ARMA INTERNATIONAL
EDUCATIONAL
FOUNDATION
RESEARCH · EDUCATION · SCHOLARSHIP

BIG BUCKETS OR BIG IDEAS?
Classification vs. Innovation
on the Enterprise 2.0 Desktop

By

Patricia Galloway, PhD, CDP
Associate Professor
School of Information
University of Texas at Austin

1

Big Buckets or Big Ideas?
Classification vs Innovation on the Enterprise 2.0 Desktop

Patricia Galloway, PhD, CDP
Associate Professor
School of Information
University of Texas at Austin

Introduction

The recent interest fostered by the U.S. National Archives in using so-called "big buckets" as a feature of their Flexible Scheduling scheme indicates some hope for a solution to classifying routinized work product not scheduled as permanent. Given that some records managers are looking toward adopting the practice for <u>all</u> records, however, it may prove to be in conflict with current research in information science on the work practices of knowledge workers, especially those whose work implicitly includes problem-solving and innovation. Conceived in another way, however, the "flexibility" provided by the concept may open the door to an application of "Enterprise 2.0" classificatory practices. This is of particular import since electronic record creation is already being supported by software systems designed to optimize flexibility for employee work practices.

Ever since the advent of the desktop metaphor, knowledge workers whose jobs include a high degree of novelty have been using virtual "piling and filing/finding and reminding" desktop arrangements and filing practices to defer classification until emergent categories become obvious and to avoid disposition of digital work "papers" that might have additional reference or inspiration value. For other workers whose jobs are more routinized, these software products have been disciplined by the imposition of required templates and interfaces, but for the increasing numbers of true knowledge workers, such constraints have been less and less applied because they have proved counterproductive. For in the case of the product of knowledge workers' work, some enterprises are beginning to find value in harvesting their work directly so as to capture enterprise intellectual property, whether classified or not, while others are going farther in appreciating the idiosyncrasies of these workers' personal classificatory activities as embodied in desktop arrangements and in unregulated tagging that has the potential to produce valuable folksonomies.

I suspect that this phenomenon will eventuate in a two-tier effect for records management, such that standardized classification activities susceptible of automation, relating to routine work involving forms and software-detectable formal instruments and mapped onto detailed job task inventories, will be fully automated; while records emerging from so-called "creative" work—that which is most difficult to classify using automated systems—will be classified by the creators themselves at their own pace and not in imposed categories, but in ways that they invent themselves, thereby capturing the knowledge they are developing in the course of their work. I suspect that this development will be supported if not driven by the emergence of ever more flexible COTS desktop productivity products and abetted by the adoption by records managers of

knowledge management practices as office activities in the developed world become increasingly part of a true "knowledge economy."

Overview: Classification and "Creatives"

Classification of digital records is a problem that took rather a long time to develop.[1] People had been creating digital records in their workplaces for quite a while before they were recognized as records at all. Initially records had to be on paper; the computer was just a glorified typewriter (a "word processor," in fact) used to create paper records. As long as this remained the case, all the constraints that restrict paper records (or affordances that characterize them)—uniqueness, incapacity to be in more than one place at a time, inability to be searched except by visual inspection—conspired to reinforce long-established filing system practices for the grouping of records in a manageable number of folders.[2] And in any case, only a very few people were even producing records digitally, and those people ("secretaries") were themselves not seen as much more than machines for the transcription of messages originated by others, others whose agency, though expressed through several layers of action they controlled, was nevertheless guaranteed by the fact that in the end it was their signature affixed to the final product that guaranteed its recordness.

Some other people also entered data into computers and created digital records, but not as the product of individual effort; they used so-called "dumb terminals," had no means of deriving paper records from those terminals, and merely added information given to them to a central resource vetted and controlled by others. These database systems were indeed the first formally recognized digital records, but they were only seen as such in the aggregate, and they were seen as easily evaluated and controlled. Ways of working with such digital records were embodied in organizational hierarchy, computers were also embedded securely in that hierarchy, such activities of recordkeeping were described vividly as "control through communication,"[3] and the records were apparently as controlled as the people. Where automated, a centralized "data processing" installation did the honors.

The story of the coming of the microcomputer is now a familiar one. In the late 1970s microprocessor chips were discovered by hobbyists and first included in kits providing a means to make them independent computers, then sold assembled for experimentation, then put to work for individual purposes. At first they were taken seriously only by a few enthusiasts, since they were difficult to use and required a real dedication to learning about electronics and even writing one's own programs. Some people imagined that they should be used more widely, especially in education and to enable individual work, and those visionaries experimented with ways to make them

---

[1] In this paper I will mean by "record" any recording of information made or received in the normal course of business; or to quote Kenneth Thibodeau of the US National Archives, "everything is a record; some just last longer than others" (overheard at a meeting of the Society of American Archivists).

[2] Although of course "long-established" in this case refers to the establishment of the vertical file at the turn of the twentieth century; see Jo Ann Yates, *Control Through Communication: The Rise of System in American Management* (Baltimore: Johns Hopkins University Press, 1989), especially Chapter 2.

[3] Yates, *Control through Communication*, describes the flow of communication through several large enterprises: Illinois Central Railroad, Scovill Manufacturing Company, and E.I. du Pont de Nemours.

easier to use. The result of their efforts was an explosion of demand, which ran so far ahead of management's understanding that in the late 1970s and early 1980s individuals took to smuggling privately-owned machines into the workplace. As is well known, the presence of individual microcomputers and their substitution for dumb terminals led to centralized LAN-based networking within enterprises; centralized networking gave way to peer-to-peer networking as the model of the ARPANET became influential; the HTTP protocol was built on top of the TCP-IP protocol, and finally the Internet was made publicly accessible and digital records that had been securely corralled within enterprises escaped to be exchanged globally. Even more threatening, national legislation was passed that made digital records that had been properly kept into permissible legal evidence in a court of law and so-called digital signatures into acceptable authenticating devices. Control of digital record-making had been lost while records managers concentrated on paper, and meanwhile the very centralized authority that had justified and supported records management was being restructured as enterprise governance hierarchies were flattened and corresponding record group hierarchies collapsed. How could control be regained?

The answer, applied once it was realized that printing out all these records was not the answer, required a better definition of "recordness" or acceptable authenticity for digital records, grounded in specific procedures of creation, care, and management. Research was carried out to determine how authenticity might be defined. The University of British Columbia (UBC) InterPARES project sought to apply to the problem the theory of diplomatics, a theory of record authenticity first developed in the sixteenth century for physical records, which eventuated in a notion of record authentication by the record's creator at the point of creation,[4] ignoring the fact that it is one thing to apply a bulky seal to a document that might constitute one of perhaps five or ten, representing three or four records categories, that a person might be responsible for in his lifetime; but it is quite another thing to ask someone who may create that number of records in a single morning to authenticate and classify them into the myriad categories representing the complexities of the large enterprises of modernity. A second project carried out at the University of Pittsburgh turned to what it referred to as "literary warrant" (legal statutes, business practices, regulations, and policies) and focused explicitly on the notion of evidence,[5] but here there was also little concern for how the actual practice of determination of record would take place. In both cases researchers were concerned to bring digital records under the discipline of management in order to guarantee their evidentiary value, but in neither case was there adequate concern for precisely how this would be done. In the event, the UBC-developed Department of Defense 5015.2 STD called for a method, and vendors were happy to oblige by offering client software that intervened at the point of creation on the user's desktop and demanded that the new creation be categorized and its status declared so that it could be stored in a

---

[4] Luciana Duranti, *Diplomatics: New Uses for an Old Science* (London: Scarecrow, 1998).

[5] School of Information Sciences, University of Pittsburgh, "Functional Requirements for Evidence in Recordkeeping: the Pittsburgh Project," retrieved 10/5/08 from http://www.archimuse.com/papers/nhprc/; also Richard Cox, "The Record: Is it evolving?" *Records and Retrieval Report* 10(3), 1994, 1-16. URLs in this paper may obviously not work forever, though I anticipate that the items to which they refer will probably have a long life. This is another feature of Web 2.0, which is driving us toward search engines and away from URLs.

document/content management system. Yet such systems have been slow to be adopted[6] and have proved problematic in use, ostensibly because the classification being asked of users was potentially too complex.

Several issues had been overlooked in the assumption that work practices could and should be simply changed by imposition of a new discipline. It is especially unrealistic to imagine that records can or should be categorized when many may still be in progress, subject to the repeated revision and reworking that is one of the most important affordances of computer-supported writing.[7] It is not surprising that people asked to work in an environment where record creation is gated by records management applications that force instant classification might try to escape this kind of nannying by turning to a web services solution like Googledocs, in effect taking the same kind of action to defend favored ways of working that they had taken in the first place by smuggling personal computers into the workplace. The fact is that the very environment in which most computer users work today, the environment in which they have in fact worked since the late 1980s, has provided them with a creative freedom that many are loath to lose and that is increasingly recognized as of value both to their work process and to the employer who owns their output.

In the world of digital recordkeeping, I want to argue, the requirement that records creators promptly classify their output is ironically in conflict with the kind of idiosyncratic arrangement and filing of materials that are overtly supported by software products designed to maximize creativity and innovation by knowledge workers. Prompt classification requires that knowledge workers superimpose a set of requirements that from their perspective may appear to be outdated or irrelevant because they may be directly orthogonal to the way the workers' own work environment segments the world. This problem has come to a head with the demands for Web 2.0 technologies to be implemented in so-called "Enterprise 2.0" networked working environments and has been recognized in some early responses from records managers looking toward a hybrid classification environment combining automated pre-coordinate with user-supplied post-coordinate classifications.[8] I want to engage here with the notion that there may be perceptible <u>value</u> to an organization in deferring user classification of the output of knowledge work while capturing the usage and eventual informal classification patterns of users, and that therefore it is worth investigating how such hybrid environments will

---

[6] DoD STD 5015.2 was introduced in 1997; even by 2005 the Federal Chief Information Officers Council expressed their concern about the "disparate processes" being used by Federal agencies to manage records and offers an RM profile according to which the choice of STD 5015.2 was only one of three possible choices. See Federal Enterprise Architecture Records Management Profile, Version 1.3, Draft, July12, 2005, accessed 9/21/2008 from http://colab.cim3.net/file/work/geocop/20050804_Posting/RM Profile.071205.v1.3.pdf

[7] Lucy Suchman, "Office Procedure as Practical Action: Models of Work and System Design," *ACM Transactions on Office Information Systems* 1(4), October 1983, 320-28. This research pointed clearly to the disconnect between the procedural paradigm of recordkeeping and what actually takes place, yet the procedural view, promoted by the application of diplomatic theory to digital recordkeeping that produced the DoD STD 5015.2, continues to be instantiated in the standard, which has in turn had its effect on such derivative standards as the EU's MOREQ and the Australian VERS.

[8] See Steve Bailey, *Managing the Crowd: Rethinking Records Management for the Web 2.0 World* (London: Facet, 2008).

emerge: I suggest that the affordances of COTS software products and Web 2.0 tools will be significant drivers in such a development.

To investigate these ideas, the essay that follows will draw from pertinent areas of four literatures:
1) classification and conventional records management practice;
2) the historical development of office work practice and the role of office software suites, based on the desktop metaphor and its extensions, in document creation, retention/use, and personal disposition, to include underlying research on user behaviors;
3) the emergent literature on the use of tagging and folksonomies in the enterprise and the response of records managers to Enterprise 2.0 phenomena;
4) the literature on business communication as situated and evolving action, affected by genre and medium as well as alterations in business aims and practices.

Knowledge workers as records creators

I will be focusing in this study on the work of so-called "knowledge workers," professionals charged with carrying out their work mostly as they see fit and in the course of it making use of computer technologies in a creative way to solve problems and come up with new ideas that bring profit or praise to the enterprise. Given the assumed application of the NARA Flexible Scheduling scheme, however, it is worth beginning with a brief discussion of the kinds of records with which most of records management labor seems to have been occupied (because it has represented the largest body of records), that of the large number of clerical workers who employ computer technologies in specific and circumscribed ways to carry out routinized tasks.[9]

In the context of routinized clerical work, the actual value of the work lies precisely in rigid adherence to established system. Routine enhances rather than devalues clerical productivity, and employees are supplied with specific genre structures to employ in creating records.[10] Most production of digital records by clerical staff consists of finished products, completed in a short time.[11] Where work consists of data input, although it is now generally performed on a personal computer connected to a networked server, it is carried out in a strictly controlled, dumb-terminal-emulating green-screen environment; employees doing this kind of work may apparently do little else. Where work consists of general clerical work such as typing, even though it may be performed using a standard suite of desktop programs, it will frequently be constrained by the use of

---

[9] The distinction between "knowledge workers" and "clerical workers" is of course an artificial one; office ethnographies have shown that clerical workers often expend equal creativity in just making inadequate systems work. The difference pointed to here is that knowledge workers are rewarded for innovation and given wide leeway in how they perform their jobs, whereas clerical workers are discouraged from innovating overtly. Additional research is needed to discover whether the "neat" use of desktops by clerical workers is due to the restrictive nature of their job requirements rather than their preferences.

[10] For the application of the rhetorical concept of genre to communicative forms, see JoAnne Yates and Wanda J. Orlikowski, "Genres of Organizational Communication: A Structurational Approach to Studying Communication and Media," *The Academy of Management Review* 17(2), April 1992, 299-326.

[11] Note that Suchman (1983) shows that such "finished products" often mask complex problem-solving tasks neither recognized nor facilitated by information systems.

compulsory templates for the type of document being created. Thus in fact most clerical workers using personal computers connected to departmental servers, producing finished products about whose underlying task elements nobody wants to know, in fact create records that are classifiable into a very limited set of records series.[12] This raises the question of why such workers would be asked to classify records at all, since the records themselves can be scanned for recognizable elements like the use of a particular template or form (thus providing genre identification) and metadata about content can also be extracted from the readable text entered. In short, although much has been made of the difficulty of training clerical staff to classify digital records, in fact automated systems should be capable of carrying out most classification of the final-product records created by such workers, once supplied with a set of business rules based on the job description and competencies of the clerical staff member and the records series normally created.[13]

How are "knowledge workers" different? In general their main work is nonroutine or professional. Because as professionals they enjoy a marked autonomy in their work practices, the problems raised by their digital records creation and use are similar to those seen in the realm of personal information management. They are required to solve problems, communicate with a range of people, call the tacit knowledge they possess into play and engage in information seeking and sensemaking behaviors. Although they may well create some well-defined record types related systematically to their job descriptions and competencies, a good deal of their use of computer technologies of all kinds—and these kinds may extend to a variety of mobile equipment in addition to desktop or laptop personal computers—will be much more informal than that of clerical workers.[14] They are responsible for gathering *data*, not entering it; for adding value to data through categorization, analysis, and summarization, thereby generating *information* from that data; and for creation of *knowledge* from that information through such complex procedures as research providing comparison, establishment of relationships with other

---

[12] In a preliminary 2001 study of the distribution of record series among work units in an elaborate functional records schedule for the Texas Railroad Commission, Galloway and Susan Cisco found this to be the case. See NSF grant proposal, "Legitimating e-mail for official government business: Automatic adaptive classification," 2001.

[13] Recent texts on records management for digital records assume that most metadata provision for records from this kind of work will be generated or harvested automatically. See Kate Cumming, "Metadata Matters," in Julie McLeod and Catherine Hare (eds.), *Managing Electronic Records*, 34-49 (London: Facet, 2005). I leave aside here the deprecated underlying tasks, although they may in fact, if ignored, lead to compliance problems in the form of the use of computer technologies to accomplish these tasks in an ad hoc manner via "new" applications like email, text messaging, etc.

[14] Suchman (1983), echoed in Elizabeth Yakel, "The Way Things Work: Procedures, Processes, and Institutional Records," *American Archivist* 59, Fall 1996, 454-64, pointed to organizational theorists' unease here: such "informal" or "unstructured" activities were thought to "undermine the organizational objectives of rationality and control" (321). This kind of concern is now embedded in law and procedure, even though organizational ideals have in some cases moved on: see Carol Choksy, *Domesticating Information: Managing Documents inside the Organization* (Lanham: Scarecrow, 2006), for the assertion that whatever might be the archivist's ideal, RIM professionals have to deal with the statutory compliance environment they find, even in a context where organizational procedures and practices are changing. Geoffrey Bowker and Susan Leigh Star, *Sorting Things Out: Classification and its Consequences* (Cambridge: MIT Press, 1999) have shown how strong such a statutory instantiation of a dead paradigm can be and have characterized the non-procedural improvisation to which it gives rise as "articulation work" meant to smooth over contingencies that arise in situated actions (p. 310).

information, determination of consequences implied by information, and discussion of information with others.[15] Some of these processes (e.g., statistical analysis) are relatively well-defined, but knowledge workers will use and must understand appropriate software to carry them out and will have the expertise to choose such tools and to define appropriate parameters for their use. Others of these processes (e.g. research) are very much open-ended, and may eventuate in many sorts of additional data, such as a set of website links, digital images gathered to document an experiment, and textual notes toward an intended report, the last containing a list of ideas, a partial outline, and telephone numbers and email addresses of individuals to contact. Although knowledge workers do carry out short-term work (frequently now in the form of emails), they more frequently produce "work product" materials that form a working archive for an ongoing project. They may also generate temporally deeper archives (of contacts, for example) that they use and update throughout their working lives. And as a number of recent business writers have observed, that part of their knowledge that is not tacit is frequently couched in forms that are not only difficult to codify for management, as has long been known, but doing so may discourage their creative activity.[16] The real problem knowledge workers create for knowledge managers as well as records managers is that it is often difficult or impossible to specify what is crucial to their work at the outset, and although records managers may decide to rule most of it out as ephemeral, knowledge managers may not agree, as increasingly businesses find that even knowledge that is created seemingly peripherally can turn out to be valuable to the organization.[17] To meet the concerns of both, is it possible that there could be a different way to approach the task of finding such knowledge?

"Big Buckets" as a solution in search of a problem

Susan Cisco began her 2007 white paper on big buckets with a useful outline of the relevant steps that have led in the direction of this scheduling method as office organization felt the impact of automation. In the 1960s and 70s, retention schedules followed the model of NARA record groups, based on departmental structure and mirroring—indeed serving as the first step for the establishment of—archival provenance. In the 1980s, as government records managers were pressured by growing quantities of records, the use of functional analysis led to the development of general schedules and hierarchical decomposition of record groups into series. But these efforts, while they were effective in bringing order to that which records managers had time to attend to, required such granular analysis that records managers everywhere were unable to keep up with paper records, never mind the new digital formats then beginning to take

---

[15] See Thomas H. Davenport and Laurence Prusak, *Working Knowledge: How Organizations Manage What They Know* (Boston: Harvard Business School Press, 2000), Chapter 1, in which the authors make the distinctions among these three.

[16] See Eric Abrahamson and David H. Freedman, *A Perfect Mess: The Hidden Benefits of Disorder* (New York: Little, Brown & Co., 2007), for an argument in favor of randomness as a source of innovation.

[17] Choksy, *Domesticating*, 160-65 has an analysis of how this came to be, which includes the placement of active digital records into a "content management" compartment and the historical confinement of records managers to semiactive or inactive records as well as paper ones, most of which are increasingly determined valueless to the organization.

over government and business. NARA's answer to this problem was Flexible Scheduling.[18]

NARA in its NARS avatar had initiated American government records management in response to the challenges of World War II recordkeeping in 1941. Codified by Theodore Schellenberg, promoted for the states by Ernst Posner, the outline above charts NARA's struggles with records management in large, while its appallingly protracted effort to initiate management of digital records beginning in the 1970s has only recently been made more understandable.[19] NARA's efforts were often dampened by the refusal of more powerful agencies to comply or regulated industries' tendencies to protest in an era where the cry of "unfunded mandate" was enough to flatten any initiative. Thus precisely in the area of digital records management, while attempting to deal reasonably with records scheduling over a possible changeover from paper to digital, NARA was repeatedly sued (as perhaps the most vulnerable agency involved) for its scheduling advice, notably GRS 20.

Such complications, most of them arising from the increasing creation of records by more and more people on networked personal computers, led to the development of a new strategic approach to records appraisal and scheduling, called "Flexible scheduling." NARA has defined this approach as "The application of appraisal criteria to multiple similar or related groupings of information across one or multiple agencies to establish a uniform retention period."[20] If this seems very like large general schedules, it is, except that the goal of the embrace of large swaths of records in a single schedule was to reduce the decision-making burden on records creators, who are now (since NARA's endorsement of DoD STD 5015.2 for all agencies) being asked not only to create records, but to classify them into a schedule. It is not clear when these all-embracing schedules came to have the nickname "big buckets," but four years after the first announcement of the flexible scheduling principle, NARA herself was using the term: "A 'big bucket' or large aggregation schedule...grouped at a level of aggregation greater than the traditional file series or electronic system..."[21]

Such "buckets" were not to be applied to everything, but close enough: they were to be limited to records considered temporary (including all records with retention periods of 20 years or fewer, estimated by NARA to include more than 90% of agency series). The goal was minimum risk and minimum difficulty of classification. But the main point was that they were to be few, because it was believed that classification was easier if there were few categories to consider, and by the turn of the twenty-first century, as suggested above, functional analysis when carefully applied meant that large agencies might have not hundreds but thousands of series, and each series might have a schedule.

---

[18] Susan Cisco, "Streamlining Retention Schedules: The Benefits of 'Big Buckets.'" *Iron Mountain White Paper*, 2007.

[19] Bruce Ambacher, ed., *Thirty Years of Electronic Records* (Lanham, MD: Scarecrow, 2003), is a compendium of essays by people who survived the trench warfare leading up to today's efforts toward the ERA.

[20] National Archives and Records Administration, "Strategic Directions: Flexible Scheduling," January 2004; retrieved 2/07/08 from http://www.archives.gov/records-mgmt/initiatives/flexible-scheduling.html

[21] National Archives and Records Administration, "Bulletin 2008-04, Guidance for flexible scheduling," April 30, 2008; retrieved 7/9/08 from http://www.archives.gov/records-mgmt/bulletins/2008/2008-04.html

But for whom were things really made easier? Claims were made about how many "buckets" should exist, based on the work of psychologist George Miller. Miller had reviewed and gathered evidence, based upon absolute judgments of stimuli in experimental conditions, that seemed to prove that the capacity of human short-term memory was seven items, plus or minus two.[22] Yet in fact Miller never made that claim in the paper usually cited and subsequent research has suggested that the claim would anyway only apply to "speakers of English attempting to remember a sequence of digits."[23] In the discussion above I have pointed out that most clerical employees create few kinds of records, so few indeed that in most cases they can remember them all with no prompting, so whatever may be the differences introduced into more complex creative tasks by contextual memory, the classification task should not be so difficult for most clerical workers; even at the granular scale of many series, few will apply to any single worker. The problem is getting from the top-level "bucket" to the very lowest levels where the records are created.

For records managers, of course, who have the perspective of knowing the full set of schedules for any one enterprise, the prospect looks very grim indeed for the poor creator-classifier, but most clerical creator-classifiers just are not making a huge number of choices when it comes to the creation of deliverables. It may be suggested that if they are having such a problem, it might be due to a clash of terminologies. Records managers, with their broader organizational perspective, tend to create and especially name categories according to functional series, concepts, retention periods, or priority/temporality, and in most cases these category names have no apparent relation to the actual task space in which the records creator works or the granular business purposes for which the records are created. In that space "local knowledge" names differently the tasks and functions to which the records are related. What is needed is either a better grasp of how the creators see their creation, together with a crosswalk to local terminology, or the introduction of a means by which creators can control their own terminological domain.

Classification and management of creative work product

We first need to take a look at the kinds of classifiers that records managers use, and here we are immediately confronted by an ingredient that records creators are very unlikely to include in their conceptualization of the categories into which their records fall. In a recent article on taxonomies in records management (note that the terms taxonomy and classification are used interchangeably), Richmond and Bruno point out that the purposes of RM taxonomies include identification, retrieval, and the enabling of the application of retention requirements, which they suggest "may result in a looser application of classification rules." Their rule of thumb is that the "top buckets" of the taxonomy should number no more than seven plus or minus two, while the hierarchical levels should be no more than four (which by my count leads to a maximum of 6561 "buckets" at the lowest level and 7380 in all for 9 (7+2) top-level categories; and a

---

[22] George A. Miller, "The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information," *The Psychological Review* 63 (1956), 81-97.

[23] Derek M. Jones, "The 7±2Urban Legend," MISRA C Conference, 2002; accessed 8/24/08 from http://www.knosof.co.uk/cbook/misart.pdf

minimum of 625 at the lowest level and 780 in all for 5 (7-2) categories, assuming that each parent has the same maximum number of children in the hierarchy—thus demonstrating how strange are the results of using Miller's figure in a way not intended).[24] Another recent article explicitly identifies the recommended four levels as function, process, records series, and documents, with a possible addition of a fifth level for business process categories and a sixth for project sets that constitute business processes, presumably inserted between process and records series and perhaps helpfully getting closer to the creator's categories.[25]

Such complex classification systems, as Star and Bowker have pointed out, reflect a way of describing bureaucratic infrastructure. Indeed the tasks of creating and maintaining classification systems constitute work that is important to keeping that infrastructure intact by reifying it and thereby making it reproducible.[26] Where such classification systems are shared and agreed, they can function in this way. Records managers are advised take a top-down view when constructing a classification system, and since clear guidelines are in place for the classification of records created by specific basic functions of enterprises (financial records, personnel records, etc.)—guidelines, it might be added, that are frequently based on the practices of the specific professionals who work in these fields—these parts of the system are relatively easily constructed by adopting preexisting standards.[27] There will apparently be a good fit because the classification is governed by the professional training of the actual workers involved, which may even already be codified in instructional manuals that describe standard "genre systems" of communicative actions related to specific professional actions and their corresponding documentary forms.[28] And "big buckets" can be used because it is already clear to the workers how the large categories subsume the smaller ones.

Where classifications attempt to categorize the work of less disciplined activities, however, problems can arise when the classification, however rational when seen from a management perspective, is unfamiliar to the records creator and requires changes in her work practice. This is especially likely to happen when the record-generating activity in question is one where the worker is left to organize her work as she sees fit, either because the activity has a very well-defined deliverable and nobody cares how it is organized or because the activity is undefined and incorporates a significant degree of initiative and improvisation on the part of the worker. In the latter case, many have argued, the leading edge of innovation is to be found, and it is therefore no wonder that the records manager finds it difficult to classify the products of such work (which have no precedent) or that the worker finds it difficult to follow a generic classification: in fact, the worker does not know at creation how the work product should be classified, and may not know for some time.[29] From the perspective of the enterprise, it is in the best interest

---

[24] Heather Richmond and Denise Bruno, "The Truth about Taxonomies," *Information Management Journal*, March 1, 2003.

[25] Carol Choksy, "8 Steps to Develop a Taxonomy," *Information Management Journal*, Nov-Dec 2006.

[26] Bowker and Star, *Sorting Things Out*.

[27] I would suggest that the vast majority of records management classification construction falls under this heading of building existing standards into the classification architecture.

[28] Yates and Orlikowski, "Genres."

[29] The notion that innovation takes place at the edge of chaos comes from the work on artificial life and

of everyone that this uncertainty be allowed to persist until further work or serendipity lead to an emergent new thing (only this kind of permissiveness allowed the barely sticky substance on the back of Post-It notes to survive to become a product defining a new category; as "glue" it was useless). Since in the "knowledge economy" work of this kind has increased in frequency, entirely new practices for nurturing this kind of uncertainty by managing knowledge product have emerged in the guise of "knowledge management," which frequently takes the form of providing tools to the individual for managing her *own* records or to groups for creating their own "archives" for the support of group projects.

At this point I want to take up the issue of what may be the effects of the so-called "2.0" functionalities that are being so widely discussed: web 2.0, desktop 2.0, enterprise 2.0. To start with the number itself, in the world of software versioning, the 2.0 version is the first big overhaul of a software product. Web 2.0, according to that 2.0 of websites, Wikipedia, is "a living term describing changing trends in the use of World Wide Web technology and web design that aims to enhance creativity, information sharing, collaboration and functionality of the web. Web 2.0 concepts have led to the development and evolution of web-based communities and hosted services, such as social-networking sites, video sharing sites, wikis, blogs, and folksonomies."[30] This very quote takes us to Desktop 2.0, because as I cut and pasted it from the Internet into this document (which will in turn be posted on the Internet), it retained its hyperlinks in my word processor and thereby made this document itself a potential window into the Internet, gave it extra dimensions beyond those of a static document should it be interpreted by software able to make such connections. And as to Enterprise 2.0, the term has been applied to those enterprises that make use of the tools of Web 2.0, whether on the Internet itself or within enterprise intranets, for the express purpose of enabling collaborative work and encouraging creativity. Part of the anticipated effects of introducing such technologies into the enterprise is expected to be the surfacing of tacit knowledge, the encouragement of "information ecologies," and the capture of unstructured information.

Enterprise 2.0 tools are going to have many significant effects on how records managers do their work, but one important impact will be on classification. In the 2.0 environment, classification emerges rather than being imposed. That is not to say that it is any easier to achieve what we now call classifications, because a folksonomy still requires agreement among a relevant community of practice before it can become useful. What it does mean is that classifications will be sought out for preservation rather than imposed, because new sensemaking efforts are potentially more valuable than reductive analyses of supposedly well-understood structures. The point is that through the various forms of social networking and informal collaboration facilitated by these tools, what people know and think is made manifest and searchable if new analytics can be developed to make sense of the information captured in them. It should be pointed out that the stratospheric success of the Google search engine has been based on the invention and exploitation of just such analytics.

---

chaos theory (see Stuart Kauffman, *At Home in the Universe* [New York: Oxford University Press, 1995], 86-92) and is now manifest in a wide range of business books like Abrahamson and Freedman, *A Perfect Mess*. See also Richard McDermott, "Why Information Technology Inspired but cannot Deliver Knowledge Management," *California Management Review* 41 (4), summer 1999, 103-17; see 109.

[30] Wikipedia, "Web 2.0," retrieved 10/5/08 from http://en.wikipedia.org/wiki/Web_2.0

But although it is trendy to talk about the impact of Web 2.0 on everything, Wikipedia reports that Tim Berners-Lee (who really did invent the Web) points out that the implications of such "new" ideas were always there, from the first implementation of the hyperlink.[31] I would like to go further and argue that the empowering of the knowledge worker had begun long before, in an environment where local network connection was assumed and individual work was to be facilitated by the new technology of the personal computer. That empowerment came from the desktop metaphor itself.

The Desktop Metaphor as a working environment for "creatives"

Many people have heard about the terrible misjudgments through which the researchers at Xerox Palo Alto Research Center (PARC) discovered and designed the virtual desktop that made such a difference in expanding the popularity of personal computers, only to have their innovative work almost completely ignored by their own company and then appropriated by first Apple (for the Macintosh) and then Microsoft (for Windows). The real story is more complex than that, but what is certain is that the original desktop interface, an immersive environment with its icons and mouse, direct virtual manipulation of assets, network connection, and laser printer, was incrementally invented and pulled together by extremely creative people for, in the first instance, *their own use*, explicitly based on psychological theory about how to assist creative thinking.[32] The first "desktop metaphor" saw the light of day on the screen of an Alto computer, which was put together in-house for PARC researchers. For these people, productivity meant anything but routine and what was valued as output was not data or information, but knowledge—new knowledge. The PARC researchers built an environment in which just such non-routine work was supported.[33] And unlike the later scenario of the personal computer with the spreadsheet as its "killer app," in fact the most important application for the Alto and its descendant desktop display machines was a graphical word-processing application called BravoX, which was the principal application installed on experimental Altos placed in 1978 at the White House, the Senate, and the House of Representatives. A year later BravoX went with one of its architects, Charles Simonyi, to Microsoft, where it evolved into Word.

When Xerox finally did implement the desktop environment in a commercial product, it was in a machine called the Star (1981), which was aimed at executives.[34]

---

[31] Reported by Nate Anderson in Ars Technica, September 1, 2006, retrieved from http://arstechnica.com/news.ars/post/20060901-7650.html on 9/25/08.

[32] Patrick J. Lynch, "Visual Design for the User Interface Part 1: Design Fundamentals," *Journal of Biocommunications* 21(1), 1994, 22-30. The researchers at PARC, like many computer and information scientists then and now, were also inspired by the famous essay on personal information management using a machine with perfect recall by Vannevar Bush, "As We May Think," *The Atlantic Monthly*, July 1945.

[33] A popular account of the PARC phenomenon is Douglas K. Smith and Robert C. Alexander, *Fumbling the Future: How Xerox Invented, then Ignored the First personal Computer* (New York: Quill, 1988); a more serious treatment can be found in Michael Hiltzik, *Dealers of Lightning: Xerox PARC and the Dawn of the Computer Age* (New York: HarperBusiness, 1999).

[34] Hiltzik, *Dealers*, 361-70. See also William L. Bewley, Teresa L. Roberts, David Schroit, and William L. Verplank, "Human Factors Testing in the Design of Xerox's 8010 'Star' Office Workstation," *Proceedings of CHI '83*, 72-77; also see David Canfield Smith, Charles Irby, Ralph Kimball, Bill Verplank, Eric Harslem, "Designing the Star User Interface," *Byte Magazine* 7(4), April 1982, 242-82.

> SSD [Xerox's Systems Science Division] intended the Star to replicate a typical office setting by employing, among other things, pictures ("icons"), lists of action choices ("windows"), and multiple screen sections ("windows") to electronically re-create the desk top, file cabinet, telephone, in- and out-boxes, wastepaperbasket, and other features familiar to office workers. In addition, the Star's software sought to keep the user's work products—files, projects, calendars, etc.—available at all times.[35]

The derivative Apple Lisa (1982) was also aimed at executives; Apple's statement about the Lisa's design aims said "Lisa must be fun to use....Special attention must be paid to the friendliness of the user interface and the subtleties that make using the Lisa rewarding and job-enriching."[36] But the real commercial success for Apple came with the Macintosh, which soon became the machine of choice for designers and artists. Finally, at Microsoft, Simonyi also participated in the early design of Windows, which not only offered the virtual desktop to all office workers but expressly aimed at enhancing the agency of the user. Thus the original work at PARC, focused on supporting the creative work of researchers and executives, lies at the heart of the persistent desktop interface.[37]

One seminal piece of the underlying research for the project at PARC, especially significant for this study, addressed how people managed their information in their own workspace and explains much about the aims realized by this series of machines. Thomas Malone carried out research to support the desktop metaphor design by looking in detail at how people worked with papers in offices.[38] The participants in his study were chosen for their concentration on desk work and included 3 secretaries, 3 research scientists, a technical manager, an administrative manager, a purchasing agent, and a physician. He found that there were two general categories of office behaviors with reference to documents: filing (characteristic of people with routinized jobs, tending to neatness; his example was a purchasing agent) and piling (characteristic of people with complex/unpredictable/creative jobs, tending to messiness: his example was a research scientist). He learned that informally arranged piles of documents were found by location, that much of desktop organization "consists of untitled piles," yet most then-existing computer systems required that a new document or folder be titled and placed in one location. He then went on to advise that automated systems to support both types (but especially to address the "piler" as the predominant case) needed to pay attention to the variability of job content, to allow piles as well as files, to classify very simple things

---

[35] Smith and Alexander, *Fumbling*, 231.

[36] Quoted from an Apple document in Hiltzik, *Dealers*, 343.

[37] Note that the desktop windowing system designed to be run on Unix workstations, X, was influenced by and in turn has influenced the dominant personal computer desktop system interfaces.

[38] Malone, Thomas W. "How Do People Organize Their Desks? Implications for the Design of Office Information Systems," *ACM Transactions ob Office Information Systems* 1(1), January 1983, 99-112. For more recent office ethnography along the same lines with larger samples (delineating "neats" and "scruffies") that argues that more information can be stored in a random system than in an ordered one, see the work of David Kirsh and his students Justin Kodama and Shailendra Rao, the Context Aware Office project: http://adrenaline.ucsd.edu/external/projects/contAware/ethnoStudies.html

automatically, but also to allow deferred classification and the placement of one thing into several classes. The abstract to this widely-cited paper is worth quoting at length:

> Two principal claims are made: (1) A very important function of desk organization is to *remind* the user of things to do, not just to help the user *find* desired information. Failing to support this function may seriously impair the usefulness of electronic office systems, and explicitly facilitating it may provide an important advantage for automated office systems over their nonautomated predecessors. (2) The cognitive difficulty of categorizing information is an important factor in explaining how people organize their desks. Computer-based systems may help with this difficulty by (a) doing as much *automated classification* as possible (e.g., based on access dates), and (b) including untitled "piles" of information arranged by physical location as well as explicitly titled and logically arranged "files."[39]

All of these features were or have since been realized in the design of the virtual desktop, and improvements in virtual desktop design are still striving to instantiate them more faithfully.[40] Malone's research had even more influential things to say about support for locating unspecifically titled documents:

> A computer-based information system can implicitly classify all the documents it handles according to the dates when they were accessed. Then the user can search for otherwise unclassified documents according to when he remembers creating, receiving, or otherwise accessing them. One can even imagine a system where users search for a document by a kind of simulated time-lapse photography of the history of their electronic desktop. They could "rewind" and "fast forward" the desktop to locate the last time the desired document was on the desk.[41]

Some ten years later, Deborah Barreau at the University of Maryland and Bonnie Nardi at Apple Computer brought together their similar research on how people used the digital desktop, which by that time was firmly established as the standard user interface; they tested a small sample of users of Apple and IBM PC systems using the respective desktop interfaces on those systems.[42] They made several findings, now echoing Malone's paper findings on a digital desktop.

---

[39] Malone, "How Do People Organize," 99. Like the Vannevar Bush essay, this piece is still being mined for insights by the Human-Computer Interaction research community.

[40] Victor Kaptelinin and Mary Czerwinski (eds.), *Beyond the Desktop Metaphor: Designing Integrated Digital Work Environments* (Cambridge, MA: MIT Press, 2007); in spite of the title, perhaps half of the essays in this book describe projects to enhance the virtual desktop design.

[41] Malone, "How Do People," 109. A similar feature is now available on the Macintosh (Snow Leopard) and Microsoft (Vista) desktops.

[42] Barreau, Deborah, and Bonnie Nardi. "Finding and Reminding: File Organization from the Desktop," *ACM SIGCHI Bulletin* 27(3), July 1995.

1) Users preferred location-based finding because of its reminding function; in fact, they preferred to search for files by hand rather than logically, because they wanted to search in a particular place. In aid of this, they placed files in locations where they expected to find them, using the virtual desktop itself for placing and grouping active files.

2) Users avoided elaborate filing schemes because they were seen as not worth the time and trouble they might take to organize; instead, they spent time devising file naming conventions, also for their reminding function while browsing.

3) Users distinguished three temporal categories of information: ephemeral, working, and archived, mirroring another early study on paper records.[43] Ephemeral information was very current, only used for a short time, often received from others, and often left on the desktop. Working information was current work, needed for weeks or months, for as long as a project lasted; the authors remarked that it "is often created by the user or is the product of the user's work group," thus pointing again to the agency of the creators observed in the study. Archived information mostly represented completed work, rarely accessed, and represented a real problem for creators in organizing it to create the archive.

4) Users archived relatively little. This finding, the authors pointed out, stemmed from the fact that they were looking at routine office activities; in contrast, they admitted that "Researchers *do* keep a good deal of information around for a long while because it is in their interest to do so." And they were more concerned with the failure of system designers to assist users with the floods of incoming ephemeral information than with the problem of "institutional memory and its potential importance." Barreau and Nardi concluded that "[u]sers file information not according to systems or keywords or carefully architected logical schemes, but according to the dictates and vagaries of the kind of work they are doing and the type of information they are dealing with."

This study was quickly taken up by a research group at Yale, Scott Fertig, Eric Freeman, and David Gelernter, who suggested that Barreau and Nardi had drawn faulty conclusions from their own research.[44] They argued that location-based finding wasn't a true user preference, but stemmed from poor implementation of logical search mechanisms and user adoption of another method as a stopgap measure. As far as archiving was concerned, they oddly cited a rare paper by an archivist (Canadian Terry Cook) to indicate that organizational archiving is important (!), and then argued that again, Barreau and Nardi's subjects failed to archive because archiving mechanisms were so poor. They blamed these observations on the desktop metaphor, then reviewed new approaches to improving search tools or desktop design that they felt would change things. One of these designs was their then-in-progress system called Lifestreams, which managed the creation of documents and automatically inserted them into a chronological image of documents on the screen that users could manipulate through logical queries. It provided for locating information because current materials were uppermost in the

---

[43] I. Cole, "Human aspects of office filing: Implications for the electronic office," *Proceedings Human Factors Society* (Seattle, 1982); quoted in Barreau and Nardi 1995. Cole's categories were action information, personal work files, and archived information.

[44] Scott Fertig, Eric Freeman, and David Gelernter, "'Finding and Reminding' Reconsidered," *ACM SIGCHI Bulletin* 28 (1), January 1996.

stream; it allowed workers to remind themselves of coming events or deadlines by emailing themselves messages to be activated in the future. It also provided for archiving (or as we might notice, a primitive kind of records management) by automatically freezing documents in their current state after some user-controlled period, enabling the user to edit the document only by cloning the original document and thus creating a new one.[45]

In 1997 Nardi and Barreau replied, reinforcing their claims about user preferences for location-based search by discussing their research in greater depth.[46] They argued that workers use spatial perceptions to organize their entire workspace—not just the virtual one—and that they indeed possess a large capacity to so organize their working knowledge; more important, perhaps, doing so gives workers a sense of control and confidence. Furthermore, they stressed, this kind of reminding function for spatial organization is richly multidimensional, depending additionally upon development of habit patterns by the individual worker as to placement of materials (which may be related to broader patterns in the worker's behavior when not at work; recent micro-ethnographic work has showed how important locational search is in everyday life situations around the home).[47] Another point they made, that for most workers archiving is not an important focus, was emphasized by a review of interview data, which indicated that workers preferred to depend on others for material that was truly out of the realm of their current and not-long-past work (again, however, the authors excluded knowledge workers). Additionally, they argued, the Lifestreams model was inappropriate for archiving materials in such a way as to be useful to many people, since the Lifestreams "diary" model did not address issues of collaborative data management.[48]

One more major effort at looking at information as organized through the desktop metaphor is worth including because it built toward formal logical document-management in ways that included both logical finding and user classification activity, and it promises eventually to have a major impact on the infrastructure of commercial systems. This was the Placeless Documents project, once again homed at Xerox PARC.[49] The project built on the Barreau and Nardi insight that users chose locational finding over the construction of conventional hierarchical file structures. Its central concept is to manage and potentially represent stored documents based on attributes (metadata) assigned to them by the creator and/or user (including categories, keywords, links to

---

[45] For more detail see Eric Freeman, *The Lifestreams Software Architecture*, Dissertation, Department of Computer Science, Yale University, 1997.

[46] Bonnie Nardi and Deborah Barreau, "'Finding and Reminding' Revisited: Appropriate Metaphors for File Organization at the Desktop," *ACM SIGCHI Bulletin* 29 (1), January 1997.

[47] See Laurel Swan, Alex S. Taylor, and Richard Harper, "Making Place for Clutter and Other Ideas of Home," *ACM Transactions on Computer-Human Interaction* 15 (2), Article 9, July 2008. See also Daniel Miller, *The Comfort of Things* (Cambridge: Polity Press, 2008), especially Portrait 6, "The Aboriginal Laptop," 67-72; and Annalee Newitz, "My Laptop," in Sherry Turkle (ed.), *Evocative Objects: Things We Think With* (Cambridge: MIT Press, 2007), 87-91.

[48] Note that one of the main presentations of the Lifestreams project was entitled "Lifestreams: A Storage Model for Personal Data" (by Eric Freeman and David Gelernter, *ACM SIGMOD Bulletin*, March 1996).

[49] Dourish, Paul, W. Keith Edwards, Anthony LaMarca, John Lamping, Karin Petersen, Michael Salisbury, Douglas B. Terry, and James Thornton, "Extending Document Management Systems with User-Specific Active Properties," *ACM Transactions on Information Systems* 18(2), April 2000, 140-170.

other items, translations) and by the system itself (conventional technical descriptive elements). These attributes would be maintained in a relational database separately from the documents in a system. The attributes could even include executable code or "active properties" that could provide a mechanism for generalized document management. Properties, not file-system representations, would be the uniform way documents could be accessed and interacted with. And the Placeless proposal provided for the creation of what were called "fluid collections" defined by both the queries and exclusion/inclusion lists that produced them.

The Placeless Documents project is particularly interesting because of the research base it has provided for long-sustained implementation work by Microsoft, certainly still the most influential single provider of user interfaces in the world. Since early in the new century Microsoft programmers have been working on a still-unreleased new file system called WinFS that shares many of the features of Placeless Documents, although adding document type schemas to allow the system to know what executable tools to apply to them rather than packaging tools as attributes of the documents themselves. WinFS as originally mooted would build on Placeless display ideas to permit different virtual views of documents through standard desktop tools like calendars and allow the differential dynamic gathering of collections based on attribute queries.[50] Microsoft's current intention is apparently to make this new file system available at some point, particularly because it solves many problems to do with storage management and interoperability of file types, but it now seems that this will be in some post-Vista future. Nevertheless it is clear that there will be significant user assignment of attributes involved in this approach, and this concern on the part of Microsoft for the individual user is underlined by their considerable personal information management initiative,[51] to which far more resources have been devoted than to their records management initiative.[52]

Whatever takes place behind the screen, however, user studies continue to evaluate and critique the usability of the desktop metaphor, especially with the emergence of the specific study of Personal Information Management. Ravasio et al. carried out a user study prefaced by a literature review (surprisingly brief) of previous such studies. They found that users did in fact archive, but found it difficult and labor-intensive to construct file hierarchies that suited their requirements. In spite of this, users persevered, since "all the efforts invested in organizing, naming and maintaining the hierarchical file system structure are aimed at (1) engraving the information's content and context into the system, and (2) providing an overview at a single glance, without having

---

[50] See the "WinFS" article in Wikipedia, accessed 10/02/08 from http://en.wikipedia.org/wiki/WinFS

[51] cite MyLifeBits and the RM blog; point out that RM will be at server level as a SharePoint "site" and it's not yet clear how client computers will be affected or how this might interact with WinFS.

[52] The Microsoft Records Management initiative appears to be a small project offering an all-Microsoft answer to the DoD STD 5015.2 records management application requirements, making it possible for users of SharePoint server to set up a virtual Records Center "site" to which specific metadata templates for employee declaration can be connected as business rules. See Brian Chee and Oliver Rist, "The great Office Server smorgasbord, Parts 1-4," *Infoworld*, August 13, August 20, August 27, and September 12, 2007; accessed 7/20/2008 from http://www.infoworld.com/article/07/08/13/33TCmoss_1.html (first instalment).

to access information until it is really needed."[53] Like Fertig et al. they excoriated system-supplied search methods as too complicated to use (desktop Google search was not yet available), but they admitted that there was efficacy for users in manual searches because it refamiliarized them with their personal archives. In general, however, the paper was a call for the addition to the desktop of functionalities that would make individual work practices more effective, mirroring new affordances of email and Internet information sources, rather than constraining them further.

A slightly different and significant voice, turning away from ad hoc search, came from another group concerned with personal information management. Jones et al. investigated the use of folders for structuring information.[54] They found quite clearly that far from folders' being a troublesome artifact of hierarchical file-system schemes, users were in fact using them interactively to decompose and organize projects, and once created, folder structures helped users visualize and understand their information. The authors concluded, "Folders may be valuable information in their own right and not just a means of organizing information." It is worth recalling that the Placeless Documents project envisioned that a dynamic collection of documents in a Placeless system would itself be a valuable document managed in the system.

Researchers continue to exploit and expand the desktop metaphor, continually in the direction of empowering users to manipulate information in multiple dynamic ways during the course of their work with it, allowing them to defer classification or use classification in emergent ways as suits the task at hand. A dramatic example is the experimental "BumpTop" interface, which uses a high-resolution screen, simulation of the physics of moving papers and piles and stacks of papers around, and a set of gestures that allow the user to stack, pile, fan out, and restack "papers" (represented by document thumbnails—which can even be "crumpled") on a realistically simulated desktop.[55] Returning to the existing literature on real desks and paper, the researchers worked to exploit the real affordances of paper-on-desktop while adding affordances (being able to look inside, search, etc.) not available in the physical environment.[56]

Similarly clever and "life-like" affordances have been added to the new desktops offered by the Vista (Microsoft) and Snow Leopard (Apple) interfaces. Further, interface designers are looking increasingly to integrating so-called "visual computing" displays and manipulation modes into the user's workspace and devising means to manage information that the user does not even possess locally using collective categorization (folksonomies). Efforts such as these show, along with our awareness that the paperless

---

[53] Pamela Ravasio, Sissel Guttormsen Schär, and Helmut Krueger, "In Pursuit of Desktop Evolution: User Problems and Practices with Modern Desktop Systems," *ACM Transactions on Computer-Human Interaction*, 11 (2), June 2004, 156-80; 170.

[54] William Jones, Ammy Jiranida Phuwanartnurak, Rajdeep Gill, and Harry Bruce, "Don't Take My Folders Away! Organizing Personal Information to Get Things Done," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1505-08 (New York: ACM Press, 2005).

[55] Anand Agarawala and Ravin Balakrishnan, "Keepin' It Real: Pushing the Desktop Metaphor with Physics, Piles and the Pen," *ACM CHI 2006*.

[56] Malone, "How do People..."; Abigail J. Sellen and Richard H.R. Harper, *The Myth of the Paperless Office* (Cambridge: MIT Press, 2002); S. Whittaker and J. Hirschberg, "The character, value, and management of personal paper archives," *ACM Transactions on Computer Human Interaction* 8(2), 2001, 150-70.

office is indeed a myth, that both real and virtual desktops and their extensions are here to stay for some time longer, simply because they are effective. Clearly the design and usability research reviewed here shows that especially knowledge workers find useful the ways in which the desktop metaphor externalizes their innovative knowledge-production processes. Instantiating knowledge contexts visually on the virtual or real desktop makes it possible for the creator to work with it in ways that were impossible before the ideas were "out on the table."

Managing the Crowd: Allowing the crowd to manage itself?

Summer 2008's new book by Steve Bailey, senior advisor on records management to the British Joint Information Systems Committee for education and research, is called *Managing the Crowd: Rethinking Records Management for the Web 2.0 World*. In it, Bailey argues that the world is now irretrievably networked; enterprise data and even project work is often already out of the house and into the cloud; and records managers will be well advised to stop disciplining data and start herding it—all of it.[57] Carol Choksy in her *Domesticating Information* also points out that records managers no longer have the option to choose what records to manage: they must manage them all—even if after the fact. Meanwhile Gartner Group is producing white papers like "The Convergence of Enterprise Content Management and Archiving" and Knowledge Management has gained traction in the enterprise.[58] Finally, the very researcher who started it all, Thomas Malone—now a professor at the MIT Sloan School of Management—is now promoting a vision of work that has shifted from the command-and-control model with which records managers are so familiar, to what he is calling a coordinate-and-cultivate management practice in which especially knowledge workers are empowered to create through the latest in information technology.[59] And Bonnie Nardi has collaborated on a widely-read book about humanizing the workplace through empowering users and "using technology with heart."[60]

So what sort of "management" makes sense for records managers in an Enterprise 2.0 world where creativity is at the heart of value? In this world standard records management practices are not only ineffective in many cases where creative work is concerned, but have lost control both of whole information processes that have been outsourced and even offshored, as well as of information now being shared and worked on entirely outside the familiar client-server model. SharePoint may still be effective in keeping a hermetically-sealed enterprise tied to the client-server model, but even here the motive is to empower the user.

---

[57] Bailey, *Managing the Crowd* (London: Facet, 2008).

[58] Ken Chin and Carolyn DiCenzo, "The Convergence of Enterprise Content Management and Archiving," presentation at Gartner Storage Summit 2006.

[59] Thomas W. Malone, *The Future of Work* (Cambridge, MA: Harvard Business School Press, 2004).

[60] Bonnie Nardi and Vickie O'Day, *Information Ecologies: Using Technology with Heart* (Cambridge: MIT Press, 1998).

What is Records Management articulation work 2.0?

Years ago David Bearman, whom archivists and records managers alike enjoy dismissing because "he isn't one of us," told archivists that in the world of digital records it was time they looked to occupying a steering role lest they remain forever chained to the rowing benches of the informational trireme.[61] He saw quite clearly that it would no longer be feasible to keep all digital archival materials under official archival roofs and that archivists would have to learn to manage, authenticate, and provide access to them from afar, using the tools of the new technology. This was before the Internet explosion, before Web 2.0 and mashups, but it was telling: archival information management may be led by the archivist, but many will participate in it, not excluding the record creator. This message is doubly important for the records manager, and it is interesting to see the degree to which Bailey's book on the impact of Web 2.0 reiterates Bearman's warnings.

The key seen by both Bearman and Bailey is the use by archivists and records managers, for their own purposes, of the very digital technologies that seem threatening. One area of research that is especially helpful is in fact that of personal information management, simply because it concentrates on the dangerously individual desktop, because it has emerged from a recognition that people no longer want to sever their private and work lives, and because it is also concerned with multimedia objects as well as text—the whole gamut of digital information—produced, however, using consumer technologies. It is especially relevant because, as Bailey points out, whereas once the only desktop computer people had access to was at work, today that is not the case. Yet the only convenient way for people to unify their work and life information seeking is to migrate the whole process onto the Internet: Bailey uses the example of a single Del.icio.us account to keep a unified set of bookmarks from home and work.[62]

Another important thread of this research is its trust of user/creators and their knowledge of their own records. Records managers have long known that the real details, particularly of specialist work, escape the enterprise taxonomy, but provision for capturing users' informal ways of categorizing their work—by folders, naming, and piles, as well as changes in these over time—can also capture original perspectival information that itself has value. Research on the desktop interface has already produced the possibility of making sequential captures of desktop arrangements; the emergent WinFS operating system will record every change in all metadata for everything in a system as long as it remains there. What is most important for records management, I think, is to exploit its own creativity so as to address the opportunities that digital affordances offer rather than attempting to prune them away.

Records management, as I pointed out earlier, has always been to some degree "articulation work," Star's term for a bridging activity between ideal and reality, past and present states of affairs. It is clear that with the acceleration of change in information creation and handling, such articulation work is even more needed than ever. Where information itself is the much-valued medium of work in our society, we need

---

[61] David Bearman and Margaret Hedstrom, "Reinventing Archives for Electronic Records: Alternative Service Delivery Options," in Hedstrom (ed.), *Electronic Record Management Program Strategies*, Pittsburgh (1998), 82-98.

[62] Bailey, *Managing*, 142-44.

articulation work on the edge of chaos, that enables the conversion of information to knowledge by learning to tolerate and even support ambiguity.

References/Readings:

Abrahamson, Eric, and David H. Freedman. *A Perfect Mess: The Hidden Benefits of Disorder*. New York: Little, Brown & Co., 2007.

Agarawala, Anand, and Ravin Balakrishnan. "Keepin' It Real: Pushing the Desktop Metaphor with Physics, Piles and the Pen." *Proceedings CHI 2006*.

Bailey, Steve. *Managing the crowd: rethinking records management for the Web2.0 world*. London, Facet: 2008.

Barreau, Deborah, and Bonnie Nardi. "Finding and Reminding: File Organization from the Desktop." *ACM SIGCHI Bulletin* 27(3), July 1995.

Bearman, David, and Margaret Hedstrom. "Reinventing Archives for Electronic Records: Alternative Service Delivery Options." In Margaret Hedstrom (ed.), *Electronic Record Management Program Strategies*, 82-98. Pittsburgh, 1998.

Bewley, William L., Teresa L. Roberts, David Schroit, and William L. Verplank. "Human Factors Testing in the Design of Xerox's 8010 'Star' Office Workstation." *Proceedings of CHI '83*, 72-77.

Bowker, Geoffrey C., and Susan Leigh Star. *Sorting Things Out: Classification and its Consequences*. Cambridge: MIT Press, 2000.

Bush, Vannevar. "As We May Think." *The Atlantic Monthly*, July 1945.

Chee, Brian and Oliver Rist. "The great Office Server smorgasbord, Parts 1-4." *Infoworld*, August 13, August 20, August 27, and September 12, 2007. Accessed 7/20/2008 from http://www.infoworld.com/article/07/08/13/33TCmoss_1.html (first instalment).

Chin, Ken, and Carolyn DiCenzo. "The Convergence of Enterprise Content Management and Archiving." Presentation at Gartner Storage Summit 2006.

Choksy, Carol E.B. *Domesticating Information: Managing Documents inside the Organization*. Lanham, MD: Scarecrow Press, 2006.

Choksy, Carol E.B. "8 Steps to Develop a Taxonomy." *Information Management Journal*, Nov-Dec 2006.

Cisco, Susan. "Streamlining Retention Schedules: The Benefits of 'Big Buckets.'" *Iron Mountain White Paper*, 2007.

Cole, I. "Human Aspects of Office Filing: Implications for the Electronic Office." *Proceedings of the Human Factors Society*. Seattle, 1982.

Cox, Richard. "The Record: Is it Evolving?" *Records and Retrieval Report* 10(3), 1994, 1-16.

Cumming, Kate. "Metadata Matters." In Julie McLeod and Catherine Hare (eds.), *Managing Electronic Records*, 34-49. London: Facet, 2005.

Davenport, Thomas H., and Laurence Prusak. *Working Knowledge: How Organizations Manage What They Know*. Boston: Harvard Business School Press, 1998.

Dourish, Paul, W. Keith Edwards, Anthony LaMarca, John Lamping, Karin Petersen, Michael Salisbury, Douglas B. Terry, and James Thornton. "Extending Document Management Systems with User-Specific Active Properties." *ACM Transactions on Information Systems* 18(2), April 2000, 140-170.

Duranti, Luciana. *Diplomatics: New Uses for an Old Science*. London: Scarecrow, 1998.

Federal Enterprise Architecture Records Management Profile, Version 1.3, Draft, July12, 2005, accessed 9/21/2008 from http://colab.cim3.net/file/work/geocop/20050804_Posting/RM Profile.071205.v1.3.pdf

Fertig, Scott, Eric Freeman, and David Gelernter. "'Finding and Reminding' Reconsidered." *ACM SIGCHI Bulletin* 28 (1), January 1996, 66-69.

Freeman, Eric. *The Lifestreams Software Architecture.* Dissertation, Department of Computer Science, Yale University, 1997.

Freeman, Eric, and David Gelernter."Lifestreams: A Storage Model for Personal Data." *ACM SIGMOD Bulletin*, March 1996.

Galloway, Patricia, Susan Cisco, and Mary Dee Harris. "Legitimating e-mail for official government business: Automatic adaptive classification." NSF Proposal, 2001.

Hiltzik, Michael. *Dealers of Lightning: Xerox PARC and the Dawn of the Computer Age* (New York: HarperBusiness, 1999).

Jones, Derek. "The 7±2 Urban Legend." MISRA C Conference, 2002. Retrieved 10/5/08 from http://www.knosof.co.uk/cbook/misart.pdf

Jones, William. "Finders, keepers? The present and future perfect in support of personal information management." *First Monday* 9 (3), March 2004. Retrieved from http://firstmonday.org/issues/issue9_3/jones/index.html on 9/30/08.

Jones, William, and Jaime Reevan (eds.). *Personal Information Management*. Seattle: University of Washington Press, 2007.

Jones, William, Ammy Jiranida Phuwanartnurak, Rajdeep Gill, and Harry Bruce. "Don't Take My Folders Away! Organizing Personal Information to Get Things Done." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1505-08. New York: ACM Press, 2005.

Kaptelinin, Victor, and Mary Czerwinski (eds.). *Beyond the Desktop Metaphor: Designing Integrated Digital Work Environments*. Cambridge: MIT Press, 2007.

Kauffman, Stuart. *At Home in the Universe: The Search for the Laws of Self-Organization and Complexity*. New York: Oxford University Press, 1995.

Lynch, Patrick J. "Visual Design for the User Interface Part 1: Design Fundamentals," *Journal of Biocommunications* 21(1), 1994, 22-30.

Malone, Thomas W. "How Do People Organize Their Desks? Implications for the Design of Office Information Systems," *ACM Transactions on Office Information Systems* 1(1), January 1983, 99-112.

Malone, Thomas W. *The Future of Work*. Cambridge, MA: Harvard Business School Press, 2004.

McDermott, Richard. "Why Information Technology Inspired but cannot Deliver Knowledge Management." *California Management Review* 41(4), summer 1999, 103-17.

McLeod, Julie, and Catherine Hare (eds.). *Managing Electronic Records*. London: Facet, 2005.

Miller, David. *The Comfort of Things*. Cambridge: Polity Press, 2008.

Miller, George A. "The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information, *The Psychological Review* 63(1956), 81-97.

Nardi, Bonnie, and Deborah Barreau. "'Finding and Reminding' Revisited: Appropriate Metaphors for File Organization at the Desktop." *ACM SIGCHI Bulletin* 29 (1), January 1997.

Nardi, Bonnie, and Vickie O'Day, *Information Ecologies: Using Technology with Heart.* Cambridge: MIT Press, 1998.

National Archives & Records Administration. "Strategic Directions: Flexible Scheduling." Accessed 2/7/2008 from http://www.archives.gov.records-mgmt/initiatives/flexible-scheduling.html

National Archives & Records Administration. *NARA Bulletin 2008-04*. Accessed 7/9/08 from http://www.archives.gov.records-mgmt/bulletins/2008/2008-04.html

Newitz, Annalee. "My Laptop." In Sherry Turkle (ed.), *Evocative Objects: Things We Think With*, 87-91. Cambridge: MIT Press, 2007.

Ravasio, Pamela, Sissel Guttormsen Schär, and Helmut Krueger. "In Pursuit of Desktop Evolution: User Problems and Practices with Modern Desktop Systems." *ACM Transactions on Computer-Human Interaction*, 11 (2), June 2004, 156-80.

Richmond, Heather, and Denise Bruno. "The Truth about Taxonomies," *Information Management Journal*, March 1, 2003.

School of Information Sciences, University of Pittsburgh. "Functional Requirements for Evidence in Recordkeeping: the Pittsburgh Project." retrieved 10/5/08 from http://www.archimuse.com/papers/nhprc/

Sellen, Abigail J., and Richard H.R. Harper. *The Myth of the Paperless Office*. Cambridge: MIT Press, 2002.

Smith, David Canfield, Charles Irby, Ralph Kimball, Bill Verplank, and Eric Harslem. "Designing the Star User Interface," *Byte Magazine* 7(4), April 1982, 242-82.

Smith, Douglas K., and Robert C. Alexander. *Fumbling the Future: How Xerox Invented, then Ignored the First personal Computer*. New York: Quill, 1988.

Suchman, Lucy A. "Office Procedure as Practical Action: Models of Work and System Design." *ACM Transactions on Office Information Systems* 1(4), October 1983, 320-28.

Swan, Laurel, Alex S. Taylor, and Richard Harper. "Making Place for Clutter and Other Ideas of Home," *ACM Transactions on Human-Computer Interaction* 15(2), Article 9, July 2008.

Turkle, Sherry (ed.). *Evocative Objects: Things We Think With*. Cambridge: MIT Press, 2007.

Whittaker, S., and J. Hirschberg. "The character, value, and management of personal paper archives." *ACM Transactions on Computer Human Interaction* 8(2), 2001, 150-70.

Wikipedia. "Web 2.0." Retrieved 11/7/2008 from http://en.wikipedia.org/wiki/Web_2.0

Wikipedia. "WinFS." Retrieved 11/7/2008 from http://en.wikipedia.org/wiki/WinFS

Yakel, Elizabeth. "The Way Things Work: Procedures, Processes, and Institutional Records." *American Archivist* 59, Fall 1996, 454-64.

Yates, JoAnne. *Control through Communication: The Rise of System in American Management* (Baltimore: Johns Hopkins University Press, 1989).

Yates, JoAnne, and Wanda Orlikowski. "Genres of Organizational Communication: A Structurational Approach to Studying Communication and Media." *Academy of Management Review* 17(1992), 299-326.

Funds for this study were provided by the

The ARMA International Educational Foundation is the non-profit, (501(c)3, affiliate of ARMA International, the primary professional association for the records and information profession in the world.

Mission
The ARMA International Educational Foundation supports education and research initiatives that promote the advancement of both information managers and the information management profession. Recorded information is the lifeblood of the modern organization, but rarely is it treated as a critical asset, primarily because there is little quality research to create the comprehensive body of knowledge required to support information management as a profession. The AIEF purpose is to answer that need by soliciting funds for this research and then providing a vehicle through which conclusions can be tested, documented and communicated to the information management community.

If you found value in this publication, please consider making a financial contribution to the Endowment Fund of the Foundation. This can be accomplished by visiting the Foundation's web site, www.armaedfoundation.org, or by contacting

Foundation Administrator
ARMA Int'l Educational Foundation
1609 Terrie Drive
Pittsburgh PA 15241
USA

Additional information about the Foundation can be found at

The National Database of Non-profit Organizations
http://www.guidestar.org/pqShowGsReport.do?gotoNext=/reports/partners/guidestar/showDpLink.jsp&ein=31-1556655
Comments about this publication and suggestions for further research are welcome.
Please direct your inquiry to the Foundation Administrator.