The future of PIM: pragmatics and potential

Alan Dix

Published online: 25 Jun 2024.

Submit your article to this journal ⬈

Article views: 224

View related articles ⬈

View Crossmark data ⬈

Taylor & Francis
Taylor & Francis Group

The future of PIM: pragmatics and potential

Alan Dix[a,b]

[a]Cardiff Metropolitan University, Cardiff, Wales; [b]Computational Foundry, Swansea University, Swansea, UK

ABSTRACT

While there are challenges in transferring personal information management (PIM) research into products, PIM research does, over time, filter through into commercial systems, and practical systems often exhibit innovation that can guide future PIM research. This paper uses two approaches to understand some potential future directions for PIM research and practice. First, it deconstructs the term "personal information management" to understand how PIM techniques and tools intersect with other academic areas. Second, it examines a small selection of popular PIM tools to see how they shed light on actual adoption of PIM systems. It uses lessons from these in a more open discussion of potential future developments.

1. Introduction

From its origins in the 1980s (M. W. Lansdale, 1988), personal information management (PIM) research has grown and is defined variously as "*finding, keeping, organising, and maintaining information*" (Jones, 2010, p.5), "*the process by which individuals curate their personal data in order to reaccess that data later*" (Bergman & Whittaker, 2016, p. 1), and "*the practice and the study of the activities people perform to acquire, organize, maintain, retrieve, use, and control the distribution of information items . . . for everyday use to complete tasks and to fulfil a person's various roles*" (Jones & Teevan, 2007, p. 3). There is a clear desire to make people's digital lives easier and better, yet, despite this vision and passion within the PIM research community, it sometimes feels as though this promise is never fully achieved.

Although most of the prototypes and research systems do not survive beyond the funded project or PhD period during which they were designed, previous study has shown that underlying ideas and concepts nevertheless do trickle through (Kljun et al., 2015). Of course, this can take considerable time and the exact provenance may be hard to disentangle. For example, automated mail threading in Gmail is a feature seen as early as 1994 in the MONA mail system (Cockburn, 1994; Cockburn & Thimbleby, 1993), as well as Lotus Notes, but it is unclear whether it arose independently in Gmail or was based on this early research.

In addition, there is considerable innovation within the tech start-up sector, with a proliferation of what are essentially PIM applications although these often use terms such as "personal knowledge management" or "personal productivity" tools rather than PIM.

This paper draws on these interacting sources of innovation and uses three approaches to understand some potential future directions for PIM research and practice. First, we will deconstruct the field name "personal information management" and use that as a thought vehicle to understand how PIM techniques and tools intersect with other areas. Second, we will look at a small selection of popular PIM tools, albeit known by other phrases, to see how they shed light on what works or does not in actual adoption of PIM systems. Finally, we will use lessons from the first two approaches

combined with themes and issues from the past PIM literature to seed a more open discussion of potential future developments. This discussion is roughly divided over two sections, the first focused more on the technology and features of PIM systems, and the second more about the human context within which PIM operates, including the organizational setting and issues of personal motivation and emotion. Of course, the technology of PIM and human use of it are intimately interwoven.

This paper is a personal viewpoint; it is neither a complete review of past and present PIM research nor a systematic analysis of all current commercial applications. Rather, it attempts to offer a selection of both that seem to offer significant insights and challenges,

2. Deconstructing PIM?

As with any term, PIM has its own meaning beyond the individual words "personal information management" that gave rise to the acronym. However, reflecting on terms such as this can give us an analytic scalpel, both to understand the core discipline better and to see how it intersects and interacts with other areas.

2.1. *Personal*

The central focus of PIM is on the individual, the information resources they create and encounter day-to-day and their own personal curation of these. However, many of these resources, such as contacts, e-mails, files, photos, are partially shared with others, including families, friends, or work-colleagues. Information is rarely entirely personal; instead, there is a gradation from the personal to the public, with movement and sometimes uncertainty along that continuum. This brings benefits as we learn from and share with others, but also may create conflicts.

One of the early examples of potential conflicts between individual and organizational goals in the CSCW (computer-supported collaborative work) literature is CRM (customer relationship management) software. It is still the case today that CRM does not adequately address well-documented issues such as the way that bonuses for sales staff require them to maintain a level of personal ownership of clients, which is compromised if they share contacts with coworkers. One of Grudin's reasons for "why CSCW applications fail" (Grudin, 1988) was cost – benefit mismatch – in the case of CRM, the benefit of better client experience through contact sharing creates a cost for those with individual sales targets.

Even when individual employees fully embrace corporate goals, there is considerable resistance, especially among senior employees, to share contacts for fear that a junior colleague, or someone from another part of the organization, who does not understand the background or sensitivities of the situation might contact a high-level client and potentially sour a relationship that has taken years to build. Despite attempts in some CRM systems to control levels of access at individual or group level, dissatisfaction with this often reduces their usage. Instead, workers adopt myriad personal solutions, none of which connect with one another, nor with the corporate infrastructure. So, in one way, this example of organization – individual mismatch could be seen as an argument for PIM rather than collaborative systems; but it is really a *failure of both* if the PIM systems cannot achieve the organizational goals and vice versa.

Similar examples can arise in informal group settings, such as among family and friends, neighborhoods, or clubs; we can think of these as *informal collaborative systems*. In our own work, we have noted that community heritage, family history, and individual identity are richly intertwined (Dix et al., 2022). However, the problem here is often that corporate systems are too heavyweight or hierarchical for such communities because their information needs sit between individual and organizational solutions. Within informal collaborative settings, access is often quite informal, with a small group of active curators and larger groups of largely passive users. Some moderation may be needed, but this may be more like social media, and indeed there is heavy use of platforms such as Facebook as well as the use of off-the-shelf software such as OneDrive folders.

This is an opportunity for PIM research to address these informal collaborative systems in several senses. First, we may be able to use knowledge of successful PIM to create design advice or specific systems that address community information needs. Second, the complement of the first, some of the lessons of informal collaborative systems may be useful in PIM. Perhaps most obvious here is the way that tagging and emergent folksonomies (Marlow et al., 2006), which were successful and well used in collaborative settings, have been applied to personal systems.

A third opportunity is that successful commercial PIM systems often owe some of this success to informal sharing. Perhaps the most obvious is browser bookmarking, which for many years was regarded as an essential and potentially powerful feature, but which is, arguably, a failure given low levels of use for retrieval (Bergman, Whittaker, et al., 2021) – a write-only system. This was transformed initially by del.icio.us and then link sharing in social media, which for many users acts as not only a means to share, but also (for short periods) for individuals to re-access found content. Note that in some ways, time-based access, naturally used in social-media feeds, is often touted as an alternative PIM access tool (Fertig et al., 1996; Freeman & Fertig, 1995; Gonçalves & Jorge, 2006; M. Lansdale & Edmonds, 1992; M. W. Lansdale et al., 1989).

Even where the information use and management are personal, the *source* of information is often public or shared. For this reason, various PIM-based tools, for example bibliography management, can be seen as a way to gather or extract such non-personal information for more personal purposes. That is, PIM can sometimes be about the *personal* management of *shared/common* information. The author's own introduction to these issues came in part in the development of the commercial onCue system during the dot-com era (Dix et al., 2000), which used data-detector technology related to that in CyberDesk (Wood et al., 1997) and found now as a background feature in many systems. However, a particular moment of revelation was as a result of Jason Marshall's dissertation study of bookmark use, where he found that many users wanted to bookmark sections of pages, not just the URL itself (Dix & Marshall, 2003). This insight led directly to the experimental Snip!t system, which allowed small "snips" of web pages to be captured and combined this with web-based data-detectors (Dix et al., 2006) (see Figure 1). Similar ideas are found later in both research such as the information scraps (Bernstein et al., 2008) and commercial systems such as Evernote and Readwise.

In summary, it may be helpful to consider the "P" in PIM as a continuum ranging from personal to public, maintain the focus on an effective individual experience, but recognizing how this is always embedded in a wider human and digital ecosystem.

2.2. *Information*

If you consider commercial PIM-like software and mobile apps, they more commonly adopt the term personal *knowledge* rather than *information*. One could argue that knowledge involves human interpretation and understanding and hence it is never more than the information (or data) that is stored in a digital system. On the other hand, it is only professional librarians, archivists, or hobbyist collectors who care about the information for its own sake, and even then they archive for others to use. Certainly, the users and providers of the software are focused more on the knowledge generated by the underlying information.

However, even knowledge seems to fall short of the real purpose of PIM. Occasionally we simply want to learn new things, but more often the information or knowledge is there to achieve an end. That is, unlike play, information management is rarely autotelic, its purpose lies outside itself – if PIM is designed *for* PIM it fails.

This is perhaps most evident in the parts of PIM that are more task, action, or event focused such as calendars, reminders, and to-do lists. It is clear from both formal studies (Bellotti et al., 2004) and observation of one's own colleagues that this task management activity is often spread over many kinds of materials, both physical and digital, from explicit to-do-lists to flagged e-mails, and even the use of a chair as an in-tray (Ramduny-Ellis et al., 2005); we will return to these issues of fragmentation later. Note too that while this would be commonly regarded as part of PIM as a field of study,
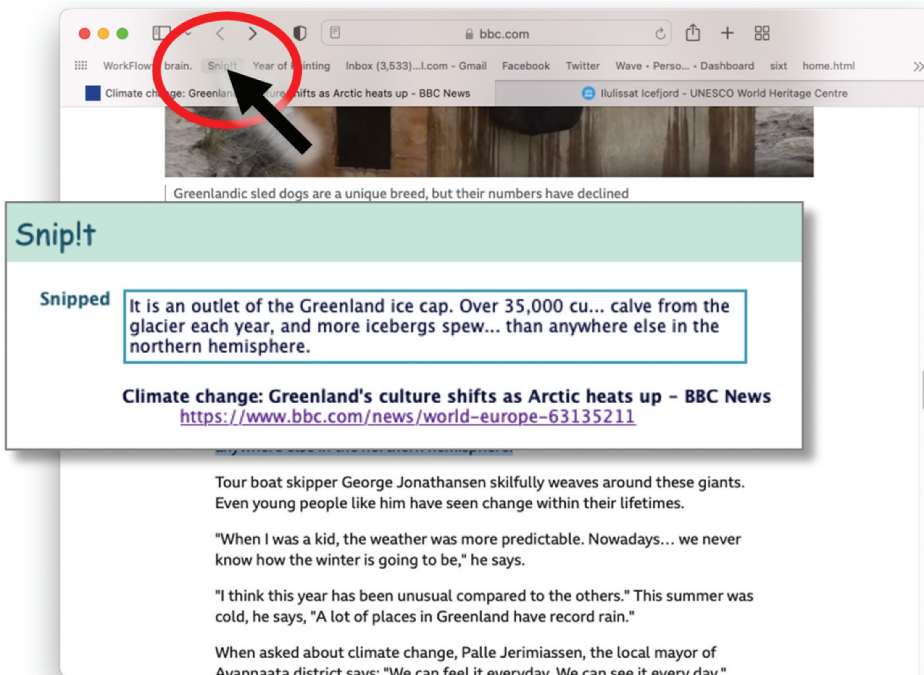
Figure 1. Snip!t in action capturing a portion of a web page (Dix et al., 2006).

the items of "information," such as an individual reminder, may be transient or may be "retrieved" only automatically to generate a notification. Of course, the actions prompted by these notifications themselves may require personal and public information, so that, ideally notifications should be timely given their purpose and the availability of the user, but also provide easy ways to access the information and other resources needed to accomplish this task (Dix & Leavesley, 2015), sadly few notification-based systems seem to achieve this.

In summary, the "I" in PIM is not just about storing data. It is important to consider the *purpose* of PIM whether to gain knowledge or insight, or to be reminded about actions and events.

2.3. *Management*

One of the first things that those new to PIM encounter in the literature is Whittaker and Sidner's no-filers, spring-cleaners, and frequent-filers (Whittaker & Sidner, 1996). One of the first things that they encounter in empirical studies is the ubiquity of users' apologies – almost before one enters their office or they open their laptop screen, users will apologize for the messy state of their physical or digital desktop and their poor filing system (or lack of it) (Katifori et al., 2008; Kaye et al., 2006). The no-filer/spring-cleaner/frequent-filer categorization has been refined to include those with a "file-on-creation strategy" (Boardman & Sasse, 2004) (or instant filers) versus those who are still frequent filers, but do so as an organizational action every few days (Katifori et al., 2008). In general, we have a picture of those who are competent and organized versus those who are a mess (even though the competent still do not see themselves as such).

In contrast, Kaye et al. (2006) found that when they asked users to locate documents, despite widely varying organizational strategies, " … *we noticed no significant difference in retrieval times between subjects who kept their bookshelves in alphabetical order and those who*

clustered their books by subject." Note that this study was of academics and when paper archives were still common allowing widespread use of "positional memory." Most of us are also aware of colleagues (or ourselves) with desks and floors covered in apparently random piles of papers, yet who can locate any desired item almost instantly. Studies of "dumping" behavior in the digital world also seemed to show that apparent "no filers" did so in a controlled manner that enables fluid work (Kamaruddin & Dix, 2010). That is, the difference between messy and organized may not make so much difference in terms of actual retrieval. Furthermore, the effectiveness of strategies depends on both personal contact and also the kind of information, notably for e-mail the effectiveness of filing for promoting retrieval is doubtful (Whittaker et al., 2011).

This is, of course, assuming that retrieval is the goal of PIM, or as William Jones (2010) phrased it, "*Keeping found things found.*" However, in a seminal paper during the early 1990s when the first hype around "knowledge work" was at its peak, Alison Kidd (1994) turned this picture on its head. Kidd's ethnographic study of knowledge workers found that while all kept an archive of some sort, it was seldom accessed. There were documents that were in use, usually in a briefcase or on the desk, but once they had passed out of active use into the archive (filing cabinet or desk drawers), they were not needed again. Indeed, Kidd wrote, "*filing information for personal re-use may actually be a redundant and resource-wasting task for knowledge workers.*" It is not that the information in the documents was unimportant, and not that it did not affect future work and actions, but the importance was in the way it transformed the knowledge worker's own knowledge and understanding, or as Kidd put it in the paper title "*the marks are on the knowledge worker.*"

There is a degree of hyperbole in this title, we all know that even if 99.9% of the items archived are never accessed again, the odd one in a thousand can be critically important, especially in professions such as finance or law. Jones' book, published fifteen years after Kidd's paper, was still needed! However, Kidd's work does still fundamentally transform the way one regards knowledge management, whether corporate or personal. In particular, it demands that we ask *why* we are managing information, or, as we asked when discussing information, what is PIM *for*? Crucially, the goal of personal information *management* is not management in itself. Indeed, as Kaye et al. (2006) note, even of those with extensive managed archives, "*while the organizational structure made retrieval possible, finding things was clearly not the main priority of this archive.*"

Many contemporary personal management techniques, such as PARA (project, area, resource, archive) (Forte, 2023) are task or project focused. Indeed, while the latter "RA" in the PARA acronym refers to information management, the first two are about shorter-term projects/tasks (the literature is very inconsistent on the use of these terms) and longer-term "areas," which are roughly roles, or what would be termed "activities" in activity theory (Kaptelinin & Nardi, 2012). Academic studies have also found the centrality of the project as an organizing principle (Copic Pucihar et al., 2016). In this light, it is interesting to look back at Henderson and Card's (1986) "Rooms" system as perhaps the first task-focused PIM work. Rooms grouped open application windows by task, so that all the resources for a task were gathered together. The original system was driven partly by the limited nature of computer-screen real estate, but also the difficulty of human context switching, a theme that has run throughout the PIM and management literature (Coviello et al., 2014; Mark et al., 2008; Rouncefield et al., 1994). PARA suggests deliberately using projects rather than topic/subject as a file organizing principle, and indeed this is often evident in users' file hierarchies. The TIC (Task Information Collections) Firefox plugin, released as part of Kljun's (2013) PhD work, went a step further allowing different kinds of resources related to a project to be gathered together in what could be seen as a web-era version of Rooms.

In summary, the "M" in PIM is not a goal itself, but like "information," it is there to serve a purpose. Seen in this light, many (but not all) apparently chaotic PIM practices are effective for their purpose, but despite this, near-universally people are dissatisfied with their level of

organization. It appears that many successful PIM practices are purpose/task oriented rather than based on intrinsic properties of information.

2.4. *PIM as a whole*

By deconstructing PIM, we are not dismantling it, but rather seeing it as a nexus of the larger ecology within which the individual finds themselves: social relationships, organizational structures, activities, and goals. Successful PIM takes these into account and thus truly *personal*, because people's lives are complex, flowing, and intertwined.

3. PIM in practice

As noted in the introduction, there is an interchange of ideas between PIM research and commercial systems. This movement of ideas is partly through the movement of people between academia and industry, for example Richard Boardman, after a PhD looking at information fragmentation (Boardman, 2004; Boardman & Sasse, 2004) moved to Google and later to other Silicon Valley companies. It is partly also because the R&D labs of larger companies are actively involved in the PIM community and aware of trends in the research literature, for example, Jones and Teevan's (2007) edited collection on Personal Information Management included contributors from across academia and industry.

Elements of PIM are found throughout the major tech industry applications, for example, mail systems, calendar, contact, and to-do list features of major office suites. Often these are based on features that were first found in smaller dedicated PIM tools such as Borland Sidekick on early PCs. There is also a proliferation of personal knowledge management and personal productivity applications, ranging from those based on overall organizing methods such as GTD (getting things done) (Allen, 2015) or PARA (Forte, 2023) and those more focused on specific aspects of PIM such as note taking or personal library management.

We will briefly look at three of the more general purpose personal knowledge management tools that are currently popular: *Workflowy*, *Notion*, and *Obsidian*. In addition, we will look at *Readwise*, a personal library app, due to its aggregation features, as this will emerge later as a key issue. Table 1 gives a summary of these, together with the popular note-taking app *EverNote*, perhaps the only PIM app that compares in impact with integrated office applications, which have a captive market. This is not a systematic selection, as noted there are many personal knowledge management and personal productivity applications. The applications selected: (a) illustrate a range of features and issues to drive and illustrate more general discussion; (b) are popular enough to be sure they have engendered real user response and are commercially successful; (c) are ones where the author had access to an enthusiastic and dedicated user to obtain an element of insider knowledge. The enthusiasm in (c) is critical as it is clear that PIM tools elicit a level of dedication among their users, even the small number of PIM early adopters who regularly swap products (E. Rogers, 2003; E. M. Rogers et al., 2014).

Table 1. Summary of selected commercial PIM products.

	Style	First website[1]	Popularity[2]
Workflowy	outliner/hierarchy	November 5, 2010	100K+
Notion	personal wiki and data	July 13, 2015	10M+
Obsidian	personal wiki	April 10, 2020	1M+
Readwise	reading highlights and annotations	September 27, 2017	100K+
EverNote	notes and capture	October 24, 2004	100M+

1. First website date is from Internet Archive (2023).
2. Popularity is based on app downloads from Google Play Store. This may not reflect overall popularity including web or iOS. It is also only a measure of initial interest, not ongoing regular use.

Consistent with the above PIM deconstruction, note that all of the three general purpose PIM tools have strong collaborative features as well as supporting individual work. All also have strong user communities (discussed in more detail later in the paper), and it is clear that all are used for a wide variety of PIM tasks, although each arose from a different initial emphasis. We will discuss these three in order of age, followed by a description of Readwise and other commercial tools.

Workflowy is the earliest of these tools and was first designed by the current company CEO to satisfy his own needs for a flexible project management tool. Its project management roots are evident in the product name, but it is in essence an outliner tool, reminiscent of early Borland Sidekick. Despite this apparent simplicity a small range of features including tagging, internal linking, sharing, and focus views have enabled it to be used in ways far beyond simple outlining or workflows.

Figure 2 (left) shows a fairly plain Workflowy screen. The hierarchical outline mode is clear. It is possible to focus on a particular branch of the hierarchy, rather like opening a folder. Each item can have rich text description, including web links, and also subitems. The interactions are deliberately designed to be easy to type a simple list using enter, tab, and shift-tab. Hashtags can be added, which are for many purposes simply plain text, but are clickable and then simply invoke the general search feature to give all items referring to the tag (Figure 2, right). Note that the search results reflect the hierarchy structure to give context. This seems a minimal set of features, but enthusiasts share many "hacks" using the features to create rich processes such as calendar management. Sub-lists can be shared to allow collaborative use.

Notion is a form of personal or organizational wiki and its "about us" page explicitly traces its inspiration to pioneers of user interfaces and hypertext:

> They dreamt a future where computers could amplify imagination (Alan Kay), augment intellect (Doug Engelbart), and expand our thoughts far beyond text on paper (Ted Nelson). (Notion, 2023)

The roots of Notion are as a team wiki and organizational information system but it is also used extensively by individuals. It makes heavy use of underlying database technology as well as support for unstructured or semi-structured text such as wikis. In unstructured mode, it can be simply rich text pages linked wiki-style to one another (Figure 3, left). However, pages can also contain semi-structured information, and a collection of pages can be treated like a table in
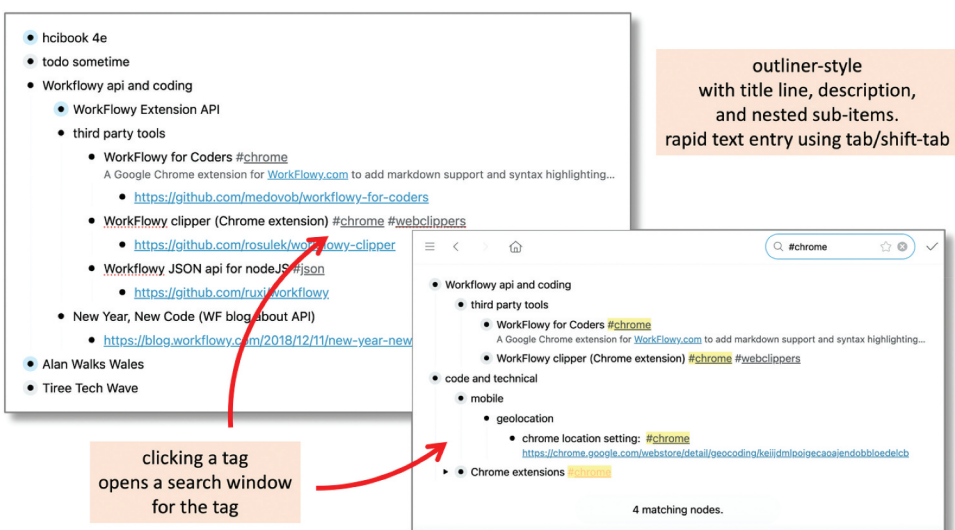

Figure 2. Workflowy – showing basic outline view and the result of clicking a hash tag.
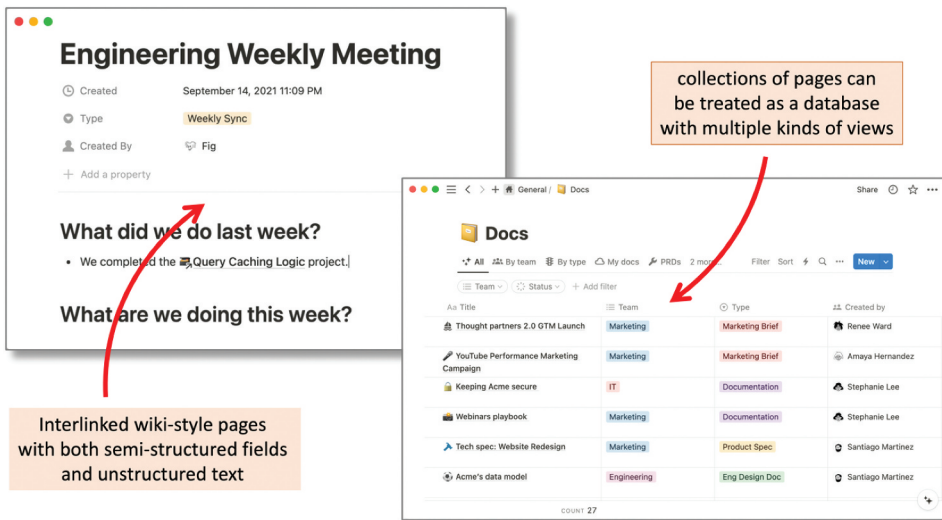
Figure 3. Notion – showing a plain page and a database view.

a database (Figure 3, right), with a variety of views including as a calendar. A rich set of datatypes and views means that it can be used for specialized purposes, such as a project management, in a way that seems almost bespoke – for example, the Notion website suggests one usage path is as an alternative to Jira. It also has an extensive API and corporate support. From personal observation, and as evident from this description, as a PIM tool Notion seems to appeal to more technically savvy users.

Obsidian is the most recent of these three. Although it is a commercial system, Obsidian is (in a manner reminiscent of the design of the Xerox Star personal computer (Canfield Smith et al., 1982)) based upon a number of guiding principles (Obsidian, 2023a): *yours*, *durable*, *private*, *malleable*, and *independent*. The central focus of the design is around a creativity mission, as evident under the Obsidian website description of the "yours" principle:

> We believe that everyone should have the tools to think clearly and organize ideas effectively. That's why our tools are free for personal use. (Obsidian, 2023a, emphasis added)

Obsidian is at heart. a personal wiki, with pages formatted using a markdown editor (Grubber, 2012), and easy interlinking (Figure 4, left). In the spirit of Unix applications, Obsidian uses the file system to manage individual pages rather than a bespoke database. Effectively, Obsidian pages are markdown files (Figure 4, right) and "vaults" (Obsidian's name for a store) are folders.

Following the creativity mission, Obsidian offers "out of the box" backlinks in an ideas panel (Figure 5) that shows all pages that link to the current focus page both explicitly, or by using the page name in plain text (that is, even if not an explicit link). As is evident, there is strong use among both formal and informal researchers to collate information. However, the flexibility of the (apparently) simple features and availability of large numbers of community plugins have enabled wide forms of use beyond this core group.

Readwise began as an add-on to Kindle, to help users keep track of their reading, highlights, and notes. Its user interface is deliberately spartan focusing on the texts highlighted (see Figure 6). Over the years, while the Kindle link is still most important, Readwise has also become an *integrator* connecting to many different reader and clipping applications, including a spreadsheet upload for non-digital reading. As well as helping to store and organize reading-related information, it also has an e-mail service, to remind you proactively about past activity. This is interesting to contrast with the common "write only" problem of browser bookmarks.
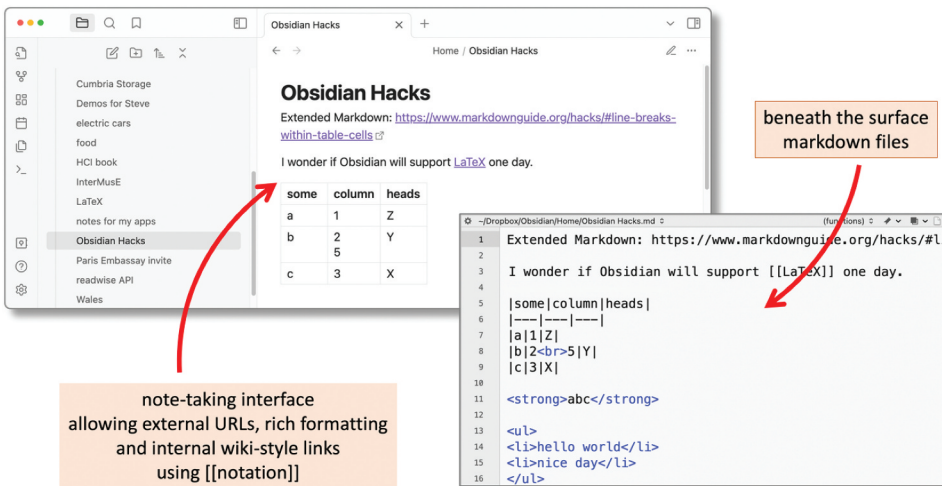
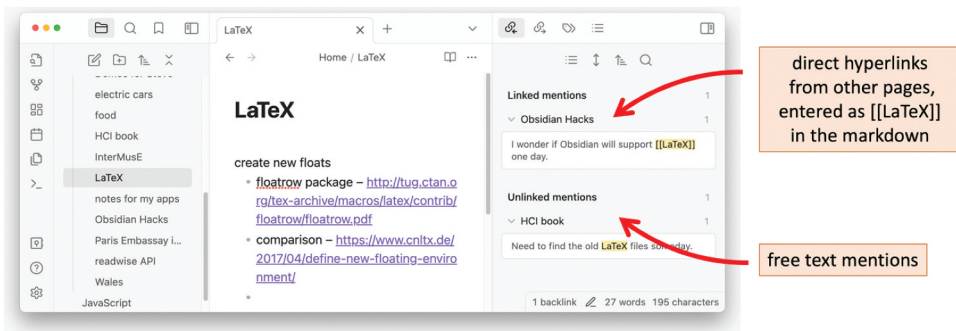**Figure 4.** Obsidian – basic mode as markdown wiki. Links.

Figure 5. Obsidian – back links in sidebar.

The large office applications, such as Microsoft Teams, offer ways to store and share information, but typically in a fashion similar to shared filesystems. These are very successful in commercial terms, but (anecdotally) few find them usable, let alone exciting. In design terms as institutional purchases, they are "baked bean" products that have to be good enough for all, rather than best for a particular person (Dix, 2010). Probably the most successful information-oriented application outside of these is *Evernote*, which has been active for nearly twenty years and has a substantial user base. It is focused on making notes both written and through media capture, with simple mechanisms to collate and tag. It also includes organization-focused task management, but in contrast to the generic PIM systems described above, this is, like the sub-applications in office suites, cross-connected but effectively a separate application.

Reasons were given earlier for focusing on *Workflowy*, *Notion*, *Obsidian*, and *Readwise* rather than the most heavily used applications. In addition, and critically for this paper, as well as being commercially successful, they are also each, in differing ways, academically interesting, including aspects of user interaction style, architectural design, and community features. As such, they are a good inspiration for those seeking to see PIM research deployed in practice.

We will use these to illustrate emerging themes in the following sections, but note a few points now. First, these systems tend to have relatively simple functionality, providing the means
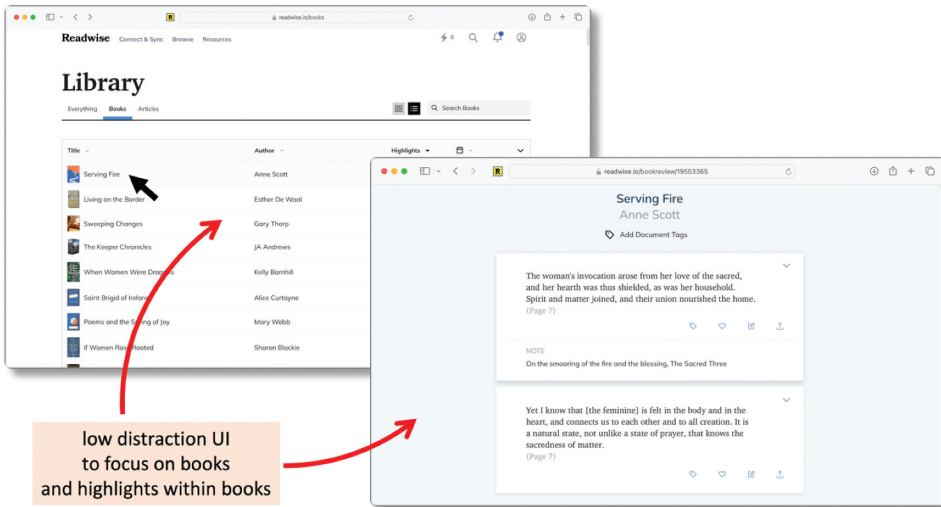
Figure 6. Readwise – showing clean no-frills content-focused interface.

for users to perform quite complex tasks through creative use of those simple features. Second, while they are very different, as noted earlier, each elicits enthusiasm and loyalty from their diverse users; unlike the generic "baked bean" office suites. Third, this individual enthusiasm is often linked to strong communities that offer "hacks," in some case of the code kind, but more critically sharing ways to use the products in innovative ways. This combination of simplicity and shared innovations allows both easy initial use for the novice and painless paths to rich expertise later.

4. Using technology

Having looked at a selection of popular PIM applications, in this and the following section, we will return to the PIM research domain connecting issues from the past and themes that have arisen in the sections above. In this section, we consider primarily the technological aspects of PIM, but of course this overlaps with the human aspects discussed in greater detail in the following section.

4.1. *Fragmentation and synchronization*

One of the constant issues in the PIM literature is fragmentation – the way that personal information is distributed over different applications. Early concerns related to the differing folder structures in files, e-mail, and bookmarks, with methods to try to homogenize the two, or at least increase consistency (Boardman & Sasse, 2004). Of course, different applications have different concerns, for example, it may make sense to have a folder connected with a person in e-mail, but not bookmarks, and this interacts with strict hierarchies that are demanded in many folder structures, meaning that, say, versions of a conference paper might be held under the publications high-level folder of a file system, whereas the correspondence about it might be in a project-related folder in e-mail. These early concerns arose well before the rise of social tagging systems (Golder & Huberman, 2006) and folksonomies (Thomas, 2007), which arguably should have made the latter problems easier, but instead led to still more fragmentation of structure between yet more applications.

Since those early days, fragmentation has increased in four senses:

- *media* – As noted, the early focus was on files, e-mail, and links, but now the list includes social media posts, instant messages, images, video and audio, numbers, and complex data.
- *applications* – The same kind of media is often distributed over multiple applications. For example, images and audio-visual media may be found in multiple photo management applications, social media posts, and sharing services such as YouTube or Flickr. Similarly, communication with a single individual or group may be split among e-mail, and numerous direct messaging applications.
- *devices* – We each typically have at least a smart phone and laptop computer, as well as various institutional and personal cloud services, including file-system-based (e.g. Dropbox, OneDrive), media-based (e.g. Google Photos, Flickr), and social-media-based (e.g. Facebook, X/Twitter).
- *people* – As noted earlier, personal information is rarely totally individual. Much of our data is intrinsically connected with others, in some cases shared with or from them, in others collaboratively edited (e.g. Google Docs, Overleaf).

The latter two add to potential synchronization issues, which have also been a long-term focus of academic study including the early CODA network filing system (Satyanarayanan et al., 1990) and more sophisticated data-type dependent approaches (Munson & Dewan, 1994). This has now become standard technology in cloud-based services, and many mobile applications, including database support such as Couchbase, but not before major hiccups in early iOS. Similarly, synchronous editing building on operational transformation methods pioneered in the GROVE shared editor (Ellis & Gibbs, 1989) is now heavily used in editors such as Google Docs. That said, there are still outstanding issues, particularly after periods of disconnection and for more complex data types – notably where there is any form of interlinking or constraints. For example, the author, who knows the literature in the area well, still encountered synchronization problems when coding for user re-ordering of to-do-list items; that is an update that involved the relationship between items and associated constraints rather than individual items in isolation.

When the author first wrote about these new fragmentation issues in an i-USEr keynote in 2011 (Dix, 2011), it looked as though the new (at the time) Windows 8 "Photo Feedr" demonstrated in beta releases might start to address these issues of application and device diversity, at least for images, by showing photos across multiple sources. However, this and other similar media-integrations are usually vendor specific or at least constrained to small range of alternatives. The author's vision at the time, was that this kind of device and application agnostic viewing of different kinds of media would become part of standard personal information environments, possibly built into operating systems. Sadly, this vision was a pipe dream and fragmentation is still a live issue.

From a commercial point of view, vendors often want to hold on to customers within their platform, and so may actively discourage sharing with other vendors' products. They may offer cross-device, and even cross-application information collation from within their own products, but with limited linkage externally, except to import. From an academic research point of view, integration applications require a lot of background work before addressing the core research issues. From personal experience, the author's attempts to integrate across bibliographic applications (including Zotero, BibSonomy, Mendeley, and LinkedIn profile publications) was dominated with managing and keeping up-to-date with many individual APIs, thus crowding out more core (and publishable) issues such as heterogeneous data integration.

That said, there are commercial integrators including IFTTT (If This Then That) for event – action-based systems; embed.ly and the open oEmbed standard for including embedded objects in any web page from across the web; and Readwise for collating reading material and annotations. However, these tend to operate in fairly niche areas, for reasons we shall explore more in the next section.

There seems to be great scope for PIM research around collation and integration applications, but this almost certainly needs coordinated efforts to create and share foundation services such as file

and data change management on individual devices/applications and API management perhaps building on Hybridauth (2023). Such a platform would enable both novel research and also pre-commercial innovation. On the whole, this is not in the immediate interests of large commercial providers, who prefer deep internal integration within their own platforms, so the impetus for open infrastructure development needs to come from the PIM research community.

4.2. *Not an island – connecting and annotating*

As well as items of PIM information being scattered over different applications and locations, these items themselves are usually regarded as isolated. Folders and tags can connect items within and, occasionally, between collections, e-mails may be gathered by thread or sender, and social media posts may be linked using "@" references and hash tags. However, often the items (messages, contacts, events) are treated as singular units.

In working with community heritage groups, we have noted how community archivists use local or cloud file stores with rich folder structures to organize their archives of images and documents. In addition, as soon as they look at an item, they start to tell you about how the item you are looking at connects with other items in the collection. Yet these connections and the stories that thread through the archive are not themselves stored (Dix et al., 2022). To a large degree, the file systems we have today have barely changed since the 1970s, with minimal support for annotation or interlinking except insofar as the application for a particular file type manages this.

Annotation of individual items is best supported, especially within a single application, such as commenting in Word or PDF documents. There are also various research and production systems to allow the recording of highlights and notes on books and articles, including those embedded within readers such as Kindle or Talis Elevate (2023) and those that operate independently such as Readwise and the author's own Booknotes (Dix et al., 2017). There are also moves to make more general external annotation of resources easier especially for web resources. This has included early hyper-text annotation systems, notably InterMedia (Garrett et al., 1986); web annotation systems, such as MADCOW (Bottoni et al., 2006); the W3C Web Annotations model (W3C, 2017); and emerging OpenAnnotationIO (2023) framework. However, these more generic attempts at annotation are still far from ubiquitous.

More complex still is interlinking. This does happen within a single application or application suite, for example Obsidian's wiki-style linking between individual notes, but rarely between applications, with the exception of URL-based linkage. The most substantial attempt to deal with this issue, was the Semantic Desktop initiative (Sauermann et al., 2005), which sought to apply the web-based RDF standard (W3C, 2004/2014) and other Semantic Web technologies to local resources such as contact lists, calendars, etc. Sadly, the full promise of the Semantic Desktop was not realized, in part as it was slightly before the growth of web-based office applications which would have made more compelling use cases. However, as we shall see, even this web use has major barriers.

One barrier to cross application linking is the lack of inter-application identifiers, comparable to URLs on the web, which can be used to refer to items within one application from another application. The Semantic Web initiative had to create these external to the applications. Amazingly, even web-based PIM applications often have no URL for individual items, which makes it hard to refer to them externally.

As an example, consider Gmail, the mail system of surely one of the most world's most web-literate technology companies. Messages initially appear to have URLs of the form "https://mail.google.com/mail/u/0/#inbox/{id}." However, this URL varies on how you view it, for example, if it has been tagged under "data," when viewed under data it would be "https://mail.google.com/mail/u/0/#label/data/{id}." This is not too difficult to correct, as "#label/data" can be rewritten to "#inbox" to generate a canonical URL. More problematic is that "/u/0/" is read as "the first user to log in on the browser" – so, if the user has more than one Gmail login (say A@gmail.com and B@gmail.com), the (apparent) URL depends on the order in which they logged in. If they logged in

the order A then B and then looked at an e-mail in the B inbox, the URL would be " ... /u/1/ #inbox/{id}.' If they then, say after a reboot, logged in next time in the order B then A, then if they tried to click a link to the URL it would look for the id under A@gmail.com inbox, not B. Finally, the id in the URL refers to the thread, not a specific message in the thread. Underlying it all, there is, in fact, a unique message id, but it is deeply hidden in the HTML.

The reason for this arcane structure is not clear. Most likely it has arisen by chance, for example the use of "/u/0/," "/u/1/," etc. as a quick hack when multiple simultaneous account logins were first enabled. The reason it has not been fixed is most likely a matter of market pull. As is evident from the closure of what, for many companies, look like successful product lines, Google regards products and features as minor if they have merely hundreds of thousands or a few million users. Use cases such as linking to an e-mail in notes or a to-do-list item must not be common enough to warrant the relatively small, development cost of fixing the e-mail structure. It is interesting to note that the web Outlook interface is even worse in that there are no linkable URLs at all (Microsoft, 2022). That said, Microsoft's internal tools (such as calendar events) do enable more *internal* cross-linking *within* the Microsoft office suite.

As is evident, if Gmail cannot manage linkable URLs for user items, there is a long way to go before even web-based interoperability is easy. This situation is if anything made worse by the growth in single-page applications, which often have a single URL for the entire application. Technically, it is possible to update the URL without refreshing the page, so that it is meaningful and serves as a deep link to application content, but it is easy for developers to forget this.

For the desktop, as noted, this is more difficult again. In early work on CSCW, the author suggested that digital support for deixis, the ability to refer and talk about things, was crucial for inter-person communication (Dix, 1994), this is of course also crucial for inter-application connections. A few years ago, the author and colleagues suggested the need for a GLUI or GLOI (globally accessible unique/object identifier) (Dix et al., 2008). This is an identifier, which can be used URL-like in applications, such as web-based calendars and guaranteed to connect to a given local resource (such as a mail message or contact) when de-referenced on the same machine on which it was created. When accessed by the user on other machines or via the web, it might also be able to dereference the link when the resource is synchronized through a cloud service, and might also (with suitable permissions) even allow other users to link through to publicly visible personal cloud resources (see Figure 7). However, this was simply a proposal, and neither it nor anything similar is commonly available.

Overall, this is an area ripe for rich interactions. Imagine being able to browse all your photos no matter whether on your own computer, in the cloud or shared in social media posts; or being able to see all communications from an individual as a thread, whether or not they came from multiple e-mail addresses, by instant messenger or text message.

4.3. *Numbers and data*

When one thinks about numbers and PIM it is probably areas such as quantified self (Lupton, 2016; Selke, 2016; Swan, 2013) or possibly financial management that come to mind. Other kinds of numeric personal information or more data-oriented analysis of other forms of PIM data seem understudied. Furthermore, with a few examples discussed below, while PIM applications offer ways to gather and preserve data, the means to explore, analyze, and hence gain insight from that data are rare.

Many classic PIM applications such as e-mails, calendar items, and contacts involve some form of semi-structured data. However, whilst this is often used as part of retrieval, it is rarely used more analytically, for example, common e-mail systems do not allow you to ask "*how many emails from person X have I replied to in the last six months*." Of course, this is precisely the goal of life-logging and quantified-self (QS) applications, which usually mix more qualitative data, with the ability to interrogate that data in order to uncover patterns of health and well-being. There are high-profile
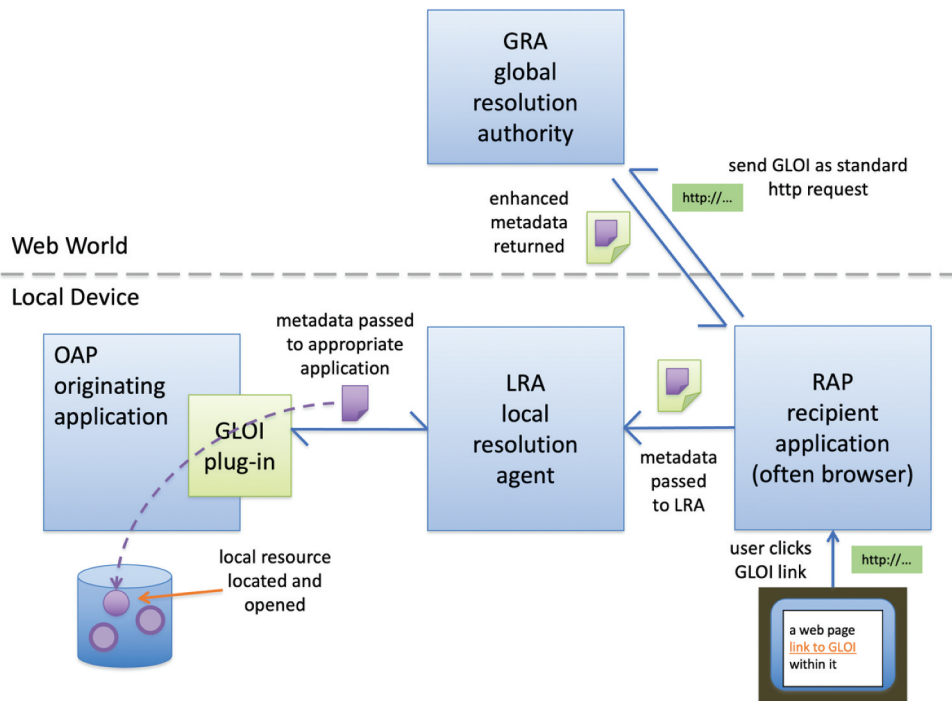
Figure 7. Potential architecture for GLOI (globally accessible unique/object identifier) (Dix et al., 2008).

examples of successful use of such analysis, especially in the health domain, for example Jessie Inchauspé, the "Glucose Goddess," who initially studied her own glucose levels (Inchauspé, 2022). There is some doubt even as to the adequacy of QS tools for more everyday insight, and evidence that in some circumstances reflection can have negative consequences as well as positive ones (Hollis et al., 2017). However, it is certainly the case that this form of analysis rarely extends as far as e-mail, files, or bookmarks. For example, what proportion of your bookmarks have you ever reused?

Numerical data are even less well established in the PIM literature, despite an early desktop calculator being one of the micro-applications in Borland Sidekick, and, of course, the ubiquity of spreadsheets. Of popular current generic knowledge management applications, Notion, being built upon database technology, does allow some numeric processing; however, it also attracts a more technology-centered user anyway. In more domain-specific areas, health, fitness, and well-being apps often have a variety of pre-built metrics often in dashboards, and of course personal financial management applications, such as YNAB (You Need a Budget) have numerous reports.

When looking at everyday numeric data, one of the barriers to innovation may be researchers themselves. On the one hand, those who are more quantitative may not understand the difficulty of those less mathematically minded, both in terms of knowledge and emotional reaction to numbers, and they therefore provide solutions that, while meaningful to themselves, are not usable or acceptable by the intended users. This in part explains that while researchers creating QS tools report novel insights, this is less evident when they are deployed to a wide audience.

On the other hand, more qualitative researchers may eschew studying quantitative phenomena entirely. Spreadsheet use is a notable exception to this from Nardi and Miller's (1990) early study, and more recently (Sarkar et al., 2020). It is worth noting that a key observation of Nardi and Miller's work was the collaborative nature of spreadsheet use, despite spreadsheets at the time being single user applications. Given the explosion in data science, there will be large numbers of

qualitative researchers studying data scientists. There is still a challenge to make the rich visualizations and analysis tools of data science and visual analytics (Keim et al., 2010) accessible to a broader audience. However, while there is a difference between studying numerical experts and more every day numeric use, it is to be hoped that some of this academic interest will find its way into numerical PIM applications.

The management of personal numeric information is critical. Much of day-to-day life involves numeric or financial decisions, but professional advice is usually expensive – this can lead to significant financial loss, especially for those whose lives are already precarious. In addition, understanding of contemporary issues from the Covid pandemic to climate change depends on forms of numeric understanding (Dix et al., 2023). These typically do not need to be precise, instead more "qualitative – quantitative" reasoning is required (Dix, 2021) – that is informal reasoning about numeric phenomena.

There are promising directions, especially in domain-specific areas, such as means to more easily visualize personal data for self-reflection and well-being (Jörke et al., 2023). In addition, NLP-based visualizations suggest the potential to make the creation of personalized visualizations open to far wider audiences (Tabalba et al., 2023) and the methods underlying Construct-A-Vis (Bishop et al., 2019) offer insights into how small children can actively engage in analyzing and visualizing data. Collecting and collating data is also important, while most web clippers are focused on text, onCue (Dix et al., 2000) included a table recognizer that could then easily copy numeric web data into spreadsheets (still not easy today!) and the iVolver system (Nacenta & Mendez, 2017) allows users to extract numeric data from published visualizations, such as pie charts, and then revisualize it in alternative ways.

As a personal attempt to address some of these issues, the author has been investigating a number of interlinked prototypes to make a smooth path from everyday arithmetic to more complex numeric applications (Dix, 2024). One is calQ, which in basic use is simply a four-function calculator with a till roll showing past calculations, just as you would have seen in old shop-counter calculators (Figure 8). Clicking on a past entry copies the value into the current sum, like a memory function, but the values are live. Old sums in the till roll can be updated and any dependent values are updated, effectively creating simple spreadsheet features.

The till roll from calQ can be exported to a full spreadsheet or to the ws2 WordPress plugin that provides spreadsheet-like features, enabling web pages where users can update values and explore
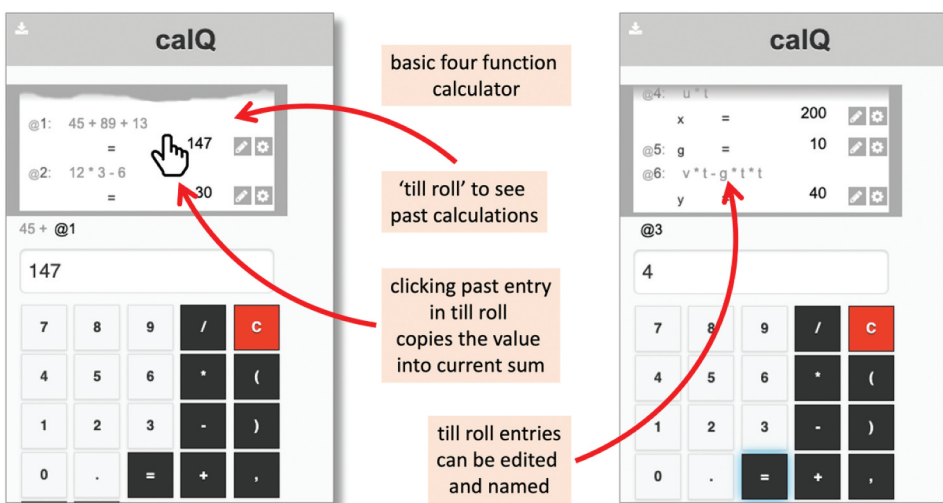

Figure 8. calQ allows easy transition from basic four-function calculator simple spreadsheet-like updateable formulae.

Figure 9. ws2 (workspace) WordPress plugin allows spreadsheet-like features including embedded tables, with values that can be updated by the reader to explore scenarios.

scenarios (see Figure 9). When connected to the number recognition in Snip!t, this creates a path for personal numeric information from capture, through creation to sharing.

4.4. UI matters

Although it should not need saying, the minutiae of user interface design is critical for successful PIM. Research-based systems are focused on novel features, so rarely have time to finely tune the other aspects of experience, but these are essential for long-term use. This is true of overall adoption, but also for the use of specific features within systems that are already being used.

As a personal example, the author continued to use the Eudora mail system for many years, well after it ceased to be supported and only executed using compatibility software on the Mac. When Ducheneaut and Bellotti (2001) published their study of e-mail use in Xerox and other companies, they reported widely different organizational strategies from a single inbox to "more than 400 folders!" (exclamation mark in the paper). At the time, the author had more than 2000 folders in Eudora ... as well as inbox of more than 6000 e-mails. Eudora eventually stopped working entirely when Apple adopted Intel's architecture, and since moving to alternative e-mail systems (Gmail, Outlook) the author has never had more than a handful of folders. This change is due in part to easier off-line use in Eudora, allowing e-mail filing whilst traveling, but mostly because of the folder management pane in Eudora. This was apparently similar to other e-mail systems (such as Apple mail at the time); but the Eudora version was easier to use and hence encouraged the creation of complex folder systems whilst not impeding rapid access.

To a small extent, better search in e-mail has filled in some of this gap, but certainly, from a personal perspective, compared with folder-filing it is far more time-consuming, less effective, and critically has greater uncertainty for both. Free text search is fast, but seems poor at ranking even for clear e-mail-related features, such as ignoring or de-emphasizing terms in e-mail footers. Structured search is both basic, and again fails at fairly fundamental e-mail issues such as multiple addresses for the same person. It should be noted that this personal view is at odds with empirical research that has found that e-mail retrieval using folder hierarchies was slower and less successful than search (Whittaker et al., 2011). This discrepancy may simply be the author's mistaken impressions: folder

seeking is more cognitively engaging and has a sense of progress so may *feel* faster; or may be due to fine differences between e-mail systems, which are certainly sufficient (in the authors' experience) to completely change retrieval behavior. Of course, if search is the preferred strategy, this makes it even more important that it is as good as possible.

The paucity of e-mail search (and indeed file search) is odd given the huge practical experience and theoretical literature in the area (Catarci et al., 2022; Ceri & Brambilla, 2011). This may be partly due to history, with some exceptions, notably Lotus Notes, e-mail, and database technology that have been distinct, and at best some elements of web search technology have found their way into e-mail. However, as with the arcane Gmail URL structure, the reasons are probably also due to a lack of market pull. For some of us, being able to retrieve past e-mails or rapidly scan past e-mail is mission critical, but for many e-mail is, either by intention or perforce, as in its earliest days, uncertain and unreliable, more like something mentioned whilst passing in a corridor.

Note while this is a single personal anecdote, it shows that the type and level of organization is not simply a matter of personal preference or skills, but also intimately tied to the detailed user interaction design choices. My filing and retrieval techniques changed, not because of a fundamental change in personality, but because of ease of interaction.

As another example, before they released Obsidian, the developers created a system Dynalist, an outliner-style alternative to Workflowy. Dynalist is in many ways more functional than Workflowy, has an official API, and has easy paths to move from Workflowy. Dynalist has developed a following, but, despite greater functionality, it has not displaced Workflowy. The reason for this seems to be that while making more complex things possible or easier, it is slightly harder to do simple tasks. For frequent tasks, especially those that users do early in their usage journey, even slight hindrance can halt adoption.

In both the personal Eudora story and the Workflow – Dynalist one, as well as the detailed user interface issues, there are also considerable personal barriers to switching between systems, even if the new system were to offer better features or user experience. Users invest considerable effort in learning how to use applications effectively, including workarounds. This is evident in major products, for example Mac and Windows users are often proud when they discover arcane key combinations that perform hidden functions, which, by any reasonable UX design standard, should have been apparent in the first place! In addition to these sunk costs, there will be cost exporting/importing data and learning new methods, all to set against the discounting of future benefit. Even the costs of upgrading within an application are felt to be too high for many users (Bergman & Whittaker, 2018) and transferring even more so. To attract users to switch, new systems need to address one or other of the desire and disaster criteria, either: avoiding *disaster* –fixing a fundamental problem in existing solutions; or eliciting *desire* –offering something that is either substantial attractive (in UX) terms or offering fundamentally new functionality (Dix, 2012).

Moving into the research domain, perhaps one of the problems with the Semantic Desktop initiative (Sauermann et al., 2005) was the lack of compelling use cases. It created a powerful infrastructure, with potential for advanced uses, but not the obvious instant win, which is so important for adoption.

4.5. *AI is the answer?*

Given the ongoing success and ensuing hype around AI, one might wonder whether this might simply obviate all future PIM research! Indeed, the potential for AI to act as an intelligent agent to help manage information has a long history within PIM, certainly back to Phone Slave (Schmandt & Arons, 1985) (which foresaw so many issues of conversational user interfaces and would be well worth reproducing in the current technology); and the 1990 envisionment video starring Tom Baker as a slightly obsequious butler-like virtual agent (Adams, 1990; Wikipedia, 2023). The inspiration for the latter was similar to that for the name of the Ask Jeeves search engine, which now seems prescient with the advent of large-language models (LLM) (Charlie Warzel, 2023).

There are problems with the direct application of LLMs and similar technology to personal information as, by definition, they require large volumes of training data, even for fine tuning of foundation models. Transfer learning is essential, that is when a model is trained on data for one application area, or group of people, but can be rapidly adapted to similar areas or people without the need for extensive new data. This may require further research to create hybrid systems that combine large-data generic systems with more structured personal data, such as contact lists, and to deal with the intrinsically subjective nature of PIM (Bergman, Beyth-Marom, et al., 2008). However, Microsoft is already rolling out a version of Copilot with Teams that can summarize missed meetings, or as the BBC phrased it "*go to meetings for you*" (Kleinman, 2023) and virtual avatars and agents are being produced for learning and wellbeing (Pataranutaporn et al., 2021, 2023). This is not so different from the early Tom Baker envisionment and, in general personal digital assistants and chatbots, have been a strong research area even before the current versions of LLMs were available (Park et al., 2022; Sarikaya, 2017). Aware of the problems of "hallucination" and confabulation, versions of LLMs are now being produced that are able to refer back to their sources, to enable explainability. It may well be that this is the turn that makes no-filing possible.

As we saw, the act of organizing is itself insightful in addition to its role in aiding future retrieval. Considering this, various forms of text analysis have also been used to automatically classify documents; perhaps most well known is the Scatter/Gather browser (Cutting et al., 1992; Pirolli et al., 1996) where automatic classification was used as part of an interactive exploration of datasets. Scatter/gather has continued to generate interest, including applications for web search (Gong et al., 2012), leading to a reprinting of the original 1992 paper in SIGIR Bulletin in 2017. Text mining techniques can struggle on small collections, but there are methods to deal with this, indeed the "large" collections of the early scatter/gather browser would probably be considered small by current standards. Back in dot-com days, the author was involved in experimental search that used the data from the Open Directory Project (or DMOZ (Wikipedia, 2024)) to create a content classifier. This classifier could then be applied to smaller collections such as image libraries with keywords, but also to cluster web search results so that, for example, in a search for "chihuahua" results associated with the place in Mexico, the dog and Taco Bell mascot would be presented in small groups with a "more like this" option (Dix et al., 2005). In some ways, this is a similar principle to foundation models (trained on large corpora, but with the potential for specialization), further emphasizing their potential to empower more personal retrieval.

Remembering that PIM is there to *do* something, there has also been a long history of task inference systems, from simple single-step data detectors to more complex programming-by example such as Eager (Cypher, 1991). Basic elements of this are already embedded into standard packages, such as the way meetings and events are recognized in e-mail messages even if there is no explicit ". ics" attachment, and the large tech companies are working on more advanced task inference (Kotler et al., 2013). There is also an active area generating tutorial videos (appropriation through sharing) based on automated task tracking (Hu et al., 2023).

5. Understanding people

We continue now to draw out threads and issues from the past research literature connecting to the deconstruction of PIM and examples of popular PIM systems. However, we shift our focus toward the human context of PIM.

5.1. *Organization: when and why*

Grudin (1988) highlighted the problems of cost–benefit mismatch between people in a collaborative application, but the same is true of cost–benefit for an individual over time. When performing organization or maintenance tasks in PIM, you are incurring cost now for the benefit of your future self, or as (Bergman and Whittaker 2016, p. 8) put it "*a solipsistic interaction between a person and*

him- or herself at two different times." In general, humans tend to discount future benefit heavily, possibly because in the distant past of early humanity we could not guarantee to be alive next year, so benefit today is most important. This does not encourage spending time on the organization of information.

Based on Kidd's "*marks are on the knowledge worker*" one could argue that organization does not matter anyway! However, as noted earlier, even though most archived items are rarely if ever accessed, the few times they are needed are often important. Of course, while some items do need to be retrieved, the majority will not; this means that the vast majority of organization time is effectively wasted. A core and continuing problem for every PIM system is how to overcome this fundamental mismatch.

Some people (the frequent filers) seem to positively enjoy the act of organizing their archives. It is often such people who produce books and techniques on time management and project organization. For the rest of us, other strategies are needed.

One solution is to obviate the need for organization using search techniques. Studies have shown that the majority of users prefer to directly access the file hierarchy, only reverting to search when the former fails (Bergman, Beyth-Marom, et al., 2008; Bergman, Israeli, et al., 2021). This is in part explained by differing verbal skill (Bergman, Israeli, et al., 2021), and independent of the quality of the search engine (Bergman et al., 2008). However, even the best desktop search has relatively poor performance due to the small volume of training data compared with web search. This even seems to be true of search in web mail services such as Gmail. Statistical big-data search algorithms have limits when dealing with smaller corpora, especially considering privacy. However, it is possible that this will change with the potential to combine general text foundation models with federated learning (Kairouz et al., 2021).

For most big vendor PIM tools, search is a tool to be invoked when needed. One of the features valued by many Obsidian users is the way in which it *proactively* shows connected notes/documents in a side window (see Figure 5), using search-technology techniques to create a latent graph. Also in
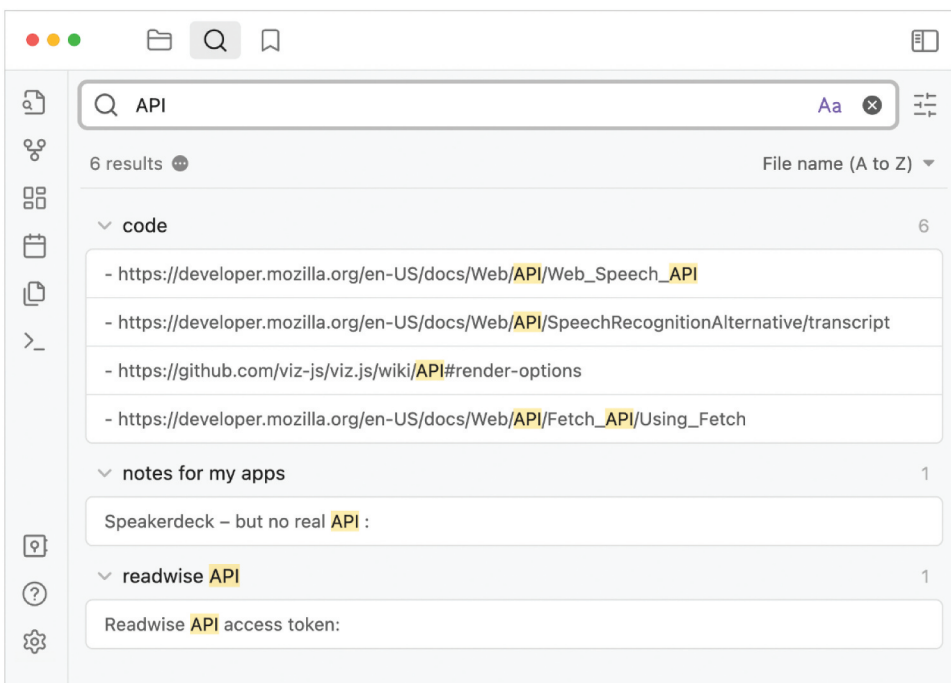

Figure 10. Obsidian search results with snippets.

many popular tools, when search is invoked explicitly, the list of results is not simply the titles of the relevant documents, but also a short snippet, in the fashion of web search (see Figure 10). This substantially increases the immediate benefit compared to typical file-system searches which only show the file name or document title.

This said, it is evident that most people with considerable stores of documents tend to impose some level of organization, even if they are dissatisfied with it. Hence, organization is still needed!

As a general design strategy, if costs and benefits can be brought closer to one another in time, we are more likely to act. This can be achieved by reducing upfront costs and enabling incremental complexity. Indeed, a recurrent feature of popular knowledge management apps, is that they initially require minimal organization, but provide ways to *incrementally* increase the structure or formality. For example, Workflowy users can start with a simple list or outline (not so different from Borland Sidekick 40 years earlier); however, one can then progress incrementally to deeply structured hierarchies, with cross-linked tags and other features. Similarly, Obsidian can initially be used as simple flat notes, but the Wiki-like links, together with various templates and add-ons, can be used to create complex information structures; again this is all in a step-by-step fashion without early over-commitment to an organizational scheme.

Alternatively, one can deliberately encourage organizational tasks at the point at which retrieval is needed. One example is tidying up the folders one encounters while searching for an item; this may include deleting old versions to make it easier to see what is relevant, so that by making the current job easier one organizes as a side effect. Another example is when you initially look for an item on one place, but then find it somewhere else; this suggests either moving the item or creating a link in the initial place. These are user practices that can be learned or taught for any system, but could be positively encouraged by the styles of interaction. Some of these human strategies, especially the second example, offer the potential for automation: if you have delved deep in your file system and then backed out before starting to work on a document elsewhere, it would be possible for the system to offer to add an alias.

Looking back again to Kidd's work, a more radical way to consider organizational work is not for retrieval, but for understanding. By sorting, organizing, grouping, and labeling we are working on the material and effectively creating fresh insights about it through interactions on it; a form of distributed cognition (Hollan et al., 2000). Mind-mapping tools are common and using these is not seen as a chore in the way that sorting files is. There are examples where this boundary between file organization and meaning making is blurred ranging from the network layout view in Notecards (Halasz et al., 1986), to Kljun's (2013) TIC and Obsidian's graph panel.

5.2. *Communities and appropriation*

A recurrent feature of popular knowledge management applications, is the way they attract a sharing community, offering tips or even writing books of "hacks" to help others to more effectively use the tools (Degenaar, 2023). Often this is about creative use of generic in-built features. For example, one Obsidian user described to the author the way they used templates to create a calendar and diary system. Some, such as Obsidian, have fully functional APIs or extension mechanisms with easy means to share these community add-ons. Other applications have less formal ways to extend functionality, for example, while Workflowy has no official API, they effectively sanction a user-written API (Satgé, 2023) and there are various browser extensions and bookmarklets.

This issue of communities intersects with end-user appropriation.

One of the guidelines for "designing for appropriation" (Dix, 2007) is "*support not control.*" That is to provide basic functionality which allows the users to create the functionality they need. This seems to be a common feature in popular systems, for example, Obsidian does not have a dedicated to-do list, but does provide easy markup for checkable boxes and indented lists. A user can thus easily create multi-level to-do lists. These enabling features are far more likely to be repurposed beyond the designer's expectations than closed features.

Another appropriation guideline is to support "*pluggability and configuration*," that is the means to create extensions via scripting, APIs or similar mechanisms, as seen in several popular applications. Indeed, some applications deliberately make use of APIs of other applications. For example, the reading/annotation collator Readwise is able to import from many applications such as Kindle and Instapaper using their APIs, and then in turn an Obsidian plugin uses the Readwise API to allow Readwise notes to be viewable from within Obsidian UI.

A third pertinent appropriation guideline is "*encourage sharing*," that is to create mechanisms both technical and social to enable users to share their own personal appropriation with each other. This is precisely what we see in the way "hacks" are shared in user communities and, for example, the Obsidian community plugin site allows more radical extensions (Obsidian, 2023b).

5.3. *People are different*

In Kaye et al.'s (2006) study of paper document organization and retrieval, they note that the previously mentioned ability of their subjects to retrieve documents despite radically different organization strategies "*suggests there is no single 'best practice' among the variety of structures we witnessed that radically influences the efficiency of information access*," or in the words of one of their subjects "*I do what works for me..*" Bergman et al. (2008) have emphasized the need for a *user-subjective approach* to PIM, taking into account the subjective nature of the kinds of information, classifications, processes, etc. used by PIM users. Our own work on personal ontologies (Dix et al., 2010) has distinguished *generic* classes that have a common meaning to everyone (such as Person or Location), *egocentric* classes (such as Friends, or Family) which have a shared meaning, but the instances of which will depend on the individual; and *idiocentric* classes that are meaningless except perhaps to close acquaintances (for example "Water Rats" for a group of sailing friends).

This variety of methods and preferences is of course evident in the various classes attributed to users including messy/neat (Malone, 1983), piler/filer (Whittaker & Hirschberg, 2001), and no filers/frequent filers/spring cleaners (Bälter, 1997; Whittaker & Sidner, 1996). At one point as part of her research on "dumping" behavior (such as leaving all new files in the desktop folder), Azrina Kamaruddin presented to a lunchtime seminar of HCI researchers. The talk described the wide variety of end-user approaches she found in the study (Kamaruddin & Dix, 2010), and yet, after the talk ended, the majority of questions were of the form "why can't they do X, it works for me" – even seasoned HCI researchers find it hard to imagine that other people think differently.

This variety is certainly very evident as one talks anecdotally to users of PIM systems or looks at the community forums. Some people find Notion is perfect for them, others use Obsidian, others plain file system hacks. The systems that offer the best appropriation often have wider usage simply because they can be used in different ways by different people. Note though this is about appropriation not necessarily customization – successful systems are useful out of the box in simple ways, but allow rich interpretation of those simple features. Where more complex customization (or deeper APIs) is provided, the successful systems typically allow these to be easily shared, so that the enthusiasts who do this are able to share their hacks with others (Dix, 2007), as was the case, for example, with the early Xerox Buttons system (MacLean et al., 1990). This variation in tool use also emphasizes the previously mentioned need for cross-tool integration, especially in organizations.

5.4. *Emotion management*

Often PIM applications are classed under "personal productivity tools" – focusing strongly on the work element of PIM. This is reminiscent of the way that the central goals of early user interface design guidelines embodied in the international standard ISO 92,401 was: "effectiveness, efficiency and satisfaction" (Bevan et al., 2015; International Organization for Standardization, 1998–2016), but the third element was often forgotten.

There are clearly many specific apps oriented toward well-being, but more generally aspects of emotion management are intimately tied to PIM practices. Studies have shown that people produce radically different folder structures when experiencing positive or negative emotions (Whittaker & Massey, 2020), and for many, the very idea of time management, to-do-lists, and tasks is stress inducing and many feel overwhelmed by mountains of documents, and messages, indeed Allen's book on the "Getting Things Done" method is subtitled "*The art of stress-free productivity*" (Allen, 2015). Email overload has been recognized as a problem since the 1990s (Whittaker & Sidner, 1996), but is certainly greater today with myriad instant messaging applications and perpetual notifications adding to the sense of "constant interruption" (Mark et al., 2008; Rouncefield et al., 1994).

As noted earlier, virtually every study of PIM practice is replete with quotes of users' dissatisfaction with their own practices, even when apparently coping. This sense of being perpetually out of control is of course a major cause of stress (Fontaine et al., 1993; Houston, 1972). PIM systems clearly not only need to work, but be *felt* to work.

In their study of busyness, Leshed and Sengers (2011), quote one of their subjects in the title of their paper "*I lie to myself that I have freedom in my own schedule.*" Here, the subject feels they need an element of (knowing) self-deception in order to feel on top of things. However, Leshed and Sengers also discuss the way some of their subjects use their PIM tools to frame personal identity, to mark the passage of the day, and to achieve a sense of accomplishment.

If we understand that emotion management is intimately tied to PIM, we can deliberately design to enhance this. When the author was preparing materials for the Interaction Design Foundation (2023) course on Creativity, he compared digital to-do lists where the item often disappears once completed, with paper-lists (and some digital ones), where the item is ticked or even crossed out in bold red marker. In the former case, the to-do list is a *not-done* list, a constant reminder of failure to achieve things intended and of pressure building. In the latter case, the to-do list is also a *done* list; a record of achievement. Some users, the author included, will often add an item and immediately tick it off, if they do something that was not on their list.

Some productivity or life-management methods deliberately try to create a sense of well-being or fun. Stephens (2023) describes her "constellation system" as "*a guide to practical magick.*" She divides tasks into more humanly meaningful areas, and then uses Latin names such as "spiritus" for "mind," "corporalis" for "body" and "divitiae" for "money"; turning financial management into something that feels magical. Some people use terms such as "quest" instead of "task" or "project," which invokes a sense of adventure instead of drudgery. There are also numerous apps that attempt to gamify time management, such as the open source Habitica (2023b; 2023b, 2023a), which includes battling monsters, magic skills, and a "Hall of Heroes," as part of day-to-day activities such as exercise and schoolwork.

6. Summary

We have covered a wide range of topics in PIM research literature and practice, and lessons from them that can guide future applications. Some are as expected, such as the importance of effective detailed user interaction design and the potential of artificial intelligence in PIM, others less commonly considered, even when highlighted early in the PIM literature.

In the deconstruction section, we saw that "personal" often includes levels of community, organizational, and public interactions, and indeed later when looking at popular applications, these are a core feature. In particular, we identified *informal collaborative systems* as a potential direction for PIM-related research. We also saw that "information management" is typically the means not the purpose of PIM, which is typically more about either gaining insight or getting tasks done. While it should not be taken as an excuse to ignore retrieval, Kidd's seminal work can help remind us that using information at hand is often more important than storing it for later.

These issues of purpose and task/project centeredness threaded through later discussions, not least the need to bring together cost and benefit during the organizational stages of PIM activity,

such as supporting sorting/tidying whilst finding. Crucially, if we understand that the benefit of organizing is often the insight it brings, this offers direction for more insight-focused information management interfaces.

Dealing with increasing fragmentation is of perhaps greater importance now than in the early days of PIM, and we saw a need both for better means to collate information of the same type from many platforms and services, and also the need to be able to annotate and interlink – a step beyond 1970s file systems. However, doing this effectively, both in research settings and also innovation start-ups would be helped enormously by foundational infrastructure to access diverse services. There is clear potential for large-scale cross-field collaboration, albeit, as noted previously, the bulk of this needing to come from the academic PIM community. Support for deixis is crucial, most likely extending URLs to be able to deal with local documents and information fragments, as suggested in GLOIs.

We also saw that because PIM tools are personal, they need to take into account both the different and common aspects of human emotion and behavior. Successful tools, either deliberately or accidently, follow many of the principles for appropriation by design, not least because, when we are dealing with diverse users, adoption is only possible through appropriation. While we all differ in how we react emotionally to our day-to-day domestic and work tasks, universally PIM is emotional work; taking this into account can both reduce the stress of PIM, and also improve its effectiveness.

As noted, this paper is not a systematic review of all PIM research and practice, and there are many gaps. In particular, notification-based systems have only been mentioned briefly in passing, despite being perhaps one of the curses of modern life. This is in part because notifications are not central features of the three popular tools discussed. The myth of multi-tasking has now been long bust, and there is widespread recognition of the problems of "constant interruption" (Mark et al., 2008; Rouncefield et al., 1994), even if we struggle to deal with it. Some years ago, when at a Dagstuhl Seminar in 2012 on "Interaction Beyond the Desktop," the author suggested that one of the future challenges for human computer interaction in a world dominated by communication was "*designing for solitude*." As we think about PIM as emotion management, this is still a goal with considering.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Notes on contributor

Alan Dix (alan@hcibook.com) is a HCI researcher with a background in mathematics, who has worked across many areas in HCI, CSCW, AI and creativity, often using eclectic methods including a one thousand mile walk around Wales; he is a Professorial Fellow at Cardiff Metropolitan University, Wales and Emeritus Professor at Swansea University, Wales.

References

Adams, D. (1990). Hyperland. *BBC Documentary*, *1990*. https://archive.org/details/DouglasAdams-Hyperland
Alan Dix. (2007). Designing for appropriation. *In Proceedings of BCS HCI 2007, People and Computers XXI* (Vol. 2). BCS eWiC. https://www.alandix.com/academic/papers/HCI2007-appropriation/
Allen, D. (2015). *Getting things done: The art of stress-free productivity*. Penguin.
Bälter, O. (1997). Strategies for organising email. *People and Computers XII: Proceedings of HCI'97*, Bristol, UK (pp. 21–38). Springer London.

Bellotti, V., Dalal, B., Good, N., Flynn, P., Bobrow, D. G., & Ducheneaut, N. (2004, April). What a to-do: Studies of task management towards the design of a personal task list manager. *In Proceedings of the SIGCHI conference on Human factors in computing systems*, Vienna, Austria (pp. 735–742).

Bergman, O., Beyth-Marom, R., & Nachmias, R. (2008). The user-subjective approach to personal information management systems design: Evidence and implementations. *Journal of the American Society for Information Science and Technology*, *59*(2), 235–246. https://doi.org/10.1002/asi.20738

Bergman, O., Beyth-Marom, R., Nachmias, R., Gradovitch, N., & Whittaker, S. (2008). Improved search engines and navigation preference in personal information management. *ACM Transactions on Information Systems (TOIS)*, *26*(4), 1–24. https://doi.org/10.1145/1402256.1402259

Bergman, O., Israeli, T., & Benn, Y. (2021). Why do some people search for their files much more than others? A preliminary study. *Aslib Journal of Information Management*, *73*(3), 406–418. https://doi.org/10.1108/AJIM-08-2020-0250

Bergman, O., & Whittaker, S. (2016). *The science of managing our digital stuff*. MIT Press.

Bergman, O., & Whittaker, S. (2018, January). The cognitive costs of upgrades. *Interacting with Computers*, *30*(1), 46–52. https://doi.org/10.1093/iwc/iwx017

Bergman, O., Whittaker, S., & Schooler, J. (2021). Out of sight and out of mind: Bookmarks are created but not used. *Journal of Librarianship and Information Science*, *53*(2), 338–348. https://doi.org/10.1177/0961000620949652

Bernstein, M., Van Kleek, M., Karger, D., & Schraefel, M. C. (2008). Information scraps: How and why information eludes our personal information management tools. *ACM Transactions on Information Systems (TOIS)*, *26*(4), 46. https://doi.org/10.1145/1402256.1402263

Bevan, N., Carter, J., & Harker, S. (2015). ISO 9241-11 revised: What have we learnt about usability since 1998? *In Human-Computer Interaction: Design and Evaluation: 17th International Conference, HCI International 2015*, August2-7, 2015, Part I17, Los Angeles, CA, USA (pp. 143–151). Springer International Publishing.

Bishop, F., Zagermann, J., Pfeil, U., Sanderson, G., Reiterer, H., & Hinrichs, U. (2019). Construct-a-vis: Exploring the free-form visualization processes of children. *IEEE Transactions on Visualization and Computer Graphics*, *26*(1), 451–460. https://doi.org/10.1109/TVCG.2019.2934804

Boardman, R. (2004). *Improving tool support for personal information management* [PhD thesis]. Imperial College London.

Boardman, R., & Sasse, M. A. (2004). "Stuff goes into the computer and doesn't come out": A cross-tool study of personal information management. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04)* (pp. 583–590). Association for Computing Machinery. https://doi.org/10.1145/985692.985766

Bottoni, P., Levialdi, S., Labella, A., Panizzi, E., Trinchese, R., & Gigli, L. (2006). MADCOW: A visual interface for annotating web pages. *In Proceedings of the working conference on Advanced visual interfaces (AVI '06)* (pp. 314–317). Association for Computing Machinery. https://doi.org/10.1145/1133265.1133331

Canfield Smith, D., Irby, C., Kimball, R., & Harslem, E. (1982). The star user interface: An overview. *In Proceedings of the June 7-10, 1982, national computer conference (AFIPS '82). Association for Computing Machinery* (pp. 515–528). https://doi.org/10.1145/1500774.1500840

Catarci, T., Dix, A., Kimani, S., & Santucci, G. (2022). *User-centered data management*. Springer Nature.

Ceri, S., & Brambilla, M. (Eds.). (2011). *Search computing: Trends and developments*. Springer.

Cockburn, A. (1994). *Groupware design: Principles, prototypes, and systems* [PhD Thesis]. University of Stirling.

Cockburn, A., & Thimbleby, H. (1993). Reducing user effort in collaboration support. *In Proceedings of the 1st international conference on Intelligent user interfaces (IUI '93). Association for Computing Machinery* (pp. 215–218). https://doi.org/10.1145/169891.169989

Copic Pucihar, K., Kljun, M., Mariani, J., & Dix, A. J. (2016). An empirical study of long-term personal project information management. *Aslib Journal of Information Management*, *68*(4), 495–522. https://doi.org/10.1108/AJIM-02-2016-0022

Coviello, D., Ichino, A., & Persico, N. (2014). Time allocation and task juggling. *American Economic Review*, *104*(2), 609–623. https://doi.org/10.1257/aer.104.2.609

Cutting, D. R., Karger, D. R., Pedersen, J. O., & Tukey, J. W. (1992). Scatter/Gather: A cluster-based approach to browsing large document collections. *Reprinted in ACM SIGIR Forum*, *51*(2), 148–159. https://doi.org/10.1145/3130348.3130362

Cypher, A. (1991). EAGER: Programming repetitive tasks by example. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '91)* (pp. 33–39). ACM. https://doi.org/10.1145/108844.108850

Degenaar, F. (2023). *Do way, way more in WorkFlowy v2.0*. WorkFlowy Academy. https://www.productivitymashup.com/do-way-way-more-in-workflowy/

Dix, A. (1994). Computer supported cooperative work: A framework. In D. Rosenberg, & C. Hutchison (Eds.), *Design issues in CSCW* (pp. 9–26). Springer London.

Dix, A. (2010). Human-computer interaction: A stable discipline, a nascent science, and the growth of the long tail. *Interacting with Computers*, *22*(1), 13–27. https://doi.org/10.1016/j.intcom.2009.11.007

Dix, A. (2011). The personal information environment beyond the personal computer. In S. Alam & S. Jaya (Eds.), *Keynote* i-USEr 2010. https://alandix.com/academic/talks/iUser-2011-keynote-personal-inf-env/

Dix, A. (2012). Desire and disaster. *Interfaces*, *91*, 22. https://hcibook.com/e3/online/desire-and-disaster/

Dix, A. (2021). Qualitative–quantitative reasoning: Thinking informally about formal things. *In Theoretical Aspects of Computing–ICTAC 2021: 18th International Colloquium, Virtual Event,*September8–10, 2021, LNCS 12819, Nur-Sultan, Kazakhstan, (pp. 18–35). Springer International Publishing.

Dix, A. (2024). *Alan Labs*. Retrieved January 3, 2024, from https://alandix.com/labs/calq/WS2%20%E2%80%93work space:https://alandix.com/labs/ws2/

Dix, A., Beale, R., & Wood, A. (2000). Architectures to make simple visualisations using simple systems. *In Proceedings of the working conference on Advanced visual interfaces (AVI '00). Association for Computing Machinery* (pp. 51–60). https://doi.org/10.1145/345513.345250

Dix, A., Bond, R., & Caraban, A. (2023). Why pandemics and climate change are hard to understand and make decision-making difficult. *Interacting with Computers*, *35*(5), 744–761. https://doi.org/10.1093/iwc/iwad043

Dix, A., Catarci, T., Habegger, B., Loannidis, Y., Kamaruddin, A., Katifori, A., Lepouras, G., Poggi, A., & Ramduny-Ellis, D. (2006). Intelligent context-sensitive interactions on desktop and the web. *In Proceedings of the international workshop in conjunction with AVI 2006 on Context in advanced interfaces (CAI '06)* (pp. 23–27). Association for Computing Machinery. https://doi.org/10.1145/1145706.1145710

Dix, A., Jones, E., Neads, C., Davies, V., Cowgill, R., Armstrong, C., Ridgwell, R., Downie, J. S., Twidale, M., Reagan, M., & Bashford, C. (2022, July 11–13). Tools and technology to support rich community heritage. *In Proceedings of British HCI Conference (BHCI2022),Keele, UK.* https://alandix.com/academic/papers/BHCI2022-community/

Dix, A., Karapetis, F., & Kefalidou, G. (2005). Concept classification for decentralised search. *Unpublished working paper.* Lancaster University. https://alandix.com/academic/papers/Concept-Classification-for-Decentralised-Search-2005/

Dix, A., Katifori, A., & Lepouras, G. (2008, July 04). Globally accessible local URIs - Discussion ideas. Retrieved October 24, 2023, from https://alandix.com/academic/projects/TIM/docs/gloi/

Dix, A., & Leavesley, J. (2015). Learning analytics for the academic: An action perspective. *Journal of Universal Computer Science (JUCS)*, *21*(1), 48–65.

Dix, A., Lepouras, G., Katifori, A., Vassilakis, C., Catarci, T., Poggi, A., Ioannidis, Y., Mora, M., Daradimos, I., Akim, N. M., Humayoun, S. R., & Terella, F. (2010). From the web of data to a world of action. *Journal of Web Semantics*, *8*(4), 394–408. https://doi.org/10.1016/j.websem.2010.04.007

Dix, A., & Marshall, J. (2003). At the right time: When to sort web history and bookmarks. *In Volume 1 of Proceedings of HCI International 2003*, Crete, Greece (pp. 758–762).

Dix, A., Tamblyn, R., & Leavesley, J. (2017). From intertextuality to transphysicality: The changing nature of the book, reader and writer (unpublished). *In Extended version of presentation at: Future of Books and Reading in HCI. Workshop at NordiCHI 2016 (Oct 2016, Gothenburg Sweden).* https://alandix.com/academic/papers/future-books-nordichi-2016/

Ducheneaut, N., & Bellotti, V. (2001). E-mail as habitat: An exploration of embedded personal information management. *Interactions*, *8*(5), 30–38. https://doi.org/10.1145/382899.383305

Ellis, C. A., & Gibbs, S. J. (1989). Concurrency control in groupware systems. *In Proceedings of the 1989 ACM SIGMOD international conference on Management of data (SIGMOD '89). Association for Computing Machinery* (pp. 399–407). https://doi.org/10.1145/67544.66963

Fertig, S., Freeman, S., & Gelernter, D. (1996). Lifestreams: An alternative to the desktop metaphor. *In Proceedings of CHI 1996*, Vancouver, BC, Canada (pp. 410–411).

Fontaine, K. R., Manstead, A. S. R., & Wagner, H. (1993). Optimism, perceived control over stress, and coping. *European Journal of Personality*, *7*(4), 267–281. https://doi.org/10.1002/per.2410070407

Forte, T. (2023). *The PARA method: Simplify, organise and master your digital life*. Profile Books.

Freeman, E., & Fertig, S. (1995, November). Lifestreams: Organizing your electronic life. *In AAAI Fall Symposium: AI Applications in Knowledge Navigation and Retrieval*, Cambridge, MA, USA (pp. 38–44).

Garrett, N. L., Meyrowitz, N., & Smith, K. E. (1986). Intermedia: Issues, strategies, and tactics in the design of a hypermedia system. *In Proceedings of the Conference on Computer-Supported Cooperative Work*, Austin Texas. ACM.

Golder, S. A., & Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, *32*(2), 198–208. https://doi.org/10.1177/0165551506062337

Gonçalves, D., & Jorge, J. A. (2006). Quill: A narrative-based interface for personal document retrieval. CHI'06 Extended Abstracts on Human Factors in Computing Systems, Montréal, Québec, Canada (pp. 327–332).

Gong, X., Ke, W., & Khare, R. (2012). Studying scatter/gather browsing for web search. *Proceedings of the American Society for Information Science and Technology*, *49*(1), 1–4. https://doi.org/10.1002/meet.14504901328

Gruber, J. (2012). *Markdown. Daring fireball website*. Retrieved October 30, 2023, from http://daringfireball.net/projects/markdown/

Grudin, J. (1988). Why CSCW applications fail: Problems in the design and evaluation of organizational interfaces. *In Proceedings of the 1988 ACM conference on Computer-supported cooperative work*, Portland, Oregon, USA (pp. 85–93).

Habitica. (2023a). *GitHub*. Retrieved October 24, 2023, from https://github.com/HabitRPG/habitica

Habitica. (2023b). *Habitica – gamify your life*. Retrieved October 24, 2023, from https://habitica.com/

Halasz, F. G., Moran, T. P., & Trigg, R. H. (1986). NoteCards in a nutshell. *ACM SIGCHI Bulletin*, *17*(SI), 45–52. https://doi.org/10.1145/30851.30859

Henderson, D. A., Jr., & Card, S. (1986). Rooms: The use of multiple virtual workspaces to reduce space contention in a window-based graphical user interface. *ACM Transactions on Graphics (TOG)*, *5*(3), 211–243. https://doi.org/10.1145/24054.24056

Hollan, J. D., Hutchins, E. L., & Kirsh, D. (2000). Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on CHI*, *7*(2), 174–196. https://doi.org/10.1145/353485.353487

Hollis, V., Konrad, A., Springer, A., Antoun, M., Antoun, C., Martin, R., & Whittaker, S. (2017). What does all this data mean for my future mood? Actionable analytics and targeted reflection for emotional well-being. *Human-Computer Interaction*, *32*(5–6), 208–267. https://doi.org/10.1080/07370024.2016.1277724

Houston, B. K. (1972). Control over stress, locus of control, and response to stress. *Journal of Personality and Social Psychology*, *21*(2), 249–255. https://doi.org/10.1037/h0032328

Hu, X., Huang, Y., Liu, B., Wu, R., Hu, Y., Quigley, A. J., Fan, M., Yu, C., & Shi, Y. (2023). SmartRecorder: An IMU-based video tutorial creation by demonstration system for smartphone interaction tasks. *In Proceedings of the 28th International Conference on Intelligent User Interfaces (IUI '23)* (pp. 278–293). Association for Computing Machinery. https://doi.org/10.1145/3581641.3584069

Hybridauth. (2023). *Hybridauth, open source social login PHP library*. Retrieved October 27, 2023, from https://hybridauth.github.io/index.html

Inchauspé, J. (2022). *Glucose Revolution: The life-changing power of balancing your blood sugar*. Simon and Schuster.

Interaction Design Foundation. (2023). *Creativity: Methods to design better products and services*. Retrieved October 24, 2023, from https://www.interaction-design.org/courses/creativity-methods-to-design-better-products-and-services

International Organization for Standardization. (1998–2016). *Ergonomics of human-system interaction (ISO standard No. 9241-11)*. Retrieved October 24, 2023, from https://en.wikipedia.org/wiki/ISO_9241

Internet Archive. (2023). *Wayback machine*. Retrieved October 30, 2023, from https://archive.org/web/

Jones, W. (2010). *Keeping found things found: The study and practice of personal information management*. Morgan Kaufmann.

Jones, W., & Teevan, J. (2007). *Personal information maNAGement*. University of Washington Press.

Jörke, M., Sefidgar, Y. S., Massachi, T., Suh, J., & Ramos, G. (2023). Pearl: A technology probe for machine-assisted reflection on personal data. *In Proceedings of the 28th International Conference on Intelligent User Interfaces (IUI '23). Association for Computing Machinery* (pp. 902–918). https://doi.org/10.1145/3581641.3584054

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Nitin Bhagoji, A., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R. G. L., Eichner, H., El Rouayheb, S., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B. … Yu, H. (2021). Advances and open problems in federated learning. *Foundations & Trends® in Machine Learning*, *14*(1–2), 1–210. https://doi.org/10.1561/2200000083

Kamaruddin, A., & Dix, A. (2010). To 'dump'or not to 'dump': Changing, supporting or distracting behavior. *In 2010 International Conference on User Science and Engineering (i-USEr)*, Shah Aalm, Malaysia (pp. 49–54). IEEE.

Kaptelinin, V., & Nardi, B. (2012). *Activity theory in HCI: Fundamentals and reflections*. Morgan & Claypool Publishers.

Katifori, A., Lepouras, G., Dix, A., & Kamaruddin, A. (2008). Evaluating the significance of the desktop area in everyday computer use. *In First International Conference on Advances in Computer-Human Interaction* (pp. 31–38). https://doi.org/10.1109/ACHI.2008.27

Kaye, J. J., Vertesi, J., Avery, S., Dafoe, A., David, S., Onaga, L., Rosero, I., & Pinch, T. (2006). To have and to hold: Exploring the personal archive. *In Proceedings of the SIGCHI conference on Human Factors in computing systems*, Montréal, Québec, Canada (pp. 275–284).

Keim, D., Kohlhammer, J., Ellis, G., & Mansmann, F. (2010). *Mastering the information age solving problems with visual analytics*. Eurographics Association. https://www.visual-analytics.eu/book/

Kidd, A. (1994, April). The marks are on the knowledge worker. *In Proceedings of the SIGCHI conference on Human factors in computing systems*, Boston, MA, USA (pp. 186–191).

Kleinman, Z. (2023). *Microsoft's new AI assistant can go to meetings for you*. BBC news 18 March2023. https://www.bbc.co.uk/news/technology-67103536

Kljun, M. (2013). *The information fragmentation problem through dimensions of software, time and personal projects* [Ph.D. dissertation]. Lancaster University.

Kljun, M., Mariani, J., & Dix, A. (2015, March). Transference of PIM research prototype concepts to the mainstream: Successes or failures. *Interacting with Computers*, *27*(2), 73–98. https://doi.org/10.1093/iwc/iwt059

Kotler, M. J., Friend, N. B., Kikin-Gil, E., Park, C. W., Satterfield, J. C., & Zaika, I. (2013). Context-based task generation. *U.S. Patent No. 8,375,320*. U.S. Patent and Trademark Office.

Lansdale, M. W. (1988). The psychology of personal information management. *Applied Ergonomics*, *19*(1), 55–66. https://doi.org/10.1016/0003-6870(88)90199-8

Lansdale, M., & Edmonds, E. (1992). Using memory for events in the design of personal filing systems. *International Journal of Man-Machine Studies*, *36*(1), 97–126. https://doi.org/10.1016/0020-7373(92)90054-O

Lansdale, M. W., Young, D. R., & Bass, C. A. (1989). MEMOIRS: A personal multimedia information system. In G. Salvendy & M. Smith (Eds.), *Designing and using human-computer interfaces and knowledge based systems* (pp. 845–852). Elsevier Amsterdam.

Leshed, G., & Sengers, P. (2011, May). "I lie to myself that I have freedom in my own schedule" productivity tools and experiences of busyness. *In* Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada (pp. 905–914).

Lupton, D. (2016). *The quantified self*. John Wiley & Sons.

MacLean, A., Carter, K., Lövstrand, L., & Moran, T. (1990). User-tailorable systems: Pressing the issues with buttons. *In* Proceedings of the SIGCHI conference on Human factors in computing systems, Seattle, USA (pp. 175–182).

Malone, T. W. (1983). How do people organize their desks? Implications for the design of office information systems. *ACM Transactions on Information Systems (TOIS)*, *1*(1), 99–112. https://doi.org/10.1145/357423.357430

Mark, G., Gudith, D., & Klocke, U. (2008). The cost of interrupted work: More speed and stress. *In Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, Florence, Italy (pp. 107–110).

Marlow, C., Naaman, M., Boyd, D., & Davis, M. (2006). HT06, tagging paper, taxonomy, Flickr, academic article, to read. *In Proceedings of the seventeenth conference on Hypertext and hypermedia (HYPERTEXT '06)*. Association for Computing Machinery (pp. 31–40). https://doi.org/10.1145/1149941.1149949

Microsoft. (2022). *How to create shareable outlook email link with others*. Microsoft Community. Retrieved March 16, 2022, from https://answers.microsoft.com/en-us/outlook_com/forum/all/how-to-create-shareable-outlook-email-link-with/7b53a20a-acea-470e-9cb9-40a261e83909

Munson, J. P., & Dewan, P. (1994). A flexible object merging framework. *In Proceedings of the 1994 ACM conference on Computer supported cooperative work (CSCW '94)*. (pp. 231–242). Association for Computing Machinery. https://doi.org/10.1145/192844.193016

Nacenta, M. A., & Mendez, G. G. (2017). iVolver: A visual language for constructing visualizations from in-the-wild data. *In Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*, New York, United States (pp. 438–441).

Nardi, B. A., & Miller, J. R. (1990). An ethnographic study of distributed problem solving in spreadsheet development. *InProceedings of the 1990 ACM conference on Computer-supported cooperative work (CSCW '90)* (pp. 197–208). Association for Computing Machinery. https://doi.org/10.1145/99332.99355

Notion. (2023). About notion: A story of tools and the future of work. Retrieved October 30, 2023, from https://www.notion.so/about

Obsidian. (2023a). *About obsidian: Manifesto*. Retrieved October 30, 2023, from https://obsidian.md/about

Obsidian. (2023b). *Community plugins*. Retrieved October 23, 2023, from https://help.obsidian.md/Extending+Obsidian/Community+plugins

OpenAnnotationIO. (2023). Retrieved October 27, 2023, from https://openannotation.io

Park, D.-M., Jeong, S.-S., & Seo, Y.-S. (2022). Systematic review on chatbot techniques and applications. *Journal of Information Processing Systems*, *18*(1), 26–47. https://doi.org/10.3745/JIPS.04.0232

Pataranutaporn, P., Danry, V., Blanchard, L., Thakral, L., Ohsugi, N., Maes, P., & Sra, M. (2023). Living memories: AI-Generated characters as digital mementos. *In Proceedings of the 28th International Conference on Intelligent User Interfaces (IUI '23)*. (pp. 889–901). Association for Computing Machinery. https://doi.org/10.1145/3581641.3584065

Pataranutaporn, P., Danry, V., Leong, J., Punpongsanon, P., Novy, D., Maes, P., & Sra, M. (2021). AI-generated characters for supporting personalized learning and well-being. *Nature Machine Intelligence*, *3*(12), 1013–1022. https://doi.org/10.1038/s42256-021-00417-9

Pirolli, P., Schank, P., Hearst, M., & Diehl, C. (1996). Scatter/gather browsing communicates the topic structure of a very large text collection. *In Proceedings of the SIGCHI conference on Human factors in computing systems*, Montréal Québec Canada (pp. 213–220).

Ramduny-Ellis, D., Dix, A., Rayson, P., Onditi, V., Sommerville, I., & Ransom, J. (2005). Artefacts as designed, artefacts as used: Resources for uncovering activity dynamics. *Cognition, Technology & Work*, *7*(2), 76–87. https://doi.org/10.1007/s10111-005-0179-1

Rogers, E. (2003). *Diffusion of Innovations* (5th Edition. Simon and Schuster. Free Press of Glencoe, 1962.

Rogers, E. M., Singhal, A., & Quinlan, M. M. (2014). Diffusion of innovations. In D. Stacks, & M. Salwen (Eds.), *An integrated approach to communication theory and research* (pp. 432–448). Routledge.

Rouncefield, M., Hughes, J. A., Rodden, T., & Viller, S. (1994). Working with "constant interruption": CSCW and the small office. *In Proceedings of the 1994 ACM conference on Computer supported cooperative work (CSCW '94)* (pp. 275–286). Association for Computing Machinery. https://doi.org/10.1145/192844.193028

Sarikaya, R. (2017). The technology behind personal digital assistants: An overview of the system architecture and key components. *IEEE Signal Processing Magazine*, *34*(1), 67–81. https://doi.org/10.1109/MSP.2016.2617341

Sarkar, A., Borghouts, J., Iyer, A., Khullar, S., Canton, C., Hermans, F., Gordon, A., & Williams, J. 2020. Spreadsheet use and programming experience: An exploratory survey. *In Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–9). https://doi.org/10.1145/3334480.3382807

Satgé, J. (2023). *WorkflowyPHP. First release 2015.* Retrieved October 23, 2023, from https://github.com/johansatge/workflowy-php

Satyanarayanan, M., Kistler, J. J., Kumar, P., Okasaki, M. E., Siegel, E. H., & Steere, D. C. (1990). Coda: A highly available file system for a distributed workstation environment. *IEEE Transactions on Computers*, *39*(4), 447–459. https://doi.org/10.1109/12.54838

Sauermann, L., Bernardi, A., & Dengel, A. (2005). Overview and outlook on the semantic desktop. *In Proceedings of the 1st Workshop on The Semantic Desktop at the ISWC 2005 Conference*, Galway, Ireland (Vol. 175. pp. 1–19).

Schmandt, C., & Arons, B. (1985). Phone Slave: A graphical telecommunications interface. *Proceedings of the Society for Information Display*, *26*(1), 79–82.

Selke, S. (Ed.) (2016). *Lifelogging: Digital self-tracking and lifelogging-between disruptive technology and cultural transformation.* Springer.

Stephens, R. (2023). *A quickstart guide to the constellation system.* Retrieved January 24, 2023, from https://www.rachaelstephen.com/constellation

Swan, M. (2013). The quantified self: Fundamental disruption in big data science and biological discovery. *Big Data*, *1*(2), 85–99. https://doi.org/10.1089/big.2012.0002

Tabalba, R. S., Kirshenbaum, N., Leigh, J., Bhattacharya, A., Grosso, V., DiEugenio, B., Johnson, A. E., & Zellner, M. (2023). An investigation into an always listening interface to support data exploration. *In* Proceedings of the 28th International Conference on Intelligent User Interfaces (IUI '23) (pp. 128–141). Association for Computing Machinery. https://doi.org/10.1145/3581641.3584079

Talis. (2023). *Talis elevate.* Retrieved October 27, 2023, from https://talis.com/talis-elevate/

Thomas, V. W. (2007). *Folksonomy coinage and definition.* Retrieved October 27, 2023, from https://www.vanderwal.net/folksonomy.html

Warzel, C. (2023). The vindication of ask jeeves. *The Atlantic.* Retrieved March 3, 2023, from https://www.theatlantic.com/technology/archive/2023/03/ask-jeeves-chatgpt-bing-ai-chatbot-google-search/673275/

W3C. (2004/2014). *Resource Description Framework (RDF). RDF working group. Dated 2014-02-25, first version 2004-02-10.* Retrieved October 27, 2023, from https://www.w3.org/RDF/

W3C. (2017, February 23). *Three recommendations to enable annotations on the web. Web annotation working group.* Retrieved October 27, 2023, from https://www.w3.org/news/2017/three-recommendations-to-enable-annotations-on-the-web/

Whittaker, S., & Hirschberg, J. (2001). The character, value, and management of personal paper archives. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *8*(2), 150–170. https://doi.org/10.1145/376929.376932

Whittaker, S., & Massey, C. (2020). Mood and personal information management: How we feel influences how we organize our information. *Personal and Ubiquitous Computing*, *24*(5), 695–707. https://doi.org/10.1007/s00779-020-01412-4

Whittaker, S., Matthews, T., Cerruti, J., Badenes, H., & Tang, J. (2011). Am I wasting my time organizing email? A study of email refinding. *In* Proceedings of the SIGCHI conference on human factors in computing systems, Vancouver BC Canada (pp. 3449–3458).

Whittaker, S., & Sidner, C. (1996). Email overload: Exploring personal information management of email. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '96). Association for Computing Machinery* (pp. 276–283. https://doi.org/10.1145/238386.238530

Wikipedia. (2023). *Hyperland. Entry dated 4/9/2023.* Retrieved February 8, 2024, from https://en.wikipedia.org/wiki/Hyperland

Wikipedia. (2024). *DMOZ. Entry dated 4/2/2024.* Retrieved February 8, 2024, from https://en.wikipedia.org/wiki/DMOZ

Wood, A., Dey, A., & Abowd, G. D. (1997). CyberDesk: Automated integration of desktop and network services. *In Proceedings of the ACM SIGCHI Conference on Human factors in computing systems (CHI '97)* (pp. 552–553). Association for Computing Machinery. https://doi.org/10.1145/258549.259031