An empirical study of long-term personal project information management

Klen Copic Pucihar and Matjaž Kljun
FAMNIT, University of Primorska, Koper, Slovenia
John Mariani
School of Computing and Communications, Lancaster University, Lancaster, UK, and
Alan John Dix
School of Computer Science, University of Birmingham, Birmingham, UK

Abstract

Purpose – Personal projects are any kind of projects whose management is left to an individual untrained in project management and is greatly influenced by this individual's personal touch. This includes the majority of knowledge workers who daily manage information relating to several personal projects. The authors have conducted an in-depth qualitative investigation on information management of such projects and the tacit knowledge behind its processes that cannot be found in the organisational structures of current personal information management (PIM) tools (file managers, e-mail clients, web browsers). The purpose of this paper is to reveal and understand project information management practices in details and provide guidelines for personal project management tools.

Design/methodology/approach – Semi-structured interviews similar to that in several other PIM exploratory studies were carried out focusing on project fragmentation, information overlap and project context recreation. In addition, the authors enhanced interviews with sketching approach not yet used to study PIM. Sketches were used for articulating things that were not easily expressed through words, they represented a time stamp of a project context in the projects' lifetime, uncovered additional tacit knowledge behind project information management not mentioned during the interviews, and were also used to find what they have in common which might be used in prototype designing.

Findings – The paper presents first personal project definition based on the conceptualisations derived from the study. The study revealed that the extensive information fragmentation in the file hierarchy (due to different organisational needs and ease of information access) poses a significant challenge to context recreation besides cross-tool fragmentation so far described in the literature. The study also reveals the division of project information into core and support and emphasises the importance of support information in relation to project goals. Other findings uncover the division of input/output information, project overlaps through information reuse, storytelling and visualising information relations, which could help with user modelling and enhancing project context recreation.

Research limitations/implications – On of the limitations is the group of participants that cannot represent the ideally generalised knowledge worker as there are many different kinds of knowledge workers and they all have different information needs besides different management practices. However, participants of variety of different backgrounds were observed and the authors converged observations into points of project information management similarities across the spectrum of different professions. Nevertheless, its observations and conceptualisations should be repeatable. For one, some of the issues that emerged during this work have been to different extents discussed in other studies.

Practical implications – The empirical findings are used to create guidelines for designing personal project information management tools: support the selective focus on information with the division into core and supportive information; visualise changes in project information space to support narratives

The authors thank the study volunteers for their participation and the universities for support and sponsorship.

Aslib Journal of Information Management
Vol. 68 No. 4, 2016
pp. 495-522

2050-3806
DOI 10.1108/AJIM-02-2016-0022

for context recreation; overcome fragmentation in the file system with selective unification; visualising project's information relationship to better understand the complexity of project information space; and support navigating in project information space on two axes: time and between projects (overlaps through information).

Originality/value – The study presents a longitudinal insight into personal project information management. As such it provides a first formal definition of personal project from the information point of view. The method used in the study presented uses a new approach – sketching in which participants externalised and visualised personal information and projects they discussed. The insights derived from the study form design implications for personal project management tools for knowledge workers.

Keywords Project management, Personal information management, Information fragmentation, Personal projects, Project information overlap, Project visualization

Paper type Research paper

1. Introduction

In the personal information management (PIM) community a task is often described as something to put on a to-do list, while a project is an undertaking composed of numerous sub-projects and tasks (Czerwinski *et al.*, 2004; Jones *et al.*, 2005, 2007). Projects can be understood at different levels: on the corporate or institute-level projects are usually well defined and highly structured, while on the personal-level things are usually very different. While a considerate amount of work has been published about formal and corporate project management (Kerzner, 2013), there are not many studies on how (even formal) projects and their related information are managed on the personal level.

Managing projects on a personal level is an indispensable part of the daily assignments of knowledge workers. Such project management is personal in a sense that it is (most commonly) up to knowledge workers to decide on how to manage them. The core resource to be managed is personal information – information that an individual manages to satisfy their needs, requirements and fulfil their roles (Jones, 2015, p. 5). However, it is still not clear what the characteristics of personal project management are. The empirical study presented in this paper focuses on the (tacit) knowledge behind personal project information management that is not captured by current PIM applications (e.g. level of fragmentation, project stages, context recreation, etc.). The results of our study help understanding the problems behind project information management and indicate possible solutions.

2. Literature review

Task and project-related information management has long been of interest to the research community. The majority of empirical studies in this area focus on short tasks (e.g. sending an e-mail), interruptions and restarts. Several of these studies pre-date computers and mainly focus on the managerial level in companies (Panko, 2009). The consistent finding is that tasks can be characterised as brief and fragmented. Recent studies revealed a significant amount of task switching and interruptions, and problems users have in reinstating the context of long-term projects (Altmann *et al.*, 2014; Borst *et al.*, 2015; Czerwinski *et al.*, 2004; Mark *et al.*, 2008). It has also been shown that the nature, complexity and time of interruptions as well as task goals and tasks' problem state during interruptions play a crucial role in performance and effectiveness (Altmann and Trafton, 2002; Borst *et al.*, 2015; Gillie and Broadbent, 1989). Researchers noted that users have a keen sense to prioritise both tasks and projects and refined strategies to manage them (Bellotti *et al.*, 2004; Bogunovich and Salvucci, 2011; Katidioti *et al.*, 2014). Users also relatively quickly forget a lot about their short-term

tasks and memory aids can significantly improve awareness for such tasks and their sequence (Czerwinski and Horvitz, 2002).

Compared to task studies, project management has received considerably less attention. Interest in project information management has led to studies about how much of information is project related. In the digital world this is very noticeable through the naming conventions (Barreau and Nardi, 1995; Boardman and Sasse, 2004; Jones *et al.*, 2007, 2015). A study of 28 knowledge workers showed that in the file hierarchy 34 per cent of folder names were related to projects and 9 per cent to on-going activities associated with life roles (e.g. parent, teacher); in e-mail 22 per cent of folder names were role-related, 20 per cent project-related; bookmarks contained 6 per cent of project-related folders while the majority were topic/interests oriented (Boardman and Sasse, 2004). Another study reported an even higher number: 79 per cent of folders were named by projects and when describing their information space, participants most commonly referred to projects (70 per cent of the cases) (Bergman *et al.*, 2006). Boardman and Sasse (2004) observed overlap in folders between the file and e-mail hierarchy (12.5 per cent of folders), file/bookmarks (5.2 per cent) and e-mail/bookmarks (8.7 per cent) – most of overlapping folders were based on users' projects (40 per cent) or roles (27 per cent). Bergman *et al.* (2006) reported 20 per cent of folder overlap between hierarchies. This clearly shows that information related to projects is fragmented cross-tools, although users have different organisational needs in each tool.

For long lasting projects the hierarchical structures of project folders as a possible decomposition into sub-projects can be of a great help when reinstating the project context (Jones *et al.*, 2005). Another anchor used to remember and reinstate tasks and projects is to expose information items as reminders (Barreau and Nardi, 1995; Jones *et al.*, 2015; Malone, 1983; Ramduny-Ellis *et al.*, 2005).

It has been observed that participants organise their work into units of interrelated tasks or events that share a common goal, involve a group of collaborators, use unique resources (e.g. information items), tools and each has an individual time frame in which it evolves (González and Mark, 2004). The researchers called such units working spheres and divided them into: central (currently in focus) and peripheral, and urgent (time frame needs to be adjusted) and default (follows a set time frame). Researchers have also distinguished and classified task/project-related information in: action (to be processed in next days), working (current projects) and archived information (Cole, 1982); ephemeral, working and archived (Barreau and Nardi, 1995); hot (actively used), warm (just been processed or will be in a near future) and cold information (Sellen and Harper, 2001); and active (current activities, e.g. working and ephemeral), dormant (potentially useful), un-useful (not needed anymore) and un-assessed (not yet processed) (Boardman, 2004, p. 202).

Other categorisations have been based on how information is needed for a task/project at hand: problem information (describes the structure, properties and requirements of the problem at hand), domain information (consists of known facts, concepts, laws and theories) and problem solving information (covers the methods of problem treatment) (Barr and Feigenbaum, 1982). Another categorisation divided tasks/projects by complexity in terms of (un)established procedures and information requirements, which affect users' perceptions of information needs and search strategies (Byström and Järvelin, 1995): automatic information processing with known procedure and results; normal information processing with known procedure and slight decision making; normal decision with a more free procedure and more decision making; genuine decision tasks/projects with unknown procedure but known result;

and genuine decision tasks/projects which are unexpected, new and unstructured with unknown procedure and results.

There are many interesting and valuable findings presented in the above studies. However, none of them focused on the longitudinal aspects of information management of projects. The unexplored areas are identified as follows.

In relation to project resources, Bondarenko (2006) mentions that the set of project documents changes during the project's lifetime. However, little is known on what exactly is a common motif and how it manifests in the form of information used, produced and organised as a result of a project and how resources of a project are stored, grouped and retired in support of this process. Many studies mention information fragmentation (Bergman *et al.*, 2006; Boardman, 2004; Jones *et al.*, 2007, 2015; Karger *et al.*, 2005); although there is little known on the degree of information fragmentation, how it affects management and to what extent should information related to a project be integrated to satisfy the information fragmentation problem. Some researchers argue for unification of all information together (Karger *et al.*, 2005) while others claim that (at least in current PIM tools) different information types are acquired differently (automatically for e-mail and manually for files) and thus (currently) need different management practices (Boardman, 2001; Jones, 2007a).

Other issues that illustrate the need for more studies in this area are: how project information management of one project relates to information management of other projects (how they overlap through information); and how is the context of a project recreated when restarting work on it? For the former, even if studies have referred to project-related information as unique to one project (Bruce *et al.*, 2011; González and Mark, 2004; Jones *et al.*, 2005, 2007) – observing the evolution project folders in no relation to other information in other file folders or of other types – it is almost impossible to imagine that information management of one project would not affect others and that none of its information is not reused, exploited or recycled for different projects (Whittaker, 2011). For the latter, studies have shown that artefacts in a working environment help recreate the working context. However, little is known on what tacit knowledge (e.g. information value) behind project information management (i.e. not revealed by present PIM tools) can be externalised and computerised in support to reinstating project work (especially after long-term gaps).

3. Method

The most significant gap in the research to date is the lack of longitudinal studies focused entirely on (personal) project information management. The main goals of this study were to understand how project information is organised and accessed in digital and physical environments over longer periods of time, how users cope with fragmentation and what tacit knowledge (along with explicit artefacts) helps them recreate projects' contexts. The specific research questions that this research aims to answer are:

RQ1. How are projects understood from a user's perspective, why they are seen as projects and what is the common denominator of projects from the information management viewpoint?

RQ2. How is project information organised, what affects organisation and to what extent needs information to be integrated to satisfy the information fragmentation problem?

RQ3. How is project context reinstated and what is the tacit knowledge behind project information management that can be externalised to ease project management and context recreation?

RQ4. Are projects related, how are they related and how do they overlap through information and how does this affect information management?

The next sections describes data collection, the participants' profiles and data analysis.

3.1 Data collection

Due to the lack of prior longitudinal combined research on project fragmentation, information overlap and project context recreation, the nature of the study is necessarily exploratory. For this purpose we carried out semi-structured interviews similar to that in several other PIM exploratory studies (Barreau and Nardi, 1995; Malone, 1983; Voida *et al.*, 2011; Whittaker and Hirschberg, 2001). In addition, as described later, we enhanced interviews with sketching approach not yet used to study PIM.

Interviews were focused on two levels of project management:

(1) description of the most important projects in the last two weeks (beginning, ending, other people involved), why these projects are important and how they fit in with their work and with other projects; and

(2) description and sketching of information related to each project mentioned, its location, type of acquisition (created, received, found), why the information is valuable and how they would feel if it got lost.

Interviewing has some limitations as participants might (i) try to provide socially acceptable answers, (ii) talk only about what they think is important to researchers, and (iii) have imperfect memory about the subject of the interview (Nielsen, 2001). To capture projects, which participants felt like describing and not the ones they thought might accommodate the researcher (to address points (i) and (ii)), we did not instruct them about what a project is.

At the end of each interview we asked participants to sketch how they visualised their projects in relation to information. Sketching is ideal for articulating things that are not easily expressed through words (van der Lugt, 2005). However, it was used also for other three reasons: sketches from different interviews represented a time stamp of a project context in the projects' lifetime; to uncover any additional tacit knowledge behind project information management not mentioned during the interviews; and to find what sketches have in common which might be used in prototype designing. The last point presented a kind of participatory design of which users were unaware. By simply asking them (without any pressure) to visualise the state of the project in respect to its information, the expectations on both sides were not high. Acknowledging that users are not designers, we nevertheless hoped to uncover some common visualisation features. Nevertheless, sketches helped us to confirm consistence of participants' answers given during interviewing.

Participants were also asked if they were willing to have up to four consequent interviews. The gaps between interviews were from two weeks to four months based on the time available and willingness of each individual. The only difference between the first and consequent interviews was that participants were shown their previous project sketches and asked about possible changes and why these occurred. The consequent interviews were also used to discover what users have forgotten to mention during the previous interviews (see point (iii) above of the possible drawbacks).

3.2 Participants

We recruited participants by posting an announcement on universities' mailing lists and our social networking sites asking to spread it further if possible (a combination of convenience and snowball sampling). This announcement informed prospective participants about the aim and length of the study and stated that there were no specific criteria for participation. In all, 19 participants answered our call and agreed to participate. The number is similar to other PIM qualitative studies (e.g. Bruce *et al.*, 2011; Jones *et al.*, 2007, 2015) and was seen as adequate for the nature of the study. Participants were aged from 23 to 61 ($m = 30.5$, $\sigma = 8.8$) and included two university professors (computer science, education science), one senior researcher (psychology), one web developer, two store managers, one library manager, one event manager, one security manager, seven PhD students from five departments (of which three worked outside academia for more than three years), two freelance translators and one administrative worker. Participants will be referred to as P1-P19 to anonymise their identities. Eight participants were prepared to continue the study: four participants had four interviews and four participants had three interviews. The reasons stated for not continuing were either not being available (changing job, position or going on exchange programme, internship) or did not have not enough time. Nonetheless, the saturation point of obtaining new data were reached before the last interview with the eight participants continuing the study.

3.3 Data analysis

We carried out 37 interviews lasting on average 90 minutes (30 minutes minimum and three hours maximum) and taking place in the participants' own working environment. The sketches and transcripts of the interviews were analysed and coded with categories developing during the coding process. This allowed us to find commonalities related to project management with iterative comparison of conceptualised data. The saturation point had been reached before all transcripts were coded and we have not sought additional participants. Through the coding iterations, two large categories were formed: a category related to project description, and a category related to project information management and its flow. These categories with their subcategories form the next two sections and their subsections.

4. Results 1: defining a project, duration and importance

To our knowledge personal projects have not yet been formally defined from the information perspective. The understanding of a project in this paper is the same as understanding of the PIM community, which informally describes them as an undertaking made up of sub-projects and tasks. Participants in this study were asked to "choose the three most important projects, activities or tasks that you worked on in the past two weeks and involved information such as files, e-mail, web bookmarks". With the emphasis on "important" and the "three" projects, we tried to avoid descriptions of tasks. During the interviews participants described 45 projects as defined by PIM and four on-going activities (two e-mailing, one called "skyping" and one maintenance of a file system). The latter could not be described as projects in a PIM sense as their tasks belonged to several projects.

For this reason, a new classification was made based on whether projects were supporting other ("real") projects, and the information goal of each project:

(1) Projects with a specific outcome, which is one or more information items (41 projects). The results of these projects are created documents such as

writing papers, reports, organising events, official funded projects (production of several documents), etc.

(2) Explicit auxiliary projects (four projects). Such projects usually started because of the projects in (1), but were seen as separated projects. The projects in (1) could not start or be completed without them (networking, searching for ideas, gathering reference material, etc.).

(3) Implicit auxiliary repetitive projects (four projects). These include everyday continuous activities such as e-mailing, maintaining file system, updating calendar, etc. Many tasks of these projects were in support of (1) and (2).

By nature, this classification is close to information complexity classification (see Literature review section). The first two groups of projects were either "normal decision projects with a freer procedure and more decision making", or "in known, genuine decision projects with unknown procedure but known result". The classification and examples of projects together with information complexity scale is presented in Table I. In this paper we will discuss mainly projects from the first group, except when parts of supportive projects were related to projects of the first group.

All projects of the (1) and (2) had at least to a certain degree defined a path to the end result (or the end result itself). Even projects that did not yet exist in any information form, but had an envisioned output were seen as projects (e.g. P7's idea of a book translation). Contrariwise, if a path and the output were not clear, participants did not consider it a project. For example, a PhD was never mentioned as a project (even when asked about it) – it was too vague and unclear, while parts of it (e.g. writing a thesis) were considered as projects. By information complexity a PhD can be classified as a

Project type	Examples	Number of projects	Type of a project by information complexity
Project with a specific outcome	Writing a paper/report/ proposal Organising a workshop Translating a book Designing and building a web site Proofreading papers for a journal	41	Normal decision projects with a more free procedure and more decision making and in known, genuine decision projects with unknown procedure but known result
Explicit auxiliary project	Gathering reference material about a topic Networking on a conference for the next possible book translation project Searching the web for programming tips	4	Normal decision projects with a more free procedure and more decision making and in known, genuine decision projects with unknown procedure but known result
Implicit auxiliary project	Daily reading and writing e-mail Frequent maintenance of the file system Daily updating the calendar Using VoIP and IM software for various conversations	4	Normal information processing with known procedure and slight decision making

Table I. Classification of projects based on whether projects were supporting other ("real") projects, and based on the information goal of each project

"genuine decision project" with yet unknown procedure and final result. This vagueness of procedures and results probably contribute to not defining such undertakings as projects – participants do not see them as projects until procedures and results become clearer.

This can also be explained by the theory of action interpretation, which states that the task is identified at the highest possible level of identification. Thus "driving a car" can represent a task for an inexperienced driver who has to think about synchronising movements, while an experienced driver could identify a task at a higher level such as "drive home". The analogy with projects is similar; a project is identified at a level where a route to the end and/or a goal (in the form of information) is known to some extent. If not, a project in information space is interpreted at the level at which this can be achieved.

Some described projects were related, nevertheless participants made a clear distinction between them. For example, a conference paper and a presentation of it were seen as two different projects (three participants); organising a workshop of which outputs were later published in a magazine and still later in a book were seen as three different projects; and translation of a book and its afterword were seen as two different projects. We found that participants divided clearly related projects if information needed different handling, if the end results had different deadlines, or if the end results from the same data were different in nature as explained in this quote:

> It's another project [presentation]. It's based on the same material, but the way I presented the results is completely different to me (P2).

4.1 Project importance

One of the study aims was to understand what makes projects important. All reasons discussed are summarised below (for most projects a combination of reasons were the case):

- they presented a (personal) challenge/ambition/aspiration/motivation;
- one's knowledge/ideas/abilities/time were invested;
- end results would establish one's position, career or prove capabilities;
- company/institution/individuals would benefit from end results;
- the audience of output information (no-one, specific circle of people, experts, general public);
- criticality of projects (e.g. deadlines); and
- because they have been self-selected/part of the job/done for money.

4.2 Sub-projects, stages, milestones and tasks

The division of projects in smaller parts has manifested mainly in: the change of a project name between the interviews as project focus changed; the change of a name from a specific part of a project to a broader project name or vice versa; and in several sketches of the same project each focusing on different information item(s). Participants referred to such changes as stages and milestones. At each, participants focused on different information item(s) or different parts of one information item. Stages/milestones were also seen as local ends of the project and represented its natural

progression steps. For example, P8 divided a translation project in searching for material, applying for financing, and translating the book.

Stages as described above (having information item(s) or information chunk within an information item as a goal) are very similar to projects (outcome in the form of information) and can easily be called smaller projects or sub-projects. Sub-projects have not been defined in the PIM literature but are often referred to as smaller projects within a larger project.

Sub-projects are also very similar to so-called working spheres (see the Literature review) described as higher levels of units of work sharing a common goal, unique resources and a group of participants that people divide their work into on a daily basis. Our participants described projects as a series of sub-projects (stages/milestones) resulting in a visible end result(s) (in the form of information item(s)). However, the examples of working spheres given by authors (e.g. "the TRK stuff" or the "the R6 spec") resembled not only projects, but also sub-projects. There are other differences: majority of projects or sub-projects were not defined on a daily basis (see projects' duration below), and do not have unique resources.

4.3 Duration, end and afterlife

The duration of discussed projects spanned from three days to five years. Most of the projects lasted from a few weeks to a few months. This is in line with other studies where projects were estimated to last everything from two weeks to two years or longer and on average 34 weeks (Jones *et al.*, 2007).

Ending times of projects proved to be hard to predict for projects with no official deadline and kept shifting in the future when more important projects were prioritised or because of other people involved. This shows that participants have a keen sense of prioritising projects as observed for tasks in Bellotti *et al.* (2004), Bogunovich and Salvucci (2011), González and Mark (2004), Katidioti *et al.* (2014). Due to prioritising projects sometimes switched from active to dormant and vice versa.

Despite the keen sense of prioritising, participants are not very good at predicting the amount of required work. These observations are consistent with the planning fallacy (Buehler *et al.*, 2002; Pezzo *et al.*, 2006). The time (cost, risk) underestimation (and overestimation of benefits) of an involved actor has been observed in a variety of everyday tasks and projects such as school work, furniture assembly and computer programming which also contributes to the ad hoc approach to management as described by P7:

> I set up a due date and push myself to work on the project but it depends on how much do I care about that date […] and some other projects might be prioritised in the mean while (P7).

For projects with disseminated information, information afterlife or how the audience accepted such information was very important (one "project importance" reason was establishing one's position, career path or prove capabilities).

5. Results 2: organising and accessing project information

The categories that characterise project information management are presented in the following subsections in the order of an idealised project flow: organising information, optimising it for a project at hand, (re)create project context and identified problems in doing so.

5.1 Organising project information

As in other studies (e.g. Boardman and Sasse, 2004), participants started to describe project information spaces in the file systems, and organisational structures could only be observed here. Only one participant filed e-mail and only one filed web bookmarks (of four people who kept them). The problems with organising bookmarks are well known for nearly two decades (Abrams *et al.*, 1998) and a recent study has shown that users rarely use them – reproducing search queries and following of known paths seem to be two main ways of re-visiting web pages (Obendorf *et al.*, 2007). Similarly (large amounts of automatically acquired), e-mail forces users to opportunistically access e-mail and to rarely relying on folder structures (Whittaker *et al.*, 2011). Other organisational techniques such as tags have not been observed for any information type, which is consistent with other PIM studies (Civan *et al.*, 2009; Whittaker *et al.*, 2011).

Three main personal information collections have been defined in the literature. Jones (2007b, p. 139) describes two types: project and reference collections. The former contains project-related information, has elaborate organisation and diverse file formats. The latter includes information of usually one format, has a fairly flat organisational structure and is organised by metadata or key (time, name, topic). The third type is dump collections for files that needed some treatment before being appropriately filed (Kamaruddin *et al.*, 2013). However, our observations extend these conceptualisations.

5.1.1 Dump collections. Dump collections observed in our study were not used only as a temporary space for information items as has been suggested in the literature. They served other purposes as well: for filing files that did not fit anywhere; for automatically acquired information; and for moving in files that were cluttering the information space. For example, a dump collection of a participant has been archived and he did not have any intention to ever look at it again:

> When desktop gets too many icons I create a folder "Old Desktop" with a date and move everything in. Then I start putting new icons on the desktop (P5).

Reference collections. Some reference collections complied with theoretical properties of a classification scheme: a unique classificatory principle, exclusive categories, and completeness (Bowker and Star, 2000). Such observed reference collections can be classified into:

- files of one format organised on one key (e.g. academic papers organised by authors' names); and
- folders (containing files of one format) organised on one key (e.g. created music files in folders organised by year and month in folder names).

However, more often than not, organisations of reference collections contained file/folder names of mixed keys (e.g. a collection of some topic and some people-related folders), different file formats (a collection of folders of meetings sorted by dates (in folder names) containing files of different formats) or non-key-based reference collections of diverse file formats (a collection of personal documents (a CV, scans of social security card, passport, bank details)) or the same format but unorganised (academic papers from the internet saved with their original names). The diversity of a reference collection of academic papers is illustrated by P3's statement:

> Some of the folders might be more general and some of them are about some specific protocols […] but it's not only topics […] is people as well. This is the stuff from Carlos (P3).

In this example a huge collection was organised in folders by at least three different keys: topics, protocol names and people. These examples are in contrast with the ideal reference collection "one format, flat structure, organised by a key". Reference collections contain reference material, although organisation of information in each can vary greatly based on different factors (e.g. the context in which information is acquired).

5.1.2 Project collections. Project collections also differed greatly from project to project. Nevertheless, most observed project collections contained what we define as core and support information. Core information items were the main output of a project while support items were documents used as a source for the content of core documents, notes, etc. This division was externalised in sketches as well. Projects with more than one sketch (18 sketches) had core information almost always the same while supporting information changed between interviews (notes, support files, etc.). Participants explained that such documents were not relevant anymore. Similar changes between sketches also occurred when participants draw the current stage and the end stage (four sketches). Participant P7 also described the split of information into core and support:

> The reason why these files are important is because they are the aim of the project […] other files are only a tool to achieve it […] like correspondence, documentation […] day to day things […] while the aim is more long lasting (P7).

Collections can also be mixed. For example, a project collection containing a dump collection, a project collection containing a subset of a bigger reference collection (duplicated information) or its own reference collection (support information), a reference collection containing a dump collection of items that needed to be treated and filed, a reference collection containing some project collections, or a reference collection of finished project collections. However, whatever the mixing, one type was always a dominant one.

While reference collections looked more integral, project-related information was often very fragmented. Support information for a project was often found outside the project folder (in reference, other project and dump collections) and core information was not necessarily stored in the project collection folder. There were several reasons for this:

- items were moved out of a project collection for easier access (this is expanded further in next section);
- legacy of old organisational structures (e.g. a project folder was not yet created at the time of creating a project-related document elsewhere);
- misplaced items (e.g. a document has been misplaced when first filed and has not been moved for the fear of not finding it); and
- software enforced organisation managed within the software itself (integrated development environment, statistical software).

On several occasions a finished project evolved in another project and shared the same folder, or files of several projects were grouped in one folder based on some key (e.g. all translation of one author). Projects were sometimes also grouped based on the period when they happened (PhD folder with several non-PhD projects). Several projects were stored directly in the storage folder for personal documents (e.g. my documents) as they contained one file only or were initially stored there and later grew in size. Without a

tour, it was very difficult to establish where all project information was stored because of the above reasons (access, misplacement, legacy) and because projects overlapped, run in parallel or in serial, reused information one from another (provenance). These actions in the file hierarchy caused a high level of fragmentation in a file system alone and posed a problem for remembering all project-related information.

5.2 Optimising organisation for easier access and reminders

Users often leave information visible and at hand as reminders and for quicker access (Malone, 1983; Ravasio *et al.*, 2004). We observed this behaviour from the project information management point of view. In physical space, participants formed circles of project-related information around them. Information closest to participants and in visible sight (computer or keyboard) tended to be of the project they currently worked on and information related to less important projects tended to be further away (on the desk edge, floor around it, shelves). The visibility proved to be very important. P8, for example remembered the valuable project notes only when she turned and noticed them (after the discussion about related project was over). In order of closeness these circles represented project(s) currently worked on, current projects on a waiting list, dormant and finished projects. We coined the term personal space of information (PSI) circles to describe this organisational technique.

In a digital space PSI circles manifested in many ways. In general, participants measured the distance in number of clicks to get to the particular folder or file. However, they all used different approaches to achieve this on a macro level of a file hierarchy and micro level of a project collection folder.

On a macro level:

- Project folders moved to the desktop or higher up the hierarchy (e.g. home folder).
- Files moved out of the project folder to the desktop or higher up the hierarchy.
- Project-related files or project folders were added to favourites or sidebar of a file manager:

  > I need to go to projects folder and then to this particular project and […]. when I'm on a phone conference and I need to go to the document very fast this becomes very annoying. This is why I have it here [favourites] (P3).

- Duplicating information (from other collections):

  > I copy all documents on the meeting agenda in the meeting folder […] If we are referring to these documents during the meeting I just open them. This saves me from navigating and going to the various levels of folders (P11).

- Files are accessed through the associated application:

  > […] I open my documents in Word […] here [recently used] are all the documents I'm working on (P1).

- Leaving files in a dump folder to get attention.

Participants also used other ways to access the same information based on the current context, what software was opened on the screen, what they remembered about information and where they were located at the moment of accessing information.

On a micro level:

- files moved up the hierarchy of a project collection folder;
- files that cluttered the project collection were moved to a dump folder;
- file names were given a prefix or suffix to be on the top of the list (as in Jones, 2007b); and
- files ordered by time by default so the often–used items were on the top of the list.

Such management practices somehow differ from just leaving information (Malone, 1983) or other artefacts (Ramduny-Ellis *et al.*, 2005) in sight as reminders. Arguably, uncluttering information space of unneeded information or moving it to more accessible locations acts as reminding. However, more importantly it reduces the number of needed steps to reach target information and eases its visual finding in a less cluttered space. It has been observed that people at each navigational step use contextual information associated with a target information (Teevan *et al.*, 2004) and that navigation task difficulty increases with hierarchy depth (Cockburn and Gutwin, 2009). It is thus not surprising that some users try to reduce the cognitive and physical effort required for navigating to desired items – especially the often accessed ones. Of the observed PSI circles creation, only one is directly implemented in the file manager (favourites and sidebars) while others happen as the consequence of artefacts present in the current software and are not particularly tailored for project work.

5.3 Recreating the context

Recreating context of a project is crucial to restart working on it. Switching to projects in the background proved difficult as explained by P3:

> This is why I don't like these interviews […] because I have to think hard about another project […] when I switch between projects I lose focus on the other project […] and now I have to switch back and think what was that project and it's really annoying (P3).

Project collection folders in a file hierarchy were of a great help and participants often took a few seconds to understand them:

> I need to navigate to it [folder] […] without seeing it I can't really do it [name related information items] (P11).

The discourse about past actions and sketching additionally reinforced memories about related information and helped to better understand the current project information stage. However, the fragmentation of project-related information in the file hierarchy (PSI circles, information misplacement and software enforced organisations) and the focus on information regarding the current project stage were often the case why several times participant remembered information only after the related project was discussed, or during the following interviews. As a result, participants often returned to previously discussed projects, when finding related files around the hierarchy, while browsing for something else, and when looking at sketches.

Participants also used other ways and techniques to help them recreate context as, for example:

> Every time I need to go through all the chapters and see what I have done before […] so it's lot of reading to catch up. When I restart the project I need to reload the mental map I had before of various sources of information and one way to help me is to print documents […] is a matter for me knowing which ones were relevant or very important at the time (P3).

Consequently, the more documents involved and the more they were fragmented in the file hierarchy and between tools, the harder the context recreation. The context recreation was even harder when other people were creating or contributing to files. Participants explained that they need to go through others' people changes as well as remembering their own.

Summarising, recreating context happens on two levels (information space and information itself) and involved:

(1) information organisation level: remembering relevant information items, their positions (organisational structures) and steps taken (naming, renaming, moving, grouping, archiving, etc.) to achieve the current state of information space; and

(2) information content level: remembering content state of individual information items and steps (changes to the content) taken to achieve it.

In one of the distributed cognition models of work, artefact (in the information space) can either work as triggers or as placeholders (Dix *et al.*, 1998). These artefacts can be either internal (in people's heads), external implicit and external explicit. The state of current information space (organisational structures and information position) exemplifies the external explicit knowledge while the internal knowledge is the knowledge of how this was achieved. Both internal and external knowledge combined support context recreation and external knowledge often triggers internal knowledge to life. The latter was even reinforced by narration and sketching which helped discover a lot more information than by navigating the information space alone and could be considered in PIM tools design.

5.4 Failures in recreating the context

We observed a handful of projects for which participants could not remember the discussed anchors for context recreation and consequently failed to find information. Boardman observed and mentioned three reasons why information could not be retrieved: deleting/archiving items, clutter, and misfiling (Boardman, 2004, p. 89). However, he did no go into details behind the failures. On the contrary, this paper provides a conceptualised and more detailed account of project context recreation failures:

- Unfamiliarity with one's own structures after long periods of absence: sometimes finding and understanding files and folders was hard as participants have not accessed them for a while. Lansdale (1988), for example argued that the memory of file (folder) names (and their position) and value decay over time. The problem also lies in the fact that naming patterns change all the time (Carroll, 2007). P8 also explained that if she used more informative folder names she might find them quicker. P2, P7 and P14 even failed to find their entire project collections of dormant projects.
- Unfamiliarity with the inherited, received, enforced or found structures and names: this unfamiliarity was obvious when users received or found folders and files. Even after a few days, the content behind the file or folder name was forgotten and required opening, reading, scanning to recognise it. Files that are managed and accessed through associated software applications are also hard to find in a file hierarchy. P14 received by e-mail a compressed folder of an entire project collection and she struggled to explain the structure and content even if she looked at everything just two days earlier. P8 received by e-mail a

compressed file she kept in a project folder and guessed wrongly what was inside until she opened it. An extreme was P10 who inherited the whole computer from a prior worker and she explained how she had to open file by file and reorganise everything, as the previous organisational approach was "entirely wrong".

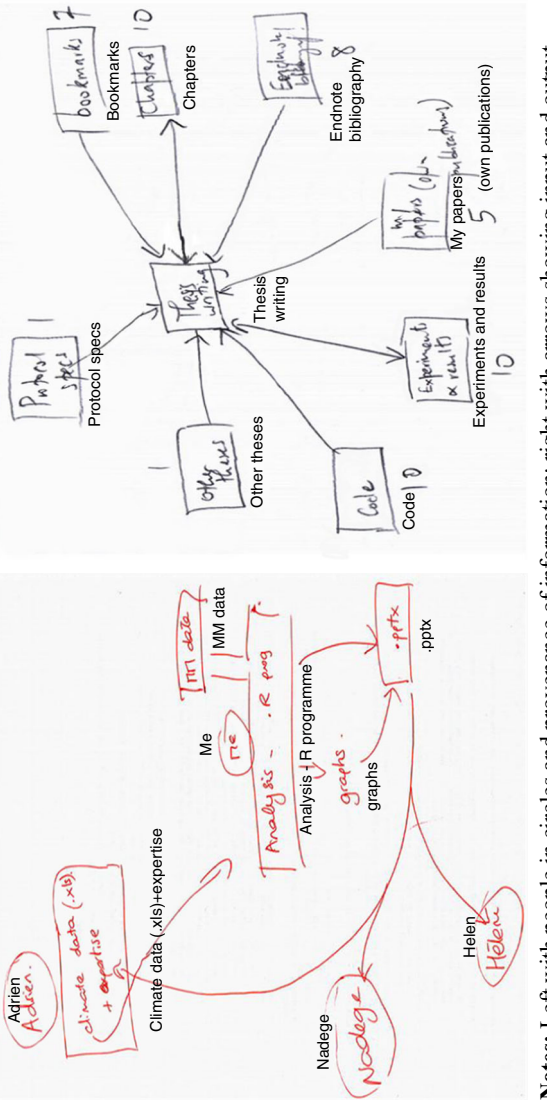
- Unfamiliarity with own broken structures: this type of unfamiliarity happened when folders or files were misplaced or have grown so big that it was hard to manage them. P3, for example had a reference collection of scientific papers with 178 subfolders in it and failed to find a file even with a desktop search engine. Other examples include e.g., a download folder of over 2,000 files, which needed to be sorted by creation date to access the recently downloaded ones (P9). Bookmark hierarchies of participants who kept them (P9, P3, P5, P14) were also just a long list of unrecognisable items: "So what is happening is my bookmarks are rapidly becoming a bit unmanageable. It takes me a long time to look in to my bookmarks and find something that I saved. I very often forget that I saved something. So a lot of very interesting things I saved I can't retrieve" (P14).
- Unfamiliarity with one's own changes in information environment: renaming or moving information to better suit ones' needs (e.g. creating PSI circles as described above) after becoming used to the old name or old location caused this type of unfamiliarity. P2, for example remembered the old file name but could not at first remember the new one, which (as she explained) often happens. While P14 always leaves accidentally misplaced files misplaced to be able to find them.
- Unfamiliarity because of information fragmentation: on some occasions participants could not remember where they kept last versions of documents. This happened to P4, P5, P8 and P15, when they worked on two computers (latest version on a thumb drive and another computer) or when they collaborated with others (last version in an e-mail as attachment not yet downloaded).

Research in task interruptions has looked into how hard it is to recreate the context of a primary task (or problem state) after an interruption. Both theories of memory for problem states (Borst *et al.*, 2015) and memory for goals (Altmann and Trafton, 2002) state that retrieving the primary's problem state (or context in which the task was left) from memory takes time and "can lead to errors when an incorrect (i.e. older) problem state is retrieved […][and] the longer the interruption, the further the activation level of the problem state will have decayed, and the longer it takes to retrieve it from memory". If tasks are a part of larger projects, storing a problem state(s) during (both information organisation and content levels) interruptions and retrieving it later is even bigger of a challenge – especially when above mentioned reasons are in place.

5.5 Visualising projects with sketches

As far as we know, sketches have not been used so far to study PIM. As mentioned throughout the paper, they revealed or supported several finding presented so far. In this section we are presenting their design only.

We ended up with 66 sketches of 45 projects; 37 sketches resembled a map like structure of a project stage with either a project name or an information item in the centre or with no apparent centre and few islands of items or locations (Figure 1). Mixing maps with other visualisations included: four sketches with a file system hierarchy, two internal document structure (the output of a project), two a chronological sequence of stages/steps of a project and one with a list of documents. Ten sketches



Notes: Left with people in circles and provenance of information; right with arrows showing input and output information

Figure 1.
Two map/stage like diagrams

visualised only a chronological sequence of projects' (past and future) steps with information entering in (Figure 2), eight sketches contained just a list of project-related information items and one had a list of items and their locations in the file hierarchy.

Besides files and folders, e-mail was also included on the sketches when it triggered an important step in the project course. P7 explained that he differentiates between the "content" of the project and e-mail as coordination of the project. On two sketches with chronological sequence of steps, e-mail was included as an on-going activity throughout the project course (Figure 2 left).

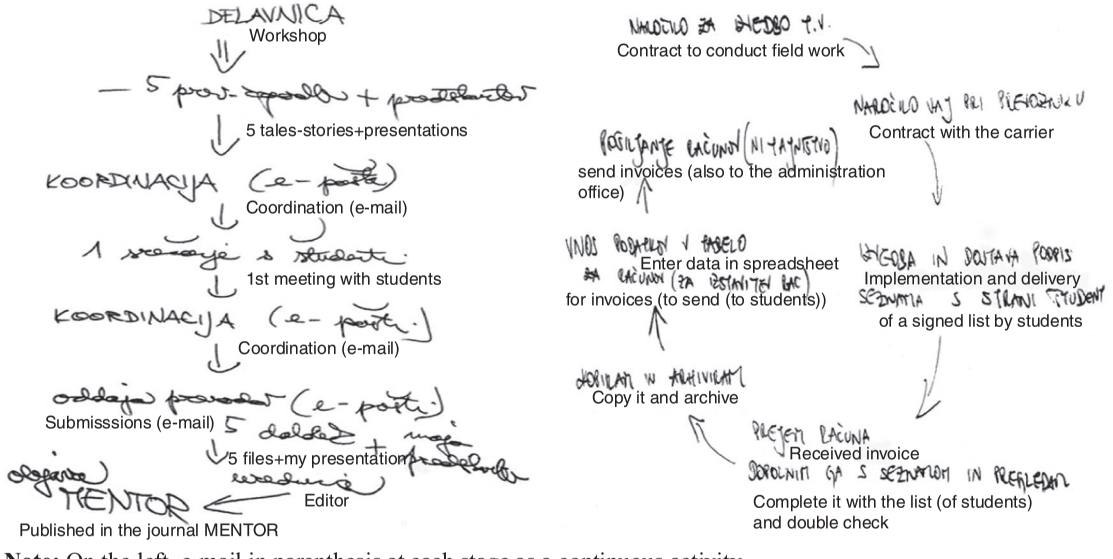
Besides information items (files, e-mail, web bookmarks) people, information relationships/provenance and information role were included in sketches. People (Figure 1 left) were associated with particular information items in the first place and only later with the project (more about this behaviour can be found in Kljun and Dix, 2012). Calendar inputs and other information never made it to the sketches. Three participants had due dates in their calendars, but as with e-mail, these were not the "content" of projects.

During the tours of project organisation, participants rarely described the relationships between information items. These became obvious only during sketching and drawing arrows between items. Participants explained that projects and information cannot be seen as related on the computer, but it is very much related and the arrows show these relations in some way. The map like sketches could be grouped into three groups based on the directions of arrows: arrows pointing from the source information to the output information (document provenance Figure 1 left); arrows pointing out from a project name in the middle to all information around it; and third, arrows pointing in and out from the project name in the middle showing projects' inputs and outputs (Figure 1 right). Other links were drawn between projects showing their relationship as when describing one project, participants sometimes started to sketch other projects that evolved from the one or run in parallel. Such sketches also confirm the information reuse or sharing between projects. There were also arrows between project stages on the chronological list of them (Figure 2).

Due to resemblance of sketches to mind-maps, an obvious question is why users do not use mind-mapping applications to manage their information and related projects (although sketches from our study are technically not "mind-maps" as described by Buzan and Buzan, 1993). It has been noted that mind-mapping applications could be particularly valuable in supporting making sense of information and information space (Jones, 2007b, p. 265). However, dragging in files and other information types to a majority of such applications results in a lost link between the original item and its version in mind-mapping application. Which further results in multiple versions of the same information item as well as multiple organisational structures to be managed.

6. Discussion

The study presented focuses on how knowledge workers manage personal project-related information. It is acknowledged that the group cannot represent the ideally generalised knowledge worker as there are many different kinds of knowledge workers and they all have different information needs besides different management practices. However, participants of variety of different backgrounds were observed and we converged observations into points of project information management similarities across the spectrum of different professions. Nevertheless, its observations and conceptualisations should be repeatable. For one, some of the issues that emerged during this work have been to different extents discussed in other studies. However, the specific, and to our knowledge



Note: On the left, e-mail in parenthesis at each stage as a continuous activity

Figure 2.
Two chronological orders of project stages with eventual documents in each stage

unique, focus on deep and long-term project understanding has brought various new issues to light. There are many implications of these findings that could help in designing project information management tools and users to recreate project contexts after short as well as long-term gaps between restarts.

6.1 Definition of a personal project

Personal projects have not yet been formally defined in the literature from the information point of view. Projects as units of analysis have been studied in several research areas such as personality science (Little, 2015), occupational theory research (Arcand-Dusseault and Egan, 2015) and nonetheless PIM (Bruce *et al.*, 2011; Jones *et al.*, 2005, 2007). The concept of a personal project is understood in the originating research area of personality science as a "set of interrelated [scheduled] acts extending over time, which is intended to maintain or attain a state of affairs [goals] foreseen by the individual" and can be short- (plant a tree) or long-termed (build a business), abstract (be a good person) or concrete (cook a dinner), and are described in terms of ends (passing the exam) and means (studying) (Little, 2015). While other research areas look at the project from a wider perspective, PIM focuses on information part. Nonetheless, personal projects are often (as also seen in this study) managed through information and need information as a mean to achieve an end (also in the form of information) (Bergman *et al.*, 2006; Bruce *et al.*, 2011; Jones *et al.*, 2005, 2007). The current consensus in PIM states that personal projects are undertakings composed of several sub-projects and tasks, and the latter are something to put on a to-do list (Czerwinski *et al.*, 2004; Jones *et al.*, 2005, 2007). However, personal projects from information perspective are characterised by many other features as revealed by this study.

Projects described by participants all had a clear and visible goal in the form of information. The analogy can be drawn from the theory of action interpretation which states that the actions are defined at the highest possible level of identification at which people still understand the consequences and know how to complete it (Vallacher and Wegner, 1987). Similarly, projects are interpreted at the highest possible level at which participants understand and visualise a goal and have an idea at how to achieve it. If the goal cannot be understood and visualised, participants have not considered it a project and thus reinterpreted it at the level at which the goal was "visible" (e.g. a paper instead of a PhD).

Personal projects are also loosely planed because of unset due dates, prioritising and planning fallacy (underestimation of work needed). This coincides with Suchman's definition of plans (Suchman, 2006): however, planned, purposeful actions are inevitably situated actions – actions taken in the context of particular, concrete circumstances. Interpreting actions as such, she claims that our actions, although systematic, are never planned in the strong sense since they need to accommodate the unforeseeable contingencies of particular situations. For these reasons plans are necessarily vague.

Based on these findings a personal project of a knowledge worker can be described as a self-defined or given undertaking lasting from days to months that is directed towards and defined by a specific goal in the form of information or/and a path to achieve it, managed (information, time, people, equipment, budget) by an individual on a day-to-day, semi-formal manner based on the current context, this individual's ingenuity, past experiences and knowledge (of technology and information), and made up of (often loosely) planned tasks and sub-projects affected by planning fallacy and completed when remembered, when time permits or when approaching formal due dates.

6.2 The division into core and supportive information for selective focus

The goal of a project in the form of information forms a set of core information items (e.g. documents) – a set that rarely changes throughout a project's lifetime. On the contrary, supportive information (e.g. e-mail, web bookmarks, paper documents and files with little time and effort invested, reference collections) constantly changes and is put aside after the use (e.g. in a dump collection to move it out of the way). To our knowledge there has been no such division of project-related information in the literature.

This information flow has shown the freedom of defining and achieving the goal and consequently resulted in unstructured and unpredictable workflow (as described by Drucker, 2009; Kirsh, 2001). From the information viewpoint the division into core and supportive information could assist users in a more structured way of project information management (focusing on core information while allowing easier flow of supportive information). This division also shows the need for selective unification of project information. Not all project-related information needs to be managed together at the same time; thus, at each project step (or a problem state) a focus on a subset of project information (core and current support) needs to be supported.

6.3 The project information space story for context recreation

Several studies have shown that storytelling about information items (Chau *et al.*, 2008; Gonçalves and Jorge, 2008) or providing snapshots of daily activities (Czerwinski and Horvitz, 2002) can greatly improve the memory for the past events and tasks. A similar phenomenon emerged during this study when participants from the current state of information space interpreted the past actions. More they discussed about this process (including sketching project information space) more related fragmented information surfaced in the stories. The same happened when they talked about the internal document structures. Thinking about how the current content was created helped participants to remember many facts about it. This storytelling on the two discussed levels (information organisation and information content) helped participants to recreate the project context. From the PIM viewpoint, the content is not of a concern (as long as it does not support PIM) and there are other studies that focus on the document provenance (Jensen *et al.*, 2010; Muniswamy-Reddy *et al.*, 2006; Rasel and Ali, 2016; Shah *et al.*, 2007). Assisting users in understanding the changes in their PSI and project information storytelling could thus provide the aid to recreate project information context.

6.4 Fragmentation in the file system as a challenge to unification

The observed personal classification approaches to project collections can be portrayed by Barreau's description of this process, which stated that the classification is (in addition to other factors) affected by personal and situational factors surrounding the classification event (Barreau, 1995). Of a great importance of forming a new project collection is, for example the fact where users start to place initial information when it is not yet known if the undertaking will grow into a project and/or how the project collection will grow in size. Kidd claimed that after information is processed and its role known it is filed properly (Kidd, 1994). However, even if information is processed and stored, the actual project itself might not have started yet and many times initial placement of information becomes permanent for the fear of not finding it if moved.

Other reasons for project information fragmentation in each hierarchy separately included PSI circles, relation to other projects and their organisations (e.g. an initial project collection serves as a collection for future projects or projects share information from other project/reference/dump collections), a time period in which project started, mistakes made in filing, software specifics, etc. Fragmentation was not limited to the support information only (input information, scraps, to-do lists) but affected core information as well.

The study presented also gave an insight of (project, reference, dump) information collections' formulation. While on the macro level of each hierarchy a prevailing classification (piling/filing) strategy could be observed, this study shows that on a micro level (most noticeable in the file hierarchy) for nearly each project collection a different classification on a filing/piling axis approach was taken. Moreover, mixing collection types (project, reference, dump) makes the filing/piling distinction even harder as, for example a neatly filed project collection can contain reference and dump collections with piled items.

Fragmentation in the file hierarchy alone caused most difficulties in finding project-related information (especially for long-term gaps between restarts). This is supported also by the claims that memories of file names and paths decay with time (Lansdale, 1988). This is an undervalued problem in the current literature and PIM tools: while other studies emphasised the information fragmentation between tools, the data collected in this study supports the view that project information fragmentation in each tool separately has to be addressed as well.

6.5 Visualising project's information for better understanding the relations and complexity

Sketching as a new approach in studying PIM confirmed several findings such as core/supportive information, storytelling and fragmentation. Sketches rarely depicted the hierarchical nature of project hierarchy organisation. Visualisation resembling the hierarchy was noted only where a project folder was a decomposition into sub-projects as observed in (Jones *et al.*, 2005). Contrary to this view, project collections were often not folder-divided into sub-projects due to the fragmentation in the file hierarchy – sketches involved information from various folders across the hierarchy and files/folders dragged out of their associated locations for easier access. Depicting these information "islands" across tools and in each tool separately on a plane space proved more suitable than a hierarchy representation; this coincides with other studies (see Scarr, 2012 for a review of spatial memory used in user interfaces).

The sketches were mostly presenting spatially arranged map like stages of linked information. It is important to note that visualisations described a current information need of a project at that moment in time. The differences between sketches of the same project from each interview showed that the core information remains for the greatest part static, while the supportive is more fluid. The mentioned links between information items presented either provenance, or input/output information to a project. Other visualised attributes of information items were due dates, associated people, various notes and the overlap between projects (sketching one project resulted in sketching another one). The latter exposed that projects are not lone "islands" of information as depicted in several studies; they are interrelated through information and share, reuse, recycle and exploit resources. Another sketching pattern was the workflow of a project. In this way participants performed a storytelling equivalent on the paper. As mentioned, such activity significantly improved memories about projects' changes and often resulted in describing information not mentioned before.

Visualising projects furthered the understanding of their relationships with other information and other projects (projects' overlap through information (re)use), the flow of information, and consequently the understanding of PSI. Similar to "storytelling" of project information space advancement, its relation to the rest of the PSI should be supported in design of future project management tools.

6.6 Navigating in information space

It is clear that users do not solely navigate in the current state of information space. Explicit or observable navigation – also called orienteering (O'Day and Jeffries, 1993; Teevan *et al.*, 2004) – is described as taking small steps when navigating in the current information space and at each step deciding what the next step will be based on state of information space. At the same time there are at least two implicit and unobservable types of navigation users can rely on:

(1) navigating in time: navigating back in time and thinking about steps made to reconstruct the current information space helps users to understand the state of information and then take new decisions; and

(2) navigating between projects: navigation between projects has surfaced in the sketches when users navigated from one project information space to another project information space.

While orienteering is directed towards specific information, the navigation between projects takes a broader information space view into account and such information is not necessarily positioned on the orienteering path. Providing support to time and project navigation could help users recreate the project context and better understand their information space and relations or information overlaps between projects.

7. Conclusions

It is easy to consider the information space as a messy environment dependent on users' perception, time, context, caring, etc. However, Dourish and Bell encouraged researchers, designers and developers to embrace this messiness and perceive it as inspiring, productive, generative and engaging (Dourish and Bell, 2011). In their view mess is not something to be fixed, tamed or removed. Rather, mess is characterised as dynamic, adaptive, fluid and open. This is also how personal projects and related information management have been characterised in this paper. "Mess" in the information space is here to stay and is a consequence of fine-grained routines, intersecting practices, dynamics of workflow and artefacts. It is also observable and understood by its users and it might not always be perceived as mess. The challenge posed is to design project information management tools in support of described practices, workflow, artefacts and routines. There are several implications derived from this study for project information management tools design that were mentioned in Discussion and are summarised below:

(1) Information unification (as a yet unsupported feature of present PIM tools) was perceived as very desirable by users. Several studies (including this one) have shown that people employ different practices in current PIM tools to support information need(s). For example, a study folder, code folder and paper folder might be on different parts of the file hierarchy as this makes sense to a user. Information unification should thus support such creative usage without interfering with current practices.

(2) Users have shown that reminiscing about the life of a project helps them to remember more about project-related information. Often users found valuable information that has been forgotten in the hierarchy. Showing the story (steps taken and changes made) of a project information space could improve context recreation.

(3) Throughout the project lifespan, participants focused on different information items that were perceived as important at the current project state. Project information management tools should provide a focus on a subset of information items while still providing a broader view of project information space.

(4) Sketches and interviews revealed tacit knowledge behind project information management that is not captured nor can be provided by users in current tools. For example, the role of information for a project at hand (input/output), information importance or value, information provenance, people associated with information items, etc. This knowledge could be gathered automatically (calculate importance based on time spent, file size changes, found due dates on related web pages, associate people base on files e-mailed, find associations through copy/past, find the role based on whether information changes or not, etc.), visualised to users and provide some basic project information modelling.

(5) There is more to observable navigation than just taking small steps towards target information. This study revealed at least two implicit or unobservable navigation techniques: navigating in time and navigating between projects through information overlaps. Supporting the former could help users in project context recreation while support for the latter would enhance understanding of the personal information space.

This study provides design implications for software we are currently building. It will allow users to organise together their project-related and fragmented information in a map like spatial interface, navigate in time and across projects by overlapping information, automatically model information importance and demote information (Bergman *et al.*, 2015) not needed at the point in time. We aim to release it as an open-source cross-platform software, and collect real-life longitudinal data about project information management to quantitatively confirm the findings of the present study and further the understanding of personal project information management. In addition, we plan to install it on computers of participants of the current study and conduct weekly short interviews to encourage them to use the software.

References

Abrams, D., Baecker, R. and Chignell, M. (1998), "Information archiving with bookmarks", in Karat, C.-M., Lund, A., Coutaz, J. and Karat, J. (Eds), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Press, New York, NY, pp. 41-48.

Altmann, E.M. and Trafton, J.G. (2002), "Memory for goals: an activation-based model", *Cognitive Science*, Vol. 26 No. 1, pp. 39-83.

Altmann, E.M., Trafton, J.G. and Hambrick, D.Z. (2014), "Momentary interruptions can derail the train of thought", *Journal of Experimental Psychology. General*, Vol. 143 No. 1, pp. 215-226.

Arcand-Dusseault, C. and Egan, M. (2015), "Occupations and personal projects: a comparison of the concepts", *British Journal of Occupational Therapy*, Vol. 78 No. 5, pp. 303-310.

Barr, A. and Feigenbaum, E.A. (1982), *The Handbook of Artificial Intelligence*, Vol. 2, Heuris Tech Press, William Kaufmann Inc., and Pitman, Stanford, CA, Los Altos, CA, and London.

Barreau, D. and Nardi, B.A. (1995), "Finding and reminding", *ACM SIGCHI Bulletin*, Vol. 27 No. 3, pp. 39-43.

Barreau, D.K. (1995), "Context as a factor in personal information management systems", *Journal of the American Society for Information Science*, Vol. 46 No. 5, pp. 327-339.

Bellotti, V., Dalal, B., Good, N., Flynn, P., Bobrow, D.G. and Ducheneaut, N. (2004), "What a to-do: studies of task management towards the design of a personal task list manager", in Elizabeth, D.-E. and Tscheligi, M. (Eds), *Proceedings of the Conference on Human Factors in Computing Systems – CHI '04*, ACM Press, New York, NY, pp. 735-742.

Bergman, O., Beyth-Marom, R. and Nachmias, R. (2006), "The project fragmentation problem in personal information management", in Grinter, R., Rodden, T., Aoki, P., Cutrell, E., Jeffries, R. and Olson, G. (Eds), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Press, New York, NY, pp. 271-274.

Bergman, O., Elyada, O., Dvir, N., Vaitzman, Y. and Ben Ami, A. (2015), "Spotting the latest version of a file with Old'nGray", *Interacting with Computers*, Vol. 27 No. 6, pp. 630-639.

Boardman, R. (2001), "Multiple hierarchies in user workspace", in Tremaine, M. (Ed.), *CHI '01 Extended Abstracts on Human Factors in Computing Systems – CHI '01*, ACM Press, New York, NY, pp. 403-404.

Boardman, R. (2004), *Improving Tool Support for Personal Information Management*, Imperial College London, University of London, London.

Boardman, R. and Sasse, M.A. (2004), " 'Stuff goes into the computer and doesn't come out': a cross-tool study of personal information management", in Elizabeth, D.-E. and Tscheligi, M. (Eds), *Proceedings of the 2004 Conference on Human Factors in Computing Systems – CHI '04*, ACM Press, New York, NY, pp. 583-590.

Bogunovich, P. and Salvucci, D. (2011), "The effects of time constraints on user behavior for deferrable interruptions", in Tan, D., Fitzpatrick, G., Gutwin, C., Begole, B. and Kellogg, W.A. (Eds), *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems – CHI '11*, ACM Press, New York, NY, pp. 3123-3126.

Bondarenko, O. (2006), "Task switching in detail: task identification, switching strategies and influencing factors", in Manolopoulos, Y., Filipe, J., Constantopoulos, P., Cordeiro, J., Papadopoulos, G. and Filipe, J. (Eds), *ICEIS '06: Doctoral Consortium at the International Conference on Enterprise Information Systems*, INSTICC, Lisboa, pp. 72-83.

Borst, J.P., Taatgen, N.A. and van Rijn, H. (2015), "What makes interruptions disruptive? A process-model account of the effects of the problem state bottleneck on task interruption and resumption", in Begole, B., Kim, J., Inkpen, K. and Woo, W. (Eds), *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems – CHI '15*, ACM Press, Vol. 1, New York, NY, pp. 2971-2980.

Bowker, G.C. and Star, S.L. (2000), *Sorting Things Out: Classification and its Consequences*, The MIT Press, Cambridge, MA.

Bruce, H., Wenning, A., Jones, E., Vinson, J. and Jones, W. (2011), "Seeking an ideal solution to the management of personal information collections", *Information Research*, Vol. 16 No. 1, p. 462, available at: http://InformationR.net/ir/16-1/paper462.html

Buehler, R., Griffin, D. and Ross, M. (2002), "Heuristics and biases: the causes and consequences of optimistic time predictions", in Gilovich, T., Griffin, D. and Kahneman, D. (Eds), *Heuristics and Biases: The Psychology of Intuitive Judgment*, Cambridge University Press, New York, NY, pp. 250-270.

Buzan, T. and Buzan, B. (1993), *The Mind Map Book How to Use Radiant Thinking to Maximize your Brains Untapped Potential*, 1st ed., BBC Books, London.

Byström, K. and Järvelin, K. (1995), "Task complexity affects information seeking and use", *Information Processing & Management*, Vol. 31 No. 2, pp. 191-213.

Carroll, J.M. (2007), "Learning, using and designing filenames and command paradigms", *Behaviour & Information Technology*, Vol. 1 No. 4, pp. 327-346.

Chau, D.H., Myers, B. and Faulring, A. (2008), "What to do when search fails", in Czerwinski, M., Lund, A. and Tan, D. (Eds), *Proceeding of the Twenty-Sixth Annual CHI Conference on Human Factors in Computing Systems – CHI '08*, ACM Press, New York, NY, pp. 999-1008.

Civan, A., Jones, W., Klasnja, P. and Bruce, H. (2009), "Better to organize personal information by folders or by tags?: The devil is in the details", *Proceedings of the American Society for Information Science and Technology – ASIST*, *Vol. 45 No. 1*, pp. 1-13.

Cockburn, A. and Gutwin, C. (2009), "A predictive model of human performance with scrolling and hierarchical lists", *Human-Computer Interaction*, Vol. 24 No. 3, pp. 273-314.

Cole, I. (1982), "Human aspects of office filing: implications for the electronic office", *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, HFES, Seattle, WA, pp. 59-63.

Czerwinski, M. and Horvitz, E. (2002), "An investigation of memory for daily computing events", in Faulkner, X., Finlay, J. and Détienne, F. (Eds), *Proceedings of HCI 2002. People and Computers XVI – Memorable Yet Invisible*, Springer London, London, pp. 229-245.

Czerwinski, M., Horvitz, E. and Wilhite, S. (2004), "A diary study of task switching and interruptions", in Elizabeth, D.-E. and Tscheligi, M. (Eds), *Proceedings of the 2004 Conference on Human Factors in Computing Systems – CHI '04*, ACM Press, New York, NY, pp. 175-182.

Dix, A., Wilkinson, J. and Ramduny, D. (1998), "Redefining organisational memory", in Fields, B. and Wright, P. (Eds), *Workshop on Understanding Work and Designing Artefacts*, British HCI Group, York, pp. 83-86.

Dourish, P. and Bell, G. (2011), *Divining a Digital Future: Mess and Mythology in Ubiquitous Computing*, Mit Press, Cambridge, MA, available at: http://mitpress.mit.edu/books/divining-digital-future

Drucker, P. (2009), *Landmarks of Tomorrow: A Report on the New Post-Modern World*, Transaction Publishers, Piscataway, NJ.

Gillie, T. and Broadbent, D. (1989), "What makes interruptions disruptive? A study of length, similarity, and complexity", *Psychological Research*, Vol. 50 No. 4, pp. 243-250.

Gonçalves, D. and Jorge, J.A. (2008), "In search of personal information: narrative-based interfaces", in Staab, S. (Ed.), *Proceedings of the International Conference on Intelligent User Interfaces – IUI '08,* ACM Press, New York, NY, pp. 179-188.

González, V.M. and Mark, G. (2004), "Constant, constant, multi-tasking craziness: managing multiple working spheres", in Dykstra-Erickson, E. and Tscheligi, M. (Eds), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Press, New York, NY, pp. 113-120.

Jensen, C., Lonsdale, H., Wynn, E., Cao, J., Slater, M. and Dietterich, T.G. (2010), "The life and times of files and information: a study of desktop provenance", in Mynatt, E., Fitzpatrick, G., Hudson, S., Edwards, K. and Rodden, T. (Eds), *Proceedings of the 28th International Conference on Human Factors in Computing Systems – CHI '10*, ACM Press, New York, NY, pp. 767-776.

Jones, W. (2007a), "Personal information management", *Annual Review of Information Science and Technology*, Vol. 41 No. 1, pp. 453-504.

Jones, W. (2007b), *Keeping Found Things Found: The Study and Practice of Personal Information Management*, 1st ed., Morgan Kaufman, Burlington, MA.

Jones, W. (2015), "Building a better world with our information: the future of personal information management, part 3", *Synthesis Lectures on Information Concepts, Retrieval, and Services*, Vol. 7 No. 4, pp. 1-203.

Jones, W., Bruce, H., Foxley, A. and Munat, C.F. (2007), "Planning personal projects and organizing personal information", *Proceedings of the American Society for Information Science and Technology – ASIST*, *Vol. 43 No. 1*, pp. 1-24.

Jones, W., Phuwanartnurak, A.J., Gill, R. and Bruce, H. (2005), "Don't take my folders away!: organizing personal information to get things done", in van der Veer, G. and Gale, C. (Eds), *CHI '05 Extended Abstracts on Human Factors in Computing Systems – CHI '05*, ACM Press, New York, NY, pp. 1505-1508.

Jones, W., Capra, R., Diekema, A., Teevan, J., Pérez-Quiñones, M., Dinneen, J.D. and Hemminger, B. (2015), " 'For Telling' the present: using the Delphi method to understand personal information management practices", in Begole, B., Kim, J., Inkpen, K. and Woo, W. (Eds), *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems – CHI '15*, *Vol. 1*, ACM Press, New York, NY, pp. 3513-3522.

Kamaruddin, A., Admodisastro, N. and Dix, A. (2013), "Before and after: user's knowledge maturity within personal information management", *International Journal of Scientific & Engineering Research*, Vol. 4 No. 5, pp. 165-170.

Karger, D.R., Bakshi, K., Huynh, D., Quan, D. and Sinha, V. (2005), "Haystack: a customizable general-purpose information management tool for end users of semistructured data", *CIDR '05: Proceedings of the Conference on Innovative Database Research, CDIR*, pp. 13-26.

Katidioti, I., Borst, J.P. and Taatgen, N.A. (2014), "What happens when we switch tasks: pupil dilation in multitasking", *Journal of Experimental Psychology: Applied*, Vol. 20 No. 4, pp. 380-396.

Kerzner, H. (2013), *Project Management: A Systems Approach to Planning, Scheduling, and Controlling*, Wiley, New York, NY.

Kidd, A. (1994), "The marks are on the knowledge worker", in Adelson, B., Dumais, S. and Olson, J. (Eds), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Press, New York, NY, pp. 186-191.

Kirsh, D. (2001), "The context of work", *Human-Computer Interaction*, Vol. 16 No. 2, pp. 305-322.

Kljun, M. and Dix, A.J. (2012), "Collaboration practices within personal information space", in Capra, R. and Teevan, J. (Eds), *Proceedings of the Workshop on Personal Information Management (PIM), Computer Supported Cooperative Work 2012, PIM Workshop*, *Seattle, WA, 11-15 February*, available at: http://pimworkshop.org/2012/pdf/kljun_2012_collaboration.pdf (accessed 15 November 2015).

Lansdale, M.W. (1988), "The psychology of personal information management", *Applied Ergonomics*, Vol. 19 No. 1, pp. 55-66.

Little, B.R. (2015), "The integrative challenge in personality science: personal projects as units of analysis", *Journal of Research in Personality*, Vol. 56, pp. 93-101, available at: www.sciencedirect.com/science/article/pii/S009265661400110X

Malone, T.W. (1983), "How do people organize their desks?: implications for the design of office information systems", *ACM Transactions on Information Systems*, Vol. 1 No. 1, pp. 99-112.

Mark, G., Gudith, D. and Klocke, U. (2008), "The cost of interrupted work: more speed and stress", in Czerwinski, M., Lund, A. and Tan, D. (Eds), *Proceeding of the Twenty-Sixth Annual CHI Conference on Human Factors in Computing Systems – CHI '08*, ACM Press, New York, NY, pp. 107-110.

Muniswamy-Reddy, K.K., Holland, D.A., Braun, U. and Seltzer, M. (2006), "Provenance-aware storage systems", in Adya, A. and Nahum, E. (Eds), *Proceedings of the USENIX Annual Technical Conference – USENIX '06*, USENIX Association, Boston, MA, pp. 43-56.

Nielsen, J. (2001), "First rule of usability? Don't listen to users", Nielsen Norman Group, Fremont, CA, August, available at: www.nngroup.com/articles/first-rule-of-usability-dont-listen-to-users/ (accessed 11 November 2014).

O'Day, V.L. and Jeffries, R. (1993), "Orienteering in an information landscape", in Ashlund, S., Henderson, A., Hollnagel, E., Mullet, K. and White, T. (Eds), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems – CHI '93*, ACM Press, New York, NY, pp. 438-445.

Obendorf, H., Weinreich, H., Herder, E. and Mayer, M. (2007), "Web page revisitation revisited: implications of a long-term click-stream study of browser usage", in Rosson, M.B. and Gilmore, D. (Eds), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems – CHI '07*, ACM Press, New York, NY, pp. 597-606.

Panko, R.R. (2009), "Managerial communication patterns", *Journal of Organizational Computing*, Vol. 2 No. 1, pp. 95-122.

Pezzo, M.V., Litman, J.A. and Pezzo, S.P. (2006), "On the distinction between yuppies and hippies: individual differences in prediction biases for planning future tasks", *Personality and Individual Differences*, Vol. 41 No. 7, pp. 1359-1371.

Ramduny-Ellis, D., Dix, A., Rayson, P., Onditi, V., Sommerville, I. and Ransom, J. (2005), "Artefacts as designed, artefacts as used: resources for uncovering activity dynamics", *Cognition, Technology & Work*, Vol. 7 No. 2, pp. 76-87.

Rasel, A.A. and Ali, M.E. (2016), "UProve2: privacy-aware, scalable, ubiquitous provenance to enhance file search", *2016 International Conference on Networking Systems and Security (NSysS), IEEE*, pp. 1-5.

Ravasio, P., Schär, S.G. and Krueger, H. (2004), "In pursuit of desktop evolution: user problems and practices with modern desktop systems", *ACM Transactions on Computer-Human Interaction*, Vol. 11 No. 2, pp. 156-180.

Scarr, J. (2012), "Supporting and exploiting spatial memory in user interfaces", *Foundations and Trends® in Human-Computer Interaction*, Vol. 6 No. 1, pp. 1-84.

Sellen, A.J. and Harper, R.H.R. (2001), *The Myth of the Paperless Office*, MIT press, Cambridge, MA.

Shah, S., Soules, C.A.N., Ganger, G.R. and Noble, B.D. (2007), "Using provenance to aid in personal file search", in Chase, J. and Seshan, S. (Eds), *USENIX '07: Proceedings of the USENIX Annual Technical Conference*, USENIX Association, Berkeley, CA, pp. 13:1-13:14.

Suchman, L. (2006), *Human-Machine Reconfigurations: Plans and Situated Actions*, 2nd ed., Cambridge University Press, Cambridge.

Teevan, J., Alvarado, C., Ackerman, M.S. and Karger, D.R. (2004), "The perfect search engine is not enough: a study of orienteering behaviour in directed search", in Elizabeth, D.-E. and Tscheligi, M. (Eds), *Proceedings of the 2004 Conference on Human Factors in Computing Systems – CHI '04*, ACM Press, New York, NY, pp. 415-422.

Vallacher, R.R. and Wegner, D.M. (1987), "What do people think they're doing? Action identification and human behavior", *Psychological Review*, Vol. 94 No. 1, pp. 3-15.

van der Lugt, R. (2005), "How sketching can affect the idea generation process in design group meetings", *Design Studies*, Vol. 26 No. 2, pp. 101-122.

Voida, A., Harmon, E. and Al-Ani, B. (2011), "Homebrew databases: complexities of everyday information management in nonprofit organizations", in Jones, M.J. and Palanque, P. (Eds), *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems – CHI '11*, ACM Press, New York, NY, pp. 915-924.

Whittaker, S. (2011), "Personal information management: from information consumption to curation", *Annual Review of Information Science and Technology*, Vol. 45 No. 1, pp. 1-62.

Whittaker, S. and Hirschberg, J. (2001), "The character, value, and management of personal paper archives", *ACM Transactions on Computer-Human Interaction*, Vol. 8 No. 2, pp. 150-170.

Whittaker, S., Matthews, T., Cerruti, J., Badenes, H. and Tang, J. (2011), "Am I wasting my time organizing email?: a study of email refinding", in Jones, M.J. and Palanque, P. (Eds), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Press, New York, NY, pp. 3449-3459.

Corresponding author
Matjaž Kljun can be contacted at: matjaz.kljun@upr.si