

Book review

Personal Information Management. By W. Jones and J. Teevan. University of Washington Press (October 30, 2007) ISBN: 978-0-295-98737-8

Personal information – information that we own or have access to or information that other people or organisations have access to about us – is of increasing relevance in today's information-oriented society. Personal Information Management (PIM) as a research field deals with understanding the behaviours and problems associated with managing personal information of various kinds. This book is intended to be the first definitive reference for the field. It has four major aims: 1) to provide a grounding for researchers of different backgrounds interested in studying PIM. 2) to outline the scope of the problem area 3) to summarise the work that has already been performed and 4) to outline the many challenges facing the field in the future. These are bold aims considering the breadth of the research domain. However, with this book, Jones and Teevan achieve their aims admirably.

The book is comprised of 17 chapters written by scholars working in various areas relevant to PIM. These chapters are organised into four parts, with the first three parts focusing on personal information in the sense of information in an individual's possession. In the final part, the focus shifts from individuals and the management of their personal information to the management of personal information in a shared or group context, to situations where external actors are in control of information about other individuals. In this review I will address each chapter in turn, providing a short summary and critique for each before concluding with some general comments and an overview of the book's contributions.

Following an introductory chapter where the editors provide relevant definitions and outline the aims and structure of the book, part 1 deals with PIM as it is experienced and performed now. In the first chapter of this part (chapter 2), Teevan et al. discuss information finding activities, drawing on literature from the fields of HCI and information science, where there is a considerable body of work on information seeking. The act of finding new information is distinguished from *re-finding* information that has been previously created or accessed. The authors summarise the outcomes of studies that have investigated how people re-find information and explain clearly the different factors that have been shown to affect re-finding behaviour. The chapter concludes with future challenges, including discovering how new tools will impact on the ways that people re-find.

In chapter 3, Jones reviews research regarding the strategies people employ to keep and manage information. This is the area which has received the most research attention to date and there is a lot of relevant literature to cover. However, Jones skilfully presents his review in a way that not only demonstrates a thorough knowledge of the field, identifying and explaining the core findings, but in a way that does not overburden the reader with terminology and facts. Although a lot of work has already been performed on this area, this chapter illustrates that there is still much to do. Jones proposes broadening research efforts by learning lessons from related fields such as cognitive science and library science to discover better ways of organising, as well as improving representations of the organisations that people create.

In chapter 4, Marshall broadens the focus further still by considering the challenges involved in managing information over a life time. PIM is usually studied by investigating keeping and re-finding behaviours without considering potential changes in the technical and social contexts surrounding these activities. However, Marshall illustrates that when studying information management over longer periods these contextual changes must be taken into account and this introduces extra challenges. Not only the cognitive challenges of determining what information will be useful in 10, 20 or 50 years time, and technical challenges, such as changing information

formats, software applications and the fallibility of storage media, but legal and social issues must be addressed too. After outlining and discussing the challenges, Marshall illustrates the difficulties by presenting a seemingly simple case study of preserving the email communication between two people over a period of six years. It soon becomes clear that the example is not as simple as the reader first thinks. This chapter represents an intelligent and thought-provoking discussion of PIM in a wider context than is usually considered.

In chapter 5, Naumer and Fischer review naturalistic approaches to studying PIM. With the same thorough approach as the other chapters in part 1, they explain the motivation for using naturalistic methods, detail the fields from which the approaches originate and clearly outline the limitations and difficulties entailed in employing such techniques.

In the second part of the book, the focus is on potential solutions to the problems outlined in part 1. In each chapter from chapter 6-10, a set of authors argues for a particular solution based on their research. In chapter 6, Tan et al. advocate a “save all” approach where all of the information a user interacts with is stored for potential later use. The chapter is mainly centred on the authors' experiences with the MyLifeBits project. The goals of the project are outlined and some of the applied technologies explained. The chapter concludes with a short discussion of the legal and social challenges to this approach. Although this chapter lacks the thorough writing of the earlier chapters, the work described is ground-breaking, opening up new directions for research. The authors describe one of the many potential avenues for exploration with a case study to improve the quality of life of an individual with acute short-term memory problems.

Chapter 7 argues for a “structure all” approach to PIM. Catarci et al. propose that personal information should be structured in such a way that relationships are defined between different kinds of information objects. This chapter examines PIM from a novel perspective, introducing relevant work outside of the main body of PIM literature. Mainly drawing on database research, the authors explain the benefits of structured personal data sets. One problem with this approach is that relationships have to be defined and previous work suggests that users will not define these relationships themselves. The second part of the chapter deals with this issue, describing the research challenge of examining how relationships can be automatically harvested from data collections.

Chapter 8 argues for a strategy of unification. One of the main problems of PIM as it is currently performed is that people have different kinds information, in different formats and different locations. Karger reviews a variety of methods of unifying fragmented information including at a visual level, at a data level and at the level of meta-data. The meta-data level arguments very much align with those presented in chapter 7, but in this chapter they are presented through literature more familiar to the general PIM community e.g. the haystack project. Karger's case for unification is well argued and it sets the stage for the following two chapters which propose particular methods of unification.

In chapter 9, Russell and Lawrence argue for a “search-all” approach to unification. They describe the advantages of constructing a unified index of an individual's whole personal information collection and allowing the user to search over this index with free-text queries. Three desktop search systems are demonstrated and concepts such as scoping are well explained. The authors argue a strong case in favour of desktop search engines. However, what this chapter lacks is a detailed discussion of the limitations of search interfaces as they are at the moment. There is no discussion of the vocabulary mismatch problem, the fact that search interfaces do not offer any reminder functions, nor that previous studies have shown that in some cases people prefer to browse for information. Instead the authors discuss future challenges from an indexing perspective, such as indexing multi-media content such as images and videos.

In chapter 10, Whittaker et al. propose email as a means to unify PIM collections. Email usage has been extensively studied and this chapter provides a detailed review of literature, demonstrating that email is used not only as a communication tool, but for information, task and content management purposes. The authors discuss how email is used, why it is used for different purposes and the limitations of current email clients for these tasks. The chapter also outlines some of the directions taken to improve email management and discusses how potential changes to the email model such as charging to send emails may impact on how email is used as an information management tool. The chapter is well written, introducing the reader to the considerable body of work on email management and information management within email clients.

In the final chapter in part 2 the focus shifts from PIM solutions to solutions for evaluating PIM tools. The difficulties in performing evaluations are well documented and relatively few of the developed PIM prototypes have been evaluated. Kelly and Teevan present four approaches to evaluation: naturalistic, longitudinal, case-study and laboratory. Citing previous work to exemplify different methods, the authors provide an accomplished overview of what can be learned from applying different techniques. Kelly and Teevan argue for a mixed approach to evaluation, combining different methods to learn about various aspects of behaviour and tool performance. This chapter demonstrates that effectively evaluating PIM is still an open challenge. Nevertheless, this chapter is an excellent resource for anyone considering evaluating a PIM tool.

Part 3 of the book is titled “PIM and the individual” and consists of two chapters. The first is a novel review of PIM behavioural literature by Gwizdka and Chignell. The authors present PIM literature in such a way that illustrates how different factors impact on peoples' information management strategies. First, external factors such as profession are discussed. This is followed by a review of factors that differentiate behaviour within groups of people, such as the filing strategies employed. Finally, the authors identify and discuss various factors that have been shown to impact on how an individual will manage their information, e.g. whether they are at work or at home, their goals and objectives and factors such as individual cognitive abilities. This chapter was a fascinating read and provided a novel and insightful presentation of core PIM literature.

In chapter 13, Moen discusses the management of personal healthcare information. She argues that current health systems place a great burden on patients to manage their health information for different aspects of their care. Again, information fragmentation is highlighted as a key problem with patients requiring to manage different types of information in different formats. The main focus of the chapter is on how health information is managed in the home. Describing the findings of one illuminating study in which she was involved, Moen reveals what information management behaviours people exhibit, how these behaviours are performed and what the findings mean in terms of future tool support for personal healthcare information management (PHIM). This chapter revealed many similarities between the difficulties and behaviour exhibited for health and more general information management. However, there were also some striking differences between the observed PHIM behaviour and the findings of the core PIM studies, such as the participants reluctance to manage information digitally. Moen argues a strong case for further research into how healthcare information is managed and how current behaviours can be better supported.

A key finding in chapter 13 was that in the case of health information management, there is usually one person who takes responsibility for the management of information within a household. Building on this, chapter 14 relates to situations where personal information is shared and managed in group contexts. Examining group information management (GIM) behaviour means considering aspects not usually studied in PIM, such as varying roles, goals and incentives across different members of a group. Drawing on computer-supported cooperative work (CSCW) literature and using several examples, including GIM in the home, Lutters et al. explain the factors impacting on

and the challenges involved in managing information within groups. Like chapter 4, this chapter broadens the focus of traditional PIM research and opens up new directions and new challenges for researchers.

The final two chapters concentrate on personal information held by third parties. Chapter 15 examines this from a security aspect. It argues that as technology and the ways that it integrates into our lives are changing, there is a greater need to divulge information to third parties. This means that there is a requirement for better policies and measures to protect users' privacy and prevent misuse of personal data. Central to the authors' argument is that there is a relationship between a user's privacy, security and trust of the system. Reviewing studies of privacy and usability, it is demonstrated that a user's trust is engendered by a sense of control over what information is released and how that information is used. The chapter concludes with a call for researchers to investigate usability issues with the creation and implementation of privacy policy.

Chapter 16 focuses on the privacy issues concerning personal information stored within public records. Starting with a somewhat shocking example of how personal information records are made available to the general public in some American states, Shamos explains the legal and practical aspects of managing personal information for public use. He outlines how technology has changed the ways in which information can be accessed and how this affects the privacy of individuals. He continues to explain different methods of dealing with the problem and discusses what the approaches mean for the various people involved. This chapter is a fascinating read, which illustrates plainly some of the privacy issues involved in managing personal information.

Overall this book represents a marvellous resource for PIM researchers. It covers PIM from a number of different perspectives and emphasises the many and wide ranging challenges that need to be addressed including legal, social, cultural and technical challenges. The chapters are concise and well written and the main arguments are clear and easy to understand. Although each chapter is written by a different set of authors and covers a wide range of domains, the editors have successfully taken measures to unify the presentation of the material. For example, most of the chapters exemplify their arguments with respect to a particular family that is introduced in the first chapter. Not only does this create a unified presentation, but it also re-enforces the fact that PIM is an everyday problem for most people – a fact that is not always obvious when reading PIM literature. The main strength of the book however, is the selection of authors chosen to write the chapters. These scholars are experts in a wide diversity of areas relating to PIM and have contributed not only with great knowledge and academic experience, but with no small amount of passion. This is felt throughout the book and is reflected in the reading experience.

One criticism I have is regarding the way that the content is structured in the book. As PIM relates largely to personal activities, it seems strange to have a section dedicated to “PIM and the individual” (part 3) and only have two chapters in this part. This is especially confusing as the first four chapters of the book all deal with PIM from an individual's perspective. Further, the second chapter in the “PIM and the individual” section – the chapter that dealt with health information management – explained that many PHIM activities involve managing information for other people. I also felt that both evaluation chapters (chapters 5 and 11) should have been located together. Chapter 5 did not really fit very well in the first part of the book, which discussed PIM behaviour and difficulties. However, these structural issues do not detract from the quality of the content, nor from the usefulness of the book.

To summarise, Jones and Teevan have created a wonderful resource for the PIM community. I would recommend this book to researchers, teachers and students. I would especially recommend the book to young researchers looking for ideas for research projects. Each chapter of this book suggests research challenges, many of which would make excellent PhD topics.