The Semantic Desktop as a foundation for PIM research

Leo Sauermann, Gunnar Grimnes
German Research Center for Artificial Intelligence DFKI GmbH
Kaiserslautern, Germany
{leo.sauermann,gunnar.grimnes}@dfki.de

Thomas Roth-Berghofer
DFKI GmbH and
Department of Computer Science
University of Kaiserslautern, Germany
thomas.roth-berghofer@dfki.uni-kl.de

ABSTRACT
Personal Information Management (PIM), both in science and in applications, is limited by the current approach to data management (files) and applications. The Semantic Web and the Resource Description Framework (RDF) provide a standardized way to represent data, across applications and capture the respective application semantics. In recent years we have developed the *Semantic Desktop* approach, a semantic layer of personal computers that serves as middleware to integrate applications and their data. The user uses the Semantic Desktop to create a *Personal Information Model* (PIMO), a formal representation of the mental model of the user, and this PIMO is used to integrate various elements from the workspace. *In this paper, we focus on what the Semantic Desktop can offer for research in Personal Information Management.*

Author Keywords
personal information management, pim, ontologies, semantic desktop, explanation

ACM Classification Keywords
H.1.0, H.3.0, H.5.4, I.2.4

INTRODUCTION
The need for Personal Information Management (PIM) originates from the limited capacity of the human mind and the wish for extending human memory in order to reduce the cognitive load for the knowledge worker.

> "The human mind (...) operates by association. With one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts, in accordance with some intricate web of trails carried by the cells of the brain."
> *Vannevar Bush, "As we may think", 1945* [8]

Tools supporting PIM often lack one ability: they cannot manage all information a knowledge worker needs, as they are unable to integrate data across applications and information sources. In [6] Boardman has shown the need for cross-application support for PIM. Using Boardman's definition of PIM: "Personal Information Management can be defined as the management of personal information as performed by the owning individual", we take *personal information* to mean the whole personal knowledge space of an individual. The key points being that the personal knowledge workspace embraces all data "*needed by an individual to perform knowledge work*", and this knowledge space is both independent from the way the user accesses the data, and independent from the source, format, and author of the data. Previously the cross-application support for PIM was achieved by integrated groupware applications and adapters to these, but we believe that this can only offer a partial solution, and to truly solve the problem we need a PIM framework that reaches beyond particular applications and devices. Such a framework can be created by the use of open standards for representing formalised knowledge.

Semantic Web
"The Semantic Web provides a framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF). "[1]

RDF [13] is used as a standard to express information about resources. Resources are first and foremost web resources, HTML documents and multimedia files on the web, but can also be things from the real world, such as products or persons. The only requirement for describing something using RDF is that the resource are identifiable by a Uniform Resource Identifier (URI). RDF expresses facts about resources using *statements*, each statement containing a subject resource, a relationship (or property) and an object resource. RDF data can be stored in files, databases and is designed to be published on the web.[2]

Semantic Desktop
The idea of the Semantic Desktop was first mentioned in [27, 9], a definition was given in [28]:
"A Semantic Desktop is a device in which an individual stores all her digital information such as documents, multimedia, and messages. These are interpreted as Semantic Web resources, each is identified by an URI and all data is accessible and queryable as RDF graph. Ontologies allow the user to express personal mental models and form the semantic glue interconnecting information and systems, and Semantic Web protocols are used for inter-application communication. The use of Semantic Web standards allows existing web resources to be incorporated into the personal knowledge space, and does also facilitate the sharing of knowledge with others, for example within a work-group."

[1] Taken from `http://www.w3.org/2001/sw/`
[2] Refer to [29] for URIs and [5] to learn about publishing RDF data

The Semantic Desktop approach does not require to replace existing systems, but enhances them with the capabilities of the Semantic Web ensuring compatibility amongst applications and that knowledge created in one place is reusable in different applications. The Semantic Desktop is not restricted to one user-interface metaphor, or one way of storing data, but rather a common ground to integrate best practices. Based on the semantic data storage, information retrieval can go beyond text search. The system should be able to answer questions such as "name participants of last week's meetings who did not write minutes" or "the document about accounting that was sent by a project participant before the last review". To realize this, the system also has to provide best-efforts to interrelate and integrate information for the user.

In [34] we find descriptions how to map multiple ontology layers to create *personal information applications*. Semexin [11] is another approach, as is Haystack [24] which provides a complete user interface abstraction language and an integrated application for multiple tasks. Gnowsis [30] is a prototype that provides an architecture for integration, in the same direction works the ongoing NEPOMUK project [16], which is the biggest endeavor in the area at the moment. More projects are listed in [28] and on the community wiki.[3]

Ontologies

The established definition of ontologies is given by Gruber in [17]: "An ontology is an explicit specification of a conceptualization."
The W3C has made recommendations how to express ontologies, a simple formalization of classes and properties is RDF-Schema [7], the Web Ontology Language (OWL) extends this with description-logic [22]. To represent taxonomies (hierarchies) and thesauri there is the SKOS standard [12]. With these languages, ontologies can be created to formalize a domain of interest: for example, people and their possible relationships are formalized in the FOAF ontology[4], and Dublin Core[5], which can be used for documents and their metadata. An important feature of RDF, is that data in different ontologies can be freely combined within one data-store, making data-merging from different sources much easier.

On the Semantic Desktop, ontologies are used to model several layers[6].

- Information Ontology (NIE): Represent files, e-mails, websites, multimedia data, and various other information elements in a way comparable to the standardization of MIME-types.

- Personal Information Model (PIMO): Represent the personal view of the user on his world, personal annotations, what elements is the user aware of.

- Ontologies shared among users: information about groups, domains of interest, collaborative work or organizational memories. [1].

Additional to these layers, various ontologies exist to formalize different aspects of knowledge work, such as *task management* [19], *the current work context of the user* [33], the social surroundings [18], etc.

Ontologies exhibit two desirable features: Firstly, by publishing formal representations one allows other researchers to conduct interesting work on existing data (such as Mika's award winning paper [23]). Secondly, they form a market: multiple competing formalizations (views) about the same facts can be published and discussed. Scientific discussion about facts can be expressed both in text and in formalization.

A personal Model and View on Ontologies

Ontologies are well accepted as a mean to express shared knowledge, for PIM purposes the formal notations can be reused, but allowing a subjective view. Individuals can make use of the upper-level concepts location, person, time, or topic [21, 10] to categorize documents. Users need the freedom to create personal, subjective views that can be used in addition to shared ontologies. We have defined the "Personal Information Model" as such a categorisation scheme [31]. It makes use of a formal language and defines the needed upper-level concepts. Other researchers criticised this conceptualisation as well as extended and used it [14].

Applications Augmented with the Semantic Desktop

In [33] we find examples of how the user can be supported in finding and reminding information based on on what context the user is currently doing his work. A sidebar shows topics related to the current work of the user, a note-taking application uses the current work context to annotate and retrieve documents. Semantic search applications such as Beagle++ [20] support information retrieval. Other applications are referenced on the community wiki.

OFFERINGS TO THE PIM COMMUNITY

Looking at criteria that have been applied in PIM applications, like timelines (lifestreams) [15], or spatial memory [25], or the distinction between ephemeral and archived [3], we note that metadata is a key issue to improve PIM. Time, place, and distinctions by use are all examples of metadata of the resources in question. Similar is the need to share metadata about resources across applications, as sketched above in the PIMO, and published in [6]. Hence, the representation and availability of metadata integrating various views is crucial for efficient and correct PIM.

Ontologies to Represent Metadata

Using ontologies and the Semantic Desktop, it is possible to build upon the existing identification mechanisms (URIs), the metadata storage (RDF) and the formal description of the data (ontologies). Building on this basis will help integrate the various PIM approaches.

[3] urlhttp://www.semanticdesktop.org
[4] `http://xmlns.com/foaf/spec/`
[5] `http://dublincore.org/documents/dcmi-terms/`
[6] see an overview at `http://www.semanticdesktop.org/ontologies`

In the EPOS project, and in NEPOMUK, we noticed that the effectiveness of research was increased by using ontologies. Each ontology can reuse modeling effort of the others: e.g., a researcher working on "user's work context" can focus on the issues surrounding representation of *attention* and *context*, and can simply reuse the fundamental concepts of the domain as modelled by others.

Ontologies to Discuss Research Results

In PIM, various fields come together: cognitive science, psychology, user interface design, ubiquitous computing, information management, databases and knowledge representation. Research results from each field can be expressed in articles, implementations and demos, and as the languages used are differing (for example: context means something different in psychology and in artificial intelligence). Like in mathematics, where researches used individual notations until a language evolved to express proofs and formulas in the 18th century, and was used to communicate amongst researchers. As an addition to scientific articles, we propose to use ontology languages (RDFS, OWL, and SKOS) to express results.

Semantic Web to Store and Share Information

For some aspects of PIM, such as ubiquitous computing, knowledge about the physical location of the user and surrounding objects and people have to be taken into consideration. The Semantic Web can be used to connect information to physical objects by assigning them URIs as unique identifiers . Via RDF and HTTP, it is then possible to get descriptions of the physical entities. For example, the model and capabilities of a device can be stored, or the name and organizational data about a person.

Explanation support

Explanations, in general, are answers to questions and are the most common method used to support humans in decision making [32]. Tim Berners-Lee's "Oh yeah?" button [4] is an early acknowledgement of the need for questioning results of a Semantic Web application, but the idea is of course equally useful for any knowledge-based system. The integration of information from various sources, of various quality, into a PIM system adds additional requirements for explaining the provenance and trust-worthiness of information.

The task of information management is a communication task among humans. Thus, *personal* information management needs to include PIM tools into the communication processes. Providing explanations is very good way to improve on interactions between humans and knowledge-based applications [32, 26, 2]. The available explicit knowledge on the Semantic Desktop in form of the PIMO provides excellent opportunities for explaining personal PIM processes and to subsequently learn more about PIM processes in general.

CONCLUSION

We believe that *Personal Information Management* is currently limited by the underlying technology. As shown by Boardman, effective PIM must be cross-application and facing the disappearing desktop, cross-device. We have shown that Semantic Web technologies, especially as used in the Semantic Desktop field, can help solve these problems. Using RDF to represent personal knowledge allows each user to create his personal view on the world, integrating information from all their existing applications; and does also facilitate collaborative work, both within closed groups and for the web as a whole.

Acknowledgements

This work was supported by the European Union IST fund (Grant FP6-027705, project NEPOMUK).

REFERENCES

1. Andreas Abecker, Ansgar Bernardi, Knut Hinkelmann, Otto Kühn, and Michael Sintek. Toward a Technology for Organizational Memories. *IEEE Intelligent Systems*, June 1998.
2. Daniel Bahls and Thomas Roth-Berghofer. Explanation support for the case-based reasoning tool mycbr. In *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence. July 22–26, 2007, Vancouver, British Columbia, Canada.*, pages 1844–1845. The AAAI Press, Menlo Park, California, 2007.
3. Deborah Barreau and Bonnie A Nardi. Finding and reminding: File organization from the desktop. 1995.
4. Tim Berners-Lee. Cleaning up the user interface, 1997. `http://www.w3.org/DesignIssues/UI.html` [Last access: 2007-10-18].
5. Christian Bizer, Richard Cyganiak, and Tom Heath. How to publish linked data on the web. http://sites.wiwiss.fu-berlin.de/suhl/bizer/pub/LinkedDataTutorial/. Tutorial.
6. Richard Boardman. *Improving Tool Support for Personal InformationManagement*. PhD thesis, Department of Electrical and Electronic Engineering Imperial College London University of London, July 13 2004.
7. D. Brickley and R.V. Guha. Rdf vocabulary description language 1.0: Rdf schema. w3c recommendation 10 february 2004. `http://www.w3.org/TR/rdf-schema/`.
8. Vannevar Bush. As we may think. *The Atlantic Monthly*, 176(1):p101–108, July 1945.
9. Stefan Decker and Martin Frank. The social semantic desktop. In *Proc. of the WWW2004 Workshop Application Design, Development and Implementation Issues in the Semantic Web*, 2004.
10. Andreas R. Dengel. Six thousand words about multi-perspective personal document management. In *Proc. EDM IEEE Workshop*. IEEE, Oct 2006.

11. Xin Dong and Alon Y. Halevy. A platform for personal information management and integration. In *Proc. of the CIDR Conference*, pages 119–130, 2005.

12. Alistair Miles (edt). Simple knowledge organisation system (skos). `http://www.w3.org/2004/02/skos/`, Feb 2004.

13. Editors F. Manola, E. Miller. Rdf primer. W3c recommendation, 2004, W3C, http://www.w3.org/TR/2004/REC-rdf-primer-20040210/, 10 February 2004.

14. Norberto Fernandez-Garcia, Leo Sauermann, Luis Sanchez, and Ansgar Bernardi. Pimo population and semantic annotation for the gnowsis semantic desktop. In *Proceedings of the Semantic Desktop and Social Semantic Collaboration Workshop at the ISWC*, volume 202 of *CEUR-WS*, 2006.

15. Eric Freeman and David Gelernter. Lifestreams: A storage model for personal data. *SIGMOD Record (ACM Special Interest Group on Management of Data)*, 25(1):pp80, 1996.

16. Tudor Groza, Siegfried Handschuh, Knud Moeller, Gunnar Grimnes, Leo Sauermann, Enrico Minack, Cedric Mesnage, Mehdi Jazayeri, Gerald Reif, and Rosa Gudjonsdottir. The nepomuk project - on the way to the social semantic desktop. In Tassilo Pellegrini and Sebastian Schaffert, editors, *Proceedings of I-Semantics' 07*, pages pp. 201–211. JUCS, 2007.

17. Thomas R. Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2):199–220, 1993.

18. Dominik Heckmann, Tim Schwartz, Boris Brandherm, Michael Schmitz, and Margeritta von Wilamowitz-Moellendorff. Gumo - the general user model ontology. In *10th International Conference on User Modeling (UM 05), Edinburgh*, 2005.

19. Harald Holz, Heiko Maus, Ansgar Bernardi, and Oleg Rostanin. From Lightweight, Proactive Information Delivery to Business Process-Oriented Knowledge Management. volume 0, pages 101–127, 2005.

20. Tereza Iofciu, Christian KohlschÃ¼tter, Wolfgang Nejdl, and Raluca Paiu. Keywords and rdf fragments: Integrating metadata and full-text search in beagle++. In Stefan Decker, Jack Park, Dennis Quan, and Leo Sauermann, editors, *Proc. of Semantic Desktop Workshop at the ISWC, Galway, Ireland, November 6*, volume 175, November 2005.

21. Khalid Latif and A Min Tjoa. Combining context ontology and landmarks for personal information management. In *Proceedings of International Conference on Computing and Informatics (ICOCI)*, Kuala Lumpur, Malaysia, June 2006.

22. D. L. McGuinness and F. Harmelen. Owl web ontology language overview w3c recommendation 10 february 2004. http://www.w3.org/TR/owl-features/.

23. Peter Mika, Michel Klein, , and Radu Serban. Semantics-based publication management using rss and foaf. In Stefan Decker, Jack Park, Dennis Quan, and Leo Sauermann, editors, *Proc. of Semantic Desktop Workshop at the ISWC, Galway, Ireland, November 6*, volume 175, November 2005.

24. Dennis Quan, David Huynh, and David R. Karger. Haystack: A platform for authoring end user semantic web applications. In *International Semantic Web Conference*, pages 738–753, 2003.

25. George Robertson, Mary Czerwinski, Kevin Larson, Daniel C. Robbins, David Thiel, and Maarten van Dantzich. Data mountain: using spatial memory for document management. In *UIST '98: Proceedings of the 11th annual ACM symposium on User interface software and technology*, pages 153–162, New York, NY, USA, 1998. ACM Press.

26. Thomas R. Roth-Berghofer. Explanations and Case-Based Reasoning: Foundational issues. In Peter Funk and Pedro A. González-Calero, editors, *Advances in Case-Based Reasoning*, pages 389–403. Springer-Verlag, September 2004.

27. Leo Sauermann. The gnowsis-using semantic web technologies to build a semantic desktop. Diploma thesis, Technical University of Vienna, 2003.

28. Leo Sauermann, Ansgar Bernardi, and Andreas Dengel. Overview and outlook on the semantic desktop. In Stefan Decker, Jack Park, Dennis Quan, and Leo Sauermann, editors, *Proceedings of the 1st Workshop on The Semantic Desktop at the ISWC 2005 Conference*, 2005.

29. Leo Sauermann, Richard Cyganiak, and Max Völkel. Cool uris for the semantic web. Technical Memo TM-07-01, DFKI GmbH, Deutsches Forschungszentrum f ür Künstliche Intelligenz GmbH Postfach 2080 67608 Kaiserslautern, February 2007. Written by 29.11.2006.

30. Leo Sauermann, Gunnar Aastrand Grimnes, Malte Kiesel, Christiaan Fluit, Heiko Maus, Dominik Heim, Danish Nadeem, Benjamin Horak, and Andreas Dengel. Semantic desktop 2.0: The gnowsis experience. In *Proc. of the ISWC Conference*, pages 887–900, Nov 2006.

31. Leo Sauermann, Ludger van Elst, and Andreas Dengel. Pimo - a framework for representing personal information models. In Tassilo Pellegrini and Sebastian Schaffert, editors, *Proceedings of I-Semantics' 07*, pages pp. 270–277. JUCS, 2007.

32. Roger C. Schank. *Explanation Patterns: Understanding Mechanically and Creatively*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1986.

33. Sven Schwarz. A context model for personal knowledge management applications. In Thomas Roth-Berghofer, Stefan Schulz, and David B. Leake, editors, *Modeling and Retrieval of Context, Second International Workshop, MRC 2005, Edinburgh, UK, July 31 - August 1, 2005, Revised Selected Papers*, volume 3946 of *Lecture Notes in Computer Science*, pages 18–33. Springer, 2006.

34. Huiyong Xiao and Isabel F. Cruz. A multi-ontology approach for personal information management. In Stefan Decker, Jack Park, Dennis Quan, and Leo Sauermann, editors, *Proc. of Semantic Desktop Workshop at the ISWC, Galway, Ireland, November 6*, volume 175, November 2005.