The Metadata is the Message

Paul André, m.c. schraefel, Max L. Wilson and Daniel A. Smith

School of Electronics and Computer Science
University of Southampton, SO17 1BJ, UK

{pa2, mc}@ecs.soton.ac.uk

ABSTRACT

The question "*What is Web Science*" is still frequently asked - even by authors of papers about Web Science. In this position paper we consider what part of the Web Science cycle makes this cycle emblematically "Web Science" rather than another form of either Law and Technology or Sociology and Technology or Computer Science and HCI. Based on our research developing and evaluating Semantic Web / Web 2.0 applications, and observations of current practice, we suggest that the particularity of Web Science is strongly correlated to a focus on human repurposing of particular Web technologies to support ever more rapid types of increased social contact. Based on this analysis, we ask how Web Science may help understand and shape this phenomenon, and what the implications may be for embracing this focus as a necessary criteria for assessing Web Science relevance of research work.

Categories and Subject Descriptors

K.4.2 [Computers and Society]: Social Issues, H.1 Information Systems [Models and Principles], H.5.m [Information Interfaces and Presentation].

Keywords

micro social interaction, web science, web evolution.

1. INTRODUCTION

Despite several key articles about Web Science, the most frequently asked question at Web Science seminars and workshops seems to be "*What is Web Science?*" Indeed, one of the founders of the Web Science Research Initiative has said at such meetings, "we like everything about the term except the word 'web' and the word 'science'". "Web" could be too limited to a current set of protocols, perhaps, and "science" could be seen to be too exclusive of voices in the Arts and Humanities. If nothing else is understood, Web Science insists it must draw on interdisciplinary expertise to understand the evolution of the Web and develop future work; to consider it from multiple vantage points blending science, engineering, law and the social sciences, as well as, we would argue, arts and the humanities.

One of the most compelling attributes of Web Science is the 'life cycle' (fig. 1) that includes human and social engagement. At a recent Web Science discussion, the question "*could what I do be Web Science?*" was frequently asked. Where one member of the WSRI board said "*yes it is*", another said "*no.*" The point of contention was around the question "does your work involve specific consideration of human needs or participation?"

In this paper we propose that the answer to that question is critical for understanding Web Science, that we are *all* being challenged to cross our disciplinary comfort zones and to ask about requirements and effects in other domains. Scientists and engineers need to consider human interaction with systems, and

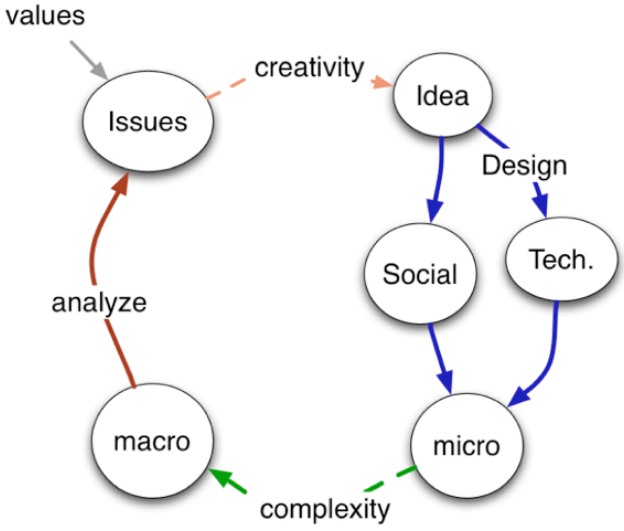

Figure 1. The Web Science "life cycle".

humanists need to consider technology *if* the interest is Web Science, rather than more of the same science being deployed in a web context. This is in no way disparaging that work, rather that it does not foreground what is or perhaps can be special or particular about Web Science.

Why is it important to make this distinction between what is and is not Web Science? We think the field is laudable for raising awareness of social aspects of technology and vice versa, but argue that there is value in emphasizing the human/technology connection, rather than making it an either/or domain; we already have domains that are exclusively either technology or humanist focused. In this paper we propose that by foregrounding a nexus of human/web interaction, we (a) see we are already well into a new era of data exchange (that adopts a post-blogosphere mindset), and that (b) this next mode of exchange is even more link-data friendly than the previous one, and as such, may well predicate a paradigm shift in how we engage not just with each other but with the machines that facilitate that communication.

To model what we mean by this paradigm shift and a potential next step in the evolution of the Web, in the following sections we consider lessons learned from previous projects entwined in the Web Science lifecycle, elaborate on observations of a new era of micro-data exchange and how this may influence a macro-effect, and how from this we might understand requirements for both a new type of computing interaction and a focus for Web Science.

2. CASE STUDY: EXPLORATORY BROWSING

In this section we present our experiences and lessons learned with a Web 2.0/Semantic Web application, mSpace, in the context of the Web Science lifecycle. We then consider a newer

experiment with linked data that may actually be breaking the Web 2.0 model, and from such, we have been reconsidering our notions of linked data applications in the context of the meaning of Web Science.

mSpace is both a framework for integrating heterogeneous data sources for real time exploration and a UI to support information seeking activities such as browsing, learning, re-organising and searching [9]. Facets of a dataset are presented as columns, and fig. 2 shows mSpace applied to a news footage archive, where the columns present metadata like decade, year, theme and subject. As part of our investigation into how mSpace is evolving to support a broader range of search tactics, new design ideas are regularly created, prototyped and studied. This pattern is in-line with the web science lifecycle; when we have investigated wide-spread use, unexpected behaviours have been seen that have identified new requirements for exploratory search and new designs to support them. A recent longitudinal study [15] of the largest installation of mSpace to date (over 30 facets and 65,000 items) investigated use over a month long period. While the main explorer hosted a variety of tweaks to make it more effective, we quietly added a set of social features for tagging, grouping and commenting on elements in the data set. Many users were logged using the online service anonymously, and 11 users were involved in regular communication about the design and software.

The aim of the study was to see how the use of the exploratory features, such as the browsing columns, were used over time compared to familiar keyword searching. Our hypothesis, based on related work in exploratory search [14] and our own micro-level experience with users and mSpace, was that in the first visits to the site the columns would be used, and that users would keyword search more often thereafter. Macro-level use told a very different story and our hypotheses were rejected. Instead, we saw that most users used the keyword search in their first session, but as they became comfortable with how the columns were reacting to their searches and what support they were providing, users began to use the columns to both explore and produce much more expressive constraints over the dataset. This is a surprising finding and important for thinking about how to improve access to and permit reuse of linked data.

Another behaviour from the investigation with macro-level use was the emphasis that users put on social interactions to support exploratory search. One example that was regularly given by the known participants is that they wanted to know what other people were doing, in particular what their peers, or experts in their field, had already found and had thought about that material. Our logs showed evidence that people were making an effort to be a part of the social network (tagging, commenting) but found it very hard to get benefit from what others were doing. Though this is somewhat similar to sites that provide recommendation, the reputation and trust (of classmates, your professor) are even more key here, as well as not just the final decision, but about the path to get there -- the process as well as the outcome. This was a surprise to the design team, who saw these features as a simple support to the overall browsing experience. Instead, behavioural evidence and participant feedback seem to indicate that the involvement of personal and social data, along with the publicly available dataset, is increasingly important to users and their exploration.

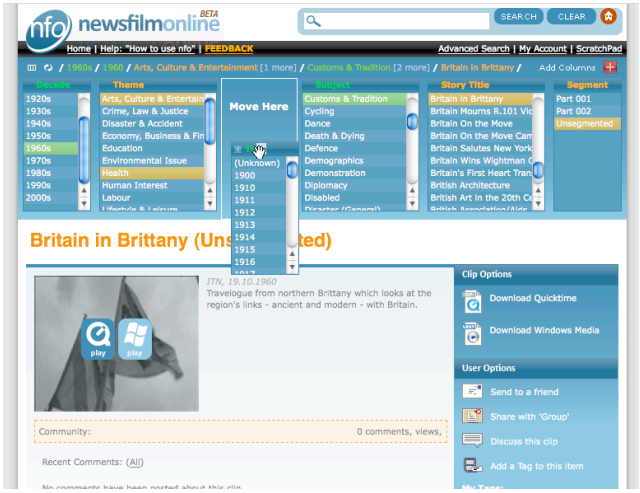

Figure 2. The mSpace explorer in NewsFilmOnline.

From the identified and unexpected behaviours of mSpace at macro-level use, we are starting a new evolution of the interface design to support this kind of social exchange specifically to enhance exploratory search. One example of this is RichTags [11], a project that builds on the tagging, commenting and sharing functionality utilised in mSpace, for cross-site browsing and exploration of digital repositories. RichTags aggregates metadata from multiple repositories, and derives additional metadata based on document contents, to enable cross-repository linking and browsing by category. Based on our experiences with mSpace, social interaction features have been foregrounded, and are available via an API to enable future work to add plugins to pull comments back to the individual repositories.

2.1 Lessons Learned

We have seen how mSpace supports the notion of the web science life cycle. In testing the latest incarnation of mSpace, we were examining the user interface, and found (a) direct observation: exploration is not overtaken by specific keyword search, but enhances it, so enabling the relations among linked data to be explored is beneficial to improving value of resources beyond keyword search alone. (b) explicit comment: we also found that tagging (regarded as a simple add-on) was triggering explicit response, as well as a desire to see what other people thought about clips. These social interactions have been made prominent in the RichTags project.

These observations highlight the importance of a Web Science focus on the social aspects of technology. As human-computer interaction researchers, the authors are perhaps somewhat more involved in the human aspects of computer science than others, but as is evident here, often only at a micro-level -- while we implemented tagging and social features, it took a longitudinal study for us to realise the emphasis users would put on these features, and even this is at a small scale compared to the macro-view of Web evolution. This is true of a majority of this discipline; HCI researchers tend to focus on specific tasks or applications, and though this work is important, Web Science seeks to look beyond local tasks and asks questions about larger interactions, pushing this community, among others, to consider the greater effects of these local contexts. In the next section we examine another aspect of social community interaction: recent phenomena in what we term "micro social interaction" -- small bits of data at a personalised interaction level -- before considering how this may result in a greater macro effect.

3. OVERVIEW OF MICRO SOCIAL INTERACTION

To date, the web science life cycle has been used in examples from e-mail use to wikipedia, and an oft-cited example for the

need to study web science is the "blogosphere". The rise of blogging was an emergent phenomenon, aided by technologies such as trackbacks -- notification when you have been linked to -- emphasising the importance of the social community. With aspects of participation, openness and interaction, blogging was seen as characteristic of the Web 2.0 concept [8]. We term this as the first wave of web science. We envision the second wave to be more localized, and argue that in the fast moving world of the Web, we are already seeing it. Rather than updating blogs for the entire world to see, the second wave is an entire system of micro-exchange: merging concepts of personal and public data in terms of updating Twitter status, uploading music listening habits to last.fm, social networking actions, or applying personal rules to public data feeds. There is also much interest in the creation of protocols, such as XMPP[1] to share metadata between sites to enable the de-coupling of oneself from specific services, so that, for example, users of one photo sharing service can comment and reference cross-site. Protocols such as OpenID are also of interest in this area, to permit cross-site web-scale authentication and trust, again decoupling the user identification layer from individual sites, and going web-wide.

3.1 Social Emphasis: Ephemeral or Enduring?

Could it be that these ideas of localization and micro social interaction are just a transient wave, or are they more deep-rooted? A spate of news stories arose in early 2008 over figures from market researcher ComCast reporting that while the total number of people on social networking sites is raising, the average amount of time people spend on them is declining. While the sites themselves are inevitably subject to plateauing popularity, it is evident that people relish social aspects of community and collaboration. Even if social networking and sites like del.icio.us, flickr and upcoming disappear, they have opened our eyes to a new possibility of personal, social and public data interaction, and that paradigm has value reaching further than any single site. We elaborate on this paradigm and the potential it has for the future of the Web and Web Science in a later section.

Further examples of social successes can be found in texting (SMS messaging), another example of micro-data. The number of texts sent has experienced year-on-year growth -- statistics from the UK indicate 200 million messages were sent in January 2000, to 6.1 billion in December 2007)[2]. Technorati report a doubling in size every 5 months of the number of blogs tracked between 2003 and 2005[3]. Kumar et al. [4] observed that up to 2003 (the limit of their data), 'burstiness' in blog communities increased, implying that local community structure and community-level interactions are being reinforced as the blogosphere grows.

3.2 Social Affects and Effects

We are increasingly seeing instances of not just Web 2.0 or 'social' web sites, applications, or companies, but of a larger social evolution in practices on the Web.

Citizen Journalism. The public, traditionally only consumers of journalism, are now creating content through blog posts, critiquing and fact-checking traditional media, or uploading photos and videos of events. The New York Times has recently run stories on frontline blogging from the war in Iraq[4], and micro-data has also become prevalent: the New York Times also reports on twittering the US Election Campaign[5], and CNN, Reuters and other media welcome news stories, images and video from the public.

Human Computation. Humans are better than computers at a range of tasks, and 'human computation' is designed to leverage that. Initially as a way to ensure robots weren't signing up for e-mail accounts (CAPTCHAs), more recently social-based 'games' utilise micro social interactions to label images, locate objects in images, and collect 'common-sense' facts[6]. This is just one example of a kind of social computation that is entirely new.

Messaging. Networked services like Twitter and Jabber are increasingly popular, allowing the transfer of micro-data, increasing social awareness. The Jabber protocol is even used in services like Jaiku to not only instant message contacts, but to receive, comment and post web feeds.

In the next section, we consider looking at a wider effect - the exchange of this data in micro social interactions in order to consider how this may drive a paradigm shift from task specific data and applications to more automated, context rich interactions.

4. BLENDING PERSONAL AND PUBLIC DATA SOURCES

We have seen how micro-data such as twitter updates, listening habits, and social network actions are becoming more common, and how this is blending the concepts of personal, social and public data. In recent work we have been pushing on what happens when this data becomes prevalent, and the affordances provided by a blend of public and personal information are explored in our prototyping of a tool we call AtomsMasher [12] to use RSS feeds as a context to inform automatic actions. AtomsMasher is able to combine information from heterogeneous sources about a person, as well as about things in the world -- people, places, books, concerts -- allowing people to create rich reactive behaviours. The tool is a proof of concept to explore how semantics enables a blend of micro-bits of personal and public data in order to take control of, or delegate the handling of, the high volume of information we produce or process. We present a short scenario to demonstrate the potential benefits of blending this data. If aspects of this scenario seem to echo previous intelligent agent scenarios, it is for good reason; but rather than being just imagination or largely dependent on artificial intelligence, we see them as achievable through user interaction and linked data.

Xaria has a meeting with clients in London this afternoon. As noon approaches, her meeting rises to the top of her to-do list. When she leaves her office for the train station, the location aware software on her mobile phone detects that her location has changed, and posts this event to her Plazer RSS location feed. AtomsMasher identifies this action -- leaving the office prior to a scheduled event at a different location -- and sets her away state on her IM at the firm to "Away from the office; en route to Charing Cross for Meeting with Baroque Architects, I will be checking e-mail periodically", taking location and description fields from the upcoming meeting entry in her calendar. (*Xaria previously defined that she would like her IM status changed to reflect her calendar appointments when she has client visits, she will have periodic e-mail access*).

[1] http://www.readwriteweb.com/archives/xmpp_web.php
[2] http://www.text.it/mediacentre/sms_figures.cfm
[3] http://www.sifry.com/alerts/archives/000298.html
[4] http://www.nytimes.com/2008/01/21/business/21iraqblogger.html
[5] http://www.nytimes.com/2008/01/21/technology/21link.html
[6] http://www.cs.cmu.edu/~biglou/

Arriving late at the station, Xaria misses her train but jumps on the next one. On the train, she flips open her mobile phone, and, habitually late as she is, hits the previously defined 'e-mail because late' action. The system infers from her calendar who to address the e-mail to, and proffers a field for her new ETA. She then switches to reading the news. A concert announcement has appeared in her personalised Arts and Entertainment column on the front page. *(AtomsMasher let it through because it fit a triggering rule that Xaria previously defined*:) the band has recently been appearing on her friends' last.fm "I listened to this" RSS feed, that friend's calendar RSS feed shows them free during the concert, and Xaria's own calendar shows her at a meeting just two tube stops from the concert, finishing just an hour before the concert begins. As her train arrives at its final station, Xaria's phone alerts her that the Tube line to the meeting location is experiencing delays, and suggests an alternative route. (*Xaria previously stated that meetings in London involve the Tube; AtomsMasher uses this and her work address -- her last known location -- to consult the London Transport Journey Planner web service for announcements that pertain to any of the lines she might be taking to her destination*).

This scenario demonstrates both subtle (ranking a to-do list/event feed) and explicit (away state setting, transport alert) use of reactive behaviours towards automating the provision and dissemination of desirable information, using a blend of personal (personal calendar and current location), public (news events and show listings), and semi-public information (friends' music listening activity) from web data feeds. Had Xaria had the time to check her and her friends' listening habits, cross-reference that with bands playing, and further cross-reference that with her various locations, this could be done manually, but by automating such a repetitive and tedious task, Xaria can instead concentrate on the important review meeting. This work is still mostly at a local scale, and experimental. In the next section we look to the future of the potential of these public, private and social micro-data bits, at a macro-scale.

5. THE POTENTIAL OF A PARADIGM SHIFT

There are two aspects to what we have been terming a paradigm shift. First is the 'new era of data exchange' mentioned in the Introduction; a shift in our concept of personal or private data, as well as in the way that data is made available. Rather than data being entirely personal, or entirely public, we increasingly see personal information available on social networking sites, our personal listening habits uploaded to the Web, as well as the ability to access other similar social information about our friends. In the previous section we saw some of the affordances blending this data could give us.

The second aspect is one of interaction. As we gain context through a blend of personal and public data, we imagine possibilities in changing the way we interact with computers facilitating engagement with other humans. Compared to the interactive, co-operative computing visions set out in futuristic sci-fi scenarios, or even in the literature [5][13][1], our world of computing today is still one of human pro-activity and initiation; there is little sense of machine co-operation or helpfulness. We posit that in the same way a personal assistant collaborates and refines tasks with their client, we can translate those functions to a personal *digital* assistant, enabling a more co-operative computing experience [10]. Maes [6] previously suggested a metaphor of a personal assistant, though relied heavily on machine learning. We believe that while machine learning and other knowledge-based AI techniques are key to this work, we can gain immediate value from explorations into end user customisation (partly addressed in AtomsMasher and in planned work), and by the first aspect of our paradigm shift -- context through the blending of personal and public information.

This vision is potentially a whole other way of interacting with the computer, and we are exploring what we are able to do with these personal, public and social data bits if we facilitate their interaction better, and combining this with research in off-desktop computing. If we have context and linked data we no longer need to use monolithic passive applications, but data can be persistently available to all in a type of 'Data Sea', reflective of the scale and diversity of elements to be found in both the personal and public space. In the following section we examine the implications that this macro-effect of micro social interaction may have on the future of the Web, and of Web Science.

6. A WEB SCIENCE FUTURE

Let us recap briefly. We observed in our own work a social-focused Web Science life cycle in action, and that on a wider Web scale the micro-exchange of personal, public and social data is prolific, and point to a number of practices indicating the Web is facilitating social evolutions. We describe our early work in blending those personal and public data sources (using AtomsMasher) in order to give value to a user. Finally, we outline what we see as the potential of those micro-data bits and what could happen if we facilitate their interaction better. In this section, we further explore the meaning of micro-data bits, the potential Web Science has in furthering and understanding these, and the implications these observations have for Web Science as a field.

6.1 The Metadata is the Message

In 1964, Marshall McLuhan wrote "*the medium is the message*" [7], warning that too often we are distracted and focus on content, proposing the media carrying the content should itself be the focus of study to aid in understanding how the medium affects our culture, cognition, and society. Web Science seeks to understand the evolution of our global village, and in this paper we have focused on how micro-bits of data, metadata about a person or event, may convey information.

Though some aspects of Web 2.0 are largely page-based (blogs, wikis), another aspect is the social side we have looked at in this paper and is evident in social networking sites. Facebook, for example, is not about pages. It is collections of aphorisms -- status updates, wall messages, photo tags -- short bursts of meaning with their own context. Though there is content, we argue these micro-bits of data are highly valued for their affect. Last.fm's 'currently listening to' might not have so much to do with the music but with recognition. These and others -- plazes (updating your location), twitter (broadcasting actions or status), social grooming -- all are about being aware, being close to someone while at a distance.

What if these messages were to get even shorter, or carry no perceptible content? The wi-fi enabled rabbit Nabaztag (nabaztag.com) gives ambient indicators of new e-mail, stock market reports, or weather forecasts through indicative glowing lights. Ludic and affective computing have also variously explored this space -- drinking vessels as a communication channel [2] or floating feathers connected to a picture frame [3] -- but we posit that what we are seeing is people already using these micro social interactions to emulate those kinds of affective computing. People are using these message fragments to

potentially convey a lot of bandwidth, of emotion; our open question is how might new applications support that?

Our prototype tool AtomsMasher, for instance, is still focused on content, with notions of personalising a newspaper, but perhaps achieves a similar effect to the above efforts through the suggestion of concerts, a social opportunity, therefore communicating affect. In that prototype we are trying to use metadata for functional linking, but the issues are not just about linking, it is about what that conveys -- how this information is important and on what level.

6.2 How Web Science Can Help

The Web Science cycle is commendable for raising awareness of the human side of technology. In bringing together the enablers and observers of effect, we look forward to great co-operation, understanding and advancement. Where does Web Science come into the work we have described above? Would AtomsMasher, for instance, be even more effective if we had a sociologist on our design team focusing on what a paradigm shift may involve and effect? Since we do not, can we call it Web Science if we are, to some extent, making it up as we go along? At a broader level, sociologists may have already recognised the popularity of social networking sites, the surgence of micro social interaction and the way people are using them for communication. As technologists, we are able to think about the implications for back-end infrastructure and interface design. With this kind of information sharing, are we able to get to a macro-level discussion suggested in the cycle?

6.3 Open Questions for Web Science

We close the paper with a number of open research questions to stimulate thought and discussion on both the work presented in this paper, and its involvement with Web Science.

Do new applications need to have some social or networked context to be of value? (As opposed to previous standalone computers and applications).

Is the next evolution of computing what we have been seeing and elaborate on in this paper -- social uses of socially constructed data? If not, will studying current phenomena help? Are aspects of affective computing being realised through the metadata?

Early web users only needed to know simple HTML to create a page. As the web got more popular, along with site design tools, we ended up with a lot of pages about cats. Then blogging sites and software gave rise to the blogosphere. The Semantic Web has not taken off in a public way. However, we are seeing a lot of social actions such as tagging. Is this social community aspect a way to leverage takeup of the Semantic Web?

Will the observed social actions change the way we design applications? a) As mentioned previously, micro social interaction may be seen as a form of affective or ambient computing, how can we best support this? And b) social networking sites store millions of people's personal information. Now that we no longer live in a single user world, will software engineers have to consider ethics as a primary function; how people are going to use, share and access theirs and others information.

7. CONCLUSION

We have observed that "micro social interactions" -- exchanges of small bits of data at a personal/social interaction level -- are increasingly prevalent, from experiences in studies of our own Web 2.0 / Semantic Web applications, through social grooming in status updates or social networking sites, to evidence of more widespread social evolution in the form of citizen journalism or human computation. From these observations arise a number of questions of how micro social interaction may come together to form a greater, gestalt macro effect. We suggest understanding that effect would be useful to explore within the context of a Web Science agenda, and that doing so may address the question with which we began: not so much what is Web Science, but what is particular about Web Science?

Through a prototype tool, AtomsMasher, we examine how the blending of personal, public and social data can result in greater value and improve in some sense quality of life, and question the potential this metadata has for both social and community interaction, as well as a paradigm shift towards co-operative interaction. We suggest that rather than the traditional observations of media post-(detrimental) effect, we are able to look to Web Science to inform, understand and shape these phenomena. As McLuhan [7] said, "*Anticipation gives the power to deflect and control force*."

8. ACKNOWLEDGMENTS

Our thanks to participants of various Web Science workshops and seminars.

9. REFERENCES

[1] Berners-Lee, T., Hendler, J., and Lassila, O. The semantic Web. Scientific American, 284 (5):28–37, 2001

[2] Chung, H., Lee, C. J., and Selker, T. Lover's cups: drinking interfaces as new communication channels. CHI '06.

[3] Gaver, B. Provocative Awareness. CSCW. 11 (3), 2002.

[4] Kumar, R., Novak, J., Raghavan, P., and Tomkins, A. 2004. Structure and evolution of blogspace. CACM 47, 12. 35-39.

[5] Licklider, J. Man-Computer Symbiosis. IRE Transactions on Human Factors in Electronics, 1(1):4–11, 1960.

[6] Maes, P. 1994. Agents that reduce work and information overload. CACM 37, 7 (1994), 30-40.

[7] McLuhan, M. Understanding Media: The Extensions of Man. 1994. MIT Press.

[8] O'Reilly, T. What Is Web 2.0. 2005. http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html

[9] schraefel, m. c., Wilson, M. L., Russell, A. and Smith, D. A. (2006) mSpace: improving information access to multimedia domains with multimodal exploratory search. CACM, 49 (4). pp. 47-49.

[10] schraefel, m. c., André, P. and Van Kleek, M. This Time It's Personal: from PIM to the Perfect Digital Assistant. CHI2008 PIM Workshop.

[11] Smith, D. A., Lambert, J. and schraefel, m. c. Rich Tags: Cross-Repository Browsing. Open Repositories Conference 2008 (OR2008).

[12] Van Kleek, M., André, P., Smith, D.A., Wilson, M.L., Bernstein, M., Karger, D. and schraefel, m.c.. AtomsMasher: PeRSSonalized information delivery and management on the web, 2007. URL http://eprints.ecs.soton.ac.uk/14827/.

[13] Weiser, M. The computer for the 21st century. Scientific American, 265(3):66–75, 1991.

[14] White, R. W., Kules, B., Drucker, S. M. and schraefel, m. c. (2006) Supporting Exploratory Search, Introduction, Special Issue, CACM 49 (4). pp. 36-39.

[15] Wilson, M. L. and schraefel, m. c. (2008) A Longitudinal Study of Exploratory and Keyword Search. In submission.