

# Who Can Claim Complete Abstinence from Peeking at Print Jobs?

Antonietta Grasso      Jean-Luc Meunier

Xerox Research Centre Europe

6, chemin de Maupertuis  
38240, Meylan France

{Antonietta.Grasso, Jean-Luc.Meunier}@xrce.xerox.com

## ABSTRACT

While systems supporting communities of practice in work organizations have been shown to be desirable many, if not all, are decoupled from daily work practices and tools. This hinders a wide collection of data about their activities, because of the additional effort that is required from the users. Therefore a pre-requisite for a system aiming at making visible the community activity is the non-intrusive collection of data about the activities that are carried on in a workplace. We present a range of personal document management services that support the construction of a collective memory of user print activities. We have internally tested the system and verified that it successfully provided personal benefit, thereby ensuring that the system receives sufficient usage for the shared memory to be useful. The system also successfully addressed privacy concerns and effectively provided large data sets about document related activities. Finally it provided information able to trigger new or to reinforce existing informal exchanges in communities of practice at a convenient moment; the print action.

## Keywords

Knowledge Sharing, Community Awareness, Document Management, Memory Support, Paper User Interface.

## INTRODUCTION

What differentiates a learning organization, an organization able to sustainably produce value and innovation? Some insight on this complex question has been brought to light by two complementary sets of concepts, which started to be developed at the beginning of the nineties: the concept of *community of practice* [1] and the concept of *ba*<sup>1</sup> as an enabling environment for the community itself [2]. In their

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CSCW'02, November 16–20, 2002, New Orleans, Louisiana, USA.

Copyright 2002 ACM 1-58113-560-2/02/0011...\$5.00.

work Lave and Wenger undertook a major rethinking of the whole learning process by re-conceptualizing how it happens. In their conceptualization, learning occurs not because of the consumption of some external “knowledge” coming from the external world to us, but through the support of mechanisms that are inherently social. Learning is proposed as being supported by the participation in a community of practice, in a first stage that is legitimately peripheral and gradually increasing in engagement and complexity.

Since then the concept has been investigated in the context of work organizations with other major insights resulting from the work of Nonaka and Takeuchi [2], who examined several industries and the way that innovative products and concepts are developed. Subsequent to these observations they proposed a model of value creation in companies based on a learning cycle involving both the individual and the organization. In their model the knowledge that increases organizational value is not an object stored in a database nor a personal set of competencies. The relevant knowledge is that which can be shared and used within communities of people. These communities have been recognized as having the same nature and characteristics of the communities of practices studied by Lave and Wenger. Their main attributes are that they are voluntarily and informally built across the organization, bound together by shared interest, shared expertise and a *passion for a joint enterprise*. While these communities are recognized to be the hidden engine that keeps an organization creative and competitive, they form a fragile structure that is principally based on the spontaneous effort of their participants.

In the networked organizations of today the links and exchanges among members partially happen through digital communication, by communicating via email, exchanging relevant documents and meeting face to face. Technology itself cannot create communities where they do not already exist (a mistake commonly observed in organizations

---

<sup>1</sup> “Ba” is a Japanese term which roughly translates into the English word “place”

nowadays), it can however make the participation in them *less costly* and therefore can be used to nurture and sustain them. A crucial aspect of these communities is in fact to be spontaneous and based upon voluntary effort. Fostering them can be undertaken by promoting their visibility yet minimizing the personal effort required to participation, especially for the buy-in phase, where personal benefits may be less clear.

In this paper we present a system that aims at improving networking among people with whom we could be sharing the same practice (e.g. the same set of techniques, knowledge, tools) by making more visible at proper moments some of their document related activities. The system is attacking the fundamental issue of being able to collect data across different activities in the organization potentially performed by people who do not know about each other. It should be noted however that systems aiming at grabbing the full set of content a person manipulates have proved to be too intrusive from a privacy standpoint, while systems relying on personal effort for publishing have proved to be not so widely used because of usage costs. In this work we propose a middle ground that permits a selective and meaningful way to publish and to receive information about other document activities currently carried on in the communities of practice of an organization.

The paper is organized in the following way: we first introduce a novel concept of document management service designed to alleviate some of the issues and costs of ineffective retrieval and classification of documents; then we present how community services can be built on top of the document management service. Finally we present the first results from the deployment of the services inside our research organization.

#### PERSONAL DOCUMENT MANAGEMENT ISSUES

With the advent of networks and PCs at the beginning of the eighties, changes occurred in the way people handled documents. In particular a lot of emphasis and hope was put on the feasibility of the so-called *paperless* office. The aim was to eliminate paper documents and fully exploit the advantages of a networked environment where every document was properly classified, searchable and made available, potentially forever. Practically speaking, the reality has been more a paradigm shift from *print-copy-distribute* to *distribute-view-print* [4] where paper has not disappeared from the scene, but appears at a different time, towards the end of the distribution cycle rather than at its beginning. In fact what has happened has been an increase in the number of total documents circulated with a corresponding increase in the number of paper documents. While the precise increase is difficult to measure, some studies reported in [4] give an estimate of current paper use; for example the situation in American companies at the end of the nineties is evaluated to be such that:

- 94 % of business information is still recorded on paper;
- 2.4 billion new sheets are placed in paper file folders each day.

On the basis of these numbers we have looked at the reality of our own work organization (a research center with around 120 people working on site with roles ranging from administration to technical support to research) and found that we consume ~600.000 sheets of paper on a yearly basis just to support black and white print jobs, and with a total of around 1 million pages if we take into account those in color. This confirms that we still need to take into account this mix of both digital and paper media when considering how to manage our documents.

In this mixed scenario, even just taking into account digital documents, there are document management costs that originate from the need to explicitly store and classify documents in the fixed hierarchical structure of a file system. The difficulties that arise when looking for these documents result in additional costs for the organization, with these costs culminating when documents cannot be found at all when needed.

Things get even more difficult when you consider the coexistence of the paper and digital worlds. Difficulties are duplicated with extra costs associated with the storage and retrieval of paper documents, with additional hidden costs of managing the link between the two worlds. For example when we have only the paper document and would like to retrieve its electronic version. Starting from the above observations, we have worked on an architecture to alleviate the issues around misclassification and non-retrieval of documents easing the co-existence of the two worlds (the paper and the digital). The basic principle we have adopted is based on a capability to trace and store the document at the moment where a digital document becomes paper-based (i.e. printing). Recently Hull et al. proposed a similar approach and studied its application in [3]. While sharing with them the basic principle of background tracing of documents based on activity around them, we propose to embody the link between the paper document and its electronically stored counterpart in the form of a so-called SmartCover cover page [17], as explained below.

In the rest of the paper we first present the document management services built on top of this architecture, then we explain how its services have been used to provide services that support community awareness of the activities of others. We then explain the SmartPrinter architecture that enables these services and finally the results and observations that we have collected from the deployment of the system in our organization.

#### A PERSONAL DOCUMENT MANAGEMENT SYSTEM

Effortless document storage and easy retrieval  
Psychologists have developed the notion of episodic memory to describe the way that humans typically

memorize past events and “that the location of the episode, who was there, what was going on, and what happened before or after, are all strong cues for recall” [14]. A direct implication of these observations is that the context itself could be used to index information and to retrieve it. If we remember the situation where some information was manipulated it could be an additional route to find it again. Precisely with this motivation, a project called “Forget-me-not” developed a system defined memory prosthesis where a portable device was instrumented to trace a number of user activities, like “attending a meeting with X and Y” or “beaming a document to someone”, and making the stored documents searchable on the base of these contextual parameters.

In the context of SmartPrinter we have borrowed the concept of memory prosthesis by providing a search

mechanism that can exploit the memory of a print action. SmartPrinter in fact creates a temporal flat memory of the documents that a person prints. This storage is additional with respect to the original one, e.g. on the desktop, it just provides an additional parameter to be exploited in the search: the memory of having printed a document. In such a way the document could be completely forgotten with respect to its title or categorization and still quickly accessible by issuing searches like “find me the documents I printed last week”. Additionally the printer characteristics, (location, name) are also valuable clues that the system provides. In Figure 1 a screenshot of a personal print memory is shown: this list can be browsed or queries can be issued on both the full text document and the metadata extracted from the memory.

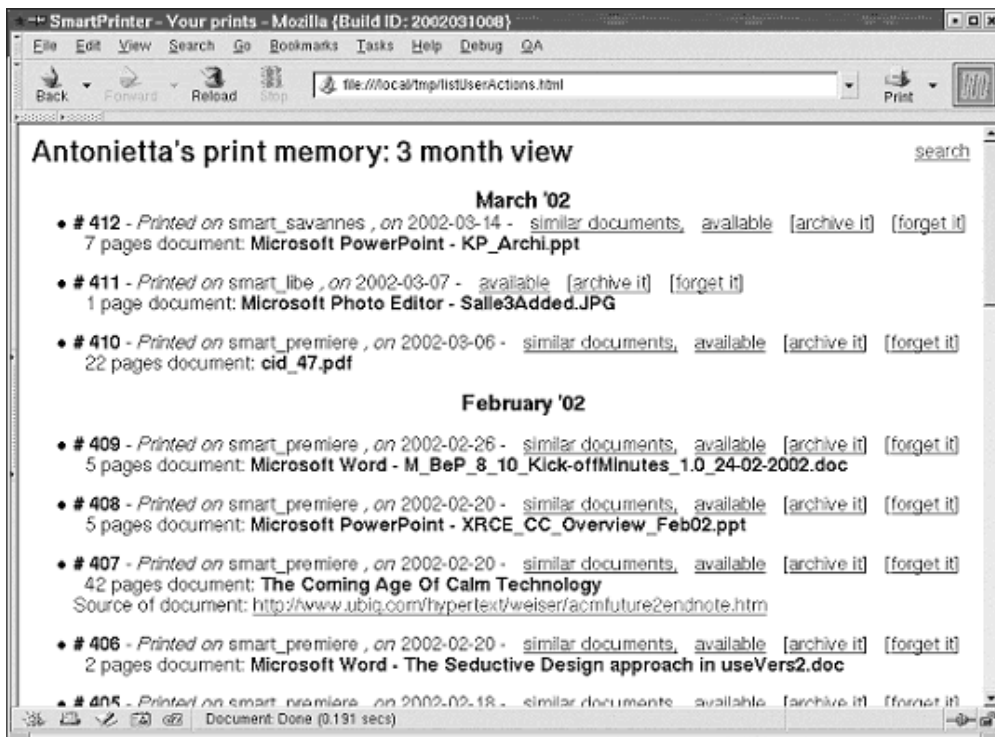


Figure 1 The SmartPrinter Web interface

#### Bridging the paper- and digital-worlds

Taking into account work practice studies [5] that studied the various affordances paper provides over electronic documents, which include easier reading and annotation, SmartPrinter also offers a paper-based user interface called SmartCover (Figure 2). It replaces the usual print separator page and allows the user to interact with SmartPrinter on paper, e.g. while reading the document.

The so-called SmartCover is generated with every printed document and acts as a container of information about the document. It also carries an encoded reference to the document, also stored into the print memory, as a means for

the user to give instructions to the system by filling it in and scanning it in a networked multi-functional device capable of processing it. A user on the go can carry a SmartCover and use it to gain access to the entire document (plus related information references and links, as we will see later) or tell the system what to do with it. There's no need to lug a computer around. The machine-readable reference is visible at the bottom of the page and is explained in the section about the system architecture.

The SmartCover currently allows the user to get information regarding the printed document. It provides a direct way to reprint the document, and an indirect one to

find it in the print memory thanks to the print automatic numbering (See in Figure 1 on the left and the top right of the SmartCover in Figure 2 below). The user can also conveniently attach hand-written annotations to it in the print memory as shown in Figure 3 below, and force the memory to forget about the document.

It also supports the provision of recommendations to a recommender system [18] in use in the center to support communities of practice, but for which a usage barrier appeared to be the difficulty to publish and recommend a document partly because this process requires printing and reading it carefully [17] with the associated cost of later retrieval of the electronic version to support sharing.

We will see in the next section that its informative ability is also exploited in support of community awareness.

### COMMUNITY SERVICES

As mentioned before, the activities of reading and reviewing have some aspects that are well supported by paper documents and we can expect to observe people continuing to print in support of these activities. This is also confirmed by the estimates around usage of paper documents that we presented in the previous section.

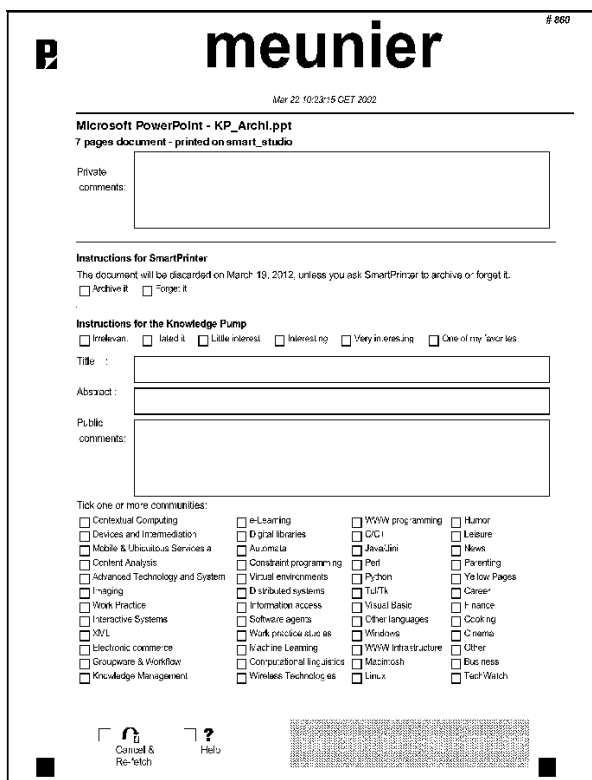


Figure 2: The SmartCover sheet

This also means that the print rooms (or shared printers) of an organization often provide a point where most of the relevant information, embodied as printed documents, passes by. But the process of printing is without memory

except, perhaps, for the print logs generated at the printer server, which are only used for administrative or accounting purposes. We propose here to exploit the personal print memories to provide additional services to the communities of practice of an organization, which include the users of one or several shared printers. In addition to the results from the user studies about usage of paper documents [5], we made some additional hypotheses regarding the print act:

- Choosing to print a document makes an implicit declaration that the document is in the scope of activities of the person printing it.
- Someone else in the community may have printed it before; the awareness of who they are can potentially trigger valuable interactions.
- There's a reasonable probability that the document is also relevant/important to other people in one or more of the communities of practice to which the person printing it belongs to.

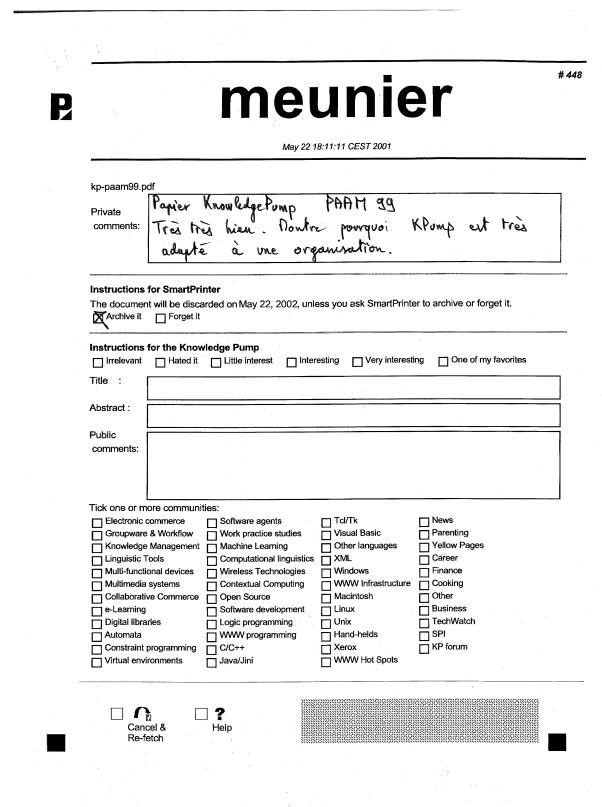


Figure 3: The SmartCover to associate notes to the printed document

On the basis of the hypothesis above, SmartPrinter has been extended to compute for each printed document a list of so-called *twin* and *friend* documents. Twins of a document are the documents that are very close to it in terms of content, typically including other versions of it, while friends are

those that are related to it. A document-to-document similarity is computed based on the document content profile computed at the moment of the printing and that will be explained in details in the architecture section.

By providing the user with the list of twins and friends of the document s/he printed ranked by degree of similarity, together with the data about who and when the document was printed, we can provide a contextualized search in line with our hypothesis. Moreover, while this information is available in the web interface of the system, it is most appropriate when the person is going to work on the document, i.e. at printing time. We therefore decided to provide this service together with the printed copy, by means of the SmartCover because it provides a convenient output space and is created together with the document.

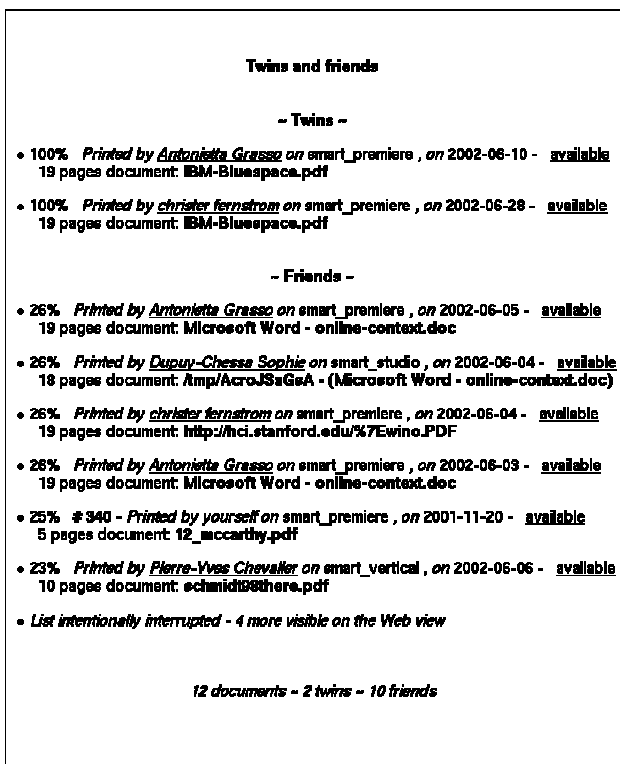


Figure 4: The verso of the SmartCover showing twins and friends of the printed document

We used the back of the SmartCover to provide the list of twins and friends of the printed document (Figure 4 above). It presents essentially the same information as that available from a Web browser, except that it comes on a convenient medium and at the moment the document is being used. It should be mentioned that the back of the SmartCover shown is currently an output only medium. In the next release, we will have checkboxes associated to each twin or friend in order to make it possible to obtain a paper print of a related document directly at the printer when collecting the original document, by scanning the paper form back in once it has been filled out.

It is worth noting that this feature also raises privacy concerns. One might come to gain knowledge of the fact that someone else printed a particular document. We believe that the act of publishing a document via the SmartPrinter is acknowledgement of the public nature of the document. However this could be refined in the future to let people benefit from the SmartPrinter services without always publishing their print activities to the whole group.

Finally, the list of related documents could also contain notes and ratings left by other people.

#### SMARTPRINTER ARCHITECTURE TO CAPTURE THE PRINT STREAM

The first choice made in the SmartPrinter architecture was about which approach to take with respect to tracing the user activity among two basic approaches: either everything going through the system is traced or the user can disable the functionality on a case by case basis. In the SmartPrinter system we have chosen a semi-automatic tracing, giving the user full control over the system behavior.

We associate with each organizational printer a virtual one where the name is prefixed by "smart\_". In this way the user can conveniently enable or disable the system at each print action. The captured documents are stored in personal print memories. The user can browse her/his print memory from a web browser (Figure 1), thus accessing a PDF version of the document, viewing personal attached notes. To help her/him to get a sense of the print context, SmartPrinter gathers all available metadata regarding the print: document title, URL, number of pages, file name, printer name, and date. And a SmartCover is produced instead of the regular print separator page.

In the following sections we explain the details of how this architecture works.

#### The SmartPrinter components

##### Augmenting the print spooler

Capturing the stream of user print jobs involves modifying the print mechanism of either the user's desktop, or the printer, or possibly the print server. We discarded the first and second alternatives because of the diversity of both the user desktops and the printers and their associated drivers. We decided instead to leverage some standard network print protocol to provide a print capture mechanism that would be independent of both the nature of the printers and the user desktops. To achieve this purpose, SmartPrinter provides a print server supporting a network print protocol, which sits somewhere on the local area network and acts as a print proxy, forwarding the print directives and responses to/from the printer or the actual print server as shown in Figure 5. For the time being, the SmartPrinter print proxy supports the Line Printing Protocol [LPR, RFC1179], as it is both widely supported by the print drivers of network printers and a fairly simple protocol to deal with from an engineering perspective.

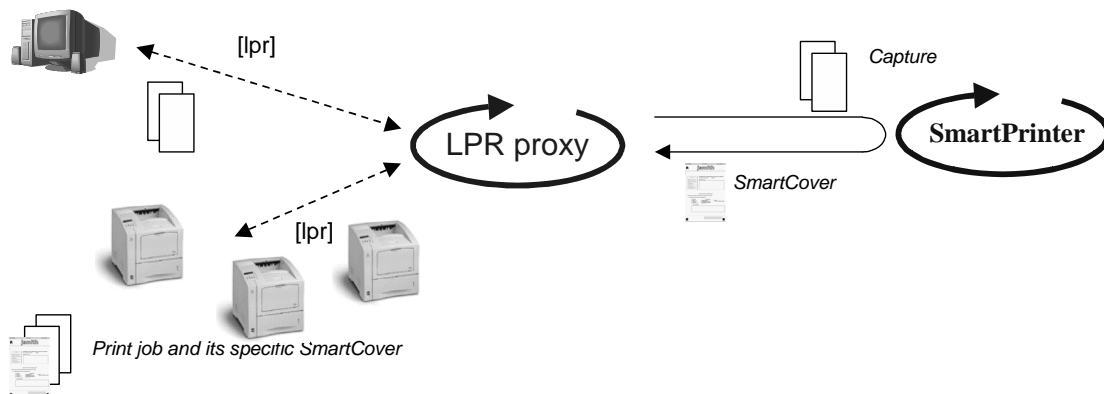


Figure 5: The SmartPrinter architecture

This architecture is multi-user, multi-printer and multi-format, as it supports several printers for several users printing in various print formats. However, the current implementation supports only PostScript level 1 and 2, which was sufficient for us to experiment with the SmartPrinter concept.

Making a network printer "smart" consists in setting up a corresponding LPR print queue on the SmartPrinter print proxy. This queue must then be registered on the user's machine, like any regular LPR queue. By convention the SmartPrinter queue is named after the printer name by prefixing it with "smart\_", so that the user can easily choose between the regular printing or the smart-one, as shown in Figure 6.

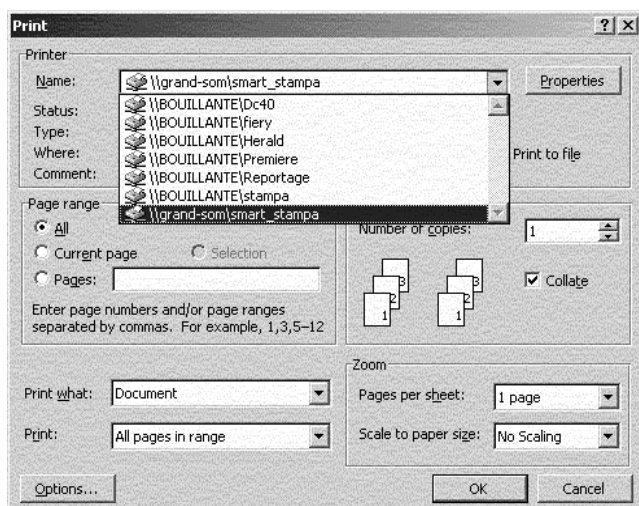


Figure 6 – Choosing between the regular *stampa* printer or its associated *smart\_stampa*

When a user is printing, the associated document is captured and a SmartCover is inserted in the print job, before it is forwarded.

The provided services involve the four steps below, which are described in the next sections:

- Extraction of metadata, such as the user identity and the document title, URL, or their number of pages.
- Storage of the document in the user's personal print memory.
- Profiling of the document content.
- Creation of a SmartCover, specific to the current document and user.

#### Metadata extractor

The three possible sources from which metadata are extracted are the LPR control data, the PostScript header and the document itself.

The LPR control data identifies the user's machine and gives the user identity in the form of the user login, as known on the user's machine. The user authentication currently relies on this information. The LPR control data may also contain a relevant title or source file name, usually useful to label the banner page, but it depends on the application and print driver that generated the print job. This is currently implemented using a series of heuristics.

From the PostScript header and comments, we reliably get the number of pages. Depending on the application and print driver that generated the postscript, we can obtain the document title and URL. Finally, common Web browsers can be configured to insert the URL of the document in the header or footer of the print. This is the only information we currently extract from the document content itself, using regular expressions. However, there is room for improvement by interpreting the PostScript and exploiting

the document layout [9] and content to calculate its title, authorship, etc.

#### Feeding the Print Memory

The SmartPrinter print proxy forwards the PostScript print job to the SmartPrinter server, which converts it to PDF. This format is more convenient for supporting the web-based document retrieval function and remains a print-ready format able to reproduce a paper document upon user demand. It is also more efficient in terms of storage space.

With the print memory comes the privacy issue. We have devised a set of features to alleviate the user's reluctance to use SmartPrinter:

- By making the decision to print on a 'smart-printer' a conscious one. The same way a user chooses the printer, s/he chooses a smart one or not (See Figure 6).
- By providing a 'forget' function that clears the memory from any information related to a captured print job. This function is available from the Web and the paper UI, i.e. the SmartCover.
- By protecting by login/password the access to individual print memories.

We have not yet included any access control mechanism for the PrintMemory more elaborate than the login/password. We regard the seamless integration with existing work practices as very important, which seems incompatible with setting access rights on print jobs. From a user perspective, a convenient solution consists in setting access rights in the print dialog box. But it requires modified print drivers, which would add complexity to the architecture of the system. Alternatively, for each printer we could designate names that define the access rights associated with print jobs. For instance declaring a `smart_X_private`, a `smart_X_mygroup` and a `smart_X_public`, to represent both the 'smart' X printer and a privacy policy for the PrintMemory.

#### Profiling the documents

To enable further information retrieval services, the SmartPrinter software computes and stores a profile for each document. An example of a profile could be the exhaustive list of keywords contained in the document, which enables the indexing of the corpus of printed documents to allow users to perform full-text searches. Instead of targeting full-search, we put a higher priority on the implementation of a different kind of profile to explore different new features in our prototype. SmartPrinter extracts terms from the document and weights them using a TFIDF (term frequency / inverse document frequency) [8], based on the average frequency of terms on the Web.

In more detail, the PostScript document is converted to text, using the well-known GhostScript tool [12]. Then the language is identified. The text is then tokenized, looked-up in a morphological lexicon and the most probable part of

speech tag for each word is calculated. This is used to produce a normalized form of each word. As this process occurs, duplicate words are discarded and a count is kept of the number of occurrences of each word to enable the calculation of weights based on word frequency. Finally, a list of stop words for the source language is used to discard frequent words that are not used for classifying the text, such as conjunctions and prepositions. All of this linguistic processing is performed by the Xerox Xelda™ technology [16]. As a result of this process, each document has associated with it a list of keywords with respective frequencies. A weight is eventually associated to each term using the TFIDF method. The weight of a term is computed as the term frequency in the document multiplied by the logarithm of the inverse frequency of the term in a reference corpus. The weight of a term increases if frequency in the document is larger than in the reference corpus. Corpora such as the British National Corpus provide a list of words, their part of speech, their frequency and the number of documents in the corpus that they appear in. We currently retain the 60 most weighted terms in the document profile, as past studies have shown that additional terms do not bring significant better results. We currently support English and French.

Similarity among documents is measured by using a Jaccard coefficient. We have set different thresholds to discriminate between *twin* documents (threshold>0.8) and *friend* documents ( $0,2 \leq \text{threshold} \leq 0,8$ ).

#### Enabling paper based interaction

The SmartCover implementation relies on the Xerox FlowPort™ [11, 13] product, which allows people to integrate the contents of hard-copy documents with e-mail, groupware and other electronic workflows. The cover sheet is a FlowPort™ document token encoded with DataGlyphs™ [10]. Check boxes become machine-readable to let users give instruction to the system and hand-written annotations can also be captured as image in dedicated zones.

## CASE STUDY OBSERVATIONS

### Settings

A preliminary version of SmartPrinter ran for the first time at our research center in February 2001, with ten users. We then presented it to the whole research center in October 2001 and gathered a dozen more users. We eventually announced, end of November 2001, a SmartPrinter lottery, where we would draw at random 3 SmartCovers out of a box of collected ones and offer bottles of champagne to the winners. This attracted 20 more users. We organized the lottery day in mid-February 2002.

XRCE is both a multicultural and multidisciplinary organization, employing about 130 people spread over 4 buildings on a single site. Most of them do either software engineering or research. The work environment favors the emergence of numerous and various communities of

practice, either well formalized by for instance projects, or less formalized and less known, such as those based on a shared center of interest. Some of them are formed by small groups (2-10 people) often co-located, while others are larger and spread across buildings and organizational structures.

#### Quantitative Observations

SmartPrinter finally had 44 users spread over 3 buildings using 6 different printers. They printed around 1300 documents in total, a number that increases almost linearly over time. A quarter of them were one page documents, but some of them were 400 or 500 pages long. The average page count is close to 11; the distribution of page numbers is close to an exponential (with a linear aspect on a log/log scale).

Hull et al. [3] reported an average age of retrieved documents of 44 days, with 38% less than one-week-old documents, and 10% more than 6 month old ones. We observed here that the average age of retrieved documents in the print memory is 66 days, 34% of retrieved documents were retrieved while less than one-week old and 15% percent while more than 6 months old. We therefore agree with their analysis stating that older documents are retrieved by users and therefore can be deemed useful.

Looking at the software log, we found that 40% of the users stopped using SmartPrinter after 1 to 5 trials, while 60% became regular users. We conducted a series of user interviews in March 2002 to understand the reasons that lead to regular usage or non-usage.

#### Interviews

We selected 10 out of the 40 users covering various positions in the organization: managers (3), project leaders (2), researchers (3), engineers (2), and covering 6 different teams. Among these 10 people, 6 were regular users: 2 managers, 1 project leader and 3 researchers. These 10 users were from 3 different buildings and used 3 different platforms: MS-Windows (4), Linux (3), Sun-Solaris (3).

The interviews were conducted as open discussions either face-to-face or over the phone. We wanted each interview to address at least the use and perceived usefulness of the print memory, the SmartCover and the Twin/Friend list, together with privacy concerns and system usability, in particular about the Web user interface, the print delay, the mechanism to select the printer in the application.

The discussion was initiated by a "Did you use SmartPrinter?" question to let the user freely give her/his perspective. If needed, the interviewer would then ask some focused questions to cover the above five aspects if not spontaneously addressed. We wanted to determine in this way the importance of these different aspects from a user perspective and leave the door open for unpredictable feedback.

All users spontaneously reported on their SmartCover usage while 9 of the 10 users did so for the twins/friends function,

all showing a positive interest. Only 4 of the 10 users spontaneously mentioned the print memory, two of them being interested in this function and the other two not. Privacy and system usability had often to be addressed by the interviewer. Half of the users complained about a lack of user documentation, regarding how to scan a SmartCover, how to access a twin or friend document, how to access the web user interface.

The interest in the SmartCover medium and the twin/friend function is encouraging given how unusual they are. This is reinforced by the lower importance attached to privacy and modality of usage.

The complaints indicate that we didn't assist our users well enough, given that only some of them attended the SmartPrinter presentation and that the notification of the lottery was made by email. We also agree with the lack of user documentation. So, given the observed system usage, we believe the reported user difficulty is due to the deployment itself and isn't intrinsic to the technology. Indeed some of our colleagues received informal explanations at various occasions and were then able to use all the features of the system without any problem.

The interviews of the 4 users who gave up use provide a mixed view:

- One tried it having almost no idea of the system's function, got no output and gave up;
- Another one tried it, but reports not having any a priori interest in any of the system functions. This person prints documents for immediate reading and then throws the document away, and is satisfied with his electronic document management habits. More importantly, he doesn't believe any twin/friend document would point him to anything relevant and conversely does not think that what he prints may be of interest to someone else.
- The two others were interested by the twins/friends function but were disappointed: one rarely prints and didn't get any twin/friend document listed after 3 trials; the other complained about the restricted information provided on the SmartCover, this after 3 trials as well, and didn't want to spend any time using the SmartPrinter Web user interface for the complete information.

It is clear here that the user benefit wasn't high enough for them. Actually, the last two users reported some disappointment, which was based on a correct understanding of the system capabilities, therefore showing some initial expectation.

From the interviews of the 6 regular users clearly emerges that:

- These users perceive the twin/friend function very favorably. All of them reported systematically looking at the SmartCover back. One said, "*I usually think*

*about the reasons for the document friendship and it helps me to establish links between topics". Three of them gave examples of relevant pointers to documents and/or people obtained in this way.*

- On the other hand, 3 of them complained that many twin/friend documents and their associated users were already known to them and actually printed by very close colleagues. They attributed this fact, as we do by the way, to the lack of critical mass of users and believe this is directly dependent on the scale of the deployment. One reported that this gave him confidence in the system algorithm.
  - All users enjoyed the SmartCover as an information medium. One said: "*You look at it while going back from the printer*". Another said: "*It has been useful a couple of times and it's effortless*".
  - On the other hand, only 2 of them actually scanned some SmartCovers back into the system. Seven of the other interviewees reported not knowing how to scan the SmartCover; one reported not having the appropriate multi-function device in his building.
  - Most of these users were not satisfied with the design of the SmartCover. One reported: "*they all look the same*", and would like to see a document thumbnail of the first page. Another complained about the lack of information regarding twins/friends and requested 2-3 line of summary for each.
  - Half of them reported benefits from using the print memory, and it was the same people who reported having used the web user interface. Actually all other 7 users reported having trouble with the web user interface: forgetting its URL, their login or password.
  - These 6 regular users reported feeling comfortable with respect to their privacy and associated this with the ease of enabling/disabling the SmartPrinter service by selecting the appropriate logical printer. Actually 8 out of the 10 interviewees said the selection of the appropriate printer was easy. Interestingly, none of them used the 'forget it' function.
  - 5 of them chose a SmartPrinter as their default printer.
- Eight users said they perceived a slight printing delay when using SmartPrinter, which was fine with them. Many reported disabling SmartPrinter when in a hurry, a commonly given example was: one must leave her/his office immediately to attend a meeting for which s/he needs a paper document.

#### Discussion

It is interesting at this point to discuss the case study from three different perspectives:

- A. *The provision of a personal document management service.*
- B. *Usage barriers.*

#### C. Community support.

There are clear positive indications for A, both quantitative (the retrieval of printed documents) and qualitative (half of the interviewed regular users), whatever the user motivation is.

Regarding B, the study indicates a low barrier to usage, especially when considering how little assistance was provided for the deployment. Despite the unusual way to activate the system (printing) and its strange output (SmartCover), users feel comfortable with it, while on the contrary they find the web user interface less convenient. This is probably because they have to take explicit action to use it, and also because of the lack of information (URL) and support (e.g. login in a persistent cookie).

The SmartCover as an input medium proved to be difficult to use and wasn't significantly used. On the contrary, its output function appeared to be very satisfactory and many users expressed wishes regarding its content and format

Privacy wasn't an issue for the users because they felt it was easy to choose the appropriate printer.

Regarding C, we found indications that point in the direction of giving different type of support to different type of communities. From our knowledge of the work setting, we can say that the users in fact belong to different types of communities:

- One well formed being co-located in the same building and belonging to the same organizational unit;
- A second one which is well formed, but located in two buildings and covering several organizational units (research and business);
- A third one, which is currently emerging at the center.

Given this situation we can see the well-established community got no value out of using SmartPrinter, but its members did use it (in confirmation of A). On the other hand, the twin/friend feature was very attractive to both the second and third community, where it fulfilled needs (either the emergent behavior or the actual physical distance) for a better visibility on activities of other people who were interested in the same practice. This would suggest that visibility of activities should be presented in different ways to members of the communities who already have strong ties, given for example by joint work on projects.

While these findings must be handled with care given both the scale of the study and the specificities of the XRCE site, a very good indication is also given by the fact that our users continue to use the system after the lottery, as witnessed by the constant rate of growth in the number of prints. The question is: to which proportion do A and C contribute to this sustained usage?

## CONCLUSION

We have presented a software system that aims at supporting communities of practice in office environments, by making visible the printing activities of their members in order to increase social networking. A particular aim of the system is to obtain this without imposing additional efforts to the users and while preserving privacy.

A medium-scale usage study together with a small number of interviews showed significant success of the system regarding usability and privacy aspects. In particular, the embedding in the daily work practices of a functionality that captures user printing activity was positively accepted, ensuring to the system the provision of data necessary to feed the community support services.

Some of the negative indications are, not surprisingly, due to the prototypical state of the system. The lack of contextual help, for example, increased the effort to access the web interface to the system. More importantly, our hypothesis regarding the convenience of the SmartCover as a link between the paper document and the information system wasn't validated, in this particular environment. However, solely the provision of a personal print memory retained the high-perceived user benefit we had aimed for.

Positively, we got the confirmation of both the ease-of-use and convenience of the SmartCover-based system output and of the user interest for the twin/friend feature. However, a more general conclusion that the twin/friend functionality is promoting social networking at the moment cannot be made. While interviewed users reported stories about triggered communications, the data and our case study were not enough to assess the benefit neither at the community nor organizational level.

Another open question remains the typology of workplaces that are addressable by a system like this one, where work practices are heavily document-based.

As a final conclusion we believe that our experience showed a successful way to address the sustainability of a collection of data, which is often the first pitfall in the design of systems aiming at the same objectives, and opens the way to further investigation.

## Acknowledgments

The authors are grateful to Natalie Gance, James Pycock and Christopher Thompson for their contribution to this work and our other colleagues at XRCE for their suggestions and feedback. We also thank the anonymous reviewers for their valuable comments.

## BIBLIOGRAPHY

1. Lave J. and Wenger E., *Situated Learning. Legitimate Peripheral Participation*, Cambridge University Press, 1991.
2. Nonaka, I. and Takeuchi, H. *The Knowledge Creating Company*, Oxford University Press, 1995.

3. Hull J. and Hart P. Toward Zero-Effort Personal Document Management. *IEEE Computer* 34, 3, 30-35.
4. Liu Z. and Stork D. Is paperless really more? *Communications of the ACM* 43, 11, 94-97.
5. Sellen, A. and Harper, R., *The myth of the paperless office*, MIT Press, 2001.
6. Newman, W., Eldridge, M. and Lamming, M., Pepys, Generating Autobiographies by Automatic Tracking, in *Proceedings of ECSCW'91* (Amsterdam The Netherlands, September 1991).
7. Want R., Hopper A., Falcao V., and Gibbons J., The active badge location system. *ACM Transactions on Information Systems* 10, 1, 91-102.
8. Salton G., Development in Automatic Text Retrieval. *Science*, 253, 974-979.
9. Giuffrida G., Shek E. and Yang J., Knowledge-Based Metadata Extraction from PostScript Files, in *Proceedings of the Fifth ACM Conference on Digital Libraries-DL'00* (San Antonio TX, June 2000), ACM Press, 77-84.
10. DataGlyphs, <http://www.dataglyphs.com>
11. FlowPort™ <http://www.xerox.com/flowport>
12. Ghostscript, an interpreter for the PostScript language and for PDF, <http://www.cs.wisc.edu/~ghost>
13. Johnson, W., Card, S. K., Jelinek, H., Klotz, L., and Rao, R.: Bridging the paper and electronic worlds: The paper user interface, in *Proceedings of INTERCHI*, ACM Press, 507-512.
14. Flynn, M. and Lamming, M. G., Forget-me-not. Intimate Computing in Support of Human Memory, in *Proceedings of FRIEND21 Symposium on Next Generation Human Interfaces* (Tokyo Japan, 1994).
15. Resnick, P. and Varian, H. R., Recommender Systems. *Communications of the ACM* 40, 3, 56-60.
16. Xelda™, Xerox linguistic engine, <http://www.xrce.xerox.com/ats/xelda>
17. Grasso A., Meunier JL. and Thompson C.: Augmenting Recommender Systems by Embedding Interfaces into Practices, in *Proceedings of HICSS'00* (Maui Hawaii, January 2000), IEEE, 63.
18. Gance N., Arregui D. and Dardenne M.: Making Recommender Systems Work for Organizations, in *Proceedings of PAAM'99* (London UK, April 1999).