Personal Information Systems: User Views and Information Categorization

[*] Dominique L. Scapin, [*] Pascal Marie-Dessoude, [#] Marco A. Winckler, and [*] Claudia Detraux
[(*)] INRIA, Rocquencourt, France
[(#)] IRIT, Toulouse, France
{dominique.scapin, pascal.marie-dessoude, claudia.detraux}@inria.fr, winckler@irit.fr

Abstract- This paper aims to improve Personal Information systems (PIMs) by understanding how people manage personal information items usually kept on notes, cards, agendas, etc., and on administration forms (paper or digital). The focus is both on what people say about information content, organization, trust, willingness to share and on how people categorize information. Preliminary studies (focus group and questionnaire) looked at how people describe their own use of information, and their views on future PIMs needs. They show a strong distrust towards such systems and reluctance to share personal information. Another study (card-sorting) looks at the way people assign individual information items to self-created categories. Results show a few variations in structure and naming, with a gender effect for category size. Detailed clustering and co-occurrence analyses show small differences between how people actually organize their personal information and our initial "theoretical" assignment. While the results suggest some modifications of the information structure and content, it supports the user-centric approach of the study, starting from user needs and associated documents, experimental testing and design iterations, which could be generalized for designing usable PIMs.

Keywords - personal information items, PIMs, semantic categories, naming, e-gov.

I. INTRODUCTION

Conducted within project PIMI (Personal Information Management through Internet), which goal is to develop a design environment and a deployment platform providing users with personal data access and services relevant to their needs, this study aims at gaining knowledge about the way users manage their information and services, how they see doing it in the future, and what should be the information content, structure and naming in a PIMI. Also, which items can be shared and other issues such as security and trust.

In recent years, computing technology (including Internet and mobiles) has increased capabilities for managing the large information sets needed in everyday life, professional or non-professional. Part of that information refers to personal data that users might decide to share or not through their relationships with other users (e.g., social networks) and applications (e.g., e-government services). Stimulated by the recent evolution of national governments policies towards the improvement of electronic services, e-government applications tend to require personal information for accessing public e-services such as in healthcare, taxes, housing, agriculture, education, social services [1].

Managing large sets of information is strongly related to the domain of Personal Information Management systems (PIMs), which corresponds [2] [3] to the research field addressing the way people manage their physical documents (books, notebooks, sheets, etc.), as well as their electronic documents (files, emails, Web pages, etc.), with the aim of designing tools that support the management of electronic documents (PIM tools). PIMs studies have mostly focused on very large data sets, such as the full content of user hard drive, and on search issues pointing out the large variability in people information search [3].

While the PIMs area usually covers many contexts and activities, in this paper we look at PIMs in a more specific way: the individual information items people keep on various notes, cards, forms, agendas, etc., the ones that are personally attached to us, that we use in our every daily life, both professional and non-professional (for administrative, social, leisure purposes, etc.). It is not the full set of available files. Besides, we look at the intuitive way people organize their personal information, with or without computer systems. Concerning personal information bits currently scattered many places, there is little research with a user-centric approach, with the view that users-based knowledge might help specifying computer-based tools.

This paper starts with a review of literature and publicly available tools. Then, preliminary studies briefly report on documents analyses and on what people say and wish about their personal information. A section describes the method, tool, procedure and participants of a card-sorting study on people intuitive organization of personal information. The results and their impact towards a future PIMI structure are presented. The conclusion summarizes the study, identifies its limits and provides insight on future research work.

II. RELATED WORK

The context is e-gov. (short for electronic government), a diffused neologism used to refer to the use of information and communication technology to provide and improve government services, transactions and interactions with citizens, businesses, and other arms of government [4].

In the area of e-gov., a literature review [5] showed little specifically on human factors in HCI (Human-Computer Interaction). Studies identified dealt mainly with user needs and accessibility (a major topic in e-gov. HCI, including studies on older people), the applicability of HCI results to e-gov., ad hoc interaction novelties (e.g., animated faces), ad hoc methods (e.g., on document exchange and scenario

 ISBN: 978-1-61208-167-0

planning), issues of user involvement and requirements, user acceptance, and patterns.

About PIMs, a more substantial body of knowledge in many different settings has begun to pile up in recent years [6] [7]. There is a lot of technical aspects such as: data synchronization across devices, version control [8], file management and applications [9], collective work and file sharing [10], novel user interface paradigms and mobility [11] [12], ontologies [13], and tools based on information association [14] [15].

There are also usability, citizen-centric studies, mainly about: privacy and security [16], hierarchical files structure issues and proposal for a tagging mechanism [17], studies following-up on [3, op. cit.], such as: further empirical investigation on ways to improve information searching [18], investigation of the role of personal notes [19], contextual use of PIMs [20], tool evaluation [21], and call for more user-centered studies, long-term studies on the evolution of user information practice from one work context to another, from a role to another [22]. In light of our goals, a few points can be selected from these studies.

One point is that hierarchical structure is the most used and preferred by users [23] [24] [25] [26]. The latter study actually shows that users built an ownership and control feeling about their data, probably due to long use of such tools. However, users have difficulties in creating consistent structures and naming items categories. More specifically, categorizing and naming new items in an existing structure seems to be difficult and represent a high cognitive load [24, op. cit.], [25, op. cit.]. Adding contextual data, such as tags, may help, but tagging may vary from one user to the other, and will not solve consistency issues, particularly when information spaces are to be shared, even partly. An interesting addition to contextual data from sensors (GPS, GSM, and movement) has shown to help find images within a collection [27].

About search strategies, users tend to first explore the structure, and use search tools only afterwards. Even though it may be explained by lack of user knowledge [26, op. cit.], the lack of flexibility of these tools may still apply [3, op. cit.].

One particularly pervasive PIM problem is information fragmentation [28], i.e., when information related to a single task is scattered across several different applications and environments. A typical example is project information, where specifications may be in a Word document, budget in an Excel file, communication with the customer and the project manager may be in emails, and other resources may reside in Intranet or the Web. A project member may have seen all these documents but may later have trouble re-accessing them. In [28, op. cit.] it is argued that grouping related information is a central PIM activity currently hindered by the artificial separation imposed by the different applications. In addition, history and versioning must be dealt with.

On the practical side, [28, op. cit.] suggest 5 factors that may hinder PIMs use: visibility, integration, co-adoption, scalability, return on investment. In absence of visibility, when the PIM is not always visible to the user, the tendency is to forget it, as well as the data already stored. Integration: when not integrated with the other tools, it can be underused. Co-adoption: for user cooperation, share and synchronization of data (e.g., appointments, agenda) is required, if not available the PIM might not be used. Scalability: the PIM tool must allow scaling (e.g., more data, projects). Return on investment: if the tool requires a large learning effort, it will not be used. Guidelines are also offered by [26, op. cit.], along three types of strategies: piling, filing, and structuring.

Concerning information transfer, results of a survey with 47 participants [30] showed that the main forms of PIM storage are computers, then external disk drives. For Web sites, the ordered storage preferences are bookmarks, email, paper. When data transfer, it is done mainly with email and memory sticks, and in some cases on web sites. The main difficulty is finding the files. The results also point out the important role of email systems in storage, sharing, search, and file exchange between computers and other devices.

In our research, the attempt is to complement these earlier findings, first on what people say (in terms of information items, of current practice, of shareability, etc.), but also focus on a novel issue: what people do intuitively with an existing set of unstructured information items.

We also looked at 15 tools [31] that claim to support personal information management. Most tools (Tools # 1, 2, 3, 4, 5, 7, 8, 11, 12, 13, 15) offer an agenda, a calendar, a contact list, a keyword based search tool, a centralized password management function, and notes editing. In addition, a few tools offer more sophisticated functions such as: a sort of mind-mapping allowing to represent user's thoughts (Tool # 2), a text-based card information management system (Tool # 6), a system for managing archives (Tool # 9), a document and pictures storage system dedicated to Android-based mobile phones (Tool # 14), and a note management system that coordinates notes (containing files) between (Mac, PC, mobiles) platforms (Tool # 10). Few of these tools are available on line (e.g., Tools # 7, 9, 15). Most others run on personal computers, except for Tool # 14 that runs only on mobiles and Tool # 10 that runs on personal computer, internet and mobile. Synchronization is provided by Tools # 10 and 15.

Most user interfaces are rather similar. Some allow tailoring, mainly on colors, window size, language, menu position. A few points concern usability issues. For instance, concepts may be difficult to grasp due to naming (e.g., Tools # 1, 4, 7). In general, no predefined structures, or default patterns are included, or are unusable for new user entries, which might make it difficult for a large public. Also, the coverage of tools, whenever items and categories are proposed, does not include many of the citizens' information such as identities, health, finance, work, etc. Some of the associated information search tools are powerful (e.g., Tools # 1, 2, 7), but a bit cumbersome (many forms and choices). Some tools offer also tags, labels, particularly for notes, contacts and events (e.g., Tools # 1, 2, 3, 6, 10, 13, 15).

Overall, useful insight can be extracted from the references in this related work section. However, few studies concern the information content of PIMs, and very few include a full user-centric process approach, such as

 ISBN: 978-1-61208-167-0

proposed in current standards, particularly ergonomics standards.

III. PRELIMINARY STUDIES: WHAT DO PEOPLE SAY ABOUT PERSONAL INFORMATION

An initial step has been the analysis of 9 administrative forms relating to international actions, students grants, solar heating incentives, livestock diseases, sport incentives, etc. available from French government and administration.

The analysis showed a very large diversity of information items (500 + 200 synonyms). 21 different topics were identified. This information content was classified and later supported the design of questionnaires and card-sorting material.

A. Focus groups

With three focus groups (a group with 6 participants from a large industrial group, and two 3rd. year university students groups, with respectively 7 and 9 participants), a study investigated the issue of electronic information storage for personal information, as well as the issue of shareability.

Besides providing a number of candidate information items for a personal information space, the main issues covered concerned: a strong distrust of electronic storage of personal information (hacking, piracy), the time needed to proceed to a single electronic storage of personal information, and their reluctance to share personal information.

B. Online survey

A questionnaire survey investigated current practice along the same issues: electronic personal information storage, shareability, etc, with the addition of people's view on their future personal information space. A self-administered questionnaire was available on Internet [32] using the tool SurveyMonkey [33]. The survey was voluntarily focused on personal information used regularly by the public.

The survey was answered by 30 participants: 14 clerks and 16 3rd. year university students. 29 respondents were women.

First, users were asked about their profile and how they dealt currently with their personal information.

Secondly, they were presented with a set of 114 personal information items organized into 9 categories and 26 sub-categories. Figure 1 shows the "theoretical" category structure, a pre-defined structure based on preliminary studies. Categories are organized in a horizontal menu whilst items within a category in the vertical menus options below.

For each sub-category users were asked to identify information items that should belong, provide alternatives names if unsatisfactory, express willingness to share that information with others, and tell if each sub-category should be part of a personal information space. Thirdly, users browsed all items, and were asked to point any category that could be missing and to comment on PIMs advantages/drawbacks, potential uses, willingness to share their PIMS with administrations, views on tools for automatic filling electronic forms.

MY PERSONAL INFORMATION SPACE								
My Identification	My Family	My Health	My Professional Activities	My Transportation means	My Finances	My Logins & Passwords	My Agenda	My Contacts
My Identity	My Family Status	My Health Coverage	My Career	My Personal transportation	My Bank	My IDs, Logins and Passwords	My Personal Agenda	My Personal contacts
My Contact Information	My Parents	My Physician	My Current Job	My Public Transportation	My Income and Social Benefits		My Professional Agenda	My Professional Contacts
My Identity documents	My Spouse / Partner	My Medical Records			My Investments			
My Biometric Data	My Children				My Loans			
					My Wealth			
					My Income Taxes			

Figure 1. Structure of the "theoretical" personal information space

Due to space, results on current and future use are not reported here, except to mention a strong distrust for such systems and reluctance to share information. Focusing on the categorization and naming issues, the results support the study material. Regarding information items categories, the participants tend to agree (79,2%) with the categories offered. For the remaining 20,8% the few categories suggested are: Hobbies", "Insurance", "Music", "Sports", "Union activities", as well as "Spending". Regarding naming, the participants did not make much suggestion. All items are well accepted, except 5 (out of 114) for which proposals are made.

IV. HOW DO PEOPLE ORGANIZE THEIR PERSONAL INFORMATION ITEMS

Another study focused on how participants organize information items. They were asked to create their own "boxes", i.e., information categories in which to insert their personal information items, using a Card-Sorting technique.

A. Material, procedure, participants

Card-Sorting is a way of gaining insight on categorization and mental models about information architecture that can be described by means of small cards [34]. It starts with writing each statement on the information architecture on a small card. Then, participants are asked to sort a set of cards with words or pictures into piles of similar cards. Participants may be asked to provide labels for the card piles they have created or they may be provided with pre-defined labels and asked to match the cards to them.

The Card Sorting tool: currently several card-sorting tools are available for enabling users to classify items on a computer instead of using paper cards. Most tools run under Web platforms, allowing card-sorting studies to be administered remotely. The results provided by online tools are quite similar to studies run using paper-based cards [35]. For this study, we used the tool WebSort [36]. The tool was initialized with our 114 personal information items (in alphabetical order, random assignment not being manageable with the tool). Users moved items from the list presented at the left side to the right area, creating groups of items they named as categories.

Material: 114 personal information items focusing on everyday life aspects involving health, banking, social welfare, citizenship administrative papers, etc.

 ISBN: 978-1-61208-167-0

Procedure: a call for participation was distributed through web sites and email lists. The study material was accessible from an Internet address. After a short definition of "Personal Information", of "Personal Information Space", and a description of the study goals, and tool functions (creating boxes, naming, renaming, etc.), participants were instructed to run their individual session at their own pace. Once the session was completed and saved, the participants were asked to fill a questionnaire (participants' profile, current ways of managing your personal information, views/ suggestions on a future personal information space).

Participants were recruited through various professional social networks and email lists. Aside 6 initial answers used for pre-testing, and 3 beta-testing answers, 56 participants responded to the card-sorting study. Due to incomplete answers (from 5 to 112 unsorted items, or no answer to the questionnaire), 13 participants were excluded from data analysis.

The characteristics of the remaining 43 are: 32 male participants (74%) and 11 female (26%), 33 in the 18-39 y. age bracket (77%), 9 in the 40-59 y. (21 %) and 1 in the 60-74 y. (2%), 9 are single (21 %) while the 34 others are not (78%) [18 married and 16 shared living], 14 participants have children (33%), while 29 do not (67%), 31 participants are employees (72%), 5 self-employed (12%), 4 students (9%), 2 unemployed (5%) and 1 retired (2%), 41 from France (95%), 1 USA and 1 Belgium.

B. Card sorting results

The average time for a full session (card-sorting and questionnaire) was 42 minutes: for the Card-Sorting part only, average duration was 35 minutes.

The next sections describe the categories (from now on called "boxes") created by the participants, their size and content, variations in naming, clustering aspects, and the analysis of the role of the participants' characteristics.

Boxes created: overall, 43 participants each classified 114 information items, for a total of 4802 items. Together, 500 boxes were created. Each participant created from 5 to 22 boxes (mean= 11.627, sd= 4.434, median= 11.000). In each box, they included 1 to 53 items per box (mean= 9.739, sd= 7.291), i.e., 34 different sizes.

Four sizes have been distinguished: “Very Large Size Boxes” > 21 items (and < 1 to 6 > boxes that size), “Large Size Boxes” < 9 to 21 items > (and < 6 to 28 > boxes that size), “Midsize Boxes” < 3 to 8 items > (and < 34 to 50 > boxes that size - *most numerous*), and “Small Size Boxes” < 3 items. Most topics are distributed across boxes, big, mid-size, and small, without much recognizable patterns.

These variations are further illustrated Figure 2 showing both (ordinate) size of boxes (i.e., number of items in boxes) and (abscissa) number of boxes each size (i.e., equal number of items in a box).

One can see that there is a rather regular parallel increase/decrease for the boxes size/number of items, respectively, until 25 items per box, while from 27 items, the curve gets more erratic for the number of large size boxes.

The next section looks more closely at the boxes content.

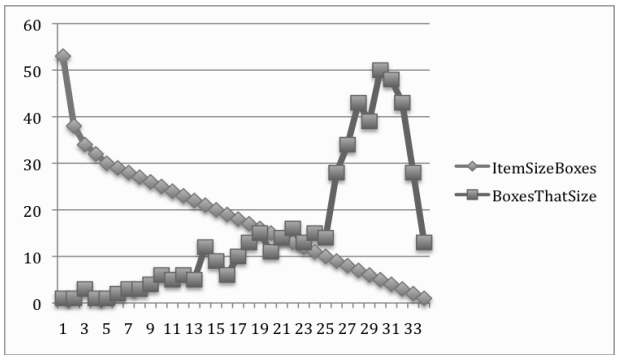

Figure 2. Curve # items per box X # boxes of each size

Lexical variations: naming of boxes will be later analyzed, particularly to help define naming of categories in the future PIMI tool. We simply mention here a few elements of variation.

- Imprecise naming, e.g., *later, confidential, others.*
- Typographic variations: caps/ lowercase, capitalized, with/ without accents, typos.
- With possessive or not, e.g., *(My)health, (My)family.*
- Singular vs. plural, e.g., *finance(s), address book (s).*
- Syntax: verb or noun or adjective, e.g., *administrative, administration.*
- Synonyms and abbreviations versus names.
- A quite large area of variation concerns the use of several terms together covering different aspects.

Conceptual variations: besides the fact that smallest size boxes cover more specific concepts than the larger ones, we looked at their scope and meaning. We distinguished Single concepts, Multiple Concepts (Explicitly Grouped by Name), and Higher-Level Concepts. The way boxes group items is quite variable in terms of concept scope, which confirms the analysis on boxes sizes.

These linguistic and conceptual variations have given some hints on default names for a future editable PIM, but also point at conceptual associations that are discussed in the clustering analyses. For instance, the frequency of identical "concept names" is quite high for boxes such as "health", "bank", "professional life".

Clustering analysis: a multiple correspondence analysis (MCA), using the SPAD tool [37] was performed with 9 classes (as in theoretical structure). Multiple Correspondence Analysis (MCA) is a method for nominal or categorical data sets (e.g., [38] [39]). The goal is to visualize a data set by representing data as points in a low-dimensional Euclidean space. This procedure is similar to principal component analysis for categorical data. MCA is also an extension of simple correspondence analysis (CA) in that it is applicable to a large set of categorical variables. In this case MCA is a CA on the Burt table formed from these variables. As in factor analysis methods, the first axis is the most important dimension, the second axis the second most important, and so on. The number of axes to be retained for analysis is determined by calculating the eigenvalues (a set of scalars associated with a linear system of equations, i.e., a matrix equation, sometimes also known as characteristic roots or

 ISBN: 978-1-61208-167-0

values, proper values, or latent roots). In our application we have retained 20 axes. On this low-dimensional Euclidian space we applied a hierarchical method. Hierarchical algorithms find successive clusters using previously clusters. This algorithm chosen is an agglomerative ("bottom-up") algorithm. This algorithm begins with each element as a separate cluster and merges them into successively larger clusters. An important step in most clustering is to select a distance measure, which will determine how the similarity of two elements is calculated. We have selected the Euclidian distance on the 20 factors selected by the MCA and the aggregative criterion used is the Ward's criterion [40].

Using such hierarchy in the Ward sense, 9 classes were detected. It shows overall a rather consistent category assignment to items, but some differences with the "theoretical assignment". The initial clustering (see Table I) started by leaving out 34 individual items, while establishing 33 initial groups (called level 1 clusters). They vary in item numbers content: 24 groups with 2 items, 7 with 3 items, and only 1 with 5 or 6 items.

TABLE I. NUMBER OF ITEMS IN INITIAL CLUSTERS

# Items	**1**	**2**	**3**	**4**	**5**	**6**	**Total**	
# Groupings		24	7	0	1	1	**34** single	**33** grouped
# Total Items	**34**	48	21	0	5	6	**114** items (80 grouped)	

Distribution of the number of items per level is shown Figure 3. Clustering varies little up to level 5 in terms of items numbers. There is a drop for levels 6 to 9 that actually tend to group item clusters from lower levels.

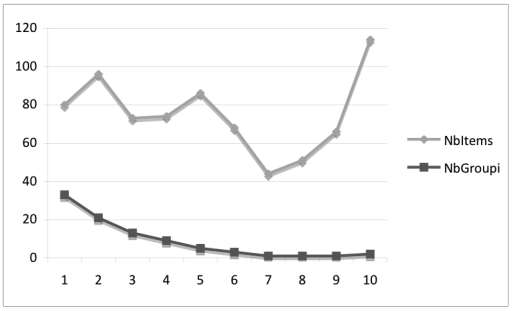

Figure 3. Curve items per level

To detail results, the 9 classes are seen at the incremental level, from "leaves" to "root" in the SPAD dendrogram. The clustering level refers to the rank at which the information items are grouped together in the clustering analysis: each time individual items or clusters are grouped, it adds a level (incrementing by 1 the previous highest level). Otherwise said, the more levels within a class, the more variations in boxes assignment. It is a hint on coherence of certain groups compared to others (e.g., many levels within the finance area, while few levels in the taxes, or health areas).

Looking more closely at the tree, an illustration is provided Figure 4 for non-health clusters and Figure 5 for health clusters. All grouped items are "boxed", individual items are not. All nodes show a level (L1 to 10). The lowest level is L1 (minimum items is 2, listed with "+" to associate them). All level 1 nodes show in addition a parenthesis referring to the level rank as used in the explanatory text below. The 9 SPAD classes are identified with "★ #" (8 non-health, 1 health). The main observations are the following.

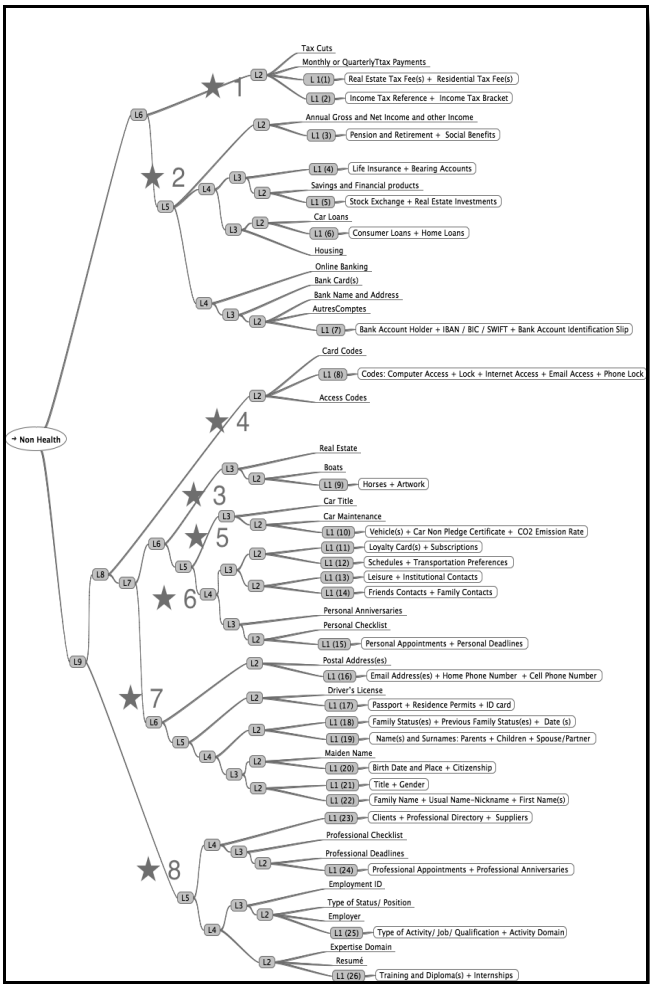

Figure 4. Non-Health Clustering Tree

A **first** class of 6 items corresponds to taxes. It is a quite consistent group, not been found together with other financial considerations. This suggests it is a specific concept for the participants.

A **second** class of 19 items deals with other financial aspects: bank (7 items), income (3 items), investments (4 items), loans (4 items), and savings (1 item). This actually contradicts the theoretical assignment by excluding patrimonial aspects.

A **third** class of 4 items relate to a "patrimonial" category, distinct from other financial aspects. Unlike theoretically, monetary savings do not belong, nor vehicles meant as collectible, which may not have been clear.

A **fourth** class of 7 items groups all codes. This cluster is distinct from the other groupings: it is only grouped with other clusters at a much higher level.

 ISBN: 978-1-61208-167-0

A **fifth** class of 5 items groups information related to car information (4 items) to which a *Vehicle* item was added (meant to be vintage/ collectible cars, but not clearly expressed in card sorting explanations). This corresponds to the theoretical "My Personal Transports" with the addition of *Vehicle*, but without *Driver's License*. This is a case of potential redundancy within a PIM, a driver's license being at the same time a vehicle, and a proof of identity.

A **sixth** class of 12 items groups a more heterogeneous set dealing with: public transportation (4 items), personal dates and appointments (4 items), and leisure (3 items), adding *Institutional contacts*, probably viewed as the ones needed in every day life. Overall, this cluster shows a large number of different clustering levels, which makes it a bit unstable compared to others. Again, it may mean that some information items should be made redundant, depending on their context of use.

A **seventh** class of 22 items groups various aspects of people identification: personal identity (8 items), contact information (4 items), identity documents (3 items), family status (3 items), information on relatives (3 items), and *Driver's License,* which seems to be considered as part of a person's identity. Also, it does not include *Health records,* which was in the theoretical assignment, but meant for children health.

An **eight** class of 16 items groups all professional information: career (4 items), job (5 items), but also professional agenda (4 items), professional/clients contacts (3 items). Participants tended to group all professional items, rather than distinguishing documents/ status items and contacts aspects.

Finally, a **ninth** class of 23 items joins all health aspects, whether biometric data (5 items), or health analyses (12 items), or administrative and health contacts (5 items), and children *Health Record*.

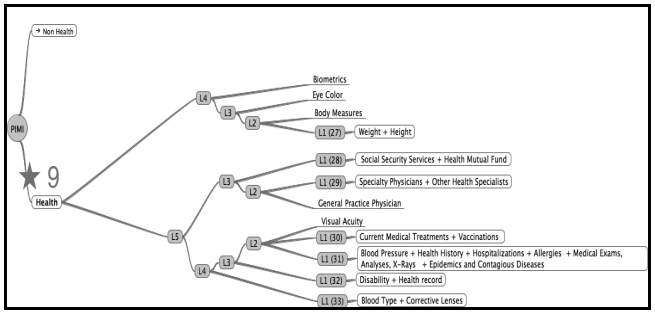

Figure 5. Health Clustering Tree

Co-occurrence of items in boxes: the analysis searched all cases ("frequent sets") for which at least 80% of the participants (i.e., at least 35) put a set of items into the same box. This resulted in 1041 sub-sets of items. To reduce that large collection, we looked at "closed frequent sets" within these "frequent sets". A "closed frequent set" is a set for which all its sub-sets are frequent, but which is not itself contained in another "frequent set". By topics, the items were grouped from largest to smallest number of participants within largest to smallest sets within a topic. The maximum of items found together is 9, minimum being 2. The analysis is consistent with previous clustering analyses. It also shows that the primary topics under which information items are being put together on a regular basis (highest co-occurrence topics, ranked first by number of participants, and within, by number of items) relate to health (13), bank/ finance (12), and identity (12), followed by work (7) and codes (6), the last ones being taxes (2), loans (1), telephone (1), agenda (1).

Differences due to participants' characteristics: the focus is here, per participant, on the number of boxes, the size of the smallest boxes, the average size of the boxes, as well as their maximum size (see Table II).

TABLE II. FISHER TEST AND ANOVA ON GENDER

VARIANCE ANALYSIS :				(Fisher) Student's T		ANOVA				
	Value	Average	Std-dev	Test		Variance decomposition		Significance level		
Nb_Boxes	W	13,8182	**5,3632**	Fisher **T**	1,9073	Source	Sum of square	Statistics	Value	Proba
	M	10,8750	**3,8834**	**df**	10/31	BSS	70,9101			
	All	11,6279	4,4348	**p-value**	0,1649	WSS	755,1364	Fisher's **F**	3,850054	**0,056556**
						TSS	826,0465			
Min_ite ms	W	**2,6364**	1,6293	Fisher **T**	5,8235	**Source**	**Sum of square**	**Statistics**	**Value**	**Proba**
	M	**4,3438**	3,9318	**df**	31/10	BSS	23,8637			
	All	3,9070	3,5511	**p-value**	**0,0056**	WSS	505,7642	Fisher's **F**	1,934522	0,171768
						TSS	529,6279			
Avg_ite ms	W	9,5776	3,8759	Fisher **T**	1,4673	**Source**	**Sum of square**	**Statistics**	**Value**	**Proba**
	M	11,9877	4,6950	**df**	31/10	BSS	47,5487			
	All	11,3712	4,5802	**p-value**	0,5339	WSS	833,5487	Fisher's **F**	2,338791	0,133867
						TSS	881,0973			
Max_i	W	23,3636	7,7236	Fisher **T**	1,0351	**Source**	**Sum of square**	**Statistics**	**Value**	**Proba**
	M	23,1563	7,8581	**df**	31/10	BSS	0,3521			
	All	23,2093	7,7323	**p-value**	1,0180	WSS	2510,7642	Fisher's **F**	0,005749	0,939928
						TSS	2511,1163			

Out of the participants' characteristics (gender, age, marital status, children, activity, location), only gender showed a significant role. It mainly means that women create more boxes than men, with a higher variability (Anova F, $p < 0.0565$, women std. dev. 5.3632, men std. dev. 3.8834). However, women tend to create less small boxes (Student T, $p < 0.0056$), women average 2.6364, while men average 4.3438. Otherwise said, women tend to use more boxes and to vary more in their size and content.

C. *New information category structure*

Results showed that our items list, initial classification, and labels were quite satisfactory for participants.

However, the analyses showed some discrepancies suggesting a new candidate structure along: Identity & Contacts (personal Identity, Identity papers, personal Contacts), Work (current Work, Career, professional Contacts), Agenda, Contacts & Transports (personal Agenda, Contacts, individual Transports, public Transports), Codes & Passwords, Finances (Income and social benefits, Investments, Loans, Bank accounts), Taxes, and Health (Social security & Mutual Funds, physicians, and medical Records).

Compared to the previous structure, there are 7 categories instead of 9. "My Family" is removed. "My Agenda", "My Contacts", and "My Transportation" are grouped, and a specific category "Taxes" is created. Agenda may be questionnable as distinguishing personal versus professional may not make sense if a future tool includes a common calendar. In addition, categories and sub-categories were reordered according to expected frequency and initial tool set-up order. Information items were reduced based on assignment stability, but more pragmatically on expected usefulness (in a new experiment, young students, may not need in their PIMI tool topics such as patrimony or children health). In terms of naming, the use of the possessive will not be pursued and a number of category and information names

 ISBN: 978-1-61208-167-0

will be changed to reflect the most frequent name generation by the study participants. Flexibility is another issue of prime importance. Despite any efforts at designing a good structure, good naming, there is probably no single solution: a future PIMI tool probably should offer a user tailorable structure and content, as well as flexible search features.

D. Questionnaire results

The participants input to the questionnaire after the card-sorting exercise is quite rich, and will need further qualitative analyses. The main quantitative results, rather consistent with the previous survey, are as follows:

- 86.5% participants use both paper and electronic storage, while 7.7% only electronic and 5.8% only paper. 91.7% use a personal computer, 27.1% a cellular phone, 10.4% an iPhone, 8.3% a smartphone, 8.3% a PDA, and 18.8% other means.
- Only 30.6% use some information protection method for their personal information while 69.4% do not.
- 46.9% currently share some of their personal information, while 53.1% do not. They share it with spouse (62.5%), administration (37.5%), social networks (34.4%), family (31.3%), friends (31.3%), employer (21.9%), colleagues (18.8%), and 9.4% others. The topics for which they show reluctance to share are codes, health-related information, and income.
- 70.6% organize their personal information into categories, while 29.4% do not. When stated, the categories correspond either to the ones proposed or to their naming of the card-sorting boxes, with the addition of a few categories such as “bills”, "bicycle", "religion", "politics".
- For a future information space, 65.2% did not offer new categories, while 34.8% added a few new ones, e.g., friends birthday, food recipes, computer IP, bills, music, photos and videos.
- 71.4% see both advantages and drawbacks in a future information space, while 22.8% see only advantages, and 5.7% only drawbacks. Those mainly concern trust and security, while advantages concern centralization and ease of access. Regarding expected functionalities, the top ones are searching, storage, filtering, and accessibility.
- 72.3% are worried about sharing their information with administrations, while 27.7% are not. The main concerns, relate again to trust and security.
- 55% are in favor of automatic form-filling, while 42.5% have mixed feelings, and 2.5% are against it. The main benefit mentioned is time saved, while the concern relates to strict selection of information to be auto-filled (security and confidentiality issues).

V. CONCLUSION

This paper reported novel user-centric work in the area of PIMS and e.gov., focusing on the personal information bits currently scattered many places, electronic or paper, rather that on disk file management issues. The methods used attempted to identified what people say and what people do about their personal information. These methods have limits related to remote experiments through Internet, to the card-sorting tool (management of redundancies, sub-categories, monitoring participants’ modification strategies).

However, the results support the user-centric approach of the study, starting from user needs and associated documents, experimental testing and design iterations, which could be generalized for designing usable PIMs. On the practical side, the results also suggest a few modifications of the information structure and content to be further tested in the PIMI project.

The design of tools for supporting citizen to use and share personal information is a complex task. Some of the issues are trust and willingness to share information which is of prime importance to users (even though many of them have lesser concerns when posting very private material on Facebook or other social networks), coping with users behavior variability and preferences, setting up proper procedures for information exchange in e-gov. contexts, reducing the information fragmentation issue, such as making sure different tools and environments allow consistent and synchronized use resources, and providing efficient search tools, with queries adapted to the users.

In our research, the next step is to test a mockup system based on the previous findings. The mockup will be tested in depth, with users being monitored, with tasks to perform, with usability measurements, both objective and subjective. One underlying idea is to explore how people can actually tailor their own information space, in terms of structure and naming, but also in terms of sharing parameters, and associated (flexible and contextualized) search tools.

Later on, the current static view (i.e., about content, information items and categories) will be extended to a dynamic view that will include procedures (i.e., using personal information items to manage one's personal space and local information transfers, as well as to fill manually or automatically e-gov. forms). This will concern service composition, building ontologies, associating information items related documents (e.g., file copy of driver's license, of administrative documents).

An additional goal will be to formalize the design process for delivering validated new e-gov. content: context of use study (e.g., documents), user needs gathering (ideas through focus groups, facts and opinions through questionnaires), concepts definition (design step), subjective testing (through questionnaires), content assessment (through card-sorting and questionnaire), prototype design and user testing.

Hopefully, in the end, this will contribute to more citizen-centric personal information systems.

VI. ACKNOWLEDGMENTS:

Thanks to Y. Lechevallier for his expert contributions and advice on statistical analyses, as well as for the numerous stimulating discussions. Preliminary studies have been funded by the ANR project "MyCitizSpace". The card-sorting study has been funded by the ANR project "PIMI".

 ISBN: 978-1-61208-167-0

VII. REFERENCES

[1] Calvary, G., Serna, A., Coutaz, J., Scapin, D. L., Pontico, F., and Winckler, M.. "Envisioning advanced user interfaces for e-government applications: a case study". In Practical Studies in E-Government: Best Practices from Around the World. Assar, S., 2010, Chapter 12, pp. 205-228.

[2] Bergman, O., Boardman, R., Gwizdka, and J., Jones, W. "Personal information management". SIG at CHI 2004, April 24–29, 2004, Vienna, Austria, pp. 1598-1599.

[3] Blanc-Brude, T. and Scapin, D. L. "What do people recall about their documents? Implications for desktop search tools", Proc. IUI 2007, Honololu, HI, USA, January 28-31 2007. pp. 102-111.

[4] http://en.wikipedia.org/wiki/E-Government

[5] Scapin, D. L. "E-government HCI: a genuine research field?" Proc. DEGAS'2009, In conjunction with INTERACT'2009, August 24, 2009, Uppsala, Sweden, pp. 33-37.

[6] Teevan, J., Jones, W., and Bederson, B.. "Special issue: personal information management". Communications of ACM 49, 1, January 2006, pp.40-43.

[7] Jones, W.and Teevan, J. "Personal information management", 2007, Seattle, WA: University of Washington Press. Jones, W., Teevan, J. (Editors), ISBN 978-0-295-98737-8.

[8] Tungare, M., Pyla, P. S., Pérez-Quiñones M,. and Harrison, S. "Personal information ecosystems and implications for design", Technical Report cs/0612081, ACM Computing Research Repository, 2006.

[9] Sauermann, L., Grimnes, G. A., and Roth-Berghofer, T. "The semantic desktop as a foundation for PIM research". PIM workshop CHI 2008, April 5-6, 2008, Florence, Italy.

[10] Collins, A. and Kay, J. "Collaborative personal information management with shared, interactive tabletops" . PIM workshop CHI 2008, April 5-6, 2008, Florence, Italy.

[11] Jetter, H. C., König, W., Gerken, J., and Reiterer, H. "ZOIL a cross-platform user interface paradigm for personal information management". PIM workshop CHI 2008, April 5-6, 2008, Florence, Italy.

[12] Woerndl, W. and Woehrl, M. "SeMoDesk: towards a mobile semantic desktop". PIM workshop CHI 2008, April 5-6, 2008, Florence, Italy.

[13] Katifori, A., Vassilakis, C., Daradimos, I., Lepouras, G., Ioannidis, Y., Dix, A., et. al. "Personal ontology creation and visualization for a personal interaction management system". PIM workshop CHI 2008, April 5-6, 2008, Florence, Italy.

[14] Chau, D. H., Myers, B., and Faulring, A. "Feldspar: a system for finding information by association". PIM workshop CHI 2008, April 5-6, 2008, Florence, Italy.

[15] Diehl, J. "Associative personal information management". CHI 2009, New York, NY, USA, pp. 3101-3104.

[16] Karat, C-M., Brodie, C., and Karat, J. "Usable privacy and security for personal information management". Communication ACM 49, 1 January 2006, pp. 56-57.

[17] Hsieh, J. L., Chen, C. H., Lin, I. W., and Sun, C. T. "A web-based tagging tool for organizing personal documents on PCs". PIM workshop CHI 2008, April 5-6, 2008, Florence, Italy.

[18] Fuller, M., Kelly, L., and Jones, G. J. F. "Applying contextual memory cues for retrieval from personal information archives". PIM workshop CHI 2008, April 5-6, 2008, Florence, Italy.

[19] Bernstein, M., Van Kleek, M., Karger, D., and Schraefel, M. C. "Information scraps: How and why information eludes our personal information management tools". ACM Trans. Inf. Syst. 26, 4, 2007, pp.1-46.

[20] Bergman, O., Beyth-Marom, R., and Nachmias, R. "The user-subjective approach to personal information management systems design: evidence and implementations". J. Am. Soc. Inf. Sci. Technol. 59, 2 (January 2008), pp. 235-246.

[21] Gonçalves, D. and Jorge, J. A. Now, "It's personal! evaluating PIM retrieval tools". PIM workshop CHI 2008, April 5-6, 2008, Florence, Italy.

[22] Tungare, M. "Understanding the evolution of users' personal information management practices". INTERACT'07, Vol. Part II. Springer-Verlag, Berlin, Heidelberg, pp. 586-591.

[23] Henderson, S. and Srinivasan, A. "An empirical analysis of personal digital document structures", Proc. HCI International 2009, Berlin, Heidelberg: Springer-Verlag, 2009, pp. 394–403

[24] Evequoz, F. and Lalanne, D. "I thought you would show me how to do it - studying and supporting PIM strategy shanges". PIM Workshop, ASIS&T 2009, Vancouver, BC Canada, Nov. 5 - 7 2009, pp. 35-42.

[25] Voit, K., Andrews, K., and Slany, W. "Why personal information management (PIM) technologies are not widespread". PIM Workshop, ASIS&T 2009, Vancouver, BC Canada, Nov. 5 - 7 2009, pp. 60-64.

[26] Henderson, S. "Guidelines for the design of personal document management user interfaces". PIM Workshop, ASIS&T 2009, Vancouver, BC Canada, Nov. 5 - 7 2009, pp. 65-72.

[27] Jakob, L. and Maciej, L. "Using mobile phone contextual information to facilitate managing image collections". PIM Workshop, ASIS&T 2009, Vancouver, BC Canada, Nov. 5 - 7 2009, pp. 73-75.

[28] Karger, D. R. "Unify everything: it's all the same to me", in Personal Information Management, (Jones & Teevan, eds.) University of Washington Press, Seattle, WA, pp. 127-152.

[29] Jones, E., Bruce, H., Klasnja, P., and Jones, W. "I give up! Five factors that contribute to the abandonment of information management strategies". ASIS&T 2008, 4,1, pp. 1-6.

[30] Capra, R. A "Survey of personal information practices". PIM Workshop, ASIS&T 2009, Vancouver, BC Canada, Nov. 5 - 7 2009, pp. 2-5.

[31] 15 Tools for personal information management: #**1** www.efficientpim.com, #**2** www.thebrain.com, #**3** www.essentialpim.com, #**4** www.winpim.com, #**5** www.lifemanagerpro.com, #**6** www.azzcardfile.com, #**7** www.pimonline.com, #**8** www.pimone.com/pimone.htm, #**9** www.myarchivebox.com, #**10** www.evernote.com, #**11** http://code.google.com/p/keynote-nf, #**12** http://www.treepad.com, #**13** www.milenix.com, #**14** www.android-software.fr/pocket-docs, #**15** www.gmail.com

[32] http://ihcs.irit.fr/winckler/MyPIM/index.html

[33] http://www.surveymonkey.com

[34] Fincher, S. and Tenenberg, J. "Making sense of card sorting data". Expert Systems, 22(3), 2005, pp. 89 – 93.

[35] Bussolon, S, Russi, B., and del Missier, F. "Online card sorting: as good as the paper version". ECCE 2006, Sept 20-22, Zurich, Switzerland, pp. 113-114.

[36] http://websort.net

[37] Morineau A. et Morin S. " Pratique du traitement des enquÍtes. Exemple d'utilisation du logiciel SPAD". Paris, 2006. (ISBN 2-9525948-0-5)

[38] Benzécri, J.-P. "L'analyse des données. Volume II. L'analyse des correspondances". Paris, France: Dunod, 1973.

[39] Greenacre, M. and Blasius, J. (editors) (2006). "Multiple correspondence analysis and related methods". London: Chapman & Hall/CRC.

[40] Ward J. H., "Hierarchical grouping to optimize an objective function", J. American Statistical Ass., n°64, 1963, pp. 236-244.

 ISBN: 978-1-61208-167-0