Places of Our Own for Digital Information: Building Structures that Work for Individuals and Small Groups

William Jones
The Information School
University of Washington
Seattle, Washington 98195 USA
williamj@u.washington.edu

Steve Whittaker
Human-Computer Interaction
University of California
Santa Cruz, CA 95064 USA
swhittak@ucsc.edu

Kenneth M. Anderson
Department of Computer Science
University of Colorado
Boulder, CO 803090430 USA
kena@cs.colorado.edu

ABSTRACT
Structuring items such as folders and tags bring many benefits. Indeed, the very act of structuring may help people to make sense of their information. But structures can also confine and clutter. Some structures are created and never used. Other structures are used differently by different members of a group or by the same person at different points in time. More serious, structures are often "locked" within specific tools. Using tools may mean the inconsistent duplication and maintenance of a structure many times over. Resulting problems of information fragmentation are made worse as personal information moves onto a variety of devices and into a "cloud" of web-based devices and services.

We argue for a methodology and supporting infrastructure to promote a cross-study investigation of information structure. Moreover, we observe that the infrastructure to support a methodology of scientific inquiry may have direct application to users as they struggle to manage their information. Research on information structure reaches towards a new age in information management wherein organizing information structures grow and change over time based on the internal needs of their owners and not the external demands of tools..

Categories and Subject Descriptors
H.5.M. [**Information Systems**]: Information Interfaces and Presentation: Miscellaneous.

General Terms
Design

Keywords
Personal information management, PIM, group information management (GIM); information structure, application integration, open hypermedia, structural computing.

1. INTRODUCTION
As people use information, they also use information structures. These structures come in many forms: folders and tags to group together documents or other information items; documents with labeled sections that provide a "place" for pictures; notes and hyperlinks; article metadata such as author, title, and publication date; the headings and subheadings of a document; the rows and columns of a spreadsheet.

Information structures bring many benefits. Structure provides a basis for navigating (browsing) to information—either as a complement to search or as a primary method of information access [1,4,25]. The very act of structuring may help people to make sense of their information [22]. Tagging a document, for example, can be regarded as an act of categorization causing people to process more deeply the item's contents or its intended purpose [9].

But structures and structuring bring costs as well as benefits. Structures are duplicated or created and never used [30]. Structures persist long after their usefulness has passed [18]. These extraneous folders add to clutter, making filing and refinding difficult. Creating effective structures requires a fundamentally difficult prediction of future contexts in which information will be needed [19,20,24,29]. Maintaining structures is also onerous [2]. And the effort of creating structures may not always pay off. For example, in studies of web refinding a user's organization of bookmarks is often not used in retrieval [15]. Likewise, studies of email refinding show that people who use email folders for retrieval are less efficient than those who use search [31].

Another major set of structural problems relate to **fragmentation**. People persistently express a desire for greater integration [8]. Instead, people end up duplicating and maintaining related structures across different tools (e.g., file managers, email clients and web browsers) [3,7]. Most of the structures people use lie buried in their tools and cannot readily be re-used or even easily examined. With the emergence of Web/mobile applications and new social media applications, we may be approaching a **structural crisis** [14].

Fragmentation problems are replicated and exacerbated for collaborating teams who have to create and maintain **shared** structures[21]. Participants need to negotiate shared organizational structures, determining how information will be classified such that everyone can find it [5,23]. In practice these problems have proved so difficult that current collaboration practices avoid these tools, instead relying on distributing shared documents via email attachments, requiring each collaborator to maintain and manage personal versions of shared documents in their own file system [12,28].

These problems have serious implications for personal and collaborative productivity. However it is striking that we as a community lack basic scientific knowledge about fundamental questions concerning structure. We argue for research to address the following key questions:

Conference'10, Month 1–2, 2010, City, State, Country.

1. **What *is* structure?** It is necessary to develop tool-independent representations for, and metrics of, structure**.**

2. **What is structure *used for*?** Once reasons have been identified, questions can be posed about the efficacy of different structures across a variety of contexts: e.g. are deep, narrow folders useful when retrieving files? Are tags or folders more useful for retrieving emails? Do standard graph theoretic measures of structure predict the ease and speed of retrieval? Does retrieval improve with in-link count? What about isolated structures? How is a person's ability to work effectively on a project compromised when project information is fragmented across separate, unconnected structures?

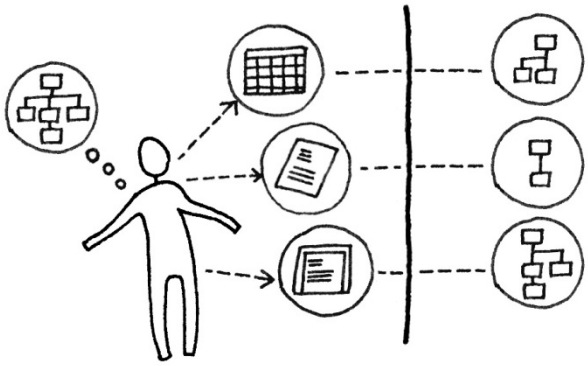

Figure 1. Users struggle with the fragmentation of their information across tools.

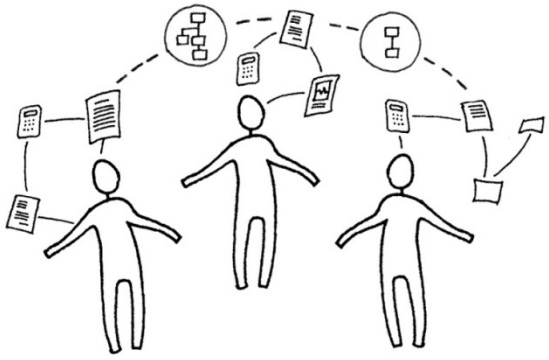

Figure 2. The structural problems encountered by individuals are exacerbated when working in groups.

3. **How do structures change *over time?*** Most research has looked at static structures captured as simple snapshots. We know little about change: when and why do structures change? What is the nature and scale of structural change?

4. **How are information structures *shared?*** How do members of a small workgroup share structure or do they operate with largely separate individual systems?

Given answers to these questions and, more important, given the means to answer these questions (through a methodology and supportive infrastructure) we as a community are in a very strong position to address a fundamental challenge of information management:

5. **How do we support better *integration* of information between tools?** How do we integrate the structures and tools people already have for managing their information—allowing them to repurpose successful structures when they come to use new tools? Can we inform people when they create maladaptive structures? Tools might allow users to create more effective structures by identifying potentially extraneous structures, such as tags that were created long ago but never used.

2. BACKGROUND

While there are initial studies addressing these questions [2,5,7,8,16,27,28,29,30], they have tended to be small scale, qualitative and often tool-specific. Despite the ubiquity of fragmentation, few studies have looked at how structures are **shared** across multiple tools [3,7,8], . Cross-tool studies have involved relatively small numbers of participants and largely qualitative data. So while we know that people are concerned about fragmentation, we do not have systematic data about the scale of this problem, about the extent to which people try to directly mirror structures across tools or about the effort expended in trying to replicate structure. Rarer still are longitudinal cross-form studies that look at **how informational structures change over time** [7].

There are also few studies that examine how structures are used **collaboratively** by small groups for GIM [21]. There are clear intuitive arguments for the utility of collaborative repositories: reduced workload as individuals no longer need to personally replicate and maintain their own versions of project information; a ready resource for onboarding new team members, and robustness to changes in personnel. Despite these putative advantages, collaborative repositories have not generally been successful [23] people still prefer to use email for sharing, detaching and managing items within their own personal system. For example, Voida et al. [28] found that almost 50% of sharing instances involve email attachments rather than collaborative repositories. Berlin et al.[5] identified reasons for the failures of shared repositories, including the inability to agree on shared labels [10], and the need for detailed metadata to promote findability by someone who was not the submitter of the information. An unwillingness to rationalize materials provided by others is shown to lead to the build-up of clutter, with decreased ability to refind information [23]. Other studies have looked specifically at sharing across peer to peer systems [11] or sharing of music [26]. However the majority of these studies are qualitative and none of these studies has looked at how structures evolve over time.

If we now consider the general movement of information from one device to a multitude of devices and onto a myriad of Web-based applications the challenges of research into information structures are magnified. For example we don't know how people exploit cloud based sharing, or how they use mobile devices for structuring information. An obvious reason for the limitations of prior work in focusing on a single tool, a single point in time, and a smaller set of participants is the cost of research. Cross-form, longitudinal studies are expensive to conduct. However that expense reflects the absence of an infrastructure collecting cross-tool structural data. On the other hand, focus on a single form of information, or a single device or application, would be less

limiting if there were more of a basis for comparing the structures that are observed and described

3. AN APPROACH

The effort to meet these challenges must interweave empirical and prototyping work:

- **Prototyping**: "meta-tools" must be developed to identify and extract the structures people impose upon their information across a diversity of devices and applications.
- **Empirical:** alternate measures of structure must be assessed for their utility.

Underlying both efforts is the need for a tool-independent metadata representation of the structure in *grouping items*. Grouping items are supported by nearly every end-user tool: A file manager such as the Finder (Mac) or the MS Windows Explorer supports folders; MS OneNote provides section and page tabs; the task management tool, "Remember the Milk", supports lists and tags; the "remember everything" tool, Evernote, supports the creation of a tagging hierarchy and provides the notion of "notebooks"; Facebook provides more content-specific grouping items such as "Groups", "Albums" and "Friend lists".

Drawing upon on our prior work [13] we argue for a metadata representation of grouping items that: 1. is modular – one "fragment" of metadata per grouping item; 2. abstracts common attributes and a basic structure of links; and 3. allows for representation of tool-specific attributes that may also influence a user's perception of structure. These desiderata are realized through the metadata representation depicted in Table 1. The explicit structure of a grouping item is simply a node plus outgoing links. This structure is further described through attributes which can be bundled by level and generality. Some attributes describe the node overall; others describe outgoing links from this node.

We underline an important distinction between the module of metadata (sometimes referred to as a "noodle" in our discussions) and the grouping item it represents. The module as metadata can be used to describe and provide a structural abstraction to any of a range of grouping items, including folders, tags, sections, "albums", and so on. The module of metadata might itself be persisted using various languages of representation and interchange (such as XML).

Note that some attributes in a module are specific to a tool. The document-like outlining tool Planz [13,17], for example, uses an attribute "isCollapsed" to determine whether the grouping item pointed to by a link should itself be shown "expanded" so that its links are in view as notes and subheadings. A mind-mapping application (such as FreeMind) might instead persist attributes needed to place links in an x-y plane.

On the other hand, other attributes such as "displayText" (or name) are generally found for all grouping items, no matter their parent application. Tool-independent attributes together with the node-link structure provide a basis for a combined linguistic and structural approach to the matching of different grouping items analogous to that described elsewhere for the matching of database schemas [6].

Table 1. The metadata of a grouping item can apply to its content or its links. Also, the metadata can be either tool-independent or tool-specific.

	Content	**Links**
Tool-independent	createdBy modificationDate …	displayText associationType associatedItem openWithDefault levelOfSynchronization …
Tool-specific	font (Planz) …	isCollapsed (Planz) … xOffset (FreeMindX) yOffset (FreeMindX)

The metadata module, then, is a start towards a representation of information structure in support of structural comparisons of grouping items, which may differ widely in external form (e.g., a folder vs. a tag).

4. CONCLUSION

A methodology and supporting infrastructure are needed to promote a tool-independent, cross-form, cross-study comparison of information structures. Efforts in this regard may yield near-term practical as well as scholarly benefits: The infrastructure that supports a methodology of scientific inquiry has, with small extension, direct application to end users as they struggle to manage their information. For example, as researchers we seek tools to identify structures that are duplicated or are no longer in use. These tools may also have application to end-users in their efforts to understand and "de-clutter" their information spaces. Beyond specific tools is a vision of personal and group information management in which many tools with many affordances might be used in support of tool-independent structures that people "own" and that grow organically to realize an oft-requested integration of personal information

5. ACKNOWLEDGMENTS

We thank Dan Suciu and David Maier for their good comments on earlier drafts of this position paper.

6. REFERENCES

1. Barreau, D. and Nardi, B.A. Finding and reminding: file organization from the desktop. *SIGCHI Bull. 27*, 3 (1995), 39-43.
2. Bellotti, V., Ducheneaut, N., Howard, M., Smith, I., and Grinter, R. Quality vs. quantity: Email-centric task-management and its relationship with overload. *Human-Computer Interaction 20*, 1-2 (2005), 89-138.
3. Bergman, O., Beyth-Marom, R., and Nachmias, R. The project fragmentation problem in personal information management. *Proceedings of the SIGCHI conference on Human Factors in computing systems*, ACM (2006), 271-274.

4. Bergman, O., Beyth-Marom, R., Nachmias, R., Gradovitch, N., and Whittaker, S. Improved search engines and navigation preference in personal information management. *ACM Trans. Inf. Syst. 26*, 4 (2008), 1-24.
5. Berlin, L.M., Jeffries, R., O'Day, V.L., Paepcke, A., and Wharton, C. Where did you put it? : issues in the design and use of group memory. *Conference on Human Factors and Computing Systems, INTERACT '93 AND CHI '93*, ACM (1993), 23-30.
6. Bernstein, P.A., Melnik, S., Petropoulos, M., and Quix, C. Industrial-strength schema matching. *SIGMOD Rec. 33*, 4 (2004), 38-43.
7. Boardman, R. and Sasse, M.A. "Stuff goes into the computer and doesn't come out" A cross-tool study of personal information management. *ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2004)*, (2004).
8. Bruce, H., Wenning, A., Jones, E., Vinson, J., and Jones, W. Seeking an ideal solution to the management of personal information collections. *Information Seeking in Context Conference -(ISIC) 2010*, (2010).
9. Craik, F.I.M. and Lockhart, R.S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior 11*, 6 (1972), 671-684.
10. Furnas, G.W., Landauer, T.K., Gomez, L.M., and Dumais, S.T. The vocabulary problem in human-system communication. *Commun. ACM 30*, 11 (1987).
11. Good, N.S. and Krekelberg, A. Usability and privacy: a study of Kazaa P2P file-sharing. *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (2003), 137-144.
12. Hicks, B.J., Dong, A., Palmer, R., and Mcalpine, H.C. Organizing and managing personal electronic files: A mechanical engineer's perspective. *ACM Trans. Inf. Syst. 26*, 4 (2008), 1-40.
13. Jones, W. and Anderson, K.M. Many Vews, Many Modes, Many Tools... One Structure: Towards a Non-disruptive Integration of Personal Information. *Proceedings of the 22nd ACM conference on Hypertext and hypermedia*, ACM (2011), 113-122.
14. Jones, W. and Maier, D. *Report from the session on personal information management. Workshop of the information and data management program*. National Science Foundation Information, Seattle, WA, 2003.
15. Jones, W., Bruce, H., and Dumais, S. How do people get back to information on the web? How can they do it better? *9th IFIP TC13 International Conference on Human-Computer Interaction (INTERACT 2003)*, (2003).
16. Jones, W., Dumais, S., and Bruce, H. Once found, what then? : A study of "keeping" behaviors in the personal use of web information. (2002).
17. Jones, W., Hou, D., Sethanandha, B.D., Bi, S., and Gemmell, J. Planz to put our digital information in its place. *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, ACM (2010), 2803-2812.
18. Jones, W., Phuwanartnurak, A.J., Gill, R., and Bruce, H. Don't take my folders away! Organizing personal information to get things done. *ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2005)*, ACM Press (2005), 1505-1508.
19. Jones, W. Finders, keepers? The present and future perfect in support of personal information management. *First Monday, http://www.firstmonday.dk/issues/issue9_3/jones/index.html*, (2004).
20. Jones, W. *Keeping Found Things Found: The Study and Practice of Personal Information Management*. Morgan Kaufmann Publishers, San Francisco, CA, 2007.
21. Lutters, W.G., Ackerman, M.S., and Zhou, X. Group Information Management. In *Personal Information Management: Challenges and Opportunities*. University of Washington Press, Seattle, WA, 2007.
22. Qu, Y. and Furnas, G.W. Sources of structure in sensemaking. (2005).
23. Rader, E. Yours, mine and (not) ours: social influences on group information repositories. *Proceedings of the 27th international conference on Human factors in computing systems*, ACM (2009), 2095-2098.
24. Sellen, A.J., Louie, G., Harris, J.E., and Wilkins, A.J. What brings intentions to mind? An in situ study of prospective memory. *Memory & Cognition 5*, 4 (1996), 483-507.
25. Teevan, J., Alvarado, C., Ackerman, M.S., and Karger, D.R. The perfect search engine Is not enough: A study of orienteering behavior in directed search. *ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2004)*, (2004), 415-422.
26. Voida, A., Grinter, R.E., Ducheneaut, N., Edwards, W.K., and Newman, M.W. Listening in: practices surrounding iTunes music sharing. *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (2005), 191-200.
27. Voida, A., Harmon, E., and Al-Ani, B. Homebrew databases: complexities of everyday information management in nonprofit organizations. *Proceedings of the 2011 annual conference on Human factors in computing systems*, ACM (2011), 915-924.
28. Voida, S., Edwards, W.K., Newman, M.W., Grinter, R.E., and Ducheneaut, N. Share and share alike: exploring the user interface affordances of file sharing. *Proceedings of the SIGCHI conference on Human Factors in computing systems*, ACM (2006), 221-230.
29. Whittaker, S. and Hirschberg, J. The character, value and management of personal paper archives. *ACM Transactions on Computer-Human Interaction 8*, 2 (2001), 150-170.
30. Whittaker, S. and Sidner, C. Email overload: exploring personal information management of email. (1996).
31. Whittaker, S., Matthews, T., Cerruti, J., Badenes, H., and Tang, J. Am I wasting my time organizing email?: a study of email refinding. *PART 5 -------- Proceedings of the 2011 annual conference on Human factors in computing systems*, ACM (2011), 3449-3458.