ABAC policies. In order to ensure long-term applicability of the defined rule set and minimize system complexity over time, a structured process for a periodic assessment and re-design of existing and new policies needs to be established. As a result, the maintenance phase deals with ensuring both, the correctness of policies and a high level of attribute quality (Activities 3.1 and 3.2, cf. Section 3, *Policy Optimization*).

Attribute Management

Regarding attribute management (Activity 3.2), we recommend the introduction of a structured monitoring process comprising two main activities, namely the periodic identification and review of quality metric violations as well as the definition of organizational agreements.

Quality measures defined during the previous phases of the migration guide form the basis for continuous attribute quality assurance. Throughout automated and periodic checks the correctness of attribute values can be investigated based on given quality measures and outlier detection methodologies. Examples for such checks can be periodic certifications of attributes by attribute owners or the detection of wrong attribute values using valid value lists. Besides such technical measures organizational agreements have to be made, e.g. in order to handle scenarios when new applications are connected to an IAM. In such cases, the IAM team has to decide whether the provided attributes fulfill the initially established constraints and attribute quality levels.

Policy Management

Besides the strategic management of attribute types and their values, the long-term maintenance of ABAC policies together with the potentially automated proposal of newly required but not yet defined policies need to be ensured. Note that both maintenance activities are highly dependent on each other. In contrast to attribute monitoring, discovering erroneous and outdated policies requires an increased level of automation. While single-valued attribute errors might be easily identified, a misconfiguration of policies granting critical access privileges can hardly be identified without tool-support. For addressing this challenge, Hummer et al. recently suggested measures and processes for strategic policy maintenance [22]. They, for instance, introduce tool-supported outlier and anomaly detection for identifying unused or outdated policies into the field of IAM.

5 Evaluation

After proposing our migration guide we now execute a naturalistic ex post evaluation covering two industry use cases based on the evaluation framework by Pries-Heje et al. [38]. The used real-life data-sets originate from companies operating in the health insurance in Switzerland (from hereinafter refereed to as 'Insucomp') as well as the research sector in Germany (from hereinafter refereed to as 'Rescomp'). All attribute values have been anonymized accordingly. While Rescomp already had a working IAM system in place, Insucomp conducted a policy development project as part of their initiative to initially implement an IAM system. The project duration was six months (Rescomp) and nine months (Insucomp) respectively, with both projects sharing the same overall goals:

1. Automatically providing new employees with correct basic access.
2. Increasing the amount of automatically distributed privileges by using dynamic provisioning policies.

In order to achieve these goals, both companies executed Phases 1 and 2 of our migration guide and facilitated our prototypical tool implementation during policy development. Insucomp additionally implemented basic measures for policy and attribute maintenance (Phase 3) while Rescomp plans to do so in future. Note that even though both use cases only aimed at policy definition based on subject attributes, our model can also be applied during the general development of policies comprising subject, object, and environmental attributes.

5.1 Insucomp

Insucomp is employing 349 external and 866 internal employees which in total own 7,777 accounts in 13 different application systems, including one AD and one SAP instance. In total, 2,297 different access rights are directly assigned to the user accounts resulting in 54,059 access rights assignments. Insucomp's variety of applications using static access privilege assignments in combination with manual provisioning processes resulted in large administrative efforts over the last years. As a result, a new IAM system based on dynamic access control policies had to be introduced between 2014 and 2015.

Preparation Phase

Throughout a kick-off workshop, Insucomp initially taught policy engineers guidelines on how to semi-automatically construct policies (Activity 1.2) while the IAM software implemented during the overall IAM project pre-defined the applied policy language (Activity 1.1). In the specific case the proprietary modeling capabilities of the Dell One Identity Manager were employed due to the reduced expected technical implementation efforts required. The system and attribute selection (Activity 1.3) took place in an iterative manner. Firstly, the HR system was defined as source for employee master data. The available attributes together with access privileges from all 13 applications were imported into our prototype. Consecutively, policy engineers and the responsible line officers agreed upon the exclusion of certain access rights from the Microsoft AD, the SAP, and the Customer Relationship Management system from further consideration. This decision was based on several reasons: Firstly, granting certain access rights in an automated manner would have resulted in an increase of license costs. Secondly, selected access privileges from the Customer Relationship Management system were classified as critical from an IT security perspective and hence excluded from automated provisioning processes. Regarding the attributes for policy development, the domain experts and IAM team selected an employee's *position* as the main HR attribute for the policy construction. Constraints and data types were defined accordingly:

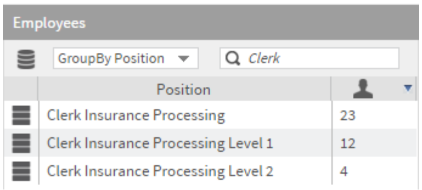
- C1: The German value for the *position* is used in policies.
- C2: A *code* is introduced for each value, referring to exactly one *position*.
- C3: A policy definition needs to contain both, a human-readable *position* as well as its respective machine-readable 4-digit *code*.
- C4: The *position* is a string value.

During attribute synchronization (Activity 1.5), violations of those constraints were identified. As an example, several languages were originally used to express an employee's position. In coordination with the HR department, the German *position* attribute value (C1) was selected as the defining attribute for later policy evaluations. Other languages were excluded from the data import and from now on are represented as translation of the main value (i.e. the German value) within a new attribute field in the HR system.

Implementation Phase

Following our migration guide a subsequent data cleansing process was conducted. Inspecting all attribute values within our prototype (Activity 2.1), Insucomp was, amongst others, able to discover ten erroneously defined *positions*. Additionally, *positions* with an inappropriate semantic granularity level were detected. For instance, initially one *position* for *Clerk Insurance Processing* existed within the HR system. However, for representing two semantically distinct insurance levels, Insucomp had to model two additional types of clerks with different access rights. As a result, the IAM team enforced the creation of more detailed *positions* and *codes* within the HR system. In the given example, two new *positions* were created in the HR system and employees were assigned accordingly (see Figure 6). Finishing the data cleansing activities a total of 253 *positions* have been available in the final attribute base.

After successful attribute cleansing, the actual detection of policy candidates within our prototype and the respective review together with domain experts



Position	👤
Clerk Insurance Processing	23
Clerk Insurance Processing Level 1	12
Clerk Insurance Processing Level 2	4

Fig. 6. Example for refactoring of employee *positions*

took place (Activity 2.2). As a side effect, Insucomp was able to discard 3,600 excessive assignments (i.e. 6.7% of all access privilege assignments) during the policy review process as our prototype highlighted additional (potentially excessive) privileges of employees assigned to a certain policy. This had a large impact on the overall project, further underlining the importance of secure provisioning and de-provisioning processes based on dynamic policies.

Finally, Insucomp exported the defined policies from our prototype and imported them within their newly set-up IAM system (Activity 2.3). They randomly selected sample policies in order to simulate correct functionality throughout various identity lifecycle processes (i.e. onboarding, change, and offboarding of employees). As a result, a total of 253 policies were put into operation. This led to the dynamic provisioning of 32% of all access rights among Insucomp's 13 connected application systems, essentially reducing the manual administrative workload while at the same time increasing the level of IT security.

Maintenance Phase

At the end of the migration project, Insucomp defined measures and quality controls in order to ensure the correctness of policies and attributes (Phase 3 of our migration guide). For conducting structured attribute management (Activity 3.2) newly introduced or changed attributes or attribute values have to be reported by the HR department to the IAM team in the future in order to adapt policies accordingly. Policy optimization has not been carried out up to now but is one element of the Insucomp IAM roadmap within the next year.

5.2 Rescomp

Rescomp already employed a working IAM system prior to the beginning of their policy definition project. Nonetheless, user management still was executed manually to a large extent for the 473 employees and the 761 different access rights (5,774 user privilege assignments in total). Rescomp's dynamic research environment requires automated and flexible access privilege provisioning in the future (e.g. for external employees like students needing temporary access to critical company data while undergoing regular organizational changes at the same time). As a result, a dIAM migration project was initiated in 2014. Similar to Insucomp, Rescomp executed the first two phases of our migration guide. Even though they have not executed maintenance activities up to now, they recently defined policy optimization as one element of their future IAM roadmap.

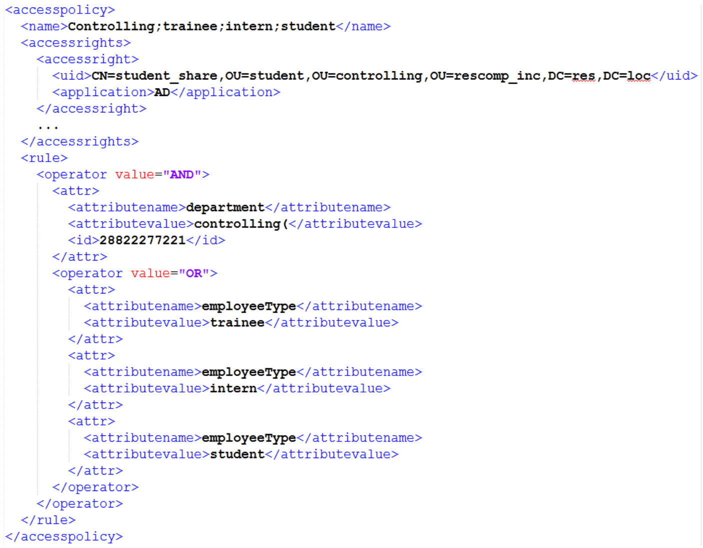
Preparation Phase

As a preparatory activity, Rescomp defined general guidelines for policy modeling (Activity 1.2). They introduced three types of valid policies, namely location-based policies, department and type-based policies, as well as function-based policies. Location-based policies represent the *physical location* of employees e.g. for granting physical access to buildings. *Department*- and *type*-based polices, in contrast, are defined based on the departmental assignment of employees in combination with their type, essentially granting access to departmental file shares for *internals*, *trainees*, *students*, or *externals*. In addition, function-based policies were defined to further refine employee's access rights according to their *job function*. Besides the three policy types, the IAM team defined a guideline regarding the definition of empty policies, i.e. policies that currently no employee is matching. In accordance with their project goals they decided to prepare such policies prior to an initial match of an employee (Activity 1.2). They, for instance, created a policy for all members of the technical service *department* whose *type of contract* is student. Students might only work within the department during their term holidays and thus the according policy might be unused for certain periods of the year but still is required during other months.

Following the migration guide, they selected two installations of their Microsoft AD for inclusion of access rights and provided the employee attributes from the HR system in place. During Activity 1.4 *department*, *type of contract*, *function*, *project* and *location* were selected as attribute base for policy definition. Similarly to Insucomp, Rescomp defined constraints and data types for these attributes. They, for instance, decided that regarding the *types of contract internals*, *apprentices*, and *students* should be treated equally in terms of their access rights.

Implementation Phase

Due to an already high attribute quality provided by the HR system, attribute cleansing was not required as no errors were identified during the attribute investigation. As a result, the IAM team subsequently conducted the policy development (Activity 2.2) in cooperation with business representatives. They started with the definition of basic *location* policies and continued with the creation of *department* and *employee type*-based policies as well as policies for employees' *function* attributes. Business representatives were asked to review the policy candidates using our prototype. In total, this process lead to the definition of 449 policies for automatic access privilege assignments, covering a total of 34.8% of all managed access privileges. Regarding the access rights, 45.9% of all initially existing privileges can now be assigned in an automatic way, i.e. they are included in at least one policy. All policies were exported from our prototype using the XML-notation and consecutively transferred into the existing LDAP-based IAM system of Rescomp using custom Python scripts (Activity 2.3). Figure 7 presents



```
<accesspolicy>
  <name>Controlling;trainee;intern;student</name>
  <accessrights>
    <accessright>
      <uid>CN=student_share,OU=student,OU=controlling,OU=rescomp_inc,DC=res,DC=loc</uid>
      <application>AD</application>
    </accessright>
    ...
  </accessrights>
  <rule>
    <operator value="AND">
      <attr>
        <attributename>department</attributename>
        <attributevalue>controlling(</attributevalue>
        <id>28822277221</id>
      </attr>
      <operator value="OR">
        <attr>
          <attributename>employeeType</attributename>
          <attributevalue>trainee</attributevalue>
        </attr>
        <attr>
          <attributename>employeeType</attributename>
          <attributevalue>intern</attributevalue>
        </attr>
        <attr>
          <attributename>employeeType</attributename>
          <attributevalue>student</attributevalue>
        </attr>
      </operator>
    </operator>
  </rule>
</accesspolicy>
```

Fig. 7. Example policy export using XML notation

a short XML export example of one department and employee type-based policy bundling students, trainees, and internships within a controlling department.

6 Conclusion

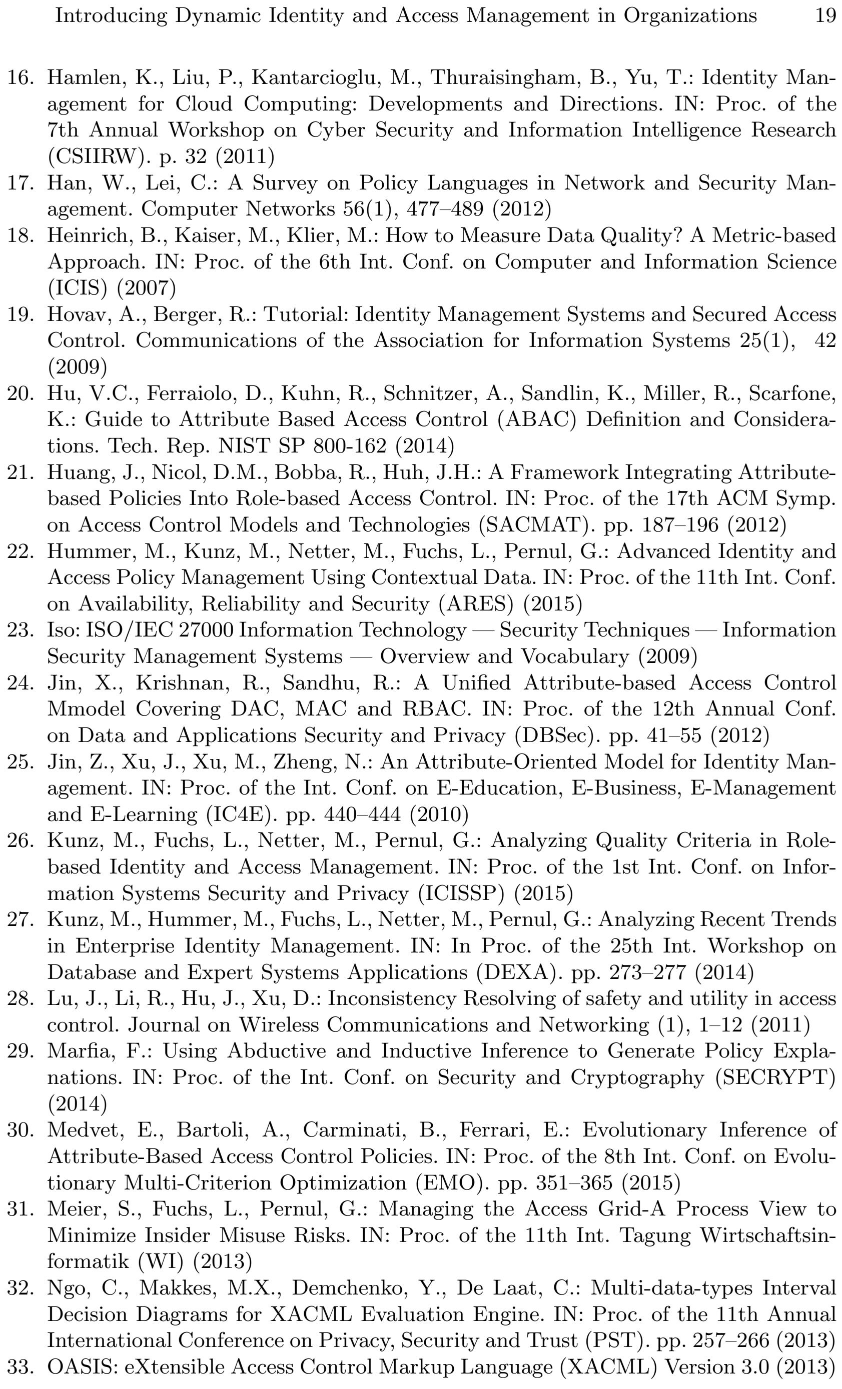
Dynamically assigning and revoking access privileges in company-wide IAM infrastructures has gained significant importance when it comes to automated and secure user management. Migrating to a dynamic IAM infrastructure based on ABAC policies can decrease manual administrative efforts while at the same time increasing the overall IT security level within companies. In order to support organizations during their required migration efforts, we proposed a novel three-step migration guide for implementing dynamic IAM based on ABAC policies in a structured manner. Up to now, no such structured process model highlighting and coordinating the respective migration tasks has been proposed. Our migration guide covers the required preparation, setup, as well as maintenance tasks and additionally offers tool-support in order to automate attribute and policy management activities. By doing so it increases the flexibility of policy engineers, reduces errors during policy modeling, and speeds-up the overall process of policy creation. Evaluating our migration guide throughout two real-life use cases we have further underlined its practical applicability.
In the future, we plan to extend our software prototype by implementing automated identity attribute monitoring activities that support companies during long-term attribute maintenance. In contrast to organizational guidelines this

would support the enforcement of quality rules for attribute management. Additionally, we plan to expand policy development and policy maintenance capabilities in order to allow for a better cooperation between the responsible domain experts and the policy engineers.

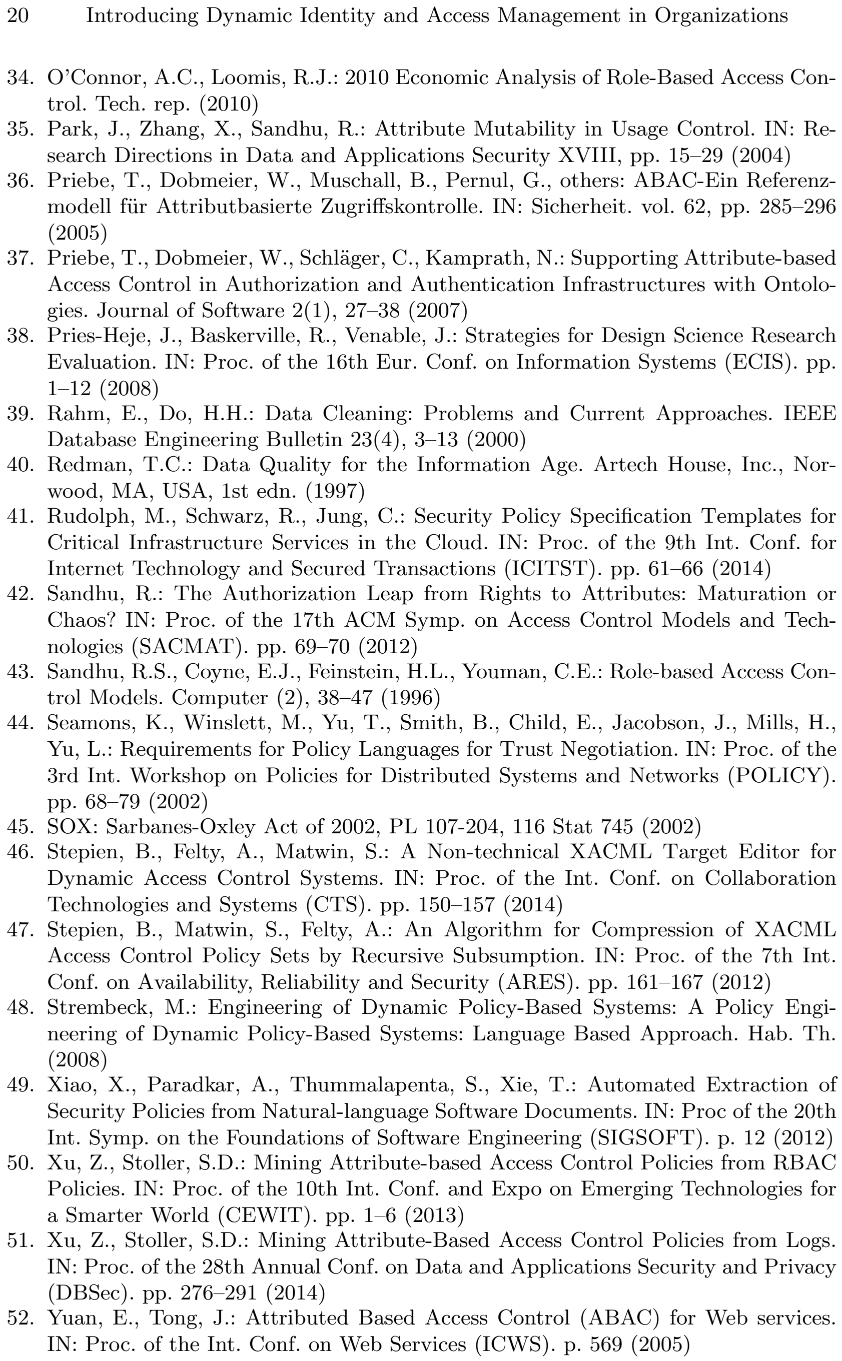
Acknowledgment. The research leading to these results was supported by the "Bavarian State Ministry of Education, Science and the Arts" as part of the FORSEC research association.

References

1. Ashley, P., Hada, S., Karjoth, G., Powers, C., Schunter, M.: Enterprise Privacy Authorization Language (EPAL 1.2). Submission to W3C (2003)
2. Aubert, J., Gateau, B., Incoul, C., Feltus, C.: SIM: An Innovative Business-Oriented Approach for a Distributed Access Management. IN: Proc. of the 3rd Int. Conf. on Information and Communication Technologies: From Theory to Applications (ICTTA). pp. 1–6 (2008)
3. Basel Committee on Banking Supervision: Basel III - A Global Regulatory Framework for More Resilient Banks and Banking Systems (2011)
4. Beckerle, M., Martucci, L.A.: Formal Definitions for Usable Access Control Rule Sets From Goals to Metrics. IN: Proc. of the 9th Symp. on Usable Privacy and Security (SOUPS). p. 2 (2013)
5. Bhatti, R., Bertino, E., Ghafoor, A.: X-FEDERATE: a Policy Engineering Framework for Federated Access Management. IEEE Transactions on Software Engineering 32(5), 330–346 (2006)
6. Bijon, K.Z., Krishman, R., Sandhu, R.: Constraints Specification in Attribute Based Access Control. Science 2(3), pp–131 (2013)
7. Buecker, A., Andrews, S., Forster, C., Harlow, N., Lu, M., Muppidi, S., Norvill, T., Nye, P., Waller, G., White, E.T.: IT Security Policy Management Usage Patterns Using IBM Tivoli Security Policy Manager. IBM Redbooks (2011)
8. Chadwick, D.W., Inman, G.: Attribute Aggregation in Federated Identity Management. IEEE Computer 42(5), 33–40 (2009)
9. Elliott, A., Knight, S.: Role Explosion: Acknowledging the Problem. IN: Proc. of the Int. Conf. on Software Engineering Research and Practice (SERP). pp. 349–355 (2010)
10. Fuchs, L., Kunz, M., Pernul, G.: Role Model Optimization for Secure Role-Based Identity Management. IN: Proc of the 22st Eur. Conf. on Information Systems (ECIS) (2014)
11. Fuchs, L., Pernul, G.: HyDRo – Hybrid Development of Roles. IN: Proc. of the 4th Int. Conf. on Information Systems Security (ICISS). pp. 287–302 (2008)
12. Fuchs, L., Pernul, G.: Qualitätssicherung im Identity- und Access Management. HMD Praxis der Wirtschaftsinformatik 50(1), 88–97 (2013)
13. Fuchs, L., Pernul, G., Sandhu, R.: Roles in Information Security - a Survey and Classification of the Research Area. Computers & Security 30(8), 748–769 (2011)
14. Gartner: Gartner IAM 2020 Predictions, `http://www.avatier.com/products/identity-management/resources/gartner-iam-2020-predictions/`
15. Gupta, P., Stoller, S.D., Xu, Z.: Abductive Analysis of Administrative Policies in Rule-based Access Control. IEEE Transactions on Dependable and Secure Computing 11(5), 412–424 (2014)



16. Hamlen, K., Liu, P., Kantarcioglu, M., Thuraisingham, B., Yu, T.: Identity Management for Cloud Computing: Developments and Directions. IN: Proc. of the 7th Annual Workshop on Cyber Security and Information Intelligence Research (CSIIRW). p. 32 (2011)
17. Han, W., Lei, C.: A Survey on Policy Languages in Network and Security Management. Computer Networks 56(1), 477–489 (2012)
18. Heinrich, B., Kaiser, M., Klier, M.: How to Measure Data Quality? A Metric-based Approach. IN: Proc. of the 6th Int. Conf. on Computer and Information Science (ICIS) (2007)
19. Hovav, A., Berger, R.: Tutorial: Identity Management Systems and Secured Access Control. Communications of the Association for Information Systems 25(1), 42 (2009)
20. Hu, V.C., Ferraiolo, D., Kuhn, R., Schnitzer, A., Sandlin, K., Miller, R., Scarfone, K.: Guide to Attribute Based Access Control (ABAC) Definition and Considerations. Tech. Rep. NIST SP 800-162 (2014)
21. Huang, J., Nicol, D.M., Bobba, R., Huh, J.H.: A Framework Integrating Attribute-based Policies Into Role-based Access Control. IN: Proc. of the 17th ACM Symp. on Access Control Models and Technologies (SACMAT). pp. 187–196 (2012)
22. Hummer, M., Kunz, M., Netter, M., Fuchs, L., Pernul, G.: Advanced Identity and Access Policy Management Using Contextual Data. IN: Proc. of the 11th Int. Conf. on Availability, Reliability and Security (ARES) (2015)
23. Iso: ISO/IEC 27000 Information Technology — Security Techniques — Information Security Management Systems — Overview and Vocabulary (2009)
24. Jin, X., Krishnan, R., Sandhu, R.: A Unified Attribute-based Access Control Mmodel Covering DAC, MAC and RBAC. IN: Proc. of the 12th Annual Conf. on Data and Applications Security and Privacy (DBSec). pp. 41–55 (2012)
25. Jin, Z., Xu, J., Xu, M., Zheng, N.: An Attribute-Oriented Model for Identity Management. IN: Proc. of the Int. Conf. on E-Education, E-Business, E-Management and E-Learning (IC4E). pp. 440–444 (2010)
26. Kunz, M., Fuchs, L., Netter, M., Pernul, G.: Analyzing Quality Criteria in Role-based Identity and Access Management. IN: Proc. of the 1st Int. Conf. on Information Systems Security and Privacy (ICISSP) (2015)
27. Kunz, M., Hummer, M., Fuchs, L., Netter, M., Pernul, G.: Analyzing Recent Trends in Enterprise Identity Management. IN: In Proc. of the 25th Int. Workshop on Database and Expert Systems Applications (DEXA). pp. 273–277 (2014)
28. Lu, J., Li, R., Hu, J., Xu, D.: Inconsistency Resolving of safety and utility in access control. Journal on Wireless Communications and Networking (1), 1–12 (2011)
29. Marfia, F.: Using Abductive and Inductive Inference to Generate Policy Explanations. IN: Proc. of the Int. Conf. on Security and Cryptography (SECRYPT) (2014)
30. Medvet, E., Bartoli, A., Carminati, B., Ferrari, E.: Evolutionary Inference of Attribute-Based Access Control Policies. IN: Proc. of the 8th Int. Conf. on Evolutionary Multi-Criterion Optimization (EMO). pp. 351–365 (2015)
31. Meier, S., Fuchs, L., Pernul, G.: Managing the Access Grid-A Process View to Minimize Insider Misuse Risks. IN: Proc. of the 11th Int. Tagung Wirtschaftsinformatik (WI) (2013)
32. Ngo, C., Makkes, M.X., Demchenko, Y., De Laat, C.: Multi-data-types Interval Decision Diagrams for XACML Evaluation Engine. IN: Proc. of the 11th Annual International Conference on Privacy, Security and Trust (PST). pp. 257–266 (2013)
33. OASIS: eXtensible Access Control Markup Language (XACML) Version 3.0 (2013)



34. O'Connor, A.C., Loomis, R.J.: 2010 Economic Analysis of Role-Based Access Control. Tech. rep. (2010)
35. Park, J., Zhang, X., Sandhu, R.: Attribute Mutability in Usage Control. IN: Research Directions in Data and Applications Security XVIII, pp. 15–29 (2004)
36. Priebe, T., Dobmeier, W., Muschall, B., Pernul, G., others: ABAC-Ein Referenzmodell für Attributbasierte Zugriffskontrolle. IN: Sicherheit. vol. 62, pp. 285–296 (2005)
37. Priebe, T., Dobmeier, W., Schläger, C., Kamprath, N.: Supporting Attribute-based Access Control in Authorization and Authentication Infrastructures with Ontologies. Journal of Software 2(1), 27–38 (2007)
38. Pries-Heje, J., Baskerville, R., Venable, J.: Strategies for Design Science Research Evaluation. IN: Proc. of the 16th Eur. Conf. on Information Systems (ECIS). pp. 1–12 (2008)
39. Rahm, E., Do, H.H.: Data Cleaning: Problems and Current Approaches. IEEE Database Engineering Bulletin 23(4), 3–13 (2000)
40. Redman, T.C.: Data Quality for the Information Age. Artech House, Inc., Norwood, MA, USA, 1st edn. (1997)
41. Rudolph, M., Schwarz, R., Jung, C.: Security Policy Specification Templates for Critical Infrastructure Services in the Cloud. IN: Proc. of the 9th Int. Conf. for Internet Technology and Secured Transactions (ICITST). pp. 61–66 (2014)
42. Sandhu, R.: The Authorization Leap from Rights to Attributes: Maturation or Chaos? IN: Proc. of the 17th ACM Symp. on Access Control Models and Technologies (SACMAT). pp. 69–70 (2012)
43. Sandhu, R.S., Coyne, E.J., Feinstein, H.L., Youman, C.E.: Role-based Access Control Models. Computer (2), 38–47 (1996)
44. Seamons, K., Winslett, M., Yu, T., Smith, B., Child, E., Jacobson, J., Mills, H., Yu, L.: Requirements for Policy Languages for Trust Negotiation. IN: Proc. of the 3rd Int. Workshop on Policies for Distributed Systems and Networks (POLICY). pp. 68–79 (2002)
45. SOX: Sarbanes-Oxley Act of 2002, PL 107-204, 116 Stat 745 (2002)
46. Stepien, B., Felty, A., Matwin, S.: A Non-technical XACML Target Editor for Dynamic Access Control Systems. IN: Proc. of the Int. Conf. on Collaboration Technologies and Systems (CTS). pp. 150–157 (2014)
47. Stepien, B., Matwin, S., Felty, A.: An Algorithm for Compression of XACML Access Control Policy Sets by Recursive Subsumption. IN: Proc. of the 7th Int. Conf. on Availability, Reliability and Security (ARES). pp. 161–167 (2012)
48. Strembeck, M.: Engineering of Dynamic Policy-Based Systems: A Policy Engineering of Dynamic Policy-Based Systems: Language Based Approach. Hab. Th. (2008)
49. Xiao, X., Paradkar, A., Thummalapenta, S., Xie, T.: Automated Extraction of Security Policies from Natural-language Software Documents. IN: Proc of the 20th Int. Symp. on the Foundations of Software Engineering (SIGSOFT). p. 12 (2012)
50. Xu, Z., Stoller, S.D.: Mining Attribute-based Access Control Policies from RBAC Policies. IN: Proc. of the 10th Int. Conf. and Expo on Emerging Technologies for a Smarter World (CEWIT). pp. 1–6 (2013)
51. Xu, Z., Stoller, S.D.: Mining Attribute-Based Access Control Policies from Logs. IN: Proc. of the 28th Annual Conf. on Data and Applications Security and Privacy (DBSec). pp. 276–291 (2014)
52. Yuan, E., Tong, J.: Attributed Based Access Control (ABAC) for Web services. IN: Proc. of the Int. Conf. on Web Services (ICWS). p. 569 (2005)

4 Adaptive Identity and Access Management - Contextual Data Based Policies

Current status:	Published
Journal:	EURASIP Journal on Information Security (JIS)
Date of acceptance:	31 July 2016
Full citation:	Matthias Hummer, Michael Kunz, Michael Netter, Ludwig Fuchs, Günther Pernul. Adaptive Identity and Access Management — Contextual Data based Policies *EURASIP Journal on Information Security (JIS), 2016, Volume 2016:19, Springer.*

Journal Description: JIS is a Springer-published open access journal with focus on IT security research. It is supported by the European Association for Signal Processing (EURASIP) and aims at bringing together experts from research and practice dealing with the general topic of information security.

Hummer *et al. EURASIP Journal on Information Security* (2016) 2016:19
DOI 10.1186/s13635-016-0043-2

EURASIP Journal on Information Security

RESEARCH Open Access

Adaptive identity and access management—contextual data based policies

Matthias Hummer[1,2]*, Michael Kunz[2], Michael Netter[1], Ludwig Fuchs[1] and Günther Pernul[2]

Abstract
Due to compliance and IT security requirements, company-wide identity and access management within organizations has gained significant importance in research and practice over the last years. Companies aim at standardizing user management policies in order to reduce administrative overhead and strengthen IT security. These policies provide the foundation for every identity and access management system no matter if poured into IT systems or only located within responsible identity and access management (IAM) engineers' mind. Despite its relevance, hardly any supportive means for the automated detection and refinement as well as management of policies are available. As a result, policies outdate over time, leading to security vulnerabilities and inefficiencies. Existing research mainly focuses on policy detection and enforcement without providing the required guidance for policy management nor necessary instruments to enable policy adaptibility for today's dynamic IAM. This paper closes the existing gap by proposing a dynamic policy management process which structures the activities required for policy management in identity and access management environments. In contrast to current approaches, it utilizes the consideration of contextual user management data and key performance indicators for policy detection and refinement and offers result visualization techniques that foster human understanding. In order to underline its applicability, this paper provides an evaluation based on real-life data from a large industrial company.

Keywords: Identity management, Policy management, Policy mining, Access control, Security management

1 Introduction

The efficient administration of employees' access to sensitive applications and data is one of the biggest security challenges for today's organizations [1]. Typically, large organizations manage millions of user access privileges across thousands of IT resources. Due to ineffective and application-specific user management, employees accumulate excessive access rights over time. As a consequence, most users are overprivileged, meaning they are assigned more permissions than necessary to perform their work. At the same time, organizational guidelines and policies can hardly be enforced in a decentralized environment. As a result, organizations implement a company-wide identity and access management (IAM) system for the centralized management of digital identities [2]. This enables organizations to implement standardized user lifecycle processes, reduce security vulnerabilities and comply with existing national and international regulations like the Sarbanes-Oxley Act [3] or Basel III [4].

In general, typical IAM systems are built on three pillars: processes, technologies and policies [5]. Core identity lifecycle processes like user (de)provisioning or access privilege management are implemented using available automation technologies. Existing products offer a variety of functionalities like identity directories for data storage, provisioning engines for user management or workflow capabilities. Both processes and technologies are controlled by a set of company-specific policies. These policies control technological aspects like data synchronization or data storage. At the same time, they are responsible for process-related aspects like access privilege management, provisioning processes, and security management within the IAM.

While available systems offer a variety of technologies and functionalities for implementing user management processes, policies have received little attention among researchers and practitioners so far. Policy management commonly still needs to be carried out manually by

*Correspondence: matthias.hummer@nexis-secure.com
[1] Nexis GmbH, Franz-Mayer-Straße 1, 93053 Regensburg, Germany
[2] University of Regensburg, Universitätsstraße 31, 93051 Regensburg, Germany

IT administrators with hardly any means for structured policy definition or ongoing policy management being available. Moreover, only static data is employed (e.g. department of an employee), letting valuable data lie fallow. As a result, only a small number of basic policies are defined and implemented in practice. These policies are commonly extracted from partly documented internal regulations and requirements and remain unchanged during system operation. This results in a situation where policies outdate over time, leading to security vulnerabilities, essentially reducing the advantages of a centralized user management. Consecutively, it is mandatory that policies evolve over time in order to reflect organizational and technological changes within a company.

In order to overcome the existing limitations, this paper introduces the dynamic policy management process (DPMP) for IAM. It provides a structured approach for policy management for IAM by applying automation technologies. On the one hand, these techniques are used in order to create a better knowledge about identity data by calculating key performance indicators (KPI) to automatically adjust policies to the current system state. On the other hand, we use them to detect new and potentially relevant policies as well as outdated policies. In contrast to existing approaches, our approach integrates the analysis of user management data as well as contextual data. The process model has been designed based on previous academic work as well as on experience gathered during our participation in several industry projects. In order to underline its applicability, we extended an existing IAM tool proposed in [6] with DPMP functionality. The tool itself provides standard IAM connectors for widely used application systems. This allowed us to facilitate available functionality and further evaluate the DPMP within a real-life use case of a large industrial company (see Section 5).

Our research methodology follows the paradigm of design science research as presented by [7] and [8]. Following the design science cycle, we derive awareness for the problem (step 1 of the design science cycle) in Section 1. In order to overcome the problems, we propose its current state of research (Section 2) together with objectives of our approach in Section 3 (2). We designed our artifact, the DPMP, in Section 4 (3). The evaluation (4) and demonstration (5) following a real-world ex-post evaluation is presented in Section 5. The adequate communication started at ARES 2015 and is further continued with this extended article in the EURASIP journal (6).

The remainder of the paper is structured as follows. In Section 2, an overview of related work is presented, and Section 3 gives a conceptual overview of current IAM systems and introduces our proposed improvement. Section 4 introduces the DPMP, while the use case based on real-world data from a large industrial company is presented in Section 5. Section 6 provides a summary and outlook for future work.

2 Related work

A large amount of research considering technological components of IAM systems and their implementation (e.g. [5, 9]), as well as their underlying access control models has been carried out [10]. However, while the importance of IAM policies in general [5] and of organizational policies in particular [11] has been acknowledged, hardly any work specifically considers the challenge of policy detection and management in large and complex environments.

In the field of policy management, researchers have proposed a variety of top-down and bottom-up policy detection approaches. Examples for discovering security policies top-down by extracting information for policy definition from existing business processes are [12–14]. Wolter et al. [12], for instance, use business process models to formulate a set of security policies using the eXtensible Access Control Markup Language. Similarly, [13] convert results from business process execution language-based processes into an role-based access control (RBAC) state [15]. Bhatti [16] specifically focus on the detection of security policies, such as separation of duty (SOD) policies. However, SOD policies only represent a small portion of the policies required in IAM systems. Bailey et al. [17] introduce a self-adaptive framework that monitors authorizations made by role- or attribute-based systems, analyzes user behavior and adapts the target systems accordingly. However, like other approaches, they focus on the detection of security policies rather than providing a guided process for comprehensive policy management in company-wide user management.

Besides the top-down approaches, several researchers have proposed bottom-up policy mining techniques [18–20]. In [20], for instance, security policies are derived from firewall and network information. Besides general policy mining approaches, the research community recently focused on mining attribute policies for attribute-based access control [21, 22] in order to ease the migration from traditional access control models such as RBAC [18, 23]. While being valuable as a technological solution, these approaches do not, among others, consider business semantics or context information required in the context of IAM to validate the correctness of suggested policies.

Additionally, these approaches focus on policy mining based on static input data. Yet, within the context of IAM, we aim to establish policy mining which uses dynamic input and thereby reduces the need for permanent policy adjustment. Due to the amount and heterogeneity of identity data, key indicators are necessary to abstract from the overall complexity and generate information about the current quality and state of the underlying IAM system.

While mining technologies are capable of finding any information within a certain set of data, this may lead to unusable output due to the improper input (“garbage in, garbage out”). By using KPIs, it is possible extract understandable and processable data out of static identity information and thus support adaptability of policies without having to change the policy itself. To our best knowledge, this part is missing in IAM policy management although it creates significant business value. Until now, IAM research mainly focus on key performance indicators for business decision support systems [24, 25]. These approaches evaluate the strategic and economic value of IAM within an enterprise and thereby compare potential benefits (e.g. reduced administration efforts or security benefits) to emerging efforts (implementation costs or operational risks). Further research aims at measuring the performance of IAM processes [26] regarding their maturity level or quality and coverage of processes.

Summing up, available bottom-up and top-down approaches mainly focus on policy detection and do not provide the structured guidance organizations requirements: (1) to implement policy discovery and recommendation mechanisms and (2) ongoing policy maintenance in IAM environments. They do not consider the integration of available context data or KPIs, decide upon the value of certain information for policy detection, or show how to transfer detected policies into daily operation. We argue that a comprehensive process model is required for structuring policy management in a company-wide IAM. Due to the complexity of IAM systems, missing support for human decision-makers reduces applicability in practical scenarios, essentially limiting the benefit of centralized user management.

3 Conceptual overview

In the following, an overview of IAM systems and their main components is provided. On this basis, we propose the extension of current IAM infrastructures using a policy mining engine for improved policy detection and recommendation. Section 4 consecutively introduces the dynamic policy management process facilitating the capabilities of the newly introduced policy mining engine throughout its structured approach for policy handling.

3.1 Identity and access management components

Typical IAM systems consist of three fundamental components (Fig. 1): IAM *data* stored in the infrastructure, tool-supported *functionalities* for executing and

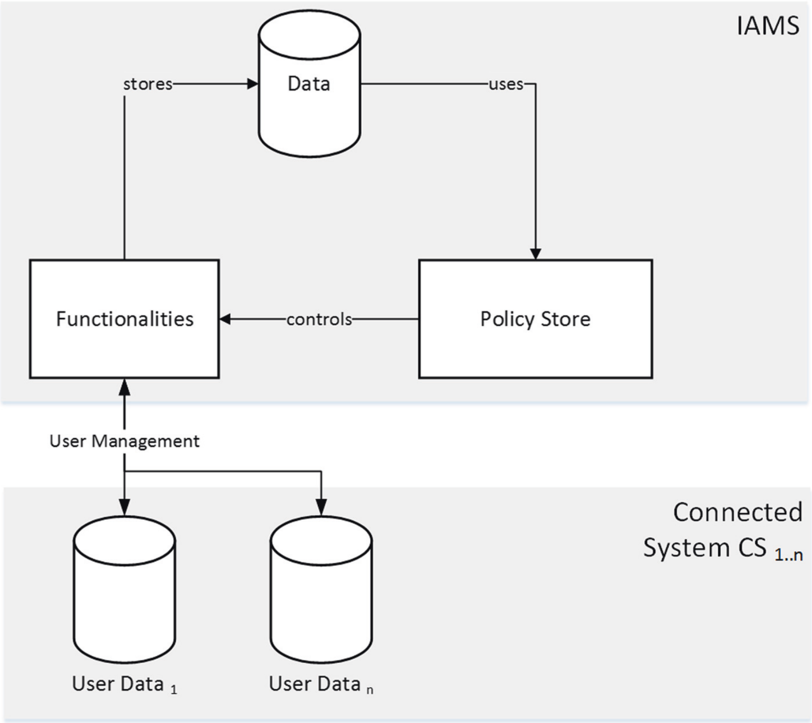


Fig. 1 Basic architecture of identity and access management systems

automating user management tasks and *policies* structuring the management of the overall IAM system itself [27].

3.1.1 IAM data

The required data within the IAM system is commonly periodically loaded from connected applications. Those can be enterprise applications having a dedicated user administration (such as the Microsoft Active Directory or SAP Enterprise-Resource-Planning (SAP ERP) systems). They, however, also can represent resources hosted by partner companies (i.e. using identity federations) or cloud-based resources. Table 1 gives a general overview of existing and used data types. Typically, one or several personnel data systems (HR system) provide employee data such as an employee's name, departmental assignment and further attributes like his or her cost center or location. At the same time, other applications provide user account information such as account identifiers and entitlement information like access privileges and related attributes (e.g. owner or description).

The IAM data coming from the various sources is linked and stored in a central database, creating new data types for a global view on identities (e.g. combining an employee's master data with his or her application-specific user accounts) and entitlements (such as business roles that group access privileges from connected applications). Both the connector technology as well as the data handling mechanisms rely on policies, e.g. for structuring the frequency of data synchronization or data correlation mechanisms.

3.1.2 Functionalities

IAM functionalities implement the logic required to operate the system and provide automated services. This includes modules for user management, access management, data handling and synchronization, or user provisioning [5, 9]. User management is concerned with managing the identity lifecycle, whereas access management provides functionality to authenticate and authorize users. Data handling and synchronization deal with integrating information from applications and exchanging data in a consistent manner. Finally, user provisioning is concerned with the allocation and revocation of user accounts and access privileges or business roles. All of these functionalities require the existence of policies guiding their mode of operation.

The last column of Table 1 underlines that current IAM systems commonly operate on the basis of information on the subject (like employee data), the object (like access privileges and applications) and the assignments between both. Thus, they are only able to process a limited static view without considering extended contextual information like an employee's activities within certain applications. In fact, most applications generate a huge amount of (audit) data such as information about a requesting entity, the affected resources, the location of access, the time and the decision of whether the request was granted or denied. Beside that, static information like assigned permissions, information about the access model or the history of an employee provides a relevant data source. Based on this source, key performance indicators may be computed e.g. for criticality or data quality which adapt the system's current state thus providing better policy input. Additionally, data such as an employee's contract status stemming from an HR system might further support policy management. We argue that considering these extended data types allows for the improved detection of access management policies.

3.1.3 Policies

Policies are used in order to define the behavior of a (software) system by using a dynamic parametrization [28]. Thus, both data and functionalities of IAM systems rely on policies for guiding their mode of operation. Among others, this has already been shown by [28] and [11], who provide an overview of various policy types and their distinct sectors of applicability. Strembeck [28] introduces three types of policies, namely authorization policies, obligation policies and delegation policies. Similarly, [11]

Table 1 Data generated within current IAMS

System	Data type	Examples	Used
HR system$_{1..n}$	Employee master data	Name, personnel number	x
	Employee context	Login state of an employee regarding different applications, vacation, criticality of entitlements	
Application$_{1..n}$	Account information	Account identifiers, account attributes (e.g. system accounts, privileged accounts)	x
	Entitlement information	Entitlement identifiers, entitlement attributes (e.g. critical entitlements)	x
	Account activity	Permission activations, activation sequences, type of permission usage, requested resources	
IAMS	Identity information	Accounts, corresponding systems, entitlements, roles	x
	Entitlement/role information	Corresponding systems, attributes	x
	Provisioning information	Requesting entities, affected resources, approving authorities, decisions	x

categorize policies into process policies, IAM policies and security policies.

The focus of authorization policies is to manage access to an object [28]. This type of policy regulates access to resources within a company and aims at increasing the security of company information and access to sensitive resources. For example, a depiction of the rule that only managers can view top-secret files falls into this category. Delegation policies are a specific set of authorization policies that allow a subject to transfer the decision-making tasks to other subjects.

Obligation policies can be divided into process policies and IAM policies. IAM policies are responsible for the design and governance of the functionality of an IAM system, whereas process policies refer to rules that describe how core business processes within organizations are executed. Examples for IAM policies are the organization's guidelines on access privilege re-certifications or provisioning policies that are used to automatically grant access to a set of resources when new employees join the company. Process policies, on the contrary, describe which permissions typically are activated together or sequentially in order to execute complete process activities.

Within the context of IAM policy management, we suggest a more application-oriented classification of policies namely explicit and implicit policies. Explicit (what can be defined as "precisely and clearly expressed or readily observable") polices are enforced by the underlying IAM system. Consequently, they cannot be bypassed by users and include a detailed definition (e.g. a script, code, rule). By default, common IAM systems already provide a broad range of implementable policies, yet these are mostly of a technical nature (like synchronization modes concerning connected applications or data storage). In order to implement more specific policies, the system itself must be customized or extended. As this is costly and requires deep technical skills, those policies are hardly changed as soon as they are implemented. Explicit policies can be categorized into security or authorization policies (actions a user is allowed to execute) which are commonly implemented in some form of access control matrix and process policies (actions which involve further interaction if the user is not directly authorized to achieve a desired result).

On the other hand, we introduce implicit (what can be defined as "implied though not directly expressed") policies which are not enforced by the IAM system itself (e.g. due to lack of suitable technical means or disproportional implementation effort). Thus, they can initially be expressed in various ways (e.g. a memo or within a dialog). Those implicit IAM policies are generally enforced by a set of stringent decisions made by operators during the lifetime of the IAM system.

Despite its importance, our experience from industry projects shows that policy management and maintenance are only rudimentary realized in practical scenarios. Policies implemented during the setup phase of an IAM system outdate over time as no technological tools or organizational guidance are available for verifying them periodically or detecting newly required policies. Defined policies are rather coarse-grained and simple. The input attributes are generally static, what can partly be attributed to the lack of available (contextual) data to identify complex polices. Another reason is the human IAM engineer's lack of understanding of how and for what applications are used by employees as well as the absence of dynamically generated data in order to allow certain policies to adapt to changes without having to change the policy itself. Additionally, scripting languages are often used to store policies. Hence, only a technically experienced personnel is able to create and refine them due to missing user interfaces.

3.2 Proposed policy management extension

In order to overcome the identified shortcomings, Fig. 2 depicts our proposed improvement. Firstly, we suggest the facilitation of currently unused contextual data for policy management. Secondly, we propose an approach to calculate policy-relevant dynamic information based on static identity data in order to improve adaptability. Thirdly, we extend policy management capabilities of IAM systems with a policy mining engine that is able to consider this contextual data during the automated detection and refinement of policies according to a structured process model (presented in Section 4).

3.2.1 Context data

According to Dey, "context is any information that can be used to characterize the situation of an entity" [29]. In today's IAM systems, almost exclusively identity and entitlement attributes are used as context data for policy decisions. Following [30], we differentiate between five types of additional context elements available in applications.

- *Activity*: Frequency and count of privilege activations as well as the amount of application data accessed.
- *Individuality*: Attributes about employees, user accounts, or access privileges data commonly available within applications (e.g. department or other attributes).
- *Relations*: Activity of similar or related employees, whereas similarity can be based on employee attributes or access privilege usage patterns.
- *Location*: The employee's location from which an activity originated. Technically, IP addresses (internal, external, VPN) are often used in this respect.
- *Time*: The date and time when a permission activation occurred, e.g. within common office hours or at night.

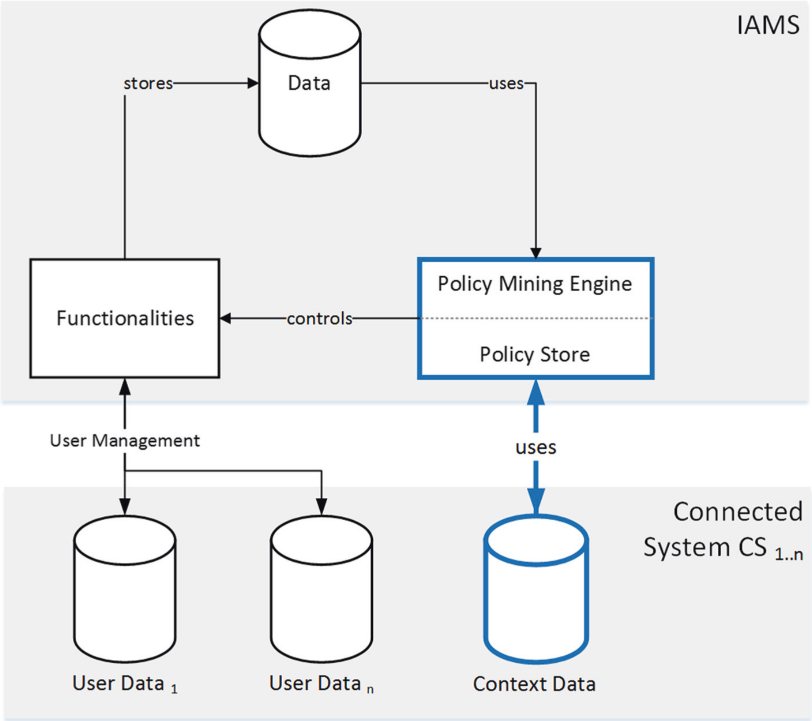


Fig. 2 Advanced architecture of identity and access management systems

3.2.2 IAM key performance indicators

The number of assignments managed by an IAM system may significantly increase over years [2]. For instance, during our evaluation (cf. Section 5.4), we analyzed an SAP ERP system with more than one million assignments of single roles to SAP user accounts, resulting in more than 36 million objects for authorization (transactions, activities, etc.). Even when such systems are carefully managed, it is hardly possible to have a detailed knowledge about every user and all of his possibilities to interact with the system based on assigned permissions. This is only one example of the growing size and complexity of modern IAM systems. While the raw data itself already is hard to comprehend due to its volume, the relations within such data are even harder to perceive. However, we argue that integrating data from the various context types explained in Section 3.2.1 can lead to a better understanding concerning the occurrence of security incidents. Due to the load of IAM data, such connections between data types need to be established in an automated way. Through detailed inspection of the integrated data, so called IAM KPIs may be defined acting as thresholds for normal behavior. Consider an example where the chief financial officer of a company is analyzing his company's net value statistics. While it is perfectly normal for him to regularly check this information, such re-occurring usage patterns integrated with the time and location of access can be good indicators for regular behavior. If such a predefined KPI reaches a value tuple that is outside of its previously common boundaries, either at runtime or ex-post, measures can be taken in order to justify abnormal behavior. Another simple KPI example are significant behavioral or entitlement changes of an employee, where there might be several reasons. Including events of the employee's history into the KPI may result in better observations about possible reasons for the changes (e.g. switched department or position, warnings) in order to generate high-quality security notifications. Thus, automatically generated information out of static data may provide an enhanced view on company events and therefore enable better automated decisions.

3.2.3 Improved policy management

To extend the policy functionality of today's IAM system, we introduce a new policy mining engine which gathers, processes and stores static and contextual data (as defined in Section 3.1.1) in order to discover KPIs and existing but not documented policies. Additionally, after monitoring and validating employees' activities for a sufficient period of time, it is able to recommend the refinement of existing policies according to the previously defined KPIs. As an example, access patterns of employees across applications can be monitored and policies for resource access can consecutively be refined based on actual usage statistics, usage times or the criticality of access privileges.

4 Dynamic policy management process

To implement our research of an improved policy management in IAM systems in complex IAM environments, a structured process model is mandatory in order to ensure applicability. In the following section, we thus introduce the dynamic policy management process supporting organizations during their policy management activities (see Fig. 3). It consists of four phases that structure the activities required for policy management.

At first, the infrastructural setup of the policy management component within the IAM system takes place (phase 1). Input data sources are identified, and policy mining mechanisms are parametrized accordingly. Consecutively, the collection of input data is carried out (phase 2). This comprises activities like data loading, data normalization and data linking required as input data might vary regarding its currency, accuracy or provided attribute dimensions. During phase 3, the data correlation and policy mining takes place in order to differentiate between normal and outlier behavior patterns hinting at potential policy definitions and policy violations. Throughout the last step (phase 4), the results are validated and presented to human IAM engineers facilitating their organizational expertise in order to model well-designed policies.

Note that phases 2–4 of the DPMP are commonly executed in a cyclic manner while the first phase must be reentered in case the system landscape changes or other strategic changes require adjustment.

The main characteristics of the DPMP are:

- Minimizing efforts to define an initial set of policies.
- Improve the quality and adaptibility of input parameters of policies.
- Providing tool support to enable human IAM engineers to execute policy modelling and refinement.
- Integrating both actual authorization usage data and business knowledge.
- Improving IT security through continuous refinement of policies based on actual employee behavior.

4.1 Infrastructure setup

Phase 1 of the DPMP is concerned with the overall pre-configuration of the infrastructure, identifying and setting up data sources, and configuring system behavior regarding policy detection and policy recommendation.

4.1.1 Data identification and connection

Prior to the actual policy mining, available sources for contextual data need to be identified. Typical data sources are applications connected to the IAM system which store contextual data in log files. Human experts (e.g. the system administrators and IAM engineers) need to decide which contextual information from a particular application should be facilitated based on the expected business value, e.g. the potential workload reduction for user management by defining new authorization policies. For the purpose of improving the provisioning processes, for instance, the number of permission activations, the time or location (e.g. in-house, through VPN, the originating country) for each application might be of relevance.

Note that this step heavily depends on the accessibility of data and their potentially temporal availability. While data from centralized applications like SAP ERP systems might be easily accessible, contextual information collection from distributed environments (like file servers in a globally operating organization) might be cumbersome. For our approach, not all data need to be synchronized but only these within the scope of policy detection.

After the identification of available contextual information, the data connection settings need to be adjusted. The goal is an automated data synchronization based on existing connectors as well as additional application connectors (e.g. in case required contextual information stems from a system not yet connected to the IAM system). Setting up the data synchronization also includes the mapping of data from applications to the entities stored in the IAM system. Contextual data such as user account activity, for instance, needs to be related to the respective user accounts and employees.

4.1.2 Policy mining settings

After successful data selection and import, the respective data analysis configuration needs to take place. This includes the weighting of input data for automated data analysis and identification of attributes relevant for key

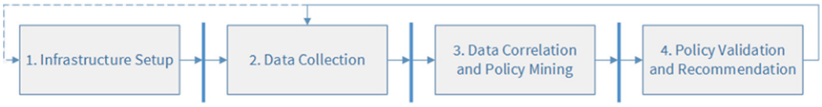

1. Infrastructure Setup
2. Data Collection
3. Data Correlation and Policy Mining
4. Policy Validation and Recommendation

Fig. 3 Proposed policy optimization process model

indicators, as well as settings regarding the system's policy recommendation behavior. Regarding the input data weighting, human IAM engineers could e.g. decide to give more weight to data values that are constantly updated, maintained and revised and thus have a high accuracy during the consecutive algorithmic analysis.

In order to provide a better understanding and provide additional input parameters, the static access model is evaluated concerning criticalities of assignments. This is achieved using data mining techniques. Using a set of defined parameters, the calculation may be calibrated and reviewed by an human IAM engineer in order to minimize false-positives and improve confidence about the data.

Additionally, the methods of policy recommendation can be parametrized according to a given organizational scenario. Similar to approaches used for the cleansing of static access privilege assignments, for example presented in [31, 32], the DPMP requires human expert interaction after the detection of potentially reasonable policies. In case the system suggests an unreasonable large number of new policies potentially including a high rate of false-positives (detected policy suggestions which are discarded after human review), it would add an additional burden rather than create value for an organization. As a result, the system's data mining techniques need to be parametrized in order to only suggest policy definitions for selected behavior patterns.

These settings commonly require the initial analysis of input data over a reasonable period of time. Imagine the correlation of access privileges usage with employees' location data. In case the investigated privilege is only used by employees from a specific location during the period of investigation, the DPMP might recommend the definition of a provisioning policy that only assigns employees from this location to the according access privilege. If the period of investigation has been set too short, employees from other locations might also request the usage of this access privilege, essentially requiring the adaption of the defined policy.

4.2 Data collection

After successful setup of the policy management system, the data collection phase takes place. During this step, the input data is loaded, normalized and linked according to the previously defined settings. The goal is a periodic and fully automated data loading process shifting from a manual administration to an automatic machine-based execution. As a result, the latest input data are available for the automated analysis at any point in time for policy management without the need for further human interaction.

In a first step, the raw data from the relevant applications is imported and normalized. Systems which create a constant data stream require a continuous import, conversion and storage of data while other applications might only support a full data export (e.g. using the CSV file format). Furthermore, data storage types might vary among applications, requiring data normalization. Examples are an ERP application providing usage data aggregated per single day and the amount of data accessed by clients in megabytes while a file service application delivers a steady stream of data and the amount of data sent to clients in bytes. During a last data collection activity, relationships among data elements stemming from different points of time are set up. For each employee, access privilege or business role, a change history is generated. This e.g. allows for the detection of activity patterns fostering the identification of user provisioning policies.

4.3 Data correlation and policy mining

During the data correlation phase, the automated policy mining takes place. The goal is to generate recommendations for relevant policies which have not been implemented up to now. At the same time, already established policies are validated for adjustment. In this paper, it is not our goal to provide a comprehensive list of pattern detection techniques but rather aim at showing that those techniques can be applied to support policy management efforts in general. For evaluation purposes, we implemented a set of analysis techniques (see Section 5). These techniques are designed as depicted in Fig. 4.

The DPMP facilitates existing data mining technologies (e.g. clustering [33] or neural networks [34]) on the basis of existing identity information, contextual data and the various data dimensions defined during the initial setup phase (see Fig. 4). Patterns of normal and outlier behavior are automatically extracted for investigated subjects. The subject may either be a single entity or a group of entities which can be uniquely identified by a set of attributes within the context of a policy. Such an entity can be an employee, a user account within an application or a role bundling access privilege from different applications. Such data are augmented by their contextual data generated, for instance, when an entity is involved in any kind of activity.

Data mining allows for a multi-dimensional analysis facilitating sets of relevant attributes of subjects (e.g. employees, user accounts, or entitlements) and objects (e.g. amount, frequency, or criticality of data accessed). The overall goal is to identify clusters of subjects that share contextual data patterns which might in turn lead to the definition of IAM policies and the detection of outliers violating the policy.

Imagine an organization that aims at ensuring the principle of least privilege [35] in order to minimize insider misuse by overprivileged employees. Employees only are allowed to have the minimum set of access privileges required by their daily work. The DPMP in this respect

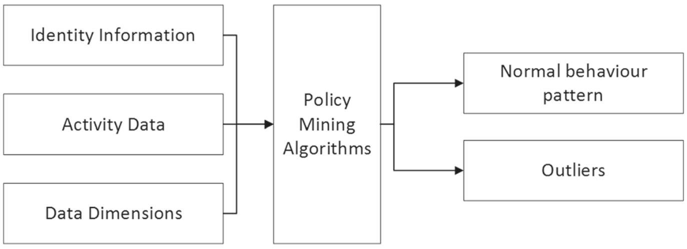


Fig. 4 Input and output of policy mining algorithms

continuously monitors existing user provisioning policies by identifying outdated access privilege assignments based on users' behavior. The example in Fig. 5 depicts the analysis of a privilege providing access to billing data within the company based on employee's location ("New York") and department ("finance"). The current provisioning policy might be refined after automated usage pattern detection identified that only employees which are assigned to the job function "clerk" actually use the respective access privileges (independent of their assigned location) while "secretaries" within the finance department in New York do not activate the access privilege at all.

Examples for the detection of anomalies (in contrast to standard usage patterns) might include entitlements for accessing financial data being activated from a VPN connection (while an according policy forbids this access) or access privileges which are used to manipulate an extraordinary amount of data.

Our approach distinguishes between the three policy types, namely security policies, process policies and IAM policies.

Every mining process is divided into three steps, namely "data construction", "data analysis" and "contextual evaluation", whereas the data analysis may go hand in hand with the contextual evaluation. During the first step, we create a suitable data structure based on availability of data and selected and weighted dimensions as well as additional information (e.g. KPIs). After that, the analysis of the data takes place. The outcome is further characterized by using available contextual data concerning involved entities. The presented approaches do not aim to present novel algorithms for the presented problems but to foster already established techniques and adjust these based on the requirements of IAM systems and their policy management components.

4.3.1 *Security policy*

Mining of security or authorization policies based on an available static access control matrix has been within the focus of research for some time (e.g. [18, 36]) especially since standards like ABAC [37] are heavily dependent on this construct, like this aim to extend these techniques

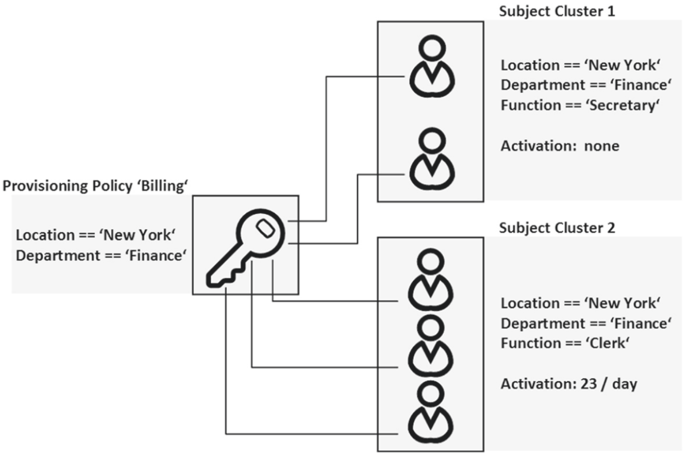


Fig. 5 Example of access privilege activation analysis

by enhancing the input data and the characterizations of output data using contextual information.

The first step is to analyze the existing access control matrix based on semantic analysis techniques and usage statistics. Semantic analysis [31, 32] introduce a classification for every assignment based on the examination in context of other entities within the given scope. A simple example would be if only one employee within the "development" department owns the entitlement "access marketing share". As a result, this permission assignment might probably be invalid and should be revoked. We use these techniques to sort out potentially invalid assignments which create noise during policy mining. Similarly, we identify and exclude unused entitlements (with an adjustable period of time) and recommend manual review by human IAM engineers.

For the authorization policy mining, we facilitate available algorithms (like those proposed at [16, 18, 21]). These algorithms operate on different types of data, e.g. log files, roles or user permission assignments, and can be adjusted according to the available access control model and available data sources.

During the last step, we analyze every mined policy according to its usage profiles. This includes attributes like location, time, consumed CPU resources or the amount of data read or written. These profiles are analyzed using classification techniques (e.g. [33]) in order to reveal normal and abnormal utilization behavior. Consider the example of an entitlement to modify data within an application. The amount of data typically modified during normal manual operation can be classified e.g. financial data are usually not modified throughout activities creating gigabytes of traffic. This enables human IAM engineers to evaluate usage profiles of entitlements or authorization policies according to compliant operation of applications.

4.3.2 Process policy

Process policies represent a subset of obligation policies. They define constraints for actions within a process which need be carried out in order to achieve a desired business value of which a user is not directly authorized to. Within context of IAM, for example, this could be the obtaining of an approval to assign a specific access right or a re-certification. In order to identify process policies, we aim to identify events which trigger such processes as well as necessary checkpoints or nodes (e.g. the head of department's approval) which need to be passed in order to achieve a desired result.

For the transformation of activities into structured processes, we use business process mining (BPM) techniques (as proposed in [38]). For data construction, we use the concept of trace clustering which divides activity logs into traceable clusters or in other words single process iterations. In the context of IAM, this could be a ticket number or a process id in relation with a specific request type. This information are subsequently mapped into a process representation (e.g. BPMN) for further analysis. Information about processes could theoretically be extracted directly from a business process management system. Yet, the goal is to create a detailed overview about the status quo within the IAM system (independent of how the system should actually be used) and possibly create a comparison to already defined processes.

During the next step, we try to derive a detailed characterization of every process node including decision-making entities as well as the respective context of the decision. We aim to identify similarities between the decisions (e.g. every re-certification was done during business hours from within the IT building by an employee within the same department and attribute "head of department") as well as outliers (e.g. every approval of this entitlement was done by an employee with the attribute "entitlement owner" during business hours, while one approval was done at 22:00 pm by an admin account). In order to achieve this, we firstly have to generalize information as far as available. For example regarding the time of actions, we may distinct between business hours and closing time, location between off- and onsite, devices could be separated into business owned devices, private devices and hybrid usage. Using this compression, we are able to reduce the number of possible definitions for each process node. After that, we classify affected entities per process step using different, individually weighted attribute permutations as input.

During step three, we create an extended process definition. For every process step, the created classifications are analyzed. If a classification which includes every entity who finished the respective step is available, its attribute combination is used as a possible definition for the step. If no suitable cluster can be identified, all clusters are used as possible definitions and thereby an extended manual review is necessary.

4.3.3 Implicit IAM policies

As mentioned above, IAM policies are responsible for the design and compliant operation of IAM systems. Yet by far not every IAM policy is technically implemented. The IT of today's companies is forced to adapt business changes and support the business in an optimal manner. Thus, directly customizing the IAM system according to every business change is hardly executed in practice because of the resulting efforts. Due to these reasons, we do not aim to exhaustively mine all possible IAM polices which are currently in place and technically enforce them as it would impose rigid restrictions to the system and potentially have a negative impact on business processes. However, we propose a mechanism which allows the system to learn current IAM policies and create recommendations regarding system operation. Consider the example of an international help desk who is in charge of processing orders regarding the IAM system (e.g. the request for assignment of an entitlement or access to a specific application) with all requests requiring manual review. By generating recommendations out of previous decisions, approvers can be supported during the review process.

Our first step is to classify users of the system according to their weighted attributes (e.g. [33]). This enables us to derive an overview of which types of users are managed by the system. If applicable, there may be several different classifications if the company is structured in a complex way. After that, we derive a set of actions (e.g. requests, authentications) carried out by those types of users. These partitioned user sets combined with affiliated contextual data and the history of activities allows us to use a context-aware recommender system (e.g. [39, 40]). Standard recommender systems in general aim at providing suggestions for items which are considered to be useful for a user [40], whereas context-aware recommender systems operate on tuples in the form of $< user, item, context, rating >$. In the context of IAM, users are the company's employees and items represent achievable resources (e.g. entitlements or files). As soon as the calculated rating extends the defined threshold, a positive recommendation is given, otherwise a negative one. As context-based frameworks for recommender systems take attributes like activity, location, user information and related resources into consideration [40], they consequently meet the requirements of an IAM policy recommendation system.

4.4 Policy validation and recommendation

After successful data correlation and policy mining, a set of potentially relevant policies (e.g. provisioning policies changing the current access control state) has been identified. As IAM systems and connected applications manage a huge amount of data, a high number of potentially relevant policies might be detected by each DPMP iteration. These policy candidates need to be validated by the policy management system before being communicated in an appropriate manner to human IAM engineers for refinement. Policy validation thus can observe the underlying rule for every detected policy over a certain period in time before it is recommended to a human IAM engineer. In case a policy suggestion is based on usage activity patterns, for instance, these patterns can be validated over a period of 1 month. In case the pattern changes during the investigation period, the policy suggestion itself can be revoked.

Policy mining is limited to generating a set of policy suggestions based on classifications of subjects together with their behavior based on contextual data and history. As a result, the focus during the last phase of the DPMP lies on the presentation of results in an intuitive and human-understandable way in order to enable the IAM engineer to easily derive appropriate actions. Visualizations can be based on techniques like charts or data tables. From our practical experience, it is essential to include the visualization of the reasons why a certain policy suggestion has been created. In case ambiguous or mutually exclusive rules have been identified, this information has to be included in the result presentation as well. A human IAM engineer might, for instance, be informed that accepting one policy suggestion might lead to the violation of another already implemented policy. He then might be able to decide whether the old policy is outdated while the new policy suggestion should be activated.

Again, it is not the goal of this paper to provide a comprehensive list of potential visualizations or rule definition scenarios but rather underline the importance of a dedicated result refinement phase including human interaction as a cornerstone of ongoing policy management in IAM.

In this section, we proposed the dynamic policy management process which enables organizations to gather a deeper understanding of its IAM, the (contextual) data and the quality of currently implemented policies as well as potential policy suggestions. Based on company-specific settings, the DPMP is able to import the necessary input data, identify patterns of standard subject behavior and support human IAM engineers during policy definition and refinement.

5 Evaluation

In this section, we evaluate the applicability of the DPMP in a real-world scenario. The evaluation is based on data stemming from the SAP ERP system and the IAM system of a globally operating manufacturing company with more than 12,000 internal and 4000 external employees. A total number of 8021 active user accounts, 3925 single roles, 762 composite roles and 1,180,962 access privilege assignments from the SAP ERP system were initially imported and anonymized. For the following evaluation, the period under observation comprised 5 weeks during which daily re-imports took place.

Increasing audit requirements force the company to improve IAM policy management. Up to now, only rudimentary provisioning and access re-certification policies have been defined due to missing tool support and knowledge about the underlying data. As a result, a policy detection project has been initiated. Its main goals are:

1. The consideration of contextual data and KPI definition from the SAP ERP system for policy generation
2. The semi-automated detection of new and potentially relevant provisioning and re-certification

policies as well as the identification of loosely defined and hence insecure existing policies

3. Providing appropriate visualizations of detected policies to support human IAM engineers

While requirement (1) corresponds to phase 1 and 2 of the DPMP, (2) relates to its data correlation and policy mining phase (phase 3). Requirement (3) deals with the presentation of discovered policies according to phase 4 of the DPMP. Even though we executed numerous policy detection activities, we focus on two specific examples for evaluation purposes in the remainder. Firstly, the analysis of access privilege activations has been compared to the static distribution generated by the current provisioning policy in the IAM system (corresponding to phase 2 of the DPMP, see Section 5.3). Secondly, detected access privilege activation frequencies were visualized in relation to the amount of data objects modified (i.e. data within the SAP ERP system) for investigation by a human IAM engineer (corresponding to phases 2 and 3 of the DPMP, see Sections 5.3 and 5.4).

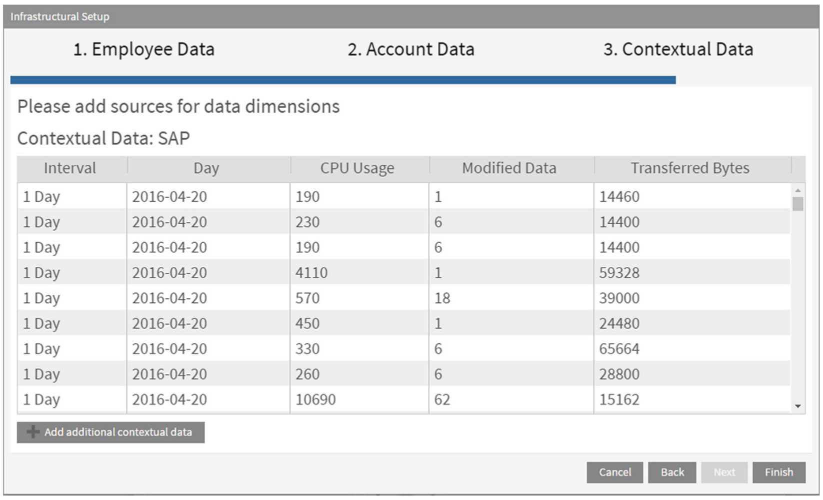
Note that a comparative evaluation of our prototype-based approach with manually executing policy detection and recommendation cannot be executed. This is due to the inapplicability of a manual examination of the several hundred thousands of access privilege assignments and the available large amount of contextual information.

5.1 Infrastructure setup

To address requirement (1), at first, context data available in the SAP ERP system was analyzed (step 1 of phase 1). Using the classification technique for context data from Section 3.2.1, the following information on user behavior was extracted and mapped: number and frequency of read and write permission usage and amount of transferred data (activity) for each account (individuality) per day (time) and the corresponding IP address (location). Subsequently, policy mining parametrization was conducted (step 2 of phase 1). Initially, the set of prototypical implemented algorithms (including data classification mechanisms and statistical distribution analysis) were applied using a default configuration. On this basis, distinctive properties of the imported data set became apparent. For instance, due to SAP ERP system limitations, user behavior can only be extracted on a daily basis. Thus, algorithms need to be configured to identify permission usage irregularities per day (e.g. suspicious permission activations on weekends) but not within the course of a single day (such as off-time activities). Furthermore, data types were weighted in cooperation with a human system expert, emphasizing the importance of data types such as IP address and employee status information during the following analyis.

5.2 Data collection

After successful configuration and parametrization, the data collection took place (phase 2 of the DPMP). We implemented a software wizard to ease the import of raw data types onto the internal data structures of the extended IAM tool (see Fig. 6) in an automated manner. The wizard shows an excerpt of available contextual data which can be extracted from an SAP ERP system. Available contextual data from applications strongly vary (e.g. contextual data for an active directory may be derived from connected share systems). The wizard shows an



Infrastructural Setup

1. Employee Data 2. Account Data 3. Contextual Data

Please add sources for data dimensions

Contextual Data: SAP

Interval	Day	CPU Usage	Modified Data	Transferred Bytes
1 Day	2016-04-20	190	1	14460
1 Day	2016-04-20	230	6	14400
1 Day	2016-04-20	190	6	14400
1 Day	2016-04-20	4110	1	59328
1 Day	2016-04-20	570	18	39000
1 Day	2016-04-20	450	1	24480
1 Day	2016-04-20	330	6	65664
1 Day	2016-04-20	260	6	28800
1 Day	2016-04-20	10690	62	15162

Fig. 6 Infrastructure setup wizard

excerpt of the broad variety of contextual data which can be extracted from applications and used for the policy mining step. Additionally, data from the SAP ERP system was mapped onto existing user management data from the IAM system. SAP user account activities, for instance, were related to the respective employees' identities. As a result, a total number of 6,214,422 records from 36 days containing contextual information as well as user management data from the SAP ERP system were gathered and mapped using our daily data import functionality.

5.3 Data correlation and policy mining

During the third phase of DPMP the actual policy mining was conducted in order to address project requirement (2). Using our implemented policy mining algorithms, we were able to detect standard usage patterns potentially leading to the definition of new policies as well as the refinement of currently implemented policies.

Concerning the first exemplary case, we computed the distribution of static assignments of access privileges among the top level departments of the company and compared these to their actual activation information. Table 2 shows the distribution of access privilege P_1 across top-level departments of the company and its actual activation in these departments. As can be seen, nearly half of the employees that are assigned to P_1 are working in the department D_3. The access privilege is almost exclusively used (99.96 %) in this department, while only a small number of activations (0.04 %) stems from department D_1. This indicates that access privilege P_1 might only be required for tasks conducted in department D_3. Thus, a refinement of the existing provisioning policy that additionally requires employees to work in department D_3 in order to obtain this access privilege is recommended. This restructuring might lead to a reduction of the number of overprivileged employees, thereby strengthening IT security.

In summary, out of the company's total 3925 single roles defined in the SAP ERP system, we identified 382 (i.e. 9.7 %) which—though being assigned to employees in a particular department—were hardly activated (activation frequency for the respective department is below 1 %). In an ongoing effort, these results are discussed with the company's IAM engineer in order to improve existing provisioning policies and refine existing SAP role definitions leading to access privilege revocation.

Table 2 Static distribution and actual use of access privilege P_1

Department	Distribution of static assignment (%)	Activation frequency (%)
D_1	21.15	*0.04*
D_2	24.39	0.00
D_3	45.08	*99.96*
D_4	4.82	0.00
D_5	0.72	0.00
D_6	1.54	0.00
D_7	2.3	0.00

For our definition of KPIs, we intensively examined a criticality value for every employee based on his currently assigned entitlements. An employee was defined as critical as soon as he owned permissions which are not common for his position within the enterprise. The more uncommon an entitlement is (e.g. because he owns multiple times as many permissions as other employees within his departments), the higher the criticality value. By calculating such value for each employee, the policies that are addressing privileged employees can be redefined. In discussions with the company, we agreed on primarily taking the employee's contextual data of permissions into account (in this case his department and other employees provisioned with similar access rights). For this effort, we used a set of parametrized anomaly detection algorithms for outlier detection algorithms [41] for the criticality determination of user permission assignments. Our applied algorithms measure the distribution of permission assignments among a defined set of users. We use different sets of input parameters ranging from a high to a low detection rate. The lower the detection rate, the higher the criticality value of the assignment. A simple example for such an algorithm would be to detect all entitlements assigned to not more than a certain percentage of employees. For each of our algorithms, we created a set of parameters varying from a severe to a slack detection rate. According to the quantity of re-occurrences of results throughout different parameter sets, we categorized the assignments of the employees on a Likert scale. This ranges from *uncritical* to *very critical*, thus providing an easy to understand overview about the current access model. Consider the following simplified example demonstrating the functionality: an employee from the finance department is switching positions within the company and is now working for the marketing department. Due to lack of correct revocation policies, the employee retains some of his old permissions. In our approach, each of these permission assignments would be detected by each run of the algorithms with all of the previously calibrated parameter sets. As even the strictest parameter set (i.e. the set that tolerates least errors in the data) together with all others flags this assignment as critical, its overall criticality is set to *very critical*. The resulting distribution for the assessment of all of the company's assignments is depicted in Table 3.

Fostering these results enables the system to automatically classify each employee concerning his criticality. For the employee categorization, we followed the maximum principle according to the BSI Grundschutz [42] which

Table 3 Criticality of assignments based on static analysis

Criticality	Number of assignments	Percentage (%)
Uncritical	1,087,939	92.12
Very low criticality	34,494	2.92
Low criticality	31,888	2.70
Critical	13,792	1.17
Very critical	12,851	1.09

states that the security level of an object should be as high as the highest of its associated resources (in our case the most critical user permission assignment):

$$\text{Criticality}_{\text{employee}} = \text{Max}\ \left(\text{criticality}_{\text{UPA}}\right).$$

The results of our criticality calculations are as follows. The data depict a very even distribution of assignments among the overall company having only 1.08 % assignments with a *high criticality* level. Nevertheless, there are 12,851 assignments affected, which may impose severe security risks upon the enterprise (according to our calculation that we established in conformance with the company). Concerned were a total number of 428 highly critical user accounts. These data can now be used in order to enhance affected policies for user management e.g. by employing more frequent re-certifications ([43]) or four-eye prinicples.

In order to address project requirement (3), we aimed to derive usage patterns of permissions based on contextual data. As mentioned above, the system currently manages over one million permission assignments (an average of about 147 assignments per user) with more than 30 million assignments possible. Therefore, we aim to qualify permission assignments based on activities carried out by users. Single roles usually conform with a well-defined set of actions which they enable a user to execute. Consequently, these actions generate a similar fingerprint concerning usage profiles (e.g. data modified or read) within a predefined period of time. We aim to identify these fingerprints by applying our security policy mechanisms. The first step is to set up a suitable data set for the analytics which consists of all user permission assignments (UPA) and a set of contextual data concerning permission activations aggregated per day (see Fig. 6). We combined these data to a set of vectors in the form of:

$$V_{\text{activity}} = \{\text{UPA}, \text{datamodified}, \text{dataread}\}.$$

For these data, we created clusters using [44]. Accordingly, we are able to classify the usage of every permission assignment on a daily level. As described above, we computed a criticality value for all user permission assignments. This enables us to combine classified usage profiles (as depicted in 7) with the criticality value of each assignment. As these results are not directly usable by human IAM engineers, suitable visualization techniques have to be applied.

5.4 Policy validation and recommendation

In order to further address project requirement (3), previously detected standard usage patterns need to be validated and visualized for human refinement. Figure 7 depicts a screenshot from our extended IAM tool which uses a bubble chart visualization in order to display detected usage patterns. The x-axis corresponds to the amount of data that has been "modified" by an employee's access privilege activations on a single day, while the y-axis denotes the amount of data being "read". During phase 1 of the DPMP, thresholds were defined for highlighting power users, i.e. employees which either read or modify large amounts of data within the SAP ERP system. In the given example, an employee has been marked as a power user (orange colored highlighting) if he either read more than 10,000 MB or modified more than 200,000 data sets

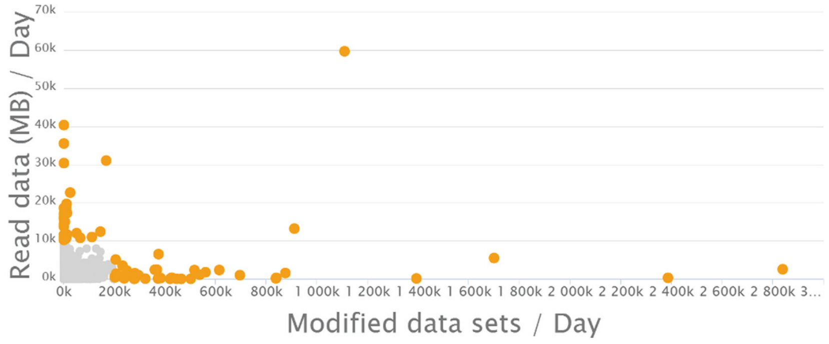


Fig. 7 Detection of SAP power users (http://www.nexis-secure.com)

per day. Bubbles in the lower left area of Fig. 7 correspond to average system users, while highlighted bubbles in the other areas correspond to power users. In the given example, 63 power users were identified for the interval of our investigation. Out of these 63 power users, we identified three users who activated assignments which have been marked critical or very critical during our KPI analysis. A human IAM engineer could use this information for defining a new re-certification policy that demands a periodic assessment of all power users' access privileges. In contrast to standard SAP users whose access privileged are re-certified once a year, power users might be re-certified more frequently in order to reflect their criticality value.

In summary, the evaluation based on data from an SAP ERP system presented in this section of the paper underlined the applicability of the DPMP for structured policy management in practice. Based on the prototypical extension of an existing IAM tool, we were able to import previously unused contextual data, identify clusters of standard as well as outlier usage behavior and visualize the gathered results. Within the company, the results increased management attention by providing in-depth insight into the current access control state and its guiding policies. At our partner's side, efforts for evaluating the application of the DPMP in a periodic manner (daily operation), the extended analysis of further applications, and the adaption of existing IAM policies are currently made.

6 Conclusions

Over the last decades, company-wide IAM systems have become a key element for controlling users' access to resources in medium to large-sized enterprises. They offer means for a centralized enforcement of standardized user management processes and policies. Despite their importance, the management of IAM policies commonly still needs to be executed manually. While current research concentrates on mechanisms for policy detection and enforcement, the complexity of user management in large environments rather requires a structured and applicable process for policy management. Human IAM engineers need to be supported with guidance and automation during the detection, implementation and refinement of IAM policies.

In order to improve the current situation, we presented the dynamic policy management process which structures the activities during policy management into four phases. It facilitates a mining engine which generates policy recommendations based on contextual data of employees and further presents gathered results to human IAM engineers. In order to underline the practical relevance and applicability of our contribution, we conducted a practical case study within a large industrial company and its ERP system managing several thousands of users and more than one million access privileges.

We showed and also experienced from our industrial projects that policy management is an important task within modern IAM architectures as it provides an anchor for the system to work properly and secure in a time where enterprises begin to realize that the traditional castle approach of their IT imposes several risks e.g. due to cloud computing, work anywhere, IoT or Industry 4.0. Due to these developments, IAM will become even more important and therefore need a stable basis to work on.

For future work, we plan to extend the DPMP in order to improve the representation and management of policy recommendations. Practical experience shows that a high amount of potentially conflicting recommendations increases manual efforts of human role engineers and requires an in-depth understanding of the underlying data. In the future, we hence aim at providing an analysis of policy interdependencies in order to overcome this limitation. We additionally aim at extending our prototype implementation and evaluate the DPMP throughout further practical use cases considering contextual data from decentralized applications.

Acknowledgements
The research leading to these results was partly supported by the "Bavarian State Ministry of Education, Science and the Arts" as part of the FORSEC research association.

Authors' contributions
Firstly, we suggest the facilitation of currently unused contextual data for policy management. Secondly, we propose an approach to calculate policy-relevant dynamic information out of static identity data in order to improve adaptibility. Thirdly, we extend policy management capabilities of IAMS with a policy mining engine that is able to consider this contextual data during the automated detection and refinement of policies according to a structured process model. Fourth we demonstrate the feasibility of our approach based on real-life data. The design was carried out by MH, MN and MK. MK fit the article into related work, MN worked on the conceptual overview. MH and MK established the DPMP along with its algorithmic approach. MH and LF carried out the evaluation. GP induced the research which lead to this article. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Received: 11 February 2016 Accepted: 2 August 2016
Published online: 15 August 2016

References

1. A Hovav, R Berger, Tutorial: identity management systems and secured access control. Commun. Assoc. Inf. Syst. **25**(1), 42 (2009)
2. A Cleven, R Winter, in *Enterprise Business-Process and Information Systems Modeling. Lecture Notes in Business Information Processing*, ed. by T Halpin, J Krogstie, S Nurcan, E Proper, R Schmidt, P Soffer, and R Ukor. Regulatory Compliance in Information Systems Research – Literature Analysis and Research Agenda, vol. 29 (Springer, Berlin Heidelberg, 2009), pp. 174–186
3. United States Code, Sarbanes-Oxley Act of 2002, PL 107-204, 116 Stat 745 (2002). https://www.sec.gov/about/laws/soa2002.pdf. Accessed 11 Aug 2016
4. Basel Committee on Banking Supervision, Basel III - A global regulatory framework for more resilient banks and banking systems (2011). https://www.bis.org/publ/bcbs189.pdf. Accessed 11 Aug 2016

5. L Fuchs, G Pernul, in *The Second International Conference on Availability, Reliability and Security, 2007: ARES 2007*. Supporting compliant and secure user handling—a structured approach for in-house identity management (IEEE Computer Society, Los Alamitos, 2007), pp. 374–384
6. L Fuchs, M Kunz, G Pernul, in *European Conference on Information Systems (ECIS)*. Role model optimization for secure role-based identity management, (2014)
7. K Peffers, T Tuunanen, CE Gengler, M Rossi, W Hui, V Virtanen, J Bragge, in *Proceedings of the First International Conference on Design Science Research in Information Systems and Technology (DESRIST 2006)*. The design science research process: a model for producing and presenting information systems research (M. E. Sharpe, Inc., Armonk, 2006), pp. 83–106
8. AR Hevner, ST March, J Park, S Ram, Design science in information systems research. MIS Q. **28**(1), 75–105 (2004)
9. D Royer, in *Proceedings of the IFIP/FIDIS summer school on "The future of identity in the information society*. Enterprise identity management—what's in it for organisations (Springer, Berlin Heidelberg, 2008), pp. 403–416
10. L Fuchs, G Pernul, R Sandhu, Roles in information security—a survey and classification of the research area. Comput. Secur. **30**(8), 748–769 (2011)
11. L Fuchs, G Pernul, Minimizing insider misuse through secure identity management. Secur. Commun. Netw. **5**(8), 847–862 (2012)
12. C Wolter, A Schaad, C Meinel, in *Web Information Systems Engineering–WISE 2007 Workshops*. Deriving XACML policies from business process models (Springer, Berlin Heidelberg, 2007), pp. 142–153
13. J Mendling, M Strembeck, G Stermsek, G Neumann, in *Enabling Technologies: Infrastructure for Collaborative Enterprises, 2004. WET ICE 2004. 13th IEEE International Workshops on*. An approach to extract rbac models from BPel4Ws processes (IEEE Computer Society, Los Alamitos, 2004), pp. 81–86
14. A Baumgrass, S Schefer-Wenzl, M Strembeck, in *IEEE*. Deriving process-related RBAC models from process execution histories, (2012), pp. 421–426
15. RS Sandhu, EJ Coyne, HL Feinstein, CE Youman, Role-based access control models. IEEE Commun. **29**(2), 38–47 (1996). doi:10.1109/2.485845
16. R Bhatti, E Bertino, A Ghafoor, X-federate: a policy engineering framework for federated access management. IEEE Trans. Softw. Eng. **32**(5), 330–346 (2006). doi:10.1109/TSE.2006.49
17. C Bailey, DW Chadwick, R de Lemos, in *IEEE 9th International Symposium on Dependable, Autonomic and Secure Computing*. Self-adaptive authorization framework for policy based RBAC/ABAC Models (IEEE Computer Society, Los Alamitos, 2011), pp. 37–44
18. Z Xu, SD Stoller, in *Data and Applications Security and Privacy XXVIII*. Mining Attribute-Based Access Control Policies from Logs (Springer, Berlin Heidelberg, 2014), pp. 276–291
19. A Baumgrass, in *Availability, Reliability and Security (ARES), 2011 Sixth International Conference On*. Deriving current state RBAC models from event logs (IEEE Computer Society, Los Alamitos, 2011), pp. 667–672
20. H Safaa, C Frédéric, C-B Nora, A Vijay, M Stéphane, in *9th International Conference on Information Systems Security*, ed. by A Bagchi, I Ray. Policy Mining: a Bottom-Up Approach Toward a Model Based Firewall Management (Springer, Berlin Heidelberg, 2013), pp. 133–147
21. J Lopez, R Oppliger, G Pernul, Authentication and authorization infrastructures (AAIs): a comparative survey. Comput. Secur. **23**(7), 578–590 (2004)
22. VC Hu, D Ferraiolo, R Kuhn, A Schnitzer, K Sandlin, R Miller, K Scarfone, Guide to attribute based access control (ABAC) definition and considerations. NIST Spec. Publ. **800**, 162 (2014)
23. Z Xu, SD Stoller, in *Emerging Technologies for a Smarter World (CEWIT), 2013 10th International Conference and Expo on*. Mining attribute-based access control policies from RBAC policies (IEEE Computer Society, Piscataway, 2013), pp. 1–6
24. D-W-ID Royer, M Meints, Enterprise identity management—towards a decision support framework based on the balanced scorecard approach. Bus. Inf. Syst. Eng. **1**(3), 245–253 (2009)
25. MC Mont, Y Beresnevichiene, D Pym, S Shiu, in *Network Operations and Management Symposium Workshops (NOMS Wksps), 2010 IEEE/IFIP*. Economics of identity and access management: providing decision support for investments (IEEE Computer Society, Piscataway, 2010), pp. 134–141
26. D Royer, M Meints, Planung und Bewertung von, Enterprise identity managementsystemen. Datenschutz und Datensicherheit-DuD. **32**(3), 189–193 (2008)
27. J Pato, OC Center, *Identity Management: Setting Context*. (Hewlett-Packard, Cambridge, 2003)
28. M Strembeck, *Engineering of Dynamic Policy-Based Systems: A Policy Engineering of Dynamic Policy-Based Systems: Language Based Approach*. (Habilitation Thesis, WU-Wien, 2008)
29. AK Dey, Understanding and using context. Pers. Ubiquit. Comput. **5**(1), 4–7 (2001)
30. A Zimmermann, A Lorenz, R Oppermann, in *Proceedings of the 6th International and Interdisciplinary Conference on Modeling and Using Context, CONTEXT'07*. An Operational Definition of Context (Springer, Berlin Heidelberg, 2007), pp. 558–571
31. L Fuchs, C Broser, G Pernul, in *Availability, Reliability and Security, 2009. ARES'09. International Conference On*. Different approaches to in-house identity management—justification of an assumption (IEEE Computer Society, Piscataway, 2009), pp. 122–129
32. A Colantonio, R Di Pietro, A Ocello, NV Verde, A new role mining framework to elicit business roles and to mitigate enterprise risk. Decis. Support. Syst. **50**(4), 715–731 (2011)
33. J MacQueen, *et al*, in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Some methods for classification and analysis of multivariate observations, vol. 1, (Oakland, 1967), pp. 281–297
34. T Kohonen, An introduction to neural computing. Neural Netw. **1**(1), 3–16 (1988)
35. JH Saltzer, MD Schroeder, The protection of information in computer systems. Proc. IEEE. **63**(9), 1278–1308 (1975)
36. Z Xu, SD Stoller, Mining attribute-based access control policies. IEEE Trans. Dependable Secure Comput. **12**(5), 533–545 (2015)
37. T Priebe, W Dobmeier, B Muschall, G Pernul, in *Sicherheit 2005: Sicherheit - Schutz und Zuverlässigkeit, Beiträge der 2. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft Für Informatik e.V. (GI), 5.-8. April 2005 in Regensburg*. ABAC - Ein Referenzmodell für attributbasierte Zugriffskontrolle (GI, Bonn, 2005), pp. 285–296
38. L García-Bañuelos, M Dumas, M La Rosa, J De Weerdt, CC Ekanayake, Controlled automated discovery of collections of business process models. Inf. Syst. **46**, 85–101 (2014)
39. K Verbert, N Manouselis, X Ochoa, M Wolpers, H Drachsler, I Bosnic, E Duval, Context-aware recommender systems for learning: a survey and future challenges. IEEE Trans. Learn. Technol. **5**(4), 318–335 (2012)
40. F Ricci, L Rokach, B Shapira, in *Recommender Systems Handbook*. Introduction to recommender systems handbook (Springer, Berlin Heidelberg, 2011), pp. 1–35
41. G Pernul, L Fuchs, Reducing the risk of insider misuse by revising identity management and user account data. J. Wirel. Mob. Netw. Ubiquit. Comput. Dependable Appl (JoWUA). **1**, 14–28 (2010)
42. B für Sicherheit in der Informationstechnik, BSI-Grundschutz Katalog (1996). https://www.bsi.bund.de/EN/Topics/ITGrundschutz/itgrundschutz_node.html. Accessed 11 Aug 2016
43. C Richthammer, M Kunz, J Sänger, M Hummer, G Pernul, *Dynamic Trust-based Recertifications in Identity and Access Management*. (IEEE Computer Society, Piscataway, 2015)
44. JA Hartigan, MA Wong, Algorithm AS 136: A k-means clustering algorithm. J. R. Stat. Soc. Ser. C (Appl. Stat.) **28**(1), 100–108 (1979)

5 Analyzing Context Data for Sustainable Identity and Access Management

Current status:	Under Review
Journal:	Sumbitted
Authors:	Matthias Hummer, Sebastian Groll, Günther Pernul
Full citation:	*Working Paper, University of Regensburg, 2018.*

Analyzing Context Data for Sustainable Identity and Access Management

MATTHIAS HUMMER, Nexis GmbH, Germany
SEBASTIAN GROLL, University of Regensburg, Germany
GÜNTHER PERNUL, University of Regensburg, Germany

The definition, implementation and adaption of a sustainable Identity and Access Management (IAM) strategy represents a crucial task for today's enterprises. While representing a key-success factor for enabling flexible company-wide IT infrastructure, strategy management currently solely relies on practical experience, due to lack of appropriate decision support methodology. This leads to a situation in which appropriate adaptions of strategical and operational IAM are not addressed as necessary, thus imposing negative consequences like additional heavy workload and increasing costs in a long-term. Within this work, we propose the a-based analysis methodology for strategic IAM decisions (TamsId) to overcome the presented shortcomings. We employ flexible IAM indicators, thereby enabling customizable monitoring. Based on simulation techniques, enterprises are empowered to gain insights concerning long-term impact of strategical IAM decisions. The approach is evaluated together with a world-wide operating manufacturing company by assessing two real-world use cases, demonstrating feasibility of the presented tool-based methodology.

CCS Concepts: • **Information systems** → *Data analytics*;

Additional Key Words and Phrases: Identity and Access Management, Strategical Identity and Access Management,Security Management

Working Paper:
Matthias Hummer, Sebastian Groll, and Günther Pernul. Analyzing Context Data for Sustainable Identity and Access Management. 17 pages.

1 INTRODUCTION

These days more than ever, organizations underlie frequent market shifts through new technological possibilities, organizational transformations, or newly arising regulatory requirements [27]. As a result, the so-called digital transformation has become a key success factor for enterprises in order to cope with these challenges while additionally increasing business outcome and overall flexibility [4, 18].

In Identity and Access Management (IAM) as a field where an organization's whole IT infrastructure is typically connected to a centralized IAM system, these challenges become apparent. Keeping pace with these new developments is a daunting task, especially for medium and large-sized companies [1], often resulting in the so called identity chaos [23]. Historically grown authorization structures, modern cloud applications, as well as custom developed and non-standardized IT systems need to be homogenized via a company-wide IAM system. Up to today, this task represents an enormously complex issue, taking into consideration that IAM needs to manage potentially thousands of users and permissions, together with millions of user permission assignments [13]. Responsible human experts are confronted with massive amounts of data both, on an operational level (e.g. by addressing employees joining or leaving the enterprise) as well as a strategical level

Authors' addresses: Matthias Hummer, Nexis GmbH, Franz-Mayer-Str. 1, Regensburg, Bavaria, 93053, Germany, matthias.hummer@nexis-secure.com; Sebastian Groll, University of Regensburg, Universitätsstr. 31, Regensburg, Bavaria, 93053, Germany, sebastian.groll@wiwi.uni-regensburg.de; Günther Pernul, University of Regensburg, Universitätsstr. 31, Regensburg, Bavaria, 93053, Germany, guenther.pernul@wiwi.uni-regensburg.de.

(e.g. management of the access model in place or introduction of new technologies and applications). Joiner/mover/leaver processes and approval workflows result in a large amount of meta information that need to be audited due to compliance requirements. The aforementioned business changes require a comprehensive IAM strategy [6, 25].

Currently, companies' IAM strategies are mostly purely based on qualitative information such as specific business or IT goals defined by management stakeholders. At the moment, none of the available meta information generated by the numerous IAM processes or human experts is fostered. We argue that this information is of significant value for every company's IAM strategy and improving current IAM processes. Recalling the fact that companies struggle when making predictions on the current as well as the future performance of their IAM system [26], there is a significant need to improve the strategic decision support in IAM by harvesting such meta information.

To the best of our knowledge, the problem of establishing a tailored IAM strategy based on both, historical as well as predictive data, has not been addressed in research. We aim at closing this gap by introducing TamsId, a tool-based IAM decision support methodology that fosters currently unused information and makes predictions on relevant IAM performance indicators. By this means, human decision makers are enabled to (1) monitor key aspects of their IAM and are able to (2) discover necessary adaptions of current processes or the overall IAM strategy.

The remainder is structured as follows: Section 2 provides an overview of related work while Section 3 introduces our research approach. Section 4 presents the various data types together with a detailed conceptual description of TamsId. Consequently, in Section 5, we evaluate our approach using real-world use cases based on data of a world-wide operating manufacturing company. Section 6 concludes our work and shows possibilities for future research.

2 RELATED WORK

Within medium and large-sized enterprises, IAM typically affects all areas which are executed or supported by information technology. Therefore the definition, implementation, and adjustment of a sustainable IAM strategy is a crucial topic within research and practice. In general, strategy means setting goals together with actions to achieve these goals, while taking limited resources and environmental factors into consideration [5]. Previous publications have already highlighted a large number of drivers and goals for IAM strategies including risk management, business facilitation, compliance requirements, costs, data and process quality, or technology and governance [12, 19, 32].

Windley et al. proposed an in-depth analysis helping organizations to receive a holistic picture of IAM [32]. The authors provide a detailed overview of maturity models as well as details regarding building blocks which may be used to define an IAM strategy (e.g. available technologies, standards, or governance topics). While providing a good starting point, practical questions like how to incorporate the introduced building blocks into a comprehensive IAM strategy in a real-life scenario remain unanswered.

Further publications are devoted to the question of how to define an individualized IAM roadmap which depicts the most important features of a company's overall strategy. For instance, they discuss how to decide which goals are relevant for an enterprise and in which way these may be achieved considering timely and economical matters [3, 20]. Yet, the provided methodologies are based on experiences of IAM operations and do not take the continuous development of IAM and its dynamic nature together with available context data as introduced in Section 1 into consideration.

Mont et al., propose an economical decision support methodology, which can be used to prioritize investments for IAM [19]. The authors define four IAM metrics which are rated by decision makers and weighted accordingly. By simulating different investment combinations, the currently optimal outcome may be achieved. While the given approach produces good results for the given metrics, it remains limited to financial decisions. Furthermore, applying it on new or individual metrics presents a challenging task.

Fuchs et al. describe how an overall IAM strategy influences the organizational and technological aspects of IAM and show which drivers and influence factors can define such a strategy [6]. While their argumentation justifies the necessity of a long term IAM planning, they do not present any support in defining such a strategy.

Additionally, various articles have been published dealing with the problem of how to strategically address different IAM goals. Royer et al. introduce a balanced scorecard approach in order to evaluate initial costs of deploying IAM in contrast to expected cost reduction [26]. They focus on analyzing if a certain IAM strategy may result in a beneficial outcome for an enterprise, effectively evaluating it before it is put into practice. Again, this publication has a strong economical focus, limiting its general applicability. In contrast, Gunter et al. use an experience-based method to constantly improve implemented workflows and access assignments for risk management and increased data quality [8]. The authors employ event analysis methods to assess a currently implemented model concerning improvements (e.g. the de-provisioning of unused permissions). However, no specific guidance or algorithms for utilizing the approach in complex real-life scenario are provided.

Summing up, no approach for continuously evaluating a company's IAM using generated IAM context data as well as predictive analyses on the basis of company-specific performance indicators has been proposed so far. Within the next Section, we are going to elaborate our research methodology, before we introduce TamsId to close this research gap.

3 RESEARCH METHODOLOGY

For our work, we utilize a series of well recognized research frameworks in the design science realm [7, 9–11, 21]. These frameworks serve as foundation of our research methodology and process. Moreover, we adapt an associated publication framework which serves as the foundation of this paper's structure [7].

According to Hevner et al. [9, 11] the performed activities of a research project are located within three inherent cycles, namely the relevance, the rigor and the design cycle. The relevance cycle establishes the link between research activity and application context, defining the research problem and the requirements of a possible solution. It assures the practical relevance of the research and serves as basis for evaluation. The rigor cycle draws expertise from the knowledge base, which provides foundations and methodologies from already existing literature and research. Carefully embracing existing knowledge ensures the innovation of the research project and prevents routine design [11]. In the design cycle one or more artifacts are created by iterating subsequently the phases *build* and *evaluate* [11]. The artifact we develop is a tool-based decision support methodology [28], which aims to address the lack of strategic decision support in IAM (relevance) by the application of our expertise in the IAM realm and existing techniques in simulation and statistics (rigor). Relating to the Design Science Research Knowledge Contribution Framework [7] we classify our contribution as an *improvement* as we seek to provide a new solution for a known problem.

Adopting this research framework to our problem domain the necessity of a strategic decision support tool initiates the design cycle and executes the build and evaluate interplay. During the build phase we use existing literature of Identity and Access Management, Data and Quality Management and simulation techniques as well as our expertise to design a decision support tool. Subsequently, we evaluate our tool within a real-world scenario from the application domain. The results of the design cycle will later contribute back to both the knowledge base as well as the application domain.

4 CONCEPTUAL OVERVIEW

In the following we present our new approach that facilitates different types of data generated by IAM systems or any IT application's user management in order to correlate pre-defined or company-specific IAM performance indicators and simulate future IAM developments. Our goal is to collect contextual data, provide a consistent data repository, and detect inherent relations among IAM performance indicators in order to predict necessary adaptions of the current IAM strategy or processes. Within the next Section, we are at first going to present a classification of the input data used by our approach, before we highlight the elements of our methodology.

4.1 Data Classification

In practice, IAM systems are synchronized with various local IT applications as well as HR-systems based on a centrally available generic data model which depicts all relevant entities. This set of contextual data is hereafter termed as *logical data*. In addition to the *logical data*, a large amount of so called *process data* (i.e. meta information concerning the execution of IAM processes) is processed by IAM systems. Examples comprise the execution time of IAM processes, the number of help desk inquiries of IAM stakeholders during IAM process execution, or synchronization statistics generated during the periodic data exchange with local IT applications.

4.1.1 Logical Data. In the context of IAM, *logical data* provide a detailed specification of a certain entity. For our work, we use the IAM data model presented by [13] (see Figure 1) to structure *logical data* entities (employees, user accounts, organizational structures, local IT applications' permissions, as well as global entitlement objects like business roles): Employees gain access to different applications through user accounts. Permissions are assigned either directly or via business-level entitlements. Entities might be assigned to an owner (e.g. a role owner for a business role) [29, 31]. Note that, even though we specifically focus on the entities provided by [13], other IAM-related entities like policies, context, or user devices could easily be considered by our approach.

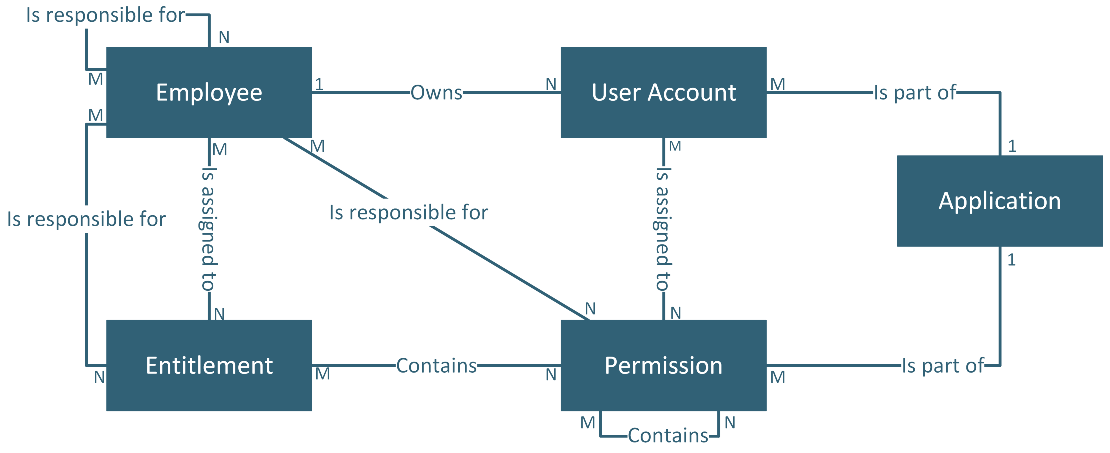


Fig. 1. Logical IAM Data Model based on [13]

Based on the data model we extract different *parameters* for the managed entities as input for our methodology. For this task, we facilitate an adapted approach as presented by Klettke et al. [16] who split entities into their different elements (*parameters*) based on their structure. Generally, elements are represented by *attributes* or *virtual attributes*. *Attributes* are contextual characteristics of a certain entity (e.g. the department or cost center attribute of an employee). *Virtual attributes* represent properties which need to be observed over a certain period of time (e.g. the number of employees created within the last month).

A conceptual representation of the used structure is depicted in Figure 2: Each entity is split up into a structure identification graph ([16]) $SG = (V, E)$, where V is a finite set of vertices or nodes and E is a set of directed edges connecting different nodes. Thereby nodes consist of a type t, value v and a set of corresponding entity identifiers $idList$ e.g. a *list* of all employee *identifiers* of a certain *attribute* value.

Nodes may be of the type $attribute$ (e.g. department or cost center attributes) or $virtualAttribute$ (e.g. the number of created and deleted entities per day). Henceforth *parameters* may be computed from nodes (e.g. using an aggregation function like $\sum idList$).

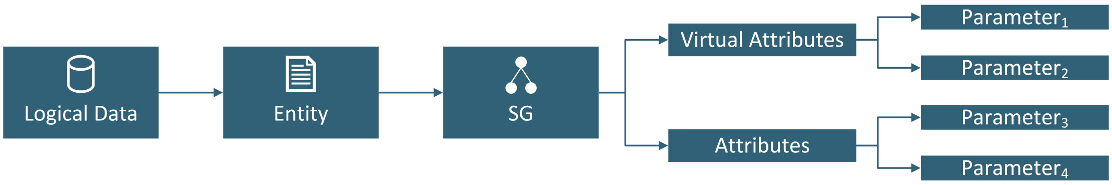


Fig. 2. Data Classification of logical IAM data

4.1.2 Process Data. Today's IAM systems provide extensive process management capabilities [22] which in turn results in the computation of related meta information. According to [24], five process perspectives $P = (F, B, OR, I, OP)$ can be considered:

- The *function* perspective F depicts what is to be executed within a certain process.
- The *behaviour* perspective B provides information regarding execution effectiveness.
- The *organization* perspective OR refers to responsibilities for process activities.
- The *information* perspective I, refers to which data are consumed and produced.
- The *operational* perspective OP refers to how the process is to be executed.

In the context of an employee requesting a permission via the IAM system, the *functional* perspective F corresponds to the IAM goal business facilitation. The request process is established to allow business users to gain access to resources with minimal administration efforts. From the *behaviour* perspective B, process efficiency is relevant. In case it takes too long to process a given request, the employee might not be able to complete daily business work. The *organization* perspective OR expresses who is in charge of processing a given request. Additional process stages (like help desk requests) might indicate a low process usability or missing knowledge of deciders. Finally, the *information* perspective I defines which data are consumed (the employee, her context, and a permission) by the process and what outcome may be achieved.

Note that within our work we do not aim at measuring the quality of the overall IAM business process model but rather focus on the evaluation of business process performance. Consequently, the *operational* perspective OP is not addressed in the following as it refers to general process model definitions.

4.1.3 Parameters and IAM indicators. Both aforementioned data types provide rich sources for supporting strategic IAM decisions. However, due to the given complexity and quantity, means of automatically processing, integrating, and analyzing the raw data need to be established.

The core element of our approach is the definition of IAM performance indicators which can be interpreted by human IAM engineers in order to evaluate the current IAM processes as well as the underlying long-term IAM strategy. We achieve this goal by firstly extracting *parameters* (defined as IAM measurements following [33]) which can be interpreted as the state changes of a certain data element over time. Examples are changes in the number of managed employees per day or the average execution time of a certain approval workflow during the last month of operation. Parameters are in turn related to *indicators* which can be understood as metrics following the definition of [33]. They represent a) bounded, b) metrically scaled, c) reliable, valid, objective, and d) context-specific metrics which are computed automatically and provide a human-understandable representation of potentially complex concerns in a numerical way.

Following [30], *indicators* may be defined in a hierarchical manner: Indicators may be composed of parameters, other indicators, or a combination of both.

The example visualization in Figure 3 shows three *parameters* p_1, p_2 and p_3 which are extracted based on existing *logical* and *process data*. p_2 and $p3$ are used to compute the *indicator* i_2. In turn, this *indicator* is used together with p_1 to compute i_1.

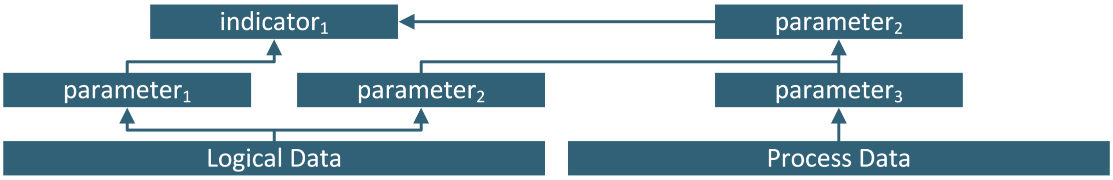


Fig. 3. Conceptual Data Classification Example

Consider the example of the number of managed employees per day. This *indicator* can be correlated with the number of business experts' help desk inquiries per managed employee, i.e. another *indicator*. Using automated simulation techniques, an enterprise might be able to judge whether the current IAM staff members are capable of handling the IAM system after a potential corporate growth, e.g. in case a merger with another enterprise is planned in the near future.

4.2 TamsId - A tool-based analysis methodology for strategic IAM decisions

Within this Section, we introduce our tool-based analysis methodology for strategic IAM decisions (TamsId). It enables companies to gather the required input data, define desired *indicators*, and derive correlations as well as conduct simulations. Different publications [13, 23] already suggest that IAM may strongly profit from such semi-automated monitoring and data analysis procedures.

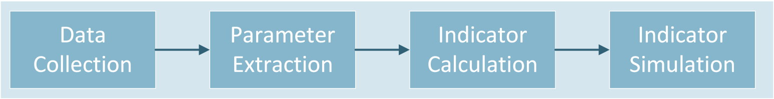


Fig. 4. Conceptual Overview of TamsId

A conceptual overview of the four main phases of our methodology is depicted at Figure 4. During the Data Collection the relevant data is gathered and stored in a consistent data repository. This data is used for the extraction of parameters serving as the foundation for the subsequently following definition, calculation and correlation of IAM indicators. By using statistical methods possible future developments of the indicators are calculated and presented.

4.2.1 Data Collection. The data collection is designed to provide a central and consistent data repository incorporating the relevant input sources (see Figure 5). The data repository uses synchronization measures to import snapshots of data states from input sources over time (e.g. daily synchronization). Using different connectors, TamsId is able to extract data from sources like files, databases (e.g. operated by IAM systems), or web services. Note, that modern IAM systems typically already provide a unified view onto relevant entities and thus are the main source of information for TamsId regarding *logical data* and *process data. Process data* in practice also might be imported from a workflow management tool in place. Additionally, built-in monitoring capabilities of local IT applications or IAM systems also can be facilitated if available.

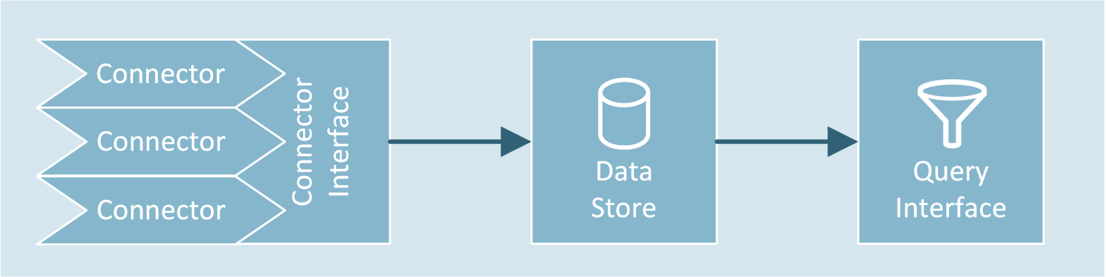


Fig. 5. Data Collection Component

4.2.2 Parameter Extraction. After the collection of raw input data, parameter extraction (see Figure 6) takes place. TamsId automatically categorizes *logical data* regarding available entity attributes (e.g. department or location of employees) as well as available *virtual attributes* (e.g. how many employees are created each day). Each *parameter* contains information regarding *(virtual) attribute* values, together with a list of identifiers of referencing entities as described in Section 4.1.1. In addition, *parameters* based on *process data* (e.g. execution time of an approval workflow) together with a list of data consumed or created by a process are extracted. The outcome is stored as *parameter* vectors for further processing. From a technical and tool-based point of view this allows us to utilize historical data without having to store full and potentially large data snapshots while at the same time minimizing the querying efforts. Note that from this point of time *logical data* and *process data* are stored and processed identically.

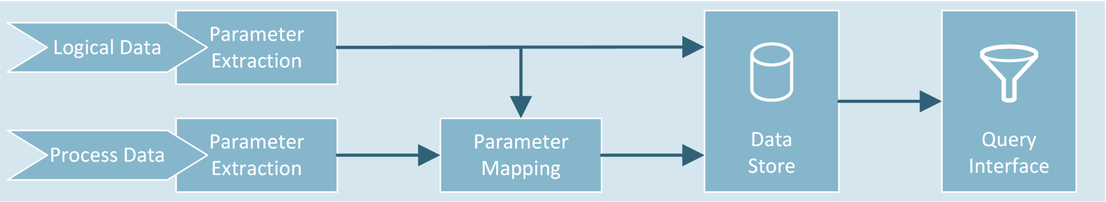


Fig. 6. Parameter Extraction Component

4.2.3 Indicator Calculation. Consuming the previously defined sets of *parameters*, TamsId subsequently calculates different *indicators* based on companies needs (see Figure 7). The initial definition of *indicators* has to be carried out on a one-time basis. Nevertheless, periodic evaluation and possible re-definitions might be required in order to ensure relevance of the computed *indicators*.

Legal requirements like the Sarbanes-Oxley Act (SOX) [29], for instance, might influence the initial indicator definition. SOX as a prime example of regulatory influence on IAM requires activities during departmental movers of employees and the adaption of affected access rights. As an example used to address this requirement, the *indicators* 'attribute department changed' $DepartmentChange$ displaying employees who switched department and 'user permission assignments (UPAs) deleted of employees which changed department' $UpaDeleted\_DepartmentChange$ might be defined. $UpaDeleted\_DepartmentChange$ can be expressed as the combination of two *parameters*, namely 'UPAs deleted' intersected with 'Attribute department changed' ($\frac{DepartmentChange \cap UpaDeleted}{DepartmentChange}$). As a result, the *indicator* 'average number of UPA changes per department change' can be derived.

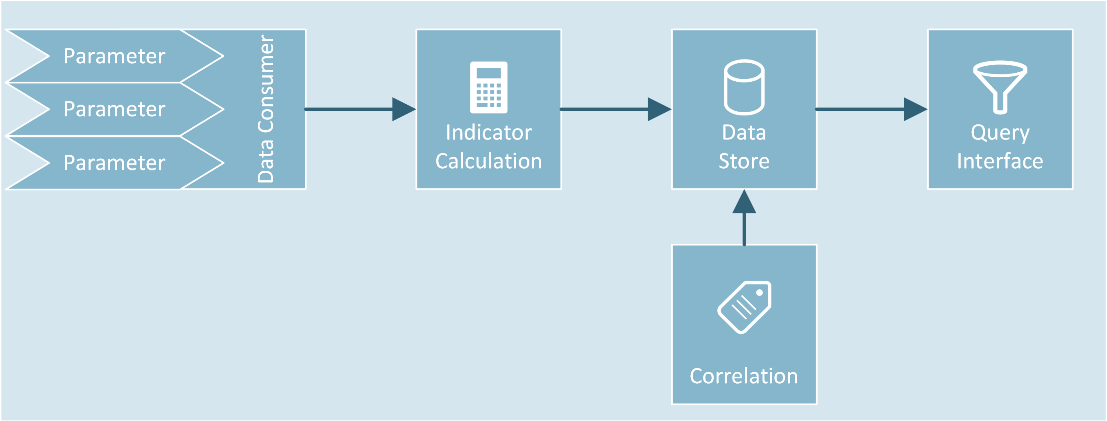


Fig. 7. IAM Indicator Computation Component

Within a next step, we derive correlations among *indicators*. For this task, we use the normalized cross correlation [17] algorithm. Unlike other common correlation algorithms, it also considers the development of data values over time. An indicator can be expressed as vector $v_{indicator} = \{i_0, .., i_t, ..i_t\}$. The comparison of two *indicator* vectors can be defined as correlation of two non-linear graphs, g_1 and g_2. The used normalized cross correlation is defined as follows:

$$\frac{\sum_{t=s}^{t=n} i_1[t] * i_2[t]}{\sqrt{\sum_{t=s}^{t=n}(i_1[t])^2 * \sum_{t=s}^{t=n}(i_2[t])^2}}$$

n is the last computed time point, i is a computed *IAM indicator* and s is the starting point of consideration.

Again, taking the mover process of an employee as an example, one expected correlated *indicator* could be the adaption rate of UPAs during process execution. In practice, such *indicator* correlation can detect hidden semantic knowledge within the available raw meta information and thus support identifying discrepancies between expected and actual states.

4.2.4 Indicator Simulation. After definition, the consecutive simulation phase tries to predict future developments of defined *indicators* and displays them to human IAM experts. In terms of visualization, imagine an *indicator* set as line graph, with the x-axis showing different points of time and the y-axis showing the related *parameter* values. Based on historical data, we aim at estimating how the graph might develop in future. In order to achieve this goal, we facilitate the Monte Carlo Simulation [2] which relies on repeated random sampling and statistical analysis. We employ a model often used within finance and economics [14] e.g. to estimate return on investments:

For each set of potentially correlated *indicators*, we use the vector $v_{indicator} = \{i_0, i_1, .., i_t\}$. i_0 represents the value of time point zero, while i_t represents the latest data value. For example, the vector $v_{activeEmployees} = \{100, 105, 107\}$ shows that in t_0 100 employees have been on the staff, 105 in t_1 and 108 in t_2. For these vectors, the standard deviation δ, variance δ^2, and the mean value μ are computed:

$$p(t+1) = p(t) * e^{(\mu-(\delta^2/2))*(RNG*\delta)}$$

with $p(t)$ representing the last value and an exponential function e. The exponent itself also contains two variables. The vector drift is depicted as $\mu - (\delta^2/2)$, while $(RNG * \delta)$ represents the random factor.

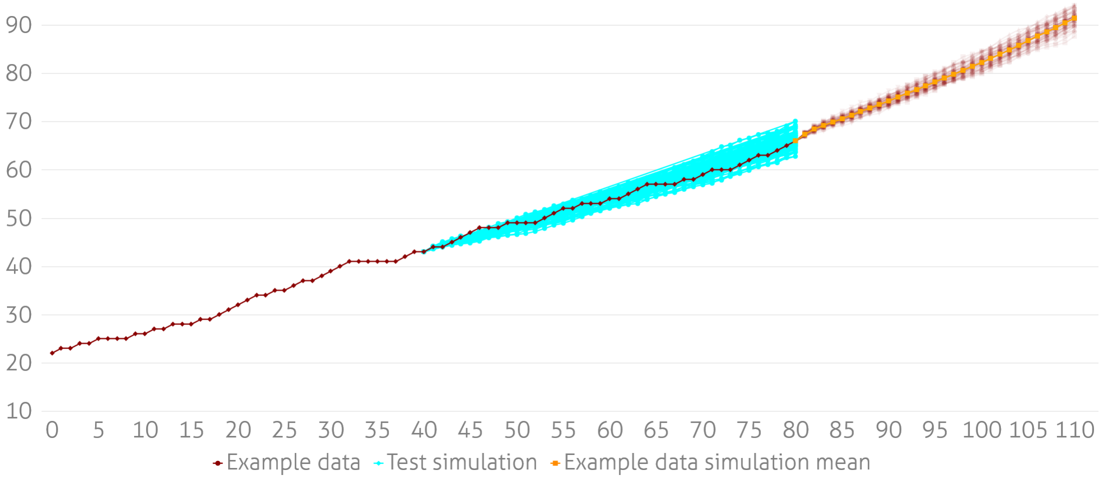


Fig. 8. Sample Monte Carlo Simulation

A visualized output example is depicted at Figure 8. The red line represents the actual vector development of the current date (e.g. the number of employees working for an organization). For predictive analysis, the red line splits up into courses computed by the Monte Carlo Simulation, marking the expected future graph behaviour. The mean value of the simulation is drawn in orange color, marking the most likely development.

The cyan blue line is representing a sample for evaluating the simulation in order to assure that it creates appropriate results. Therefore, we split up the historical real-world data into training and evaluation data. The simulation receives the training data as input and it's results are verified against the evaluation data. In case of a significant discrepancy, the simulation variables need to be adapted. We recommend several carefully executed validation steps in order to assure meaningful estimates on future developments.

Finally, a summary of all simulation results is visualized for IAM engineer interpretation. This can be done using principles of visualization for time oriented data [15]. Within our prototype implementation we use a combination of dashboard screens for quick and understandable overviews together with the visualization of the computed *indicators* as line diagrams for deeper interpretation.

5 EVALUATION

In order to evaluate our methodology and the underlying automation techniques, we cooperated with a world-wide operating manufacturing enterprise employing about 12.000 internal and 4.000 external employees. Within their core IT application (the SAP ERP system) they manage about

17.000 user accounts, 6.000 permissions, and 1.200 application roles (i.e. hierarchically aligned permissions). We decided to specifically evaluate TamsId using the SAP ERP system due to its high user management process execution rate together with its importance for compliance and risk management. About one million UPAs are managed in the SAP ERP system with about 1.9% of these assignments changing per month (about 20.000 UPA changes per month).

5.1 Data Collection

During data collection we imported three data elements on a daily basis for a period of 80 days into our data store:

- Information regarding internal employees from the IAM system using a database connector.
- Data concerning user accounts together with their mapping to employees, permissions and their attributes, as well as UPAs and application roles from the SAP ERP system.
- Analytic data regarding the risk level and usage of existing UPAs which have been provided by the SAP ERP system as well as the IAM system.

Note that due to organizational reasons during the project with our partner company, we were limited to the analysis of *logical data* for this evaluation. We argue that the existing evaluation results based on *logical data* already underline the applicability as well as relevance of our approach. However, we aim at extending the evaluation when *process data* can be provided by the enterprise.

5.2 Parameter Extraction

During the *parameter* extraction phase the project team's focus was to generate parameters related to the required administrative workload and how the SAP ERP system is managed regarding compliance requirements. Together with company IAM engineers, we proposed 34 *parameters* (see Table 1), each corresponding to a certain entity with 26 referencing attributes and 8 virtual attributes.

5.3 Simulation Test

In order to evaluate the quality of the gathered input data and the generated *parameters*, we initially performed a simulation test. One exemplary result visualization showing the development of the total number of employees managed by the IAM system is depicted in Figure 9. As input we used data values from day 0 to 40 to determine μ, μ^2 and δ. Based on these factors, we executed the Monte Carlo simulation 80 times for the days 41 to 80 and evaluated to which proportion the actual data are located within the expected (cyan blue color) range. Additionally, we marked the mean simulation values in orange, displaying the most likely data development trend.

It turned out that the simulation produces a hit rate of 82% regarding all calculated indicator vectors. This is reasoned due to the fact that real-world data contains punctual outlier values (e.g. value 71 of the depicted example graph). Yet the impact of outlier values is incorporated into training parameters, thus enabling a more accurate long-term simulation. The given exemplary visualization also shows that the data simulation produces suitable results within a certain range of time and the simulation performed very well for the given real-world data set. In the following, we execute the indicator calculation as well as the indicator simulation using two different use cases:

1. Employees joining the enterprise
2. Implementation of risk management processes

5.4 Use case 1: Employees joining the enterprise

We firstly investigated the impact of employees newly joining the company on the workload of the IAM team by calculating the following three indicators over a simulation period of three months:

Entity	Attribute	Parameters
Employee	Active	Active employees, created employees per synchronization
	Deleted	Inactive employees, deleted employees per synchronization
	Department	Parameter per department value
	Owner	Assigned to owner, not assigned to owner
User Account	Active	Active accounts, created accounts per synchronization
	Deleted	Inactive accounts, deleted accounts per synchronization
	Assigned Employee	Assigned to Employee, not assigned to employee
Permission	Active	Active permissions, created permissions per synchronization
	Deleted	Inactive permissions, deleted permissions per synchronization
	Owner	Assigned to owner, not assigned to owner
UPA	Active	Active UPAs, created UPAs per synchronization
	Deleted	Inactive UPAs, deleted UPAs per synchronization
	Risk Level	Low, medium, high, very high
	Assignment type	Direct, inherited, duplicate assignment
Permission Hierarchy	Active	Active permission hierarchies, created permission hierarchies per synchronization
	Deleted	Deleted permission hierarchies, deleted permission hierarchies per synchronization

Table 1. Overview of extracted *parameters*

1. Indicator: Number of employees created per day.
2. Indicator: Number of user accounts created for new employees.
3. Indicator: Number of created UPAs for these user accounts.

The visualization of the used indicators is depicted at Figure 10. On the x-axis, values 0 to 80 depict the development of indicators 1 - 3 throughout the 80-day period of data collection. Values 81 to 170 display corresponding simulation results for future dates, depicted as transparent areas. The red line marks the number of employees created per day while the cyan blue line marks the number of user accounts having been assigned to these employees. The dark orange line marks the number of direct UPAs which have been assigned to the given user accounts. For visualizing the predicted graph development, we used the mean values of different simulation vectors.

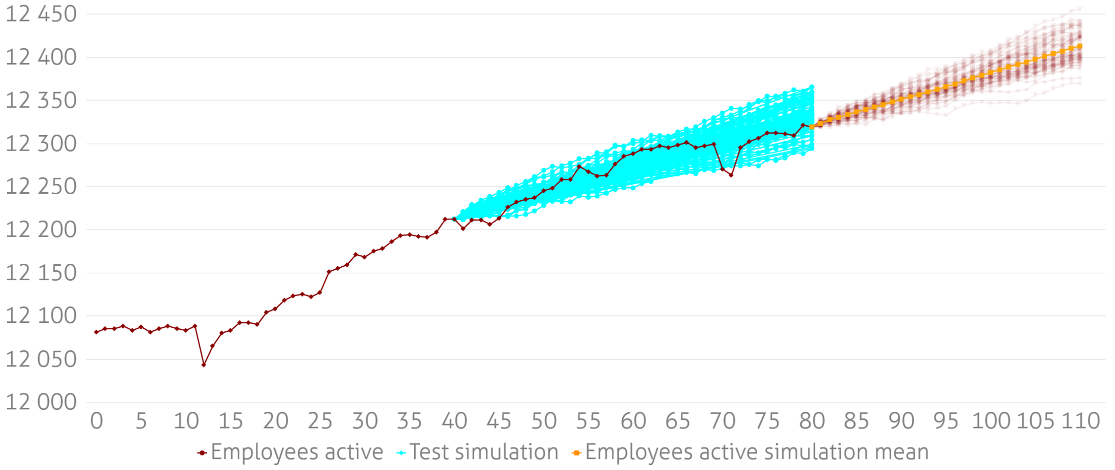


Fig. 9. Development of active employees

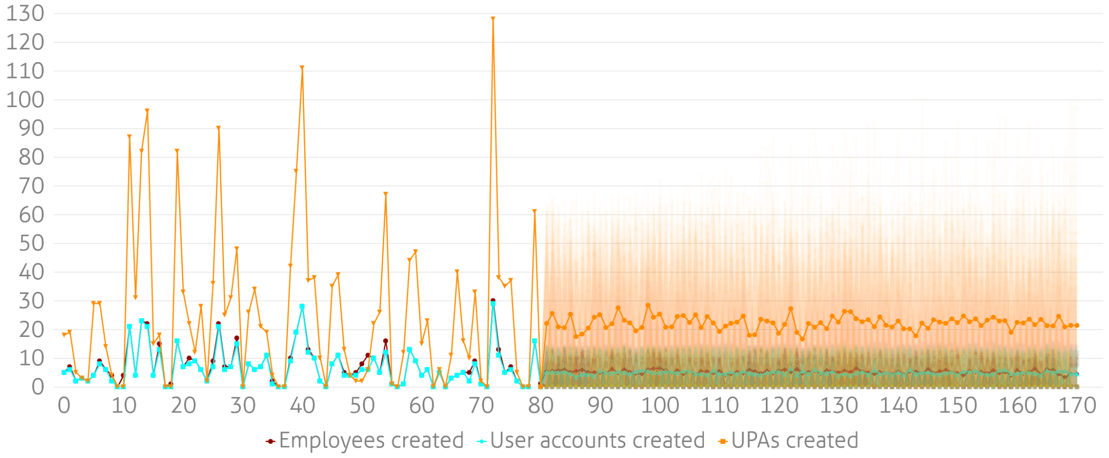


Fig. 10. Indicators of use case 1

Basic analyses revealed that

- about 7.5 employees join the enterprise per day (which corresponds to about 0.058% of the overall employee number in t_0).
- considering the number of employees leaving the company, the current employee growth rates at about 3.6 employees per day.
- on average, 96% of the new employees automatically receive a user account for accessing the SAP ERP system.

More detailed *indicator* analysis revealed that there was only a low correlation between the number of newly created employees and their number of assigned permissions. Discussions with IAM engineers revealed that newly joined employees are assigned to a minimal set of access privileges and are henceforth expected to manually request missing permissions which in turn leads to significant additional workload for both, IT and business.

On this basis, we defined 'adaptions of UPAs for created employees within one week' as new additional *indicator*. It's simulation showed that an employee's access rights must be adapted about three times in order to get ready for work (indicator value of 2.8). Due to its importance for quality evaluation, the IAM engineers decided to facilitate this *indicator* as a long term monitoring KPI. Note that revealing such a semantically relevant indicator would not have been possible without the semi-automated simulation techniques employed by our methodology. This further underlines the applicability of our tool-based approach.

Henceforth, we executed an according simulation of the given indicators. Results show an increasing trend of employee joins with an expected value of 8.1 (varying between minimal 6.5 and maximal 9.3), thus we expect about 729 employee joins within three months. Additionally, simulations of 'adaptions of UPAs for created employees within one week' show that these employees need to be assigned to 5799 UPAs within the next three months, inducing heavy workload for business and IT. As a result, the enterprise is now internally discussing means of automation regarding employee joiner processes by weigthing internal management costs in contrary to expected introduction costs of authorization automation.

5.5 Use case 2: Implementation of risk management processes

Our partner company is subject to compliance regulations which require them to control existing access risks according to the principle of the least privilege. In order to address this, the company over the last 24 months already established

- standardized access review processes for the SAP ERP system which evaluate UPAs periodically by the responsible employee or permission managers as well as
- identity analytic processes in order to automatically classify UPAs concerning their risk level (e.g. flagging potentially erroneous UPAs with a risk value 'high' based on algorithms presented in [23]).

For evaluation purposes we defined the two indicators

- 'risk level increased', depicting the UPAs newly flagged as critical and
- 'risk level decreased', depicting the UPAs with a lowered risk value e.g. due to de-provisioning of critical assignments.

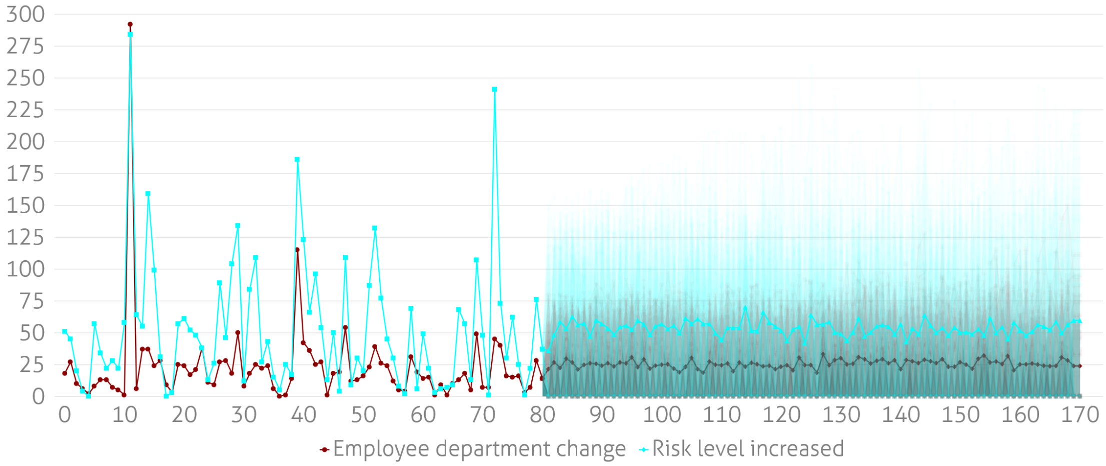


Fig. 11. IAM Indicator 'risk level increased' and 'employee department change'

Analyses showed that *indicator* 'risk level increased' correlates with the number of UPAs created. This is not surprising as the facilitated (and high number of) manual permissions assignment processes are error-prone. More interestingly, 'risk level increased' also correlates with the *indicator* 'employee department change' (correlation value 0.89) essentially expressing that the number of critical UPAs are increasing in case employees undergo a mover process. Discussions with the IAM experts revealed that UPAs in general are not automatically removed during a standard mover process. As a result those old UPAs from the employee's previous department are typically not used in the new department and hence more likely to be detected as critical.

Analysis of the collected input data also showed that the absolute number of critical UPAs shows a slightly decreasing trend (with about 2.2% less critical UPAs over the period of 80 days). We argue that this mainly stems from the implemented periodic access reviews of employees executed as part of the company's IAM strategy. This underlines the positive effect of the access reviews together with the need to continue executing them in the future and is a good example of how our methodology can be used to rate the efficiency of the currently employed IAM strategy.

Regarding the *indicator* 'risk level decreased' we initially were not able to determine any relevant correlation concerning *indicators* derived from parameters presented in 5.2. As a result, we executed a deeper analysis and defined a new composite *indicator* 'changed with owner[1]' based on the ownership-attribute of permissions (see Figure 12). This enabled differentiated investigation of the correlation of the *indicator* 'risk level decreased' in respect to permissions with and without an assigned business owner.

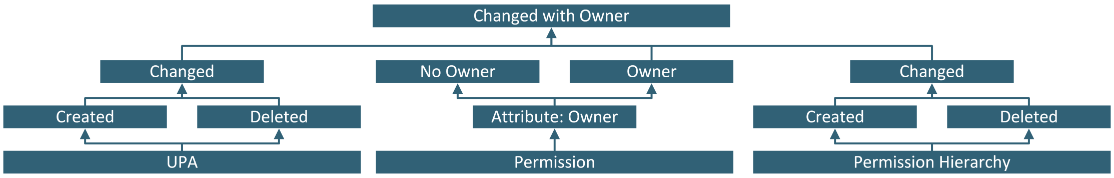


Fig. 12. IAM Indicator 'Changes of permissions which are assigned to a responsible employee'

The indicator 'changed with owner' expresses the changes of permissions with an assigned business owner and is composed from three entities, permission P, permission hierarchy PH, and UPA UPA. Formally, we used $i = P_O \cap (UPA_{Change} \cup PH_{Change})$, depicting the number of UPA and permission hierarchy changes (the superset), which refer to permissions with an owner.

Basic analyses revealed

- a high absolute number of permissions without an assigned owner (about one third of all available permissions).
- about 12.3 mover processes on average per day. Simulation showed that this value is expected to slightly increase, with about 12.6 expected changes within three months (augmentation of about 2%).
- about 80.2 daily changes (like adapting permission hierarchy or adjusting UPAs) affecting permissions with an assigned owner. Simulation showed that this value is expected to increase to about 81.7 (augmentation about 1.8%).

Additionally, our simulation techniques determined a strong correlation between the indicators 'risk level decreased' and 'changes with owner' (correlation value of 0.86, see Figures 11 and 13). This essentially means that employees which are assigned to business-owned permissions in general have less risky UPAs. Our simulation thereby underlined the importance of business ownerships for

[1] An employees who is responsible for a certain entity like further employees (as supervisor) or permissions

permissions in terms of risk management. As a result, the company now discusses the mandatory introduction and periodic re-certification of business ownership assignments for permissions and business roles in the future. At the same time they aim at reducing the currently high number of permissions without a business owner.

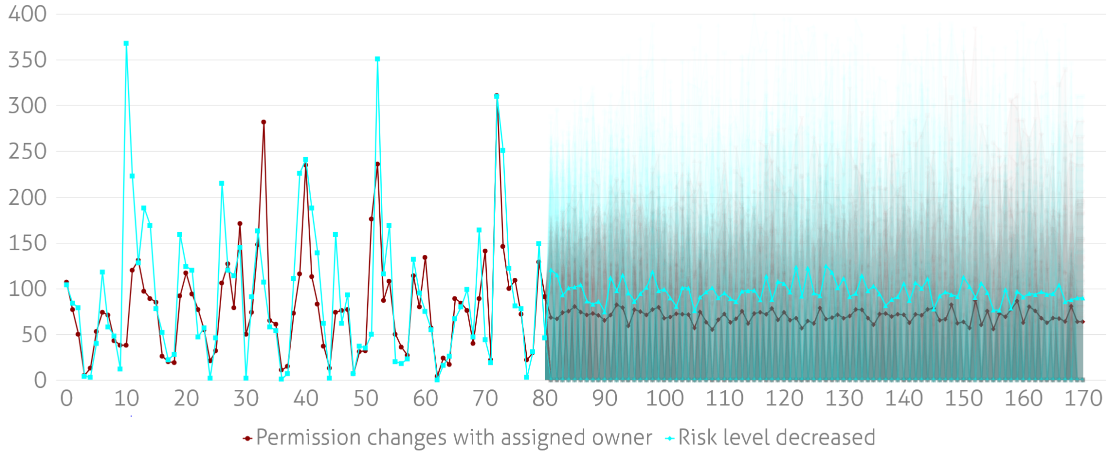


Fig. 13. IAM Indicator 'risk level decreased' and 'permission changes with assigned owner'

Summed up, our methodology and the conducted simulations and feedback sessions with IAM engineers revealed the following results. Firstly, the IAM workload can be expected to increase in a mid-term period due to the overall company growth. Secondly, we identified (and IAM engineers confirmed) deficiencies when provisioning newly joined enterprises. The company thus decided that the currently high amount of UPA changes required until an employee is assigned to all required permissions needs to be decreased in the future. This leads to an adaption of the current IAM strategy which might not have been identified without our methodology. During our second use-case, we thirdly identified that permissions with an assigned business owner are less critical in terms of automated risk analyses. This likely stems from the implemented access review processes which are not conducted accordingly for orphan permissions without an assigned business owner. Finally, the overall high amount of permissions also provides room for improvement.

6 CONCLUSIONS AND FUTURE WORK

Up to now, the implementation of a sustainable IAM strategy represents a complex and costly issue for companies. Additionally, there are no automated means for constantly revising the employed IAM strategy. Within our work, we closed this research gap by introducing TamsId, a tool-supported methodology which incorporates the analysis of existing logical and process data. While this data is, at least partially, already available on a technical level within companies, it up to now has not been facilitated for influencing strategic IAM decisions. Our methodology structures the necessary steps to define company-specific indicators as well as allows for an automated simulation for deriving future predictions for a given IAM system. We furthermore demonstrated that strategic IAM assessment may strongly benefit from employing well-known simulation techniques. Finally, we demonstrated the applicability of our approach within a real-world scenario.

During evaluation we identified two possible improvements which we are going to address in future work: Firstly, while our indicator correlation provides valuable information concerning relationships of indicators, we noticed that the high initial manual definition efforts for parameter consideration and indicator construction need to be reduced in the future. Secondly, even though we demonstrated that TamsId provides sound information, interpretation of recommended actions requires deep knowledge concerning the implemented IAM. We thus plan to develop structured guidelines for rating recommendations (e.g. by integrating best practice parameters and indicators together with guidelines for interpreting results). Lastly, we plan to re-run our real-life evaluation process including the analysis of process data after it has been made available to us.

REFERENCES

[1] Marnix Assink. 2006. Inhibitors of disruptive innovation capability: a conceptual model. *European Journal of Innovation Management* 9, 2 (2006), 215–233.

[2] Kurt Binder, Dieter Heermann, Lyle Roelofs, A John Mallinckrodt, and Susan McKay. 1993. Monte Carlo Simulation in Statistical Physics. *Computers in Physics* 7, 2 (1993), 156–157.

[3] Andras Cser and Merritt Maxim. 2017. *Build Your Identity And Access Management Strategy.* Technical Report. Forrester.

[4] Adam DeMattia and John McKnight. 2017. *How IT Transformation Maturity Drives IT Agility, Innovation, and Improved Business Outcomes.* Technical Report. Dell EMC.

[5] Lawrence Freedman. 2015. *Strategy: A history.* Oxford University Press.

[6] Ludwig Fuchs and Günther Pernul. 2007. Supporting Compliant and Secure User Handling - A Structured Approach for In-House Identity Management. In *The Second International Conference on Availability, Reliability and Security.* IEEE Computer Society, 374–384.

[7] Shirley Gregor and Alan Hevner. 2013. Positioning and presenting design science research for maximum impact. *Management Information Systems Quarterly* 37, 2 (2013), 337–356.

[8] Carl A Gunter, David Liebovitz, and Bradley Malin. 2011. Experience-Based Access Management: A Life-Cycle Framework for Identity and Access Management Systems. *IEEE Security & Privacy* 9, 5 (2011), 48–55.

[9] Alan Hevner. 2007. A Three Cycle View of Design Science Research. *Scandinavian Journal of Information Systems* 19, 2 (2007), 87–92.

[10] Alan Hevner and Samir Chatterjee. 2010. Design science research in information systems. In *Design research in information systems.* Springer, 9–22.

[11] Alan Hevner, Salvatore T. March, Jinsoo Park, and Sudha Ram. 2004. Design Science in Information Systems Research. *Management Information Systems Quarterly* 28, 1 (2004), 75–105.

[12] Matthias Hummer, Sebastian Groll, Michael Kunz, Ludwig Fuchs, and Günther Pernul. 2018. Measuring Identity and Access Management Performance - An Expert Survey on Possible Performance Indicators. In *Proceedings of the 4th International Conference on Information Systems Security and Privacy - Volume 1: ICISSP.* SciTePress, 233–240.

[13] Matthias Hummer, Michael Kunz, Michael Netter, Ludwig Fuchs, and Günther Pernul. 2015. Advanced Identity and Access Policy Management using Contextual Data. In *10th International Conference on Availability, Reliability and Security.* IEEE, 40–49.

[14] Peter Jäckel. 2002. *Monte Carlo Methods in Finance.* Wiley.

[15] Uwe Jugel, Zbigniew Jerzak, Gregor Hackenbroich, and Volker Markl. 2014. M4: A Visualization-Oriented Time Series Data Aggregation. *Proceedings of the VLDB Endowment* 7, 10 (2014), 797–808.

[16] Meike Klettke, Uta Störl, and Stefanie Scherzinger. 2015. Schema extraction and structural outlier detection for JSON-based NoSQL data stores. *Datenbanksysteme für Business, Technologie und Web* (2015), 425–444.

[17] John Lewis. 1995. Fast normalized cross-correlation. In *Vision interface*, Vol. 10. 120–123.

[18] Christian Matt, Thomas Hess, and Alexander Benlian. 2015. Digital Transformation Strategies. *Business & Information Systems Engineering* 57 (2015), 339–343.

[19] Marco Casassa Mont, Yolanta Beresnevichiene, David Pym, and Simon Shiu. 2010. Economics of identity and access management: Providing decision support for investments. In *Network Operations and Management Symposium Workshops.* IEEE/IFIP, 134–141.

[20] E. Osmanoglu. 2013. *Identity and Access Management: Business Performance Through Connected Intelligence.* Elsevier.

[21] Ken Peffers, Tuure Tuunanen, Marcus Rothenberger, and Samir Chatterjee. 2007. A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems* 24, 3 (2007), 45–77.

[22] Earl L Perkins and Ant Allan. 2005. Consider Identity and Access Management as a Process, Not a Technology. *Gartner Report* (2005).

[23] Günther Pernul and Ludwig Fuchs. 2010. Reducing the Risk of Insider Misuse by Revising Identity Management and UserAccount Data. *Journal of Wireless Mobile Networks, Ubiquitous Computing and Dependable Applications* 1, 1 (2010), 14–28.

[24] Hajo A Reijers and S Liman Mansar. 2005. Best practices in business process redesign: An overview and qualitative evaluation of successful redesign heuristics. *Omega* 33, 4 (2005), 283–306.

[25] Denis Royer. 2008. Enterprise Identity Management. In *The Future of Identity in the Information Society*. Springer, 433–446.

[26] Denis Royer and Martin Meints. 2009. Enterprise Identity Management – Towards a Decision Support Framework Based on the Balanced Scorecard Approach. *Business & Information Systems Engineering* 1, 3 (2009), 245–253.

[27] Albert Sune and Jenny Gibb. 2015. Dynamic capabilities as patterns of organizational change: An empirical study on transforming a firmâĂŹs resource base. *Journal of Organizational Change Management* 28 (2015), 213–231.

[28] Nguyen Hoang Thuan, Pedro Antunes, and David Johnstone. 2016. A Design Science Method for Emerging Decision Support Environments. *Computing Research Repository* abs/1605.04725 (2016).

[29] United States Congress. 2002. Sarbanes-Oxley Act of 2002, PL 107-204, 116 Stat 745.

[30] Vijay K. Vaishnavi, Sandeep Purao, and Jens Liegle. 2007. Object-oriented product metrics: A generic framework. *Information Sciences* 177, 2 (2007), 587 – 606.

[31] Steve G. Watkins. 2013. *An Introduction to Information Security and ISO 27001: 2013 A Pocket Guide* (2 ed.). It Governance Ltd.

[32] P.J. Windley. 2005. *Digital Identity: Unmasking Identity Management Architecture.* O'Reilly Media.

[33] Emrah Yasasin and Guido Schryen. 2015. Requirements for IT Security Metrics - An Argumentation Theory Based Approach. In *23rd European Conference on Information Systems (ECIS).*

Bibliography

[AA11] Ahmed, Khandakar Entenam U. ; Alexandrov, Vassil: Identity and Access Management in Cloud Computing. In: *Cloud Computing for Enterprise Architectures*. Springer, 2011, S. 115–133

[Ake04] Aken, Joan E v.: Management Research Based on the Paradigm of the Design Sciences: The Quest for Field-Tested and Grounded Technological Rules. In: *Journal of Management Studies* 41 (2004), Nr. 2, S. 219–246

[AKP02] Assmann, Danilo ; Kalmar, Ralf ; Punter, Teade: *Messen und Bewerten Von Webapplikationen Mit der Goal/Question/Metric Methode: Handbuch*. Fraunhofer-IESE, 2002 (IESE-Report / Fraunhofer Einrichtung Experimentelles Software Engineering)

[Bas92] Basili, Victor R.: Software Modeling and Measurement: The Goal/Question/Metric Paradigm. University of Maryland, 1992. – Forschungsbericht

[Bas11] Basel Committee on Banking Supervision: *Basel III - A Global Regulatory Framework for More Resilient Banks and Banking Systems*. Bank for International Settlements, 2011. – ISBN 9291318590

[EK10] Elliott, Aaron ; Knight, Scott: Role Explosion: Acknowledging the Problem. In: *Proceedings of the 2010 International Conference on Software Engineering Research & Practice*, 2010, S. 349–355

[Fin12] Finanzdienstleistungsaufsicht, Bundesanstalt für: Mindestanforderungen an das Risikomanagement (MaRisk). In: *Rundschreiben* 10 (2012), Nr. 2012, S. 4

[FP07] Fuchs, Ludwig ; Pernul, Günther: Supporting Compliant and Secure User Handling - A Structured Approach for In-House Identity Management. In: *The Second International Conference on Availability, Reliability and Security* IEEE, 2007, S. 374–384

[FP10] Fuchs, Ludwig ; Pernul, Günther: Reducing the Risk of Insider Misuse by Revising Identity Management and UserAccount Data. In: *Journal of Wireless Mobile Networks, Ubiquitous Computing and Dependable Applications* 1 (2010), Nr. 1, S. 14–28

[Fre15] FREEDMAN, Lawrence: *Strategy: A history*. Oxford University Press, 2015. – ISBN 1501227726

[Gil88] GILB, Tom: *Principles of Software Engineering Management*. 1988. – ISBN 0201192462

[HC10] *Kapitel* Design Science Research in Information Systems. In: HEVNER, Alan ; CHATTERJEE, Samir: *Design Research in Information Systems – Theory and Practice*. Springer, 2010. – ISBN 9781441956538, S. 9–22

[HGK$^+$18] HUMMER, Matthias ; GROLL, Sebastian ; KUNZ, Michael ; FUCHS, Ludwig ; PERNUL, Günther: Measuring Identity and Access Management Performance – An Expert Survey on Possible Performance Indicators. In: *International Conference on Information Systems Security and Privacy*, 2018, S. 233–240

[HKN$^+$15] HUMMER, Matthias ; KUNZ, Michael ; NETTER, Michael ; FUCHS, Ludwig ; PERNUL, Günther: Advanced Identity and Access Policy Management Using Contextual Data. In: *10th International Conference on Availability, Reliability and Security* IEEE, 2015, S. 40–49

[HKN$^+$16] HUMMER, Matthias ; KUNZ, Michael ; NETTER, Michael ; FUCHS, Ludwig ; PERNUL, Günther: Adaptive identity and access management—contextual data based policies. In: *EURASIP Journal on Information Security* 2016 (2016), Nr. 1, S. 19

[HKS$^+$14] HU, Vincent C. ; KUHN, David R. ; SCHNITZER, Adam ; SANDLIN, Kenneth ; MILLER, Robert ; FERRAIOLO, David F. ; VOAS, Jeffrey: Guide to Attribute Based Access Control (ABAC) Definition and Considerations. In: *NIST Special Publication* 800 (2014), S. 162

[HMPR04] HEVNER, Alan ; MARCH, Salvatore T. ; PARK, Jinsoo ; RAM, Sudha: Design Science in Information Systems Research. In: *Management Information Systems Quarterly* 28 (2004), Nr. 1, S. 75–105

[Kee88] KEEBLE, Brian R.: The Brundtland report: 'Our common future'. In: *Medicine and War* 4 (1988), Nr. 1, S. 17–25

[KFHP15] KUNZ, Michael ; FUCHS, Ludwig ; HUMMER, Matthias ; PERNUL, Günther: Introducing Dynamic Identity and Access Management in Organizations. In: *International Conference on Information Systems Security* Springer, 2015, S. 139–158

[MBPS10] MONT, Marco C. ; BERESNEVICHIENE, Yolanta ; PYM, David ; SHIU, Simon: Economics of identity and access management: Providing decision support for investments. In: *Network Operations and Management Symposium Workshops*, IEEE/IFIP, 2010, S. 134–141

[MC18] Maxim, Merritt ; Cser, Andras: Best Practices: Auswählen, Bereitstellen und Verwalten von Passwort-Managern für Unternehmen / Forrester. 2018. – Forschungsbericht

[MFL+18] Mauerer, Jürgen ; Freimark, Alex J. ; Lixenfeld, Christoph ; Reder, Bernd ; Schweizer, Michael: Studie Identity- & Access-Management 2 017 / IDG Business Media. 2018. – Forschungsbericht

[MS95] March, Salvatore T. ; Smith, Gerald F.: Design and Natural Science Research on Information Technology. In: *Decision support systems* 15 (1995), Nr. 4, S. 251–266

[ÖBF+11] Österle, Hubert ; Becker, Jörg ; Frank, Ulrich ; Hess, Thomas ; Karagiannis, Dimitris ; Krcmar, Helmut ; Loos, Peter ; Mertens, Peter ; Oberweis, Andreas ; Sinz, Elmar J.: Memorandum on design-oriented information systems research. In: *European Journal of Information Systems* 20 (2011), Nr. 1, S. 7–10

[Osm13] Osmanoglu, E.: *Identity and Access Management: Business Performance Through Connected Intelligence*. Elsevier, 2013. – ISBN 9780124081406

[PTRC07] Peffers, Ken ; Tuunanen, Tuure ; Rothenberger, Marcus A. ; Chatterjee, Samir: A design science research methodology for information systems research. In: *Journal of management information systems* 24 (2007), Nr. 3, S. 45–77

[RM05] Reijers, Hajo A. ; Mansar, S L.: Best practices in business process redesign: An overview and qualitative evaluation of successful redesign heuristics. In: *Omega* 33 (2005), Nr. 4, S. 283–306

[Roy13] Royer, Denis: *Enterprise Identity Management: Towards an Investment Decision Support Approach*. Springer Science & Business Media, 2013. – ISBN 9783642350399

[SCFY96] Sandhu, Ravi S. ; Coyne, Edward J. ; Feinstein, Hal L. ; Youman, Charles E.: Role-Based Access Control Models. In: *Computer* 29 (1996), Nr. 2, S. 38–47

[Sch17] Schwab, Klaus: *The Fourth Industrial Revolution*. Crown Publishing Group, 2017. – ISBN 9781524758868

[Sim73] Simon, Herbert A.: Applying Information Technology to Organization Design. In: *Public Administration Review* 33 (1973), Nr. 3, S. 268–278

[SL02] Slack, Nigel ; Lewis, Michael: *Operations Strategy*. Pearson Education, 2002. – ISBN 9781292017792

[Uni02] UNITED STATES CONGRESS: *Sarbanes-Oxley Act of 2002, PL 107-204, 116 Stat 745*. 2002

[VB17] VOIGT, Paul ; BUSSCHE, Axel Von d.: *The EU General Data Protection Regulation (GDPR)*. Bd. 18. Springer, 2017

[VK15] VAISHNAVI, Vijay K. ; KUECHLER, William: *Design Science Research Methods and Patterns: Innovating Information and Communication Technology, 2Nd Edition*. CRC Press, Inc., 2015. – ISBN 9781498715256

[Wat13] WATKINS, Steve G.: *An Introduction to Information Security and ISO 27001: 2013 A Pocket Guide*. 2. It Governance Ltd, 2013. – ISBN 9781849285261

[Win05] WINDLEY, P.J.: *Digital Identity: Unmasking Identity Management Architecture*. O'Reilly Media, 2005. – ISBN 0596008783

[WYSS09] WILLIAMSON, Graham ; YIP, David ; SHARONI, Ilan ; SPAULDING, Kent: *Identity Management: A Primer*. MC Press, LLC, 2009. – ISBN 9781583470930

[YS15] YASASIN, Emrah ; SCHRYEN, Guido: Requirements for IT Security Metrics - An Argumentation Theory Based Approach. In: *23rd European Conference on Information Systems (ECIS)*, 2015