Semantic Commonalities of Research Networking and PIM

David Eichmann
Institute for Clinical and Translational Science
The University of Iowa
Iowa City, Iowa 52242
david-eichmann@uiowa.edu

ABSTRACT

Research networking systems (RNSs) have recently evolved into interesting and valuable complements to personal information management (PIM) tools. Typically operating at an institutional or larger scale, research networking tools provide a ready source of information currently challenging to incorporate into PIM without substantial effort. This paper identifies a number of common areas of interest between the two approaches from a perspective of the representational semantics employed by each.

Categories and Subject Descriptors

H.3.3 **[Information Storage and Retrieval]** Miscellaneous

Keywords

research profiling, research networking, personal information management

1. INTRODUCTION

As the role of collaboration becomes increasingly critical to the conduct of science, the ability of researchers to maintain competitive visibility among their peers becomes increasingly dependent upon what information regarding a given researcher is available to those peers. Resources such as PubMed[1] are perceived as critical to the biomedical research community, but offer only a partial representation of researcher expertise and interest, viewed through the lens of the publication record.

The additional challenge of translational science, i.e. migrating scientific knowledge from one discipline to another, has led to substantial investment in infrastructure to support interchange of what is, at its core, information. The National Institutes of Health Clinical and Translational Science Awards (CTSA) program was designed to address just this challenge. Research networking and the systems to support it have emerged as a major focus for the CTSA Consortium and related projects. The relevant affinity group within the consortium as been a major driver in innovation in this domain [25].

I adopt the following definition:

> "Research Networking Systems (RNS) are systems which support individual researchers' efforts to form and maintain optimal collaborative relationships for conducting productive research within a specific context." [20]

and argue that RNSs are social frameworks in that they are intended to construct and maintain a social fabric for one or more research communities. I further make a strong distinction between social networking platforms (e.g. FaceBook), professional networking platforms (e.g., LinkedIn) and RNSs in the nature of the information being managed and particularly in its provenance. Furthermore, RNSs carry multiple senses of information as characterized by Jones [8] - control, aboutness and provisioning.

Research networking is an increasingly digital activity, as is social networking as defined in the context of this workshop. The ubiquity of ample network bandwidth has resulted in researchers collaborating in teams that span time-zones and continents. User behaviors such as those characterized in [9] are ill-suited to the tasks in this space, where tele-conferences have supplanted the conference room and papers are commonly written without the authors ever meeting face-to-face. Efficient and effective use of multiple sources of information managed by a hybrid PIM-RNS environment offers substantial potential to info-stressed researchers.

This paper considers the nature of RNSs and PIM systems from a perspective of the semantic representations used by both and the readily identified connections that can be made from that perspective. I then describe our experience in the development and operation of the Loki RNS and how, in opting for a researcher-centric model for our architecture, we have laid the groundwork for shared information management between these two domains.

2. A PIM-CENTRIC PERSPECTIVE

My previous work on the boundaries of PIM and GIM [2] focused on shared annotation, a form of 'superimposed information' [18]. Research groups (e.g., the members of a lab) frequently have a need to collaboratively manage a collection of papers and their perspectives on the research related therein. Whalen, et al [27] focus on shared creation/editing and observe the phenomenon of what they term the 'disappearing desktop' - where the sharing environment effectively becomes the functional equivalent of the desktop for a given user (e.g., mailing oneself the current document). The adoption of RNSs and eCV systems (server-based curriculum vitae management systems) are resulting in a similar lifting of researcher identity off the desktop and into the 'institutional cloud' - where one's CV is no longer a document to be physically managed but instead a multi-layered construct residing in an institutional database.

Technology innovation is accelerating the creation of private distributed (frequently mobile) environments - splintered, limited-functionality micro-contexts for PIM. I personally run OmniFocus on a

1. http://www.ncbi.nlm.nih.gov/pubmed

desktop, a laptop, a tablet and a phone, utilizing the various incarnations of the tool in platform-specific manners. I additionally have cron jobs harvest newly-downloaded PDFs from my browser caches and structure a temporally-ordered archive of all documents acquired, which is synchronized with a PDF reader on the tablet. The resulting environment still requires attention, task triage and does little to support the provisioning of PIM into the boundary with GIM. Tungare and Pérez-Quiñones' [22] exploration of cross-device access patterns and their connection to issues relating to the social sphere offers potential here, but significant behavior modification is still required to maintain information currency.

Tools should not attempt to radically change user behavior in a single step, but rather adopt an evolutionary approach [24]. Evéquoz and Lalanne [6] enumerate common user problems in PIM:

- Classification load - even with a tool available, tagging artifacts into categories is too time-consuming.
- Information fragmentation - artifacts are scattered across multiple contexts and do not propagate.
- Collaborative PIM - teaming relationships inherently lead to shared ownership of artifacts and varying manipulation of those artifacts by team members

and the resulting user needs from PIM:

- Browsing facets and information linking - varying information structuring and relationship identification requirements.
- Ubiquity versus fragmentation - the need for support of resolving mobile availability of artifacts (e.g., calendars) against the existence of a single shared authoritative artifact.
- Usage history - what has been used and when was it used?

Similar challenges relating to the 'disappearing desktop' are raised by Komninos et al [13], and leads to Kirsh's observation of the appearance of multiple personal spaces as corollary to fragmentation problem [12].

Tungare et al [21] note the role of meta-data (and loss thereof) in the ease or difficulty in identifying multiple instance of the same artifact. A key challenge for PIM lies in addressing this lossy information replication while still addressing the need to expose personal information in exchange for focused and individualized information services [16].

Groth and Eklundh [7] define a continuum of information sharing with increasing visibility:

- Extra-organizational: publications, courses, contact information
- Intra-organizational: the above, plus content
- Project: the above, plus internal information regarding project state
- Individual: the above, plus personal notes

This scheme provides an initial point of compatibility with the information management focus of RNSs, as described in the next section. There are representation gaps in Groth and Eklundh's approach, however. Enrichment of the model must account for the activities that put this information to use - i.e., process. Attention to process (task) support as well as information is critical to a robust PIM environment [1, 4, 15].

3. AN RNS-CENTRIC PERSPECTIVE

Unlike much of the PIM work in artifact management, research networking entities are typically managed only in surrogate form (e.g., a citation to a publication rather than a copy of that publication). Furthermore, unlike the PIM systems in the previous section, where information provisioning and sharing occurs at the micro level (i.e., the individual and the group), RNSs focus at the macro level (i.e., intra- and extra-organizationally).

3.1 The Research Profile

The primary modeling focus in RNSs is the researcher, or more precisely, the research profile - a representational surrogate for a researcher at the institution operating the RNS. Arising from a need to move beyond the ad hoc nature of personal web pages, research profiles are, to a significant extent, an attempt to normalize the projected information about the institution's researcher population as a whole, to reduce the effort level for publication maintenance and to improve the interlinkage of the represented population. Modeling elements are commonly restricted to researcher demographics (appointments, email address, etc.) and productivity-derived elements such as publication citations and grant award citations. The risks here relate to institutionally-driven systems with little researcher buy-in. The data can be correct, but when limited to only institution-blessed information sources (e.g., licensed citation databases and funding data from federal agency databases), with little researcher engagement beyond data validation, these data can be homogenized to such an extent that there appears little relationship to PIM. I advocate here that a researcher-centric perspective on RNS information leads to the potential for rich overlap with PIM perspectives.

3.2 An Architecture Common to RNSs

Figure 1 illustrates a commonly implemented RNS architecture. Information regarding researchers is harvested from multiple sources, some institutional-specific and -private (e.g., human resources databases) and some open (e.g., publication citation databases) and marshalled into an integrated RNS-specific data model. The monolithic nature of this architecture derives from the institutional focus in its design and implementation - such implementations run substantial risk in the platform having little or no researcher ownership of the managed data. Successful representative systems for this architecture include Profiles[2] and VIVO[3].

3.3 A PIM-friendly Modular RNS Architecture

We have adopted a more user-centric, and hence more PIM-friendly approach in our RNS research. Figure 2 illustrates the modular architecture of the Loki[4] RNS. In this approach, information ingest is kept to a minimum, focusing instead on direct connections to authoritative sources. When necessary or advantageous, as in the case of demographics data in Loki, information ingest may be employed to minimize adverse impact on those authoritative sources. We explicitly model the application in the Loki architecture as two distinct layers to accommodate modular composition of new information sources. Elements such as the MEDLINE tag library are commonly shared across multiple applications, with each application model extending its vocabulary of tags as needed through inclusion of the relevant tag libraries.

Loki has been deployed at the University of Iowa since 2007 and currently manages ~3000 profiles. Researchers are encouraged to take ownership of their profile through features such as supervised

2. http://connects.catalyst.harvard.edu/profiles
3. http://vivoweb.org
4. http://www.icts.uiowa.edu/Loki

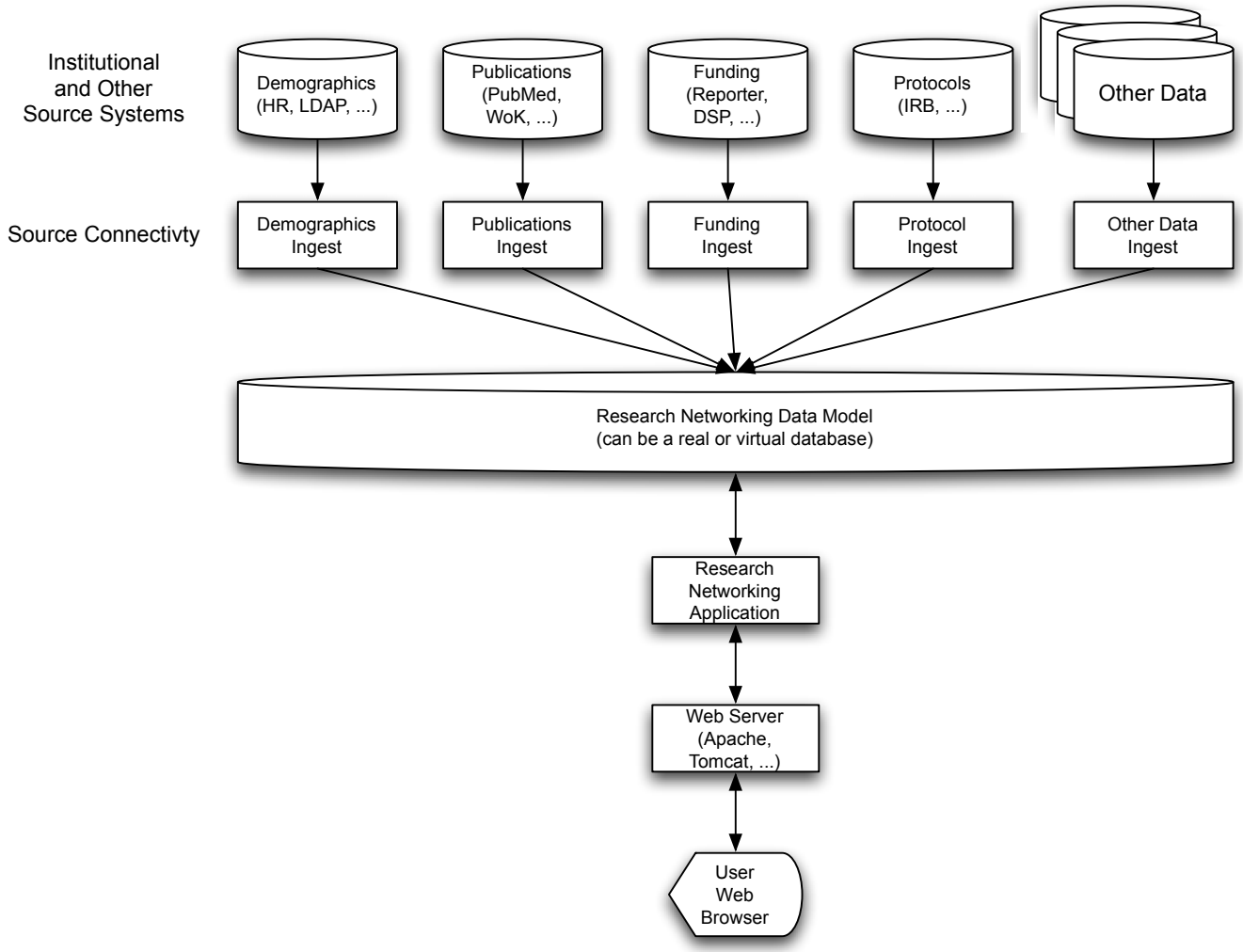

Figure 1: A Common RNS Architecture

filtering of publication streams, personalized narratives and customized awareness feeds regarding funding opportunities.

4. SEMANTIC COMMONALITY

Stepping back to consider the previous two sections, we find an interesting duality in approaches and architectures. PIM systems in their focus in the individual provide substantial benefit to a research community in aiding in management of the resulting micro-contexts. RNSs in their focus primarily in the institution and the societal aspects of research communities offer economies of scale and information completeness. Blending these strengths supports an integrative approach to federating PIM tools into the overall RNS environment, as illustrated in Figure 3. Modularity in representation supports projection of the individual (alá Kirsh) into the shared environment and projection of the societal (e.g., funding opportunities) into the personal environment. The remainder of this section presents challenges to this approach, and motivating use cases.

4.1 Knowledge Representation and Semantics

Moving from the discrete information management of a single PIM tool to a framework of systems interoperating across a broad range of granularities requires formalism that allows for innovation and localization while still supporting dynamic composition of systems. Requiring a common design and data model presents as much of a hurdle to development as such as system would likely require in learning from its users (see my earlier comments on the work of Voit et al [24]).

4.1.1 Ontologies: Personal, Group and Societal

Katifori et al [10] and Catarci et al [4] argue for personal ontology creation and management. I agree that use of such formalism, while likely challenging in design, will have the needed flexibility to manage the ad hoc integration needed with systems operating at RNS scale. Our work in extending Loki with an ontology and the mapping of that ontology into that used by VIVO leads me to believe that such semantic composition of PIM and RNS knowledge representation is not only feasible, it can be quite direct [5].

4.1.2 Semantic Web

Extending both domains (PIM and GIM) into the semantic web can overcome many of the barriers to ready composition of PIM data from disparate sources, particularly the cost, both cognitive and

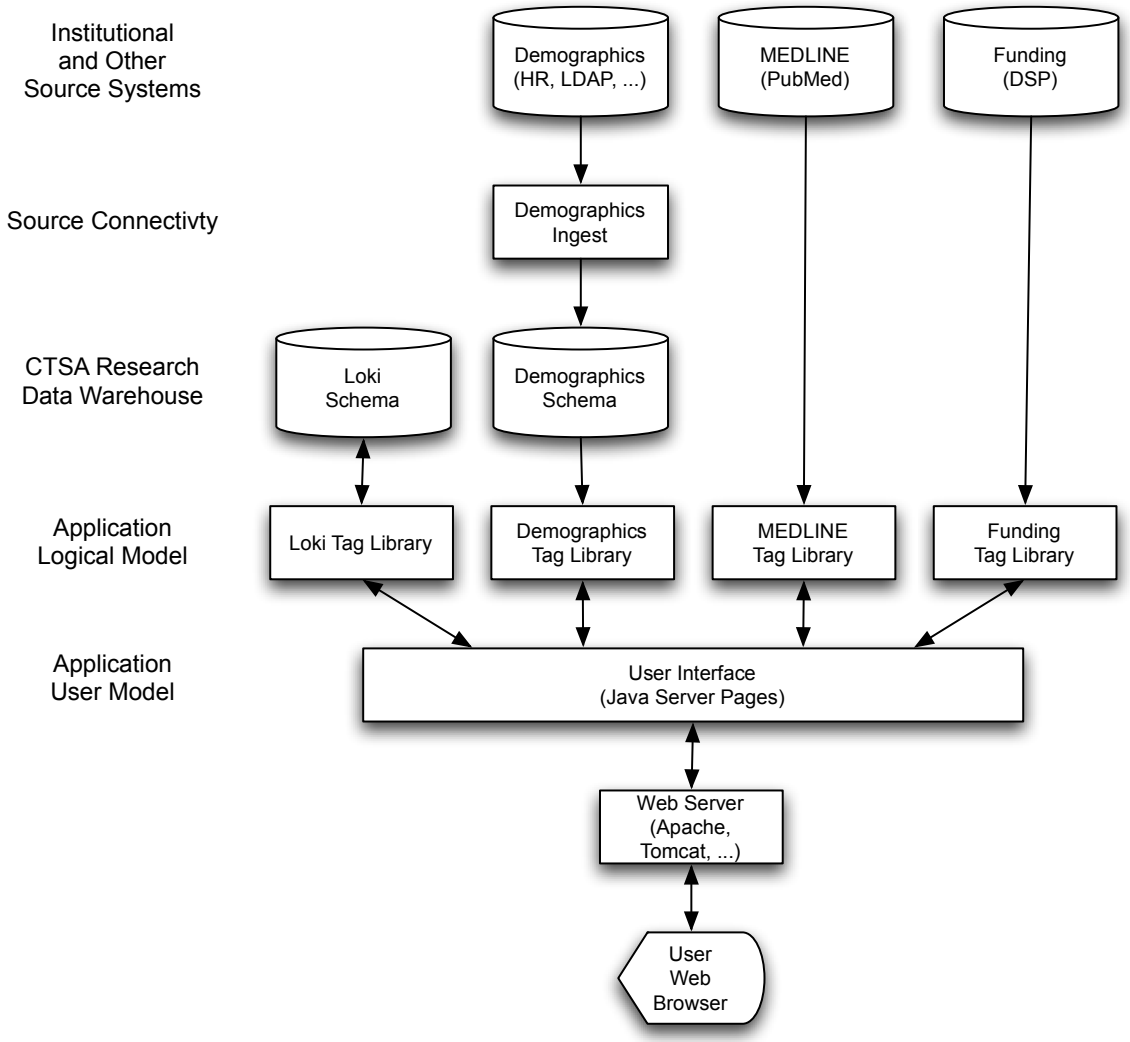

Figure 2: The Loki RNS Architecture

computational in transforming that data from its original context into the user's PIM context. Semantic Web technology offers substantial potential to mitigate these costs. Sauermann et al [19] explore the concept of a semantic desktop supporting PIM in a manner quite similar to that now being adopted by the RNS community through the VIVO collaboration [23]. The careful construction of an ontological framework is critical to maintaining distinct representation for both digital artifact and the corresponding surrogates, as well as capability to operate at varying degrees of granularity, including fine-grained knowledge artifacts [17]. The concept of a semantic file system proposed by Krishnan et al [14] holds additional potential for metaphor construction supporting users new to semantic web concepts.

4.2 RNS Modalities Supporting PIM

A number of functional elements available or currently under development for RNSs hold substantial value for users from a PIM perspective.

4.2.1 Environmental Scanning and Awareness

The maintenance of awareness of that which is new can consume much of a user's scarce attention [11]. Funding opportunities and new publications are relatively modest exercises for an RNS, when the effort expended can be amortized across a large user population. It is a natural step to support the assimilation of these data into a user's PIM environment. A key aspect of this is effort avoidance in cases where revisitation is required to support monitoring, potentially of multiple, disparate sites.

4.2.2 Collaboration facilitation

While the PIM community has long recognized the natural continuum between PIM and GIM, getting into a team science context where one needs GIM capabilities has not been addressed to any significant extent. Collaborator discovery is a core capability of RNSs. Biosketch solicitation and management for large proposals clearly overlaps the two classes of systems. Document authoring, whether as a single author or as a member of a writing team can be interpreted as a PIM to PIM/GIM scale activity, transforming research data and interpretation into artifacts (documents) to be shared with a peer community. The various RNS-supported activities comprising the

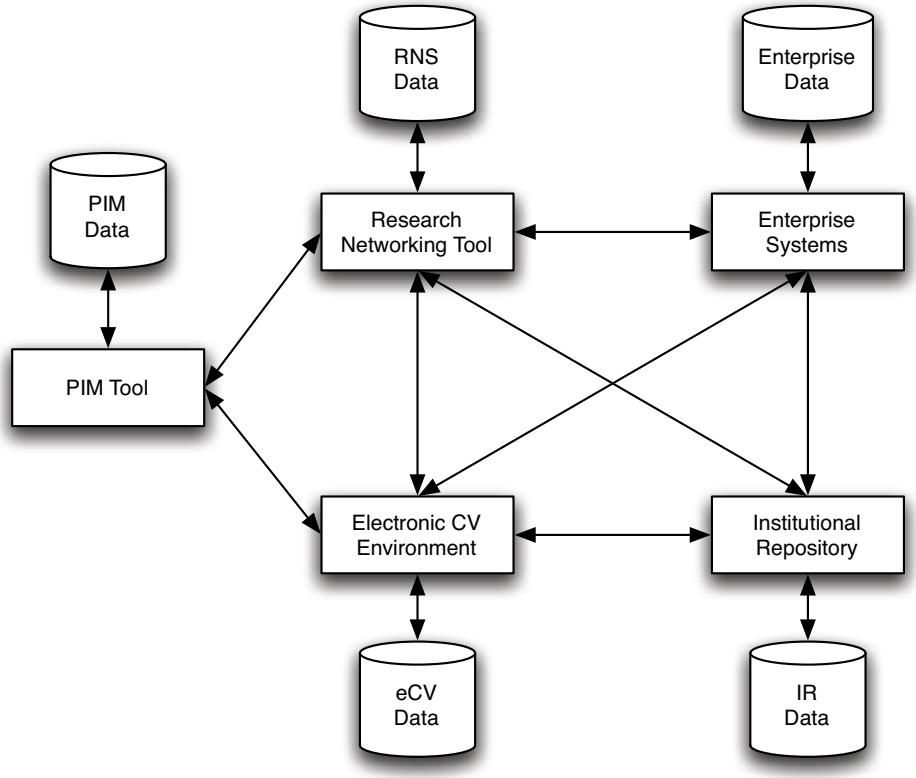

Figure 3: Integrating PIM and RNS

science life cycle from research hypothesis formation through curation of scientific results requires a range of system capabilities (see Figure 3) and PIM's utility lies in its ubiquity and personalization. Both of these features are challenging to provide when operating at RMS scale.

Spanning multiple heterogeneous research networking instances has proven to be both a technical challenge and an even greater organizational challenge. The Direct2Experts prototype [25, 26] provided a taste of the potential for such integration, but was functionally constrained by the organizational limits to what information could be shared between participating sites (basically just the number of matching persons). PIM-scale search can potentially mitigate some of these issues.

4.2.3 Administrative reporting

Few researchers enjoy the necessity of CV maintenance, annual reports, and contributing to departmental web sites. While the information required is commonly maintained to some degree in PIM scale tools, users typically only attend to these data aperiodically. Figure 3 (minus the PIM elements) illustrates the level of integration we have developed with Loki. Not only does the system feed off enterprise data sources, but our RMS is tightly coupled with the eCV tool. Connectivity with the institutional repository is already under consideration. Deployment of PIM capability in this environment could greatly reduce user overhead in maintaining and provisioning administrative data.

5. CONCLUSIONS AND FUTURE WORK

Coupling PIM and RNS technology offers clear benefit to both research communities. Ready access to large-scale data sources empowers PIM and access to PIM-managed information enhances RNS research profile specificity.

The perspective presented here is not limited to support of faculty-level researchers. We have already begun to look beyond faculty to professional staff [3]. Loki participation is open to all members of the university community, including post-doctoral fellows and graduate students.

Likewise, people and their publications are not the only entities of value worth modeling in research. Resource discovery systems (e.g., eagle-i[5]) offer complementary support for modeling and access to shared resources both tangible and intangible.

6. ACKNOWLEDGEMENTS

This work has been substantially influenced by numerous discussions with the members of the CTSA Research Networking Affinity Group and the members of the ICTS Biomedical Informatics Team. This work is supported in part by National Institute of Health (NIH) grant UL1-RR-024979.

7. REFERENCES

[1] V. Bellotti and J. Thornton, "Managing Activities with TV-ACTA: TaskVista and Activity-Centered Task Assistant," *SI-*

5. http://www.eagle-i.org

GIR Workshop on Personal Information Management, Seattle, WA, August 10-11, 2006.

[2] S. Bradshaw, M. Light, and D. Eichmann, "(Bee)Dancing on the Boundaries Between PIM and GIM," *SIGIR Workshop on Personal Information Management*, Seattle, WA, August 10-11, 2006.

[3] D. S. Brandt, M. Bosch, M. Bayless, C. A. Sinkey, K. Bodeker, K. Sprenger, K. Johnson and J. M. E. Gilmore, "A CTSA-Sponsored Program for Clinical Research Coordination: Networking, Education and Mentoring," *Clinical and Translational Science*, 4:42-47, 2011.

[4] T. Catarci, B. Habegger, A. Poggi, A. Dix, Y. Ioannidis, A. Katifori and G. Lepouras, "Intelligent user task oriented systems," *SIGIR Workshop on Personal Information Management*, Seattle, WA, August 10-11, 2006.

[5] D. Eichmann, and B. Kusenda, "Semantics and Ontologies in Research Discovery," *2nd VIVO Conference,* August 24-26, 2011, Washington, DC.

[6] F. Evéquoz and D. Lalanne, ""I Thought You Would Show Me How to Do It" - Studying and Supporting PIM Strategy Changes," *Personal Information Management: PIM 2009, ASIS&T Workshop,* November 7-8, 2009, Vancouver, B.C., Canada.

[7] K. Groth and K. S. Eklundh, "Combining Personal and Organizational Information," *SIGIR Workshop on Personal Information Management*, Seattle, WA, August 10-11, 2006.

[8] W. Jones, "How is information personal?" *Personal Information Management: PIM 2008, CHI 2008 Workshop,* April 5-6, 2008, Florence Italy.

[9] W. Jones, H. Bruce, E. Jones and J. Vinson, "Providing for Paper, Place and People in Personal Projects," *Personal Information Management: PIM 2009, ASIS&T Workshop,* November 7-8, 2009, Vancouver, B.C., Canada.

[10] A. Katifori, C. Vassilakis, I. Dradimos, G. Lepouras, Y. Ioannidis, A. Dix, A. Poggi and T. Catarci, "Personal Ontology Creation and Visualization for a Personal Interaction Management System," *Personal Information Management: PIM 2008, CHI 2008 Workshop,* April 5-6, 2008, Florence Italy.

[11] M. Kellar, "An Overview of Web-Based Monitoring: Future Directions and Challenges," *Personal Information Management: PIM 2008, CHI 2008 Workshop,* April 5-6, 2008, Florence Italy.

[12] D. Kirsh, "Personal information objects & Burden of multiple personal spaces," *SIGIR Workshop on Personal Information Management*, Seattle, WA, August 10-11, 2006.

[13] A. Komninos, L. Baillie and P. Barrie, "Holistic PIM: Managing personal information for a nomadic generation." *Personal Information Management: PIM 2008, CHI 2008 Workshop,* April 5-6, 2008, Florence Italy.

[14] A. Krishnan, M. S. Ramaiah, V. K Jaasti and T. V. Prabhakar, "Semantic 'ls' - An Approach for Personal and Private Group Information Management," *Personal Information Management: PIM 2008, CHI 2008 Workshop,* April 5-6, 2008, Florence Italy.

[15] G. Lepouras, A. Dix, A. Katifori, T. Catarci, B. Habegger, A. Poggi and Y. Ioannidis, "OntoPIM: From Personal Information Management to Task Information Management," *SIGIR Workshop on Personal Information Management*, Seattle, WA, August 10-11, 2006.

[16] Y. Li, J. Mostafa and X. Wang, "A Privacy Enhancing Infomediary for Retrieving Personalized Health Information from the Web," *SIGIR Workshop on Personal Information Management*, Seattle, WA, August 10-11, 2006.

[17] C. P. Middup and P. Johnson, "Managing the Adoption of Personal Information in Co-located Work Groups," *Personal Information Management: PIM 2008, CHI 2008 Workshop,* April 5-6, 2008, Florence Italy.

[18] S. Murthy, U. Murthy and E. A. Fox, "Using Superimposed and Context Information to Find and Re-find Sub-documents," *SIGIR Workshop on Personal Information Management*, Seattle, WA, August 10-11, 2006.

[19] L. Sauermann, G. Grimnes and T. Roth-Berghofer, "The Semantic Desktop as a foundation for PIM research," *Personal Information Management: PIM 2008, CHI 2008 Workshop,* April 5-6, 2008, Florence Italy.

[20] T. Schleyer, B. S. Butler, M. Song and H. Spallek, "Conceptualizing and Advancing Research Networking Systems," *ACM Transactions on Computer-Human Interactions,* to appear.

[21] M. Tungare, P. S. Pyla, M. Sampat and M. A. Pérez-Quiñones, "Defragmenting Information using the Syncables Framework," *SIGIR Workshop on Personal Information Management*, Seattle, WA, August 10-11, 2006.

[22] M. Tungare and M. Pérez-Quiñones, "Thinking Outside the (Beige) Box: Personal Information Management Beyond the Desktop," *Personal Information Management: PIM 2008, CHI 2008 Workshop,* April 5-6, 2008, Florence Italy.

[23] VIVO, http://vivoweb.org, last accessed Nov. 25, 2011.

[24] K. Voit, K Andrews and W. Slany, "Why Personal Information (PIM) Technologies Are Not Widespread," *Personal Information Management: PIM 2009, ASIS&T Workshop,* November 7-8, 2009, Vancouver, B.C., Canada.

[25] G. M. Weber, W. Barnett, M. Conlon, D. Eichmann, W. Kibbe, H. Falk-Krzesinski, M. Halaas, L. Johnson, E. Meeks, D. Mitchell, T. Schleyer, S. Stallings, M. Warden, M. Kahlon and Members of the Direct2Experts Collaboration, "Case report: Direct2Experts: a pilot national network to demonstrate interoperability among research-networking platforms," *JAMIA* 2011; Published Online First: 28 October 2011; doi:10.1136/amiajnl-2011-000200.

[26] G. Weber, W. Barnett, M. Conlon, D. Eichmann, W. Kibbe, H. Falk-Krzesinski, M. Halaas, L. Johnson, E. Meeks, D. Mitchell, T. Schleyer, S. Stallings, M. Warden, M. Kahlon, "Direct2Experts - Connecting VIVO and Other Research Networking Products," *2nd VIVO Conference,* August 24-26, 2011, Washington, DC.

[27] T. Whalen, E. Toms and J. Blustein, "File Sharing and Group Information Management," *Personal Information Management: PIM 2008, CHI 2008 Workshop,* April 5-6, 2008, Florence Italy.