Notes for position paper on PIM:
Personal information objects & Burden of multiple personal spaces

David Kirsh
Professor and Chair
Department of Cognitive Science
UCSD
San Diego, USA

I am currently interested in two areas of PIM broadly construed:

1. The contextual nature of many personal information objects

2. The cognitive burden associated with maintaining multiple personal information spaces – this is a corollary of the fragmentation problem

The comments that follow are not meant to be more than indications of areas in need of deeper analytic and empirical study.

Personal information objects

Although it is easiest to define information objects in terms of the representational entities that encode them, the notion of 'content I am looking for' is broader and deeper.

When I consider the 'information items' I often look for – the ones I wish I could find but which usually come in non-encapsulated form – they most often center on *ideas or thoughts.* To help me recover these thoughts later I may annotate a passage, make notes, mark things down.   The idea to be recovered is not *explicitly* encoded in the representations. But given my beliefs about my future mental state that I had at the time of keeping or creating the information, I believe the idea will be recoverable when I am in the context later. The information I most want to recover requires deeper processing.   Thus although finding an article may be the first step in recovering earlier state, the thing I really want to recapture is the idea I had when I read the article. To facilitate this process I may make notes, underline, connect passages, annotate, make sketches, leave intentional piles and so forth. These curious artifacts, encode information that is highly contextual. I think we all would regard the representational objects (notes, annotations, etc) to be personal information objects, but what about piles, and other structures whose meaning is encoded in the relation they bear to other things?   Are these legitimate types of information objects?

Here are some specific examples of curious information items I often lose: comments on someone's paper that may or may not have extensions elsewhere in the paper (such as the back of a page). A specific phrase that someone wrote in his or her paper, which I did not underline, but which I would recognize if I see it although I cannot recall it. That is, something for which I have recognition but not recall memory.   A reference cited in someone's bibliography I marked

as potentially useful for a topic I was planning to think about. Quick notes I took at a conference that relate to something I was currently working on.   Rough drafts of paragraphs that I have never used or that were for a paper I never completed. The 'unifying category' – such as 'home office expenses' – that I used to sort some of the tax papers I keep stuffed in a large box, but which I never labeled.

I think we all have lots of valuable information like this that we have trouble keeping track of. One interesting thing about them is that there is no reason to think that going more digital will really solve the problem of tracking them. The reason creating digital versions cannot be the answer, per se, is that the meaning of this material resides partly in its physical context – where it is relative to other bits of information and structure – and partly in the state of mind we had at the time when storing it. Given Tulving's well-known principle of Encoding Specificity – the more a retrieval context resembles the encoding context the more successful efforts at recall are – it is important that we get our retrieval context as close to the actual context we were in when jotting down our notes. Presumably this would apply as much to being in the same digital context for a digital item as it would be for a paper item in a paper context. But crossing contexts – moving from paper to digital – will pose a problem, since we will lose most of the context associated with look, feel and interactivity. Even more fundamentally, though, is the problem that very little is known about how much structure is needed in an encoding context to make recreating that context a powerful memory aid. Little is known about whether a context with lots of structure is better at cueing recall than a context that has little. Is it easier to remember the price of a can of peas if the marker for the can is an image of the can that has lots of brand and detail than if it is a bare icon?

These sorts of questions are relevant to PIM research because there are many occasions when the information we want to retrieve is not identical with the text we wrote or the documents we kept. The information we really want is the set of ideas or thoughts we were trying to express at the time – or the thoughts we were hoping could be recovered later – by using that text. It is not a satisfactory answer to this real problem of personal information to dissolve it by denying such semantic entities or mental states as being part of our personal information space. It is not satisfactory because the personal information I want to retrieve is the 'document' or 'information item' *plus* its encoding context. Admittedly, recovering that extra information may require knowing the task the encoder was involved in at the time of encoding. But this is not an unreasonable requirement on a PIM. Indeed in many, if not all cases, knowing the task at encoding time may be sufficient to define context. To define the context of encoding, therefore, may not require knowing what was in the encoder's mind at the time – a private event we cannot require a PIM system to ever reveal – because all the material we want is in the 'public domain' but may require appropriate representation to be effective.
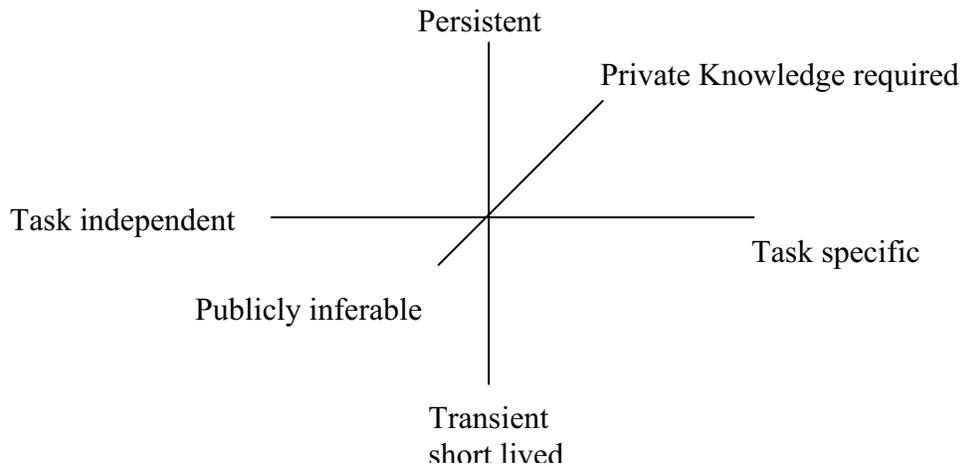
How shall we think about these sorts of information items? Typically, the more our work involves deep processing and reflection the more of these sorts of items we create.

One idea that might help to clarify their nature is to say that they are standard information objects sorted under *ad hoc* categories. These are categories that are created by individual users, typically on-the-fly, and connected to a task. For example, if I come home from shopping and unpack the groceries, laying the ingredients I intend to use for the main course on one side of the

counter, and the ones I intend for befores on the other side, I am using an organizing principle that is ad hoc. In this case, I may use the same principle of spatial separation repeatedly. So it is misleading to call it an on-the-fly category. But it is closely tied with my tasks and the way I work.

I see the project of understanding how we contextually classify information items as directly relevant to PIM research. The more we know about ad hoc categorization the easier it should be to design interactive personal retrieval systems

As a start, it may be helpful to introduce several dimensions along which ad hoc organizing principles can lie.   As an initial pass here are three dimensions.

```
                        Persistent
                               Private Knowledge required


  Task independent                              Task specific


              Publicly inferable


                        Transient
                        short lived
```

Burden of multiple personal spaces

When people talk about the fragmentation problem they are typically referring to the diverse media and physical locations where information objects may live: laptops, pda's papers, post-its, email, voice mail, contact lists in cell phones, email apps, etc. From this point of view a single filing cabinet in my office counts as a single medium in the fragmented collection, my PDA another. Is fragmentation best understood in terms of the devices where information is stored?

Personally, I find a major PIM burden stems from the mental effort associated with using different organizational systems.   Naturally it is desirable to have all my information in one physical place or device. But the notion of *place* is much deeper than the physical. Organizational structure is another way information may be fragmented. Thus if I have copied and migrated my laptop directory structure to my desktop I now have two directory systems on my home system. Both are on one system, so in a sense they are in the same place. But in a networked world there is no real change between having my laptop directory system appear as an icon on my desktop but retrieving the files over the network and having those files on the local system.   In both cases it is an effort to know *where* things are. I could, of course, merge the two directory systems, but to do so would eliminate some of the cues that allow me to find things in my laptop directory.

The problem here is related to ad hoc categories but importantly different in that now we are talking about entire organizational systems. What makes an organizational system coherent, systematic, and memorable? How hard is it to swap one scheme in and another out? How easy is to translate between the two?

We care about these schemes for a few reasons. First a major goal of PIM is to reduce cognitive overload, and maintaining several organizational systems seems to me to be a major factor in overload or at least cognitive effort. Second, retrieval is not something that is always best accomplished through search. We often rely on the cues of labels, positions, and images to trigger recall or to prompt associations that will help us to discover or recover useful material. Third, it is often desirable to create collections with organizational structure precisely because we want to heighten opportunism. Put the stuff you think is good in a single file while you are just starting your research, and although later you may question the goodness of all the elements you included, you are still likely to increase your chance of piecing together valuable nuggets and seeing useful relations. This means that the many ways we keep and organize information items in personal spaces raises additional problems for PIM systems that aim to reduce cognitive load.