

ForgetIT

Concise Preservation by Combining Managed Forgetting and Contextualized Remembering

Grant Agreement No. 600826

Deliverable D2.1

Work-package	WP2: Foundations of Forgetting and Remembering
Deliverable	D2.1: Evaluation Plan (V2 - Amended Version)
Deliverable Leader	Maria Wolters, UEDIN
Quality Assessor	Walter Allasia, EURIX
Estimation of PM spent	8
Dissemination level	PU
Delivery date in Annex I	M9 (October 2013)
Actual delivery date	20. December 2013 (v1), 1. July 2014 (v2)
Revisions	7
Status	Final Version with Amendments (V2)
Keywords:	Evaluation, Mixed Methods, Memory, Forgetting, Personal Preservation, Organizational Preservation

Disclaimer

This document contains material, which is under copyright of individual or several ForgetIT consortium parties, and no copying or distributing, in any form or by any means, is allowed without the prior written agreement of the owner of the property rights.

The commercial use of any information contained in this document may require a license from the proprietor of that information.

Neither the ForgetIT consortium as a whole, nor individual parties of the ForgetIT consortium warrant that the information contained in this document is suitable for use, nor that the use of the information is free from risk, and accepts no liability for loss or damage suffered by any person using this information.

This document reflects only the authors' view. The European Community is not liable for any use that may be made of the information contained herein.

© 2013 Participants in the ForgetIT Project

List of Authors

Partner Acronym	Authors
UEDIN	Maria Wolters, Elaine Niven, Robert Logie
UOXF	Viktor Mayer-Schönberger
DFKI	Heiko Maus
dkd	Olivier Dobberkau, Alexander Damhuis
LUH	Claudia Niederée, Nattiya Kanhabua
LTU	Jörgen Nilsson
CERTH	Vasileios Mezaris
EURIX	Francesco Gallo
IBM	Doron Chen
USFD	Mark Greenwood

Contents

List of Authors	3
Contents	4
Executive Summary	6
1 Introduction	7
2 Simulating Long-Term Preservation	10
2.1 The Festival Study	11
2.2 Developing New Experimental Paradigms	13
2.2.1 Personal Preservation: Community Archiving of Personal Digital Heritage	14
2.2.2 Organisational Preservation: Capturing Organisational Memory	16
3 Evaluating Personal Preservation using the Personal Information Management Model	18
3.1 Study 1 (Y2): Three Weeks at the Festival	18
3.2 Study 2 (Y3): Conference Travel Season	20
4 Evaluating Content Management and Preservation	21
4.1 The Press Release Use Case	21
4.2 The Framework: Action Design Research	22
5 Conclusion	24
A Documenting Success	25
B Test Sets	25
C User Studies	26
C1 Performance Assessment	27

C2	Personal Preservation: Cross-Sectional Study	28
C3	Organisational Preservation: Cross-Sectional Study	28
D	Outcomes and Success Indicators	28
	References	35

Executive summary

This document lays out the overall evaluation plan for the two main applications to be developed in ForgetIT, namely, Personal Preservation (WP9) and Organisational Preservation (WP10). The evaluation of both applications will be supported by studies of conceptual foundations and core functionality that will be conducted within WP2. There are two strands of work related to conceptual foundations, work on human memory and forgetting, which will be evaluated in experimental studies, and work towards a model of organisational memory and forgetting, which will be evaluated in three in-depth case studies of carefully selected organisations.

The two preservation applications will be evaluated along three dimensions defined in ISO 9241 [1]: (1) effectiveness, (2) efficiency, and (3) user satisfaction. In particular, we will assess whether users can see short- and medium gains that will motivate them to keep using ForgetIT solutions and lead to long-term benefits.

For each development release of the Personal and Organisational Preservation systems (D9.2–D9.4 and D10.1–D10.3), we will define evaluation tasks that are representative of the functionality implemented. Aspects of functionality that would require deployment times of a year or more will be tested in custom-designed studies within WP2.

The Personal Preservation demonstrator will be evaluated in two empirical studies, a study in Year 2 lasting around a month, and a study in Year 3 lasting around four months. The Organisational Preservation demonstrator will be evaluated in an in-depth study spanning two years with two customer sites using an action design research approach. These studies will be supplemented with standard usability testing and semi-automated end-to-end tests of functionality. The design of these evaluations and the theoretical analysis of the evaluation results will be informed by the conceptual work from WP2.

1 Introduction

ForgetIT aims to develop an approach for intelligent preservation management that is characterised by three concepts, *synergetic preservation*, *contextualised remembering*, and *managed forgetting*. Synergetic preservation is designed to bridge the gap between active use and preservation. Through contextualised remembering, objects that are retrieved from an archive are made accessible and interpretable in the current context, while managed forgetting helps in preservation decisions and ensures that active information use as well as archives are not cluttered with unnecessary detail.

Within ForgetIT, we explore two main application areas, *personal preservation* and *organisational preservation*. For example, a personal preservation system for family photos dating back to the 1860's could provide access to archival copies of photos ranging from the earliest photos, such as daguerreotypes¹, to iPhone videos, with a digital record of each photo and its main content that is regularly updated as the archiving infrastructure changes (synergetic preservation). The content of each photo is contextualised on access, so that a search for "car" across the whole archive would yield both pictures of the 2010 Toyota Prius and the 1899 Stanley Locomobile (contextualised remembering). The search also returns a few photos of damage to the previous car, a Volvo, which were used in an insurance claim. Other photos of the old Volvo have been compressed or deleted, because they are no longer of interest to the family (managed forgetting).

In WP2, we study the psychological foundations of managed forgetting and contextualised remembering, which in turn informs the design and evaluation of the personal preservation system developed in WP9.

In organisational preservation, we seek to preserve assets and artefacts that were generated by an organisation, such as press releases. However, these objects are only part of the organisational memory. In order to interpret them, we require a knowledge of the rules, procedures and structures of the organisation that produced these objects. We preserve this context for later interpretation through context-aware preservation. This preserved and maintained context will be used as part of supporting contextualized remembering. Preservation technology ensures that press releases which were stored on 5 1/4" floppy disks in 1984 are still accessible in 2014, while press releases that are no longer relevant can be summarised before deletion (managed forgetting). Synergetic Preservation, as it is developed in the ForgetIT project, makes such preservation technology more easily accessible.

Within WP2, we develop a model of conceptual memory that will be tested in a series of case studies with three organisations that have very different approaches to memory and preservation. The system developed in WP10 will be evaluated throughout the development process using a flexible, iterative approach, Action Design Research [20]. This allows us to feed new results from the model developed in WP2 into the development as

¹For more information on daguerreotypes, see for example "The Daguerreotype Medium", a page maintained by the US Library of Congress <http://www.loc.gov/pictures/collection/dag/medium.html>

they become available.

In the main body of this document, we provide an overall plan for testing the solutions that will be developed in the ForgetIT project for both use cases. For each strand of work, detailed study designs will be developed based on the two demonstrator systems described in Deliverables D9.2-9.4 and D10.1-10.3; here, we outline the general approach we will follow. The emphasis is on formative evaluation, with a short summative evaluation period in the final year of the project.

We also include a summary of our strategy for evaluating the whole project, outlined in Appendix A. Appendix B outlines the main test sets that will be developed for automatic testing. In Appendix C, we describe two studies that are designed to provide solid quantitative evidence that managed forgetting, contextualised remembering, and synergetic preservation, the three cornerstones of the ForgetIT approach, have been successfully implemented. Finally, Appendix D outlines where the success indicators described in the Description of Work are or will be documented.

Evaluating Long-Term Preservation

Long-term preservation is typically over decades (or even centuries), time spans that are well beyond the duration of the ForgetIT project. However, within WP2, we can design studies in which people access older data that were archived using a simulation of ForgetIT functionality. These studies will be informed by the concept of community archiving (personal preservation) and the theory of organisational memory (organisational preservation). A successful pilot for such a study is described in Section 2; further studies will be informed by ongoing development work and the conceptual taxonomy of remembering and forgetting to be developed in Deliverable D2.2.

The conceptual model of organisational preservation to be developed in WP2 will be validated through in-depth fieldwork and interviews at three sites that cover a range of preservation practices.

Evaluating Demonstrator Solutions

The demonstrator applications for personal preservation and organisational preservation will be tested end-to-end using both automated tests and user studies. The user studies for personal preservation will follow an observational design, where a number of users regularly interact with the system, and which is the standard for evaluating applications for managing personal photos and information [2]. We have planned two user studies, one spanning a couple of weeks, and the second spanning a couple of months, to coincide with relevant stages of application development.

For organisational preservation, we will use an approach that is tailor-made for an iterative, agile development process, Action Design Research [20], with the main summative evaluation to take place in Year 3.

Within those studies, we will collect data on three dimensions defined in ISO 9241-11 [1]: *effectiveness*, i.e. whether a task can be completed successfully, *efficiency*, i.e. the speed with which a task can be completed, and *user satisfaction*, i.e. users' views and experience of using the system. In addition, these studies will be supplemented with standard usability testing and semi-automated end-to-end tests of functionality.

This document consists of three main sections, Section 2 on evaluating simulated long-term preservation according to ForgetIT principles, Section 3 on evaluating the personal preservation solutions developed within ForgetIT, and Section 4 on evaluating organisational preservation solutions.

2 Simulating Long-Term Preservation

Within the time frame of the three-year ForgetIT project, it is not possible to test whether the ForgetIT framework successfully allows long-term preservation over several years or even decades. However, it is possible to probe how well people can access archived data that spans years or decades that has been moved to a preservation system designed along ForgetIT principles. In this section, we focus on evaluating preservation solutions over time frames that go beyond the specific application scenarios described in Sections 3 and 4.

In previous work, documented in D9.1, we gathered requirements for long-term personal and organisational preservation solutions. The personal preservation work was extended to a survey that is distributed to over a thousand people across Europe; the organisational preservation work is supplemented with a survey that is distributed to a large community of people who use content management systems such as TYPO3 (<http://typo3.org>). Preliminary outcomes of this requirements elicitation will be discussed in D2.2, where we will use them to inform our conceptual framework.

Here, we describe a pilot study that examines how the categorisation of personal photos changes over time, and outline plans for further studies around the theoretical concepts of community archiving [8] (personal preservation) and organisational memory [3] (organisational preservation).

We chose community archiving as our framework for personal preservation, because after a person's death, the archive of their personal data and items will be curated and maintained by and for a community. Typically, these communities are small (extended families), but sometimes, they can be quite large (as in the archives of former US presidents). Thus, community archives provide a good basis for simulating long-term case studies.

From the point of view of the person who owns the digital artifacts to be archived, we will investigate to what extent these artifacts can be passed on, or whether they should be destroyed in the event of death [17]. Issues of personal legacy or the digital afterlife are an important topic in personal preservation; the ongoing debate is captured in sites such as <http://www.thedigitalbeyond.com>.

Theories of organisational memory form the framework of our organisational preservation studies because as we will see below in Section 2.2.2 organisational memory is highly multifaceted. Much of the information in organisational memory is implicit. For example, the marketing department of an organisation may know who sees and edits a press release because the team knows “how things are done around here”, but there might be no material traces of who made which edits, such as an in-document version history, and no procedural manual.

2.1 The Festival Study

In August/September 2013, the Edinburgh team carried out a pilot study that addressed the following research questions:

1. How do people categorise the photos they take?
 - To what extent does this categorisation change over time?
 - Does the point in time at which people categorise photos matter when it comes to granularity and structure of categories?
2. How do people decide what to keep and what to delete, and do these decisions change over time?
3. To what extent do self-reported practices of taking photos agree with actual photo taking practice?
4. How generalisable are categorisations?
 - Are there any overarching, reliable categories that can be determined by image analysis and contextualisation algorithms developed in WPs 4 and 6?
 - Would others categorise the same set of photos in a similar way as the person who took the photos?

As the subject matter of the photos, we chose the Edinburgh Fringe Festival, more specifically the street fair on the Royal Mile in Edinburgh, UK, which is dominated by performers and artists during the Fringe (c.f. Figure 1). The Fringe Festival is one of the largest arts festivals in the world (see also http://en.wikipedia.org/wiki/Edinburgh_Festival_Fringe). While it is generally forbidden to take photos of shows, it is possible to take photos of street artists, which is why we focused on this type of performance. The Fringe is a yearly event and occurs three times in the lifetime of ForgetIT—August 2013, August 2014, and August 2015. By starting to collect material now, we can build an archive of research photos spanning multiple years that is linked by a common theme and contains material that can be distributed for research purposes.

Participants were asked to spend an hour on the Royal Mile and take a photo every three minutes using a simple Android smartphone (Huawei Ascend G300). Photos were automatically tagged with time and date by the phone software, and an interval timer was used to alert participants that it was time for a new photo. This yielded 12–20 photos per participant, because people sometimes failed to hear the timer in the noise of the street fair. Participants were divided into four groups that differed with regard to the delay between taking those photos and revisiting them. Table 1 shows an overview of the groups.

After taking their photos, participants returned to the experimenter, who downloaded their photos and asked them to recall what they had seen, without being able to use photos as cues. Groups 1–3 were then asked to go through their photos and sort them into categories. Participants were free to create as many groups as they wished. Having



Figure 1: A Street Performer at the Edinburgh Fringe on the Royal Mile. Photo by Jens Dibbern, August 1999, licensed under Creative Commons CC-BY-SA.

gone through the photos twice, participants were asked to name or describe their groups. Next, participants were asked to decide which photos to keep and which to delete. For 5 of the keep decisions and 5 of the delete decisions, participants were asked for their reason. Group 4 did not see their photos on the day they were taken.

All participants then returned again a day, a week, or a month later. 56 repeated the recall interview, sorting, deleting, and documenting processes, while the participants who had not gone through their photos already processed their photos for the first time. Participants also completed a survey about their own photography practice and about the ways in which they store, manage, access, and preserve photos.

Table 1: Groups of Participants in the Festival Study

Group	Called Back	First Sort	Number of participants
1	Day later	Day Photos Were Taken	20
2	Week later		18
3	Month later		18
4		When Called Back	18
Total			74

Preliminary results show that this study design yielded an extraordinarily rich data set, with over 1200 categorised photos. Initial analysis shows that on average, people divided their photographs into three main categories, and kept 70% of the photos taken.

We divided the categories into seven classes:

1. performance (depiction of performances and artists)
2. buildings/details of buildings
3. views and scenes
4. fringe context (such as people distributing flyers)
5. location context
6. personal
7. other

When participants sorted their photos for the second time, the number of categories in the “other” class jumped from 15 to 48, whereas the number of categories in the “performance” class fell from 63 to 44, and the categories in the “personal” class from 20 to 9.

In other words, groups that were tied to more external characteristics, such as buildings or characteristics of the location and the event appeared to remain stable, while groups that were related to personal criteria, particularly salient experiences, or other idiosyncratic categories showed considerable flux.

In further analyses, we plan to investigate changes in categories in more detail and to examine changes in keep/delete decisions. We will also perform a follow-up study with participants who have indicated that they would be willing to return in a year’s time. In order to avoid repeated exposure to the photos used in the experiments, which would affect recall, participants will not have access to the photos in the mean time.

Thus, the Festival Study allows us to track how perceived context and perceived preservation value change over time and vary between individuals, results that can be linked to studies of the Personal Photo Preservation scenarios (cf. Section 3) and to the work in WP3.

2.2 Developing New Experimental Paradigms

In addition to building on the Festival Study, we will develop new experimental paradigms that explore the theoretical basis of personal and organisational preservation.

2.2.1 Personal Preservation: Community Archiving of Personal Digital Heritage

Archiving decisions are highly personal and shaped by people's views of their own identity and of the communities to which they belong. Archiving is no longer something that professionals do with resources that they have obtained from others—instead, as Cook [8] argues, archiving is now “[...] a participatory process shared with many in society” (p. 114). Likewise, in the Canadian journal *Archivaria*, Caron and Brown [6] argue that “[w]e are beginning to understand that the construction and constitution of the civic goods of public memory are a collective, social responsibility requiring broad participation across all sectors.” (p. 20). This new understanding of archiving has shaped public policy in Canada, where the Canadian Council of Archives [5] encourages established archives to collaborate with individuals and groups that hold collections of documents and artefacts.

The paradigm of community archiving, as Cook [8] calls it, requires community members to have access to reliable archiving solutions, and for community archives to be stored and organised so that others who wish to access the collection can interpret the documents. This is particularly important for long-term archiving solutions, which allows us to preserve information in such a way that it can be interpreted by others after the death of the person who archived the information originally.

As shown in Figure 2, ForgetIT will not preserve every single detail of an event at time point A.² As an example, let us consider tickets to a comedy show. Not all tickets that were bought will be preserved, but a few sample tickets might be. Some irrelevant details will be lost at the time the event is ingested into the ForgetIT archive, which is time point B in Figure 2. In our ticket example, tickets can be preserved in photos, which removes some information about the texture and quality of the paper that the ticket was printed on. However, the events themselves will be preserved in sufficient detail to support the abstract summary that was stored in human memory. While a person might remember that they saw a particular comedian in the evening during the Fringe, the ticket will preserve information about the date, time, and venue of the show. Seeing this information might trigger further memories, such as queueing in the rain to be let in, or the high price of beer at that particular venue.

In our ticket example, the picture of the ticket will preserve information about the performer, the promotor, the venue, the exact date and time, and the price, whereas people who attended the show may only remember having paid an extortionate amount for a comedy show some time during the Edinburgh Festival in August. After a certain point in time (point C in Figure 2), ForgetIT may further compress the information stored about the ticket if its preservation value is low.

Within WP2, we plan to investigate what makes a photo an effective cue for recalling people and events that were relevant to both individuals and the wider community, building on existing work on lifelogging (e.g. [21, 15]). We will also examine ways of annotating data that reflect categories which are shared within a community. Taken together, these

²The shape of the human forgetting curve for human memory shown in Figure 2 has been experimentally validated [11, 9]

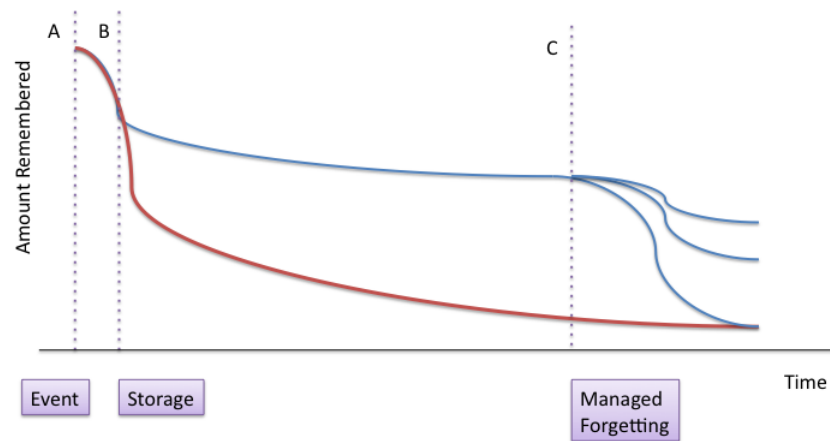


Figure 2: How ForgetIT Supports Human Recall.

Top line: recall with human memory supported by ForgetIT, bottom line: recall from human memory alone.

A: Time point of event, B: Ingestion into ForgetIT, C: Application of different additional managed forgetting strategies

results will provide input to the development of the concepts of contextualised remembering and managed forgetting. More evocative photos that support recall well should have a higher preservation value than photos that are less evocative. Choosing annotations that a target community can understand helps keep data accessible and interpretable once the person who originally archived the data has died or is no longer able to provide the context required.

Related to community archiving, we will investigate issues around sharing and passing on artifacts. When we asked users in the Festival Study why they chose to keep pictures, some said that they wanted to keep pictures for themselves, to help them remember, others decided to keep a picture because they wanted to share it. We aim to clarify the impact of privacy on forgetting algorithms using interviews, focus groups, and participatory design sessions.

The data collected in the Festival Study is a first step in this direction. Since many of our participants were long-time residents of Edinburgh, their categories provide a view of the Fringe Festival that is relevant to the community of Edinburgh and may well be different from that of a tourist, a performer, or a student who is only in Edinburgh for their degree. As the Fringe has been running since 1947, there are also many older photos and other documents for retrospective studies of the same event.

2.2.2 Organisational Preservation: Capturing Organisational Memory

In this section, we outline the WP2 related work on organisational preservation; the application of this work to the ForgetIT Organisational Preservation Use Case will be discussed further in Section 4.1.

Perhaps even more so than individual memory, organisational memory is constantly being constructed and reconstructed as the organization evolves [3]. Hence, organizational memory is substantially different from how conventionally binary data is committed to digital memory and retrieved with precision later. Thus any conceptual view of organizational memory – and by extension any attempt to create technical tools that aid and facilitate in the creation, capture, and retrieval of organizational memory – must embrace the constructive, organic and dynamic nature of it.

There are three main views of organisational memory, cognitive, behavioural, and social [10]. In the cognitive view, organisational memory is seen as a shared mental model of assets, rules, procedures, objects, and practices that are relevant to the organisation. This model is updated as the organisation changes. As a consequence, one is able to take insights gleaned from how individual humans remember and forget, and apply it to organizations and the knowledge dynamics within it.

According to the behavioural view, organisational memory is embedded in repeated behaviours, procedures, and practices within an organisation. The emphasis is on studying these practices and procedures to see whether, when and under what conditions they change.

A third perspective conceptualizes organizational memory as shaped by the social ties and connections of the members of the organization. As organizational memory and recall takes place through social ties, the creation, change, and loss of these ties shapes the overall memory available to the organization. Thus, for instance, a cognitive view may highlight the loss of organization memory embedded in an individual when that individual leaves the organization, while the social perspective emphasizes the loss of connectedness within the organization that this entails – and how this shapes the recall of existing and the creation of new organizational memory going forward.

A comprehensive model of context for contextualised remembering needs to reflect the important dimensions highlighted by and contained in these three broad views. We evaluate the conceptual model of organizational remembering (and forgetting) by studying in-depth organizational memory in three different organizations. Each of these organizations combines a number of distinct qualities of organizational remembering. Through this evaluative process we first assess whether our conceptual model is able to adequately capture all distinct qualities present or whether qualities remain fully or partially uncovered. We further evaluate the conceptual model by comparing the explanatory capabilities of how the dimensions of the model relate to each other with the qualitative data gleaned from the semi-structured interviews of the various actors in the three organizations. Finally, we look at how well the proxies set out in the conceptual model to measure and capture the remembering and forgetting dimensions are functioning in the three case

studies. The conceptual model is successful if it comprehensively captures the dimensions of organizational memory present in the case studies and is able to adequately reflect the organizational dynamic of remembering (and forgetting). Similarly, the proxies are working if they succeed in capturing the salience of the qualities they represent when tested during the case studies. We assume that this will be an iterative process that will continue throughout the duration of the case studies.

3 Evaluating Personal Preservation using the Personal Information Management Model

We now turn to the plans for evaluating the ForgetIT approach in the context of the personal preservation application developed in WP9. Within ForgetIT, the core personal preservation scenario aims to explore the benefits of managed forgetting using the Semantic Desktop ecosystem developed at DFKI [16]³Functionality for preserving personal content will be integrated into the Personal Information Management Model (PIMO) of that ecosystem. We plan two main evaluations using the Semantic Desktop that will tie in with the functionality described in Deliverable D9.2 (Study 1) and Deliverable D9.3 (Study 2).

Both evaluations will be deployment case studies, following existing best practice in evaluating applications for storing and managing photos [14, 2]. In these case studies, participants use a stable version of the system to be tested. They are interviewed extensively before and after exposure to the system, and system logs are analysed to determine usage patterns and problems. In both studies, we will measure efficiency, effectiveness, and user satisfaction using the criteria summarised in Table 2.

One of the benefits of the PIMO approach is the integration with user data via dedicated applications or plug-ins for standard applications such as browsers, address books, calendars, email, or file system. As it is not possible within the time frame of the project to provide plug-ins for the various applications, platforms, and solutions that people use for personal information management, development will focus on creating a small set of plug-ins that can then be used for field studies.

The Semantic Desktop allows extensive logs of user actions, including whether an object has been annotated or viewed, searches, archive access, and any external events that were found through integration with browser and email software (if the users enable this feature).

3.1 Study 1 (Y2): Three Weeks at the Festival

In this study, we test the main short-term benefit of the ForgetIT approach to archiving personal photos—semiautomatic, highly contextualised storage and archiving of photos. For three weeks during the Edinburgh Festival, we will ask 10–20 participants to use the ForgetIT PIMO for organizing the event and their photos. Some users will receive a PIMO that has been pre-initialised based on the results of the Festival Study. This allows us to examine whether it is possible to initialise the PIMO with a generic user model, or whether models need to be built from scratch by individual users.

Participants will use the PIMO to archive photos that they take during the Festival. The

³See D8.1 for a summary of how this component fits into the overall architecture.

Table 2: Evaluation Measures for Personal Preservation scenario. The main unit of analysis is Semantic Desktop objects; photos are special cases of these objects. Effort is defined as amount of time spent and number of steps required.

Dimension	Criteria
Effectiveness	Can users successfully add Semantic Desktop objects to the archive?
	Can stored objects be found in the archive?
	To what extent is context that is relevant for retrieval and interpretation stored?
	Can users successfully set managed forgetting strategies?
Efficiency	How much effort is required for annotating a Semantic Desktop object?
	How much effort is required to ingest an object into the archive?
	How much effort is required to find an object and retrieve it from the archive?
User Satisfaction	Would participants use the Semantic Desktop with ForgetIT functionality again?
	How quickly can users learn to use the system?
	Do users see a benefit from the ForgetIT functionality?
	How do users perceive their own task success?
	How do users perceive the efficiency of the system?

PIMO itself will be evaluated as outlined above. We hope to recall participants after a period of time (between 1 and 6 months) to evaluate different managed forgetting strategies as applied to the photos participants took during the Festival. Specifically, we will test whether photos which serve as effective memory cues are preserved better than photos that were less relevant to users.

3.2 Study 2 (Y3): Conference Travel Season

In this study, we plan to test a Semantic Desktop that is based on the original PIMO work by Maus et al [16] and uses a ForgetIT style archive. The Semantic Desktop will provide applications and plug-ins for general Personal Information Management that covers not only photo management, but also note taking, tasks, calendar, file organization, email, web browsing, and photo collection.

We will ask 3–5 researchers to use the Semantic Desktop application for 2–4 months to archive photos taken during conferences and workshops. We chose this application because many academics produce rich photo data sets during their travels, taking photos of posters and slides in talks (of professional relevance), and of local scenery and outings (personal relevance). This scenario does not require researchers to provide any information about home or family life.

All material that has been ingested into the ForgetIT system will be available through the PIMO and the ForgetIT archive. The study protocol will be designed to ensure appropriate privacy protection. If necessary, materials will be anonymised before distributing them to the consortium.

During the study, we will collect data about use and access of resources that are organised in the Semantic Desktop. These data can be used by the Managed Forgetting component to compute preservation value and memory buoyancy, two concepts that are at the core of our approach to managed forgetting [13]. Since participants will be using a research prototype, we will provide new versions with bug fixes as problems are reported and resolved.

The overall system will be evaluated briefly after each conference or workshop deployment, covering the aspects summarised in Table 2. This evaluation plan covers both ForgetIT-specific functionality and the usability of the PIMO itself, since users are less likely to take advantage of functionality that is difficult to use or access. After the end of the study period, there will be a summative evaluation of the personal preservation scenario and in-depth interviews with participants that probe benefits and problems which emerged during the study. Participants will be interviewed both as a group and individually. During those interviews, we will use suitably anonymised ForgetIT data to elicit comments on interesting patterns of data access and archiving.

We hope to run this study in spring/summer of Year 3, which would enable us to evaluate the managed forgetting module as in Study 1.

4 Evaluating Content Management and Preservation

4.1 The Press Release Use Case

The key use case for evaluating the usefulness of ForgetIT for organisational preservation will be press releases.

At first glance, a press release appears to be nothing but a simple text, but archiving it is more complex than it appears. In order to contextualise it properly, we need to link it to the key event that is being promoted (e.g., release of a web site, signing of a contract), archive drafts, document where the press release was published and how it appeared, and finally track where the press release was reported and how customers reacted to the information in the press release.

Some of these aspects can be aggregated and deleted later. For example, if a press release has been quoted by several people on Twitter, an initial step might be to archive every single tweet. Once the press release has not been linked on Twitter for a while, these tweets can be summarised and deleted from the company archive.

Thus, press releases neatly illustrate the challenges of dynamic preservation, contextualisation, and managed forgetting.

For the development and evaluation of the ForgetIT TYPO3 solution as applied to this use case, we chose action research (as discussed in D9.1). Action research was originally developed to study social change. The model we adopt here consists of four stages, identifying an issue, designing a solution, implementing the solution, and evaluating the solution through quantitative and qualitative data. These stages are repeated several times throughout the project.

The iterative action research approach makes it easy to integrate fresh advances in the study of organisational memory that come from WP2 into the practical design and formative evaluation of the demonstrator system developed within WP10. With respect to the theoretical perspectives outlined in Section 2.2.2, we see press releases as “memories in use” [3]. Loss of organisational memory may present a barrier to the interpretation of press releases that were written years or decades ago. We will use an interim version of the model as documented in D2.3 to reflect on the practices we observe in the two companies, AKO and Spielwarenmesse, that have agreed to test the WP10 ForgetIT solutions.

Action research also fits well with the agile project development methodology adopted by ForgetIT project partner dkd. Agile software development emphasises collaborating with customers and creating working software that can be discussed and changed. Most importantly, there is plenty of scope for plans to change when they no longer fit the requirements.

4.2 The Framework: Action Design Research

The variant of action research we will be using for this study is action design research [20]. In action design research, the focus is on studying a set of design principles that address a class of problems in an innovative way. In this project, we are working towards design principles for facilitating contextualised remembering, synergetic preservation, and managed forgetting in an organisational context. Our use case, press releases, and our context, web site management, is generic enough to define a class of problems that is encountered by any company with a web presence. Although some solutions will be specific to TYPO3, the development environment, analysis will focus on issues that are also relevant for more widely used web content management systems, such as Wordpress (<http://www.wordpress.com>) or Drupal (<http://drupal.org>).

The research problem, or in Action Design Research terms research opportunity, is as follows:

1. How can we enable organisations to archive press releases efficiently for later analysis or reuse? In particular, what are the relevant processes for generating a press release?
2. How can we leverage genre theory to ensure that all relevant aspects of the press release are archived? In particular, what is the relevant preservation context, and how can the necessary information be preserved?
3. How can the findings of the model of organisational memory developed in WP2 inform the design of systems and processes for preserving press releases?

Genre theory [18, 23] defines genres as types of communication with which members of a community are familiar and which they use for interacting with each other. Yoshioka et al. [23] define six aspects questions that help analysts map out genres, why (the reason for communication using a particular genre), what (typical content), where (shared culture, physical and virtual location), when (at what opportunity), who (the people who are communicating), and how (linguistic and visual form).

We are planning six evaluation cycles, which are summarised in Table 3. The first cycle concludes the requirement specification phase and will be documented in Deliverable D10.1. In D10.1, we will also specify the initial requirements for the deployed system and outline the functionality to be developed and tested in each cycle. Specifications will be refined as the project progresses, and any relevant changes will be documented in D10.2 and D10.3.

The next four cycles consist of three month-long units of software development (called sprints in the terminology of Agile project management) which culminate in a release. Before the sixth cycle, where the system is deployed, there will be one or two additional sprints to address concerns that were raised in the last round of user testing. The initial end users involved will be ACO and Spielwarenmesse, the two companies described in D9.1; if development goes well, additional dkd customers may be added.

Table 3: Development Cycles for the Organisational Preservation Use Case. For more information about functionality, see D10.1-10.3

Cycle	End	Release	Evaluator	Evaluation Method
1	M12	0.1	End-User	Wireframes / Concept presentation
2	M15	0.2	Practitioners	mixed methods testing
3	M18	0.3	End-User	mixed methods testing
4	M21	0.4	Practitioners	mixed methods testing
5	M24	0.5	End-User	mixed methods testing
6	M33	0.9	End-User	Observation of deployment of restricted system

As in WP9, mixed methods testing will involve automatic analyses of system logs that track operations of the system on a press release, and user studies with practitioners and end users that will be based on tasks derived from the D10.1 specifications. Tasks will be designed to test whether users can successfully access the functionality added by the ForgetIT framework.

For each of these tasks we will collect quantitative observations (mouse movements, key presses, video analysis, task success, number of steps required to perform task) and user feedback through interviews and structured questionnaires.

Deployment success will be monitored through analysis of system logs, ethnographic observation at the customer site(s), questionnaires, and interviews. In the qualitative work, we will focus on whether users see a benefit from the ForgetIT functionality, and what aspects of organisational memory ForgetIT can be used to store.

We will also track to what extent ForgetIT innovations are picked up and discussed by the TYPO3 community by analysing contributions in relevant TYPO3 mailing lists and forums.

5 Conclusion

In this document, we have outlined an overall evaluation plan for the ForgetIT preservation solutions to be developed in WP9 and WP10 of the project. Detailed study design will be completed based on the information in Deliverables D10.1-3 (Organisational Preservation) and D9.2-4, which will provide clear specifications for the pilot and final versions of the preservation solutions. Additional experimental work in WP2 will provide conceptual and practical underpinnings for the ongoing evaluation work.

The evaluation plan has been designed to provide timely information for application development. Preliminary evaluation results will be reported in Deliverables 10.2/10.3 and 9.3/9.4, with full summative results to be reported in Deliverable 10.4 (Organisational Preservation) and 9.5 (Personal Preservation).

Appendix

Appendix

A Documenting Success

The success of the project will be assessed using the success indicators outlined in the Description of Work. Each success indicator is mapped to a major outcome of the work plan. If outcomes have to be modified, which requires a change to a success indicator are required, this will be indicated both in the project report and in the deliverable where that success indicator is documented.

As far as possible, major outcomes will be documented both in deliverables and in peer-reviewed publications. Here, we only specify the deliverables associated with each outcome.

While *new conceptual models* are described in the relevant deliverables, *implemented functionality* is assessed through both *automated tests* and *user studies*. Automated tests will be conducted both with external benchmark test sets and data collected during the project. User studies will focus on the prototypes developed during WP9 and WP10, as these prototypes provide an end-to-end demonstration of an implemented version of the ForgetIT framework.

B Test Sets

External test sets used in the project to validate algorithms include

- TRECVID SIN for visual concept detection [19]
- TRECVID MED for event detection in videos [19]
- MediaEval SEM for synchronisation of image galleries, clustering and summarisation [7]

For evaluating individual components and models, some data sets are collected independently in individual work packages, such as the Costa Rica photo set (WP9), the Festival Tweet data set (WP3), or the Family Photo Collection data set (WP2). These data sets are described in the relevant deliverables.

In addition to the individual data sets, we will create three larger test sets designed to be used across work packages. The development of these data sets will be documented in Deliverables D2.3/D2.4 (Festival), D10.2/D10.3 (Toy) and D9.3/D9.4 (Stainer).

We expect that these data sets will be used in the cross-sectional studies and in the evaluation of solutions developed in WPs 4, 5, and 6.

The **Stainer** data set spans the life of Peter Stainer, a fictional individual, and includes work, family and travel pictures. This data set will be used in the WP9 cross-sectional evaluation.

The **Festival** data set consists of photos from the Edinburgh Festival taken in 2013 and 2014 for WP2 studies. These photos are annotated with binary preservation value (keep/delete) and content descriptors. Fully transcribed interviews and tweets (for 2013) provide further rich context.

The **Toy** data set consists of news, press photos, and press releases related to the toy business. This data set covers the organisational preservation use case and will be seeded by real press releases, news items, and photos provided by our case study partner Spielwarenmesse Nürnberg. Relevant documents and other digital assets will be defined in collaboration with WP10. The data set will be designed to cover evaluation of concepts related to the toy industry through the years.

All data sets can be split to demonstrate the ability to integrate two or more preserved data sets. For each data set, we will define an a priori split into data that is in the active system and archived data, which will allow us to test preservation functionality.

C User Studies

To ensure a comprehensive evaluation, we plan three sets of user studies:

1. Summative long-term evaluation of ForgetIT demonstrators developed in WP9 and WP10 with a small set of users (described above).
2. Summative cross-sectional evaluation of ForgetIT demonstrators developed in WP9 and WP10 (described below in Sections C2 and C3)
3. User studies based on social media platforms where it is possible to obtain longer-term data about use and access of documents, images, and videos

The social media-based user studies will be described in the Deliverables of the relevant work packages. Here, we focus on user studies that evaluate the ForgetIT system as a whole.

In the cross-sectional evaluation studies, all testers of a demonstrator system initially work with the same data set, and they will only use the systems 1–3 times. For both cross-sectional studies, we aim to recruit around 30 users from a variety of backgrounds.

The demonstrator systems will run in a test bed; storyboards, screenshots, and videos document how the components of ForgetIT were integrated in the end-to-end solutions represented by the demonstrators.

C1 Performance Assessment

To ensure compatibility across the longitudinal Personal Preservation user study (WP9), the mixed-methods evaluations of implemented prototypes (WP10), and the cross-sectional studies, the core assessment methods for all studies will be as follows.

First, we will define tasks that showcase Managed Forgetting, tasks that highlight Contextualised Remembering, and tasks that assess Synergetic Preservation. These tasks will be finalised towards the end of Year 2 and documented in Deliverable D2.3.

For each of these tasks, we will use task analysis to define the minimal number of clicks and keystrokes required for completion. For the cross-sectional evaluation studies, where all users start off with the same data set, we will also provide gold standard solutions.

Effectiveness is the user's ability to successfully complete a task. Success will be measured in four ways:

- presence / absence of crashes or system errors
- whether users quit the task
- whether users believed they had been successful
- when a gold-standard "correct" solution exists, similarity of the user's solution to the gold standard, for example number of correct items retrieved, number of items correctly interpreted

For the cross-sectional studies and some of the mixed-methods evaluations for the organisational use case, users will work with pre-specified data sets. In these cases, through careful design of tasks and data set, we can ensure that each task is associated with a gold standard solution that allows us to objectively evaluate success.

For studies where users manage their own data, it will not be possible to score users' performance against correct solutions, therefore, we need to rely on perceived success.

Efficiency measures how long it takes users to complete a task. For long-term deployments, this will be measured through time-stamped system logs. For the cross-sectional studies, we will also log the number of keystrokes and clicks / taps required for each task. We will also assess to what extent the steps taken by the user deviated from the optimal solution.

User Satisfaction is a multifaceted construct. For ForgetIT, we plan to use two well validated and widely used questionnaires, the System Usability Scale (SUS, [4]) and the NASA TLX Workload index [12]. Results from Item 1 of the SUS, "I think that I would like to use this system frequently." will be reported separately, as well, as this indicates potential for further uptake of the solutions. Normative data for the SUS has been compiled by Tullis and Albert [22].

C2 Personal Preservation: Cross-Sectional Study

Scenario: Peter Stainer has had a stroke. Stainer has filed many of his personal documents using his Semantic Desktop. Stainer had just returned from a family vacation to Edinburgh. While his own documents and photos were already archived, he still had to add his family's photos and documents to his own Semantic Desktop. There is also data on a backup medium that still needs to be integrated into the Semantic Desktop.

The baseline data set for this scenario is the Stainer data set described in Section B

Tasks in this scenario include:

- retrieve key photos and documents from Peter's life, including the last vacation, to support speech and language therapy
- add unfiled photos and documents to Peter's Semantic Desktop

Participants will complete tasks with relevant ForgetIT functionality enabled or disabled. This will help us quantify the extent to which ForgetIT reduces the time required to complete tasks which require searching through a personal archive.

C3 Organisational Preservation: Cross-Sectional Study

Scenario: You recently started your new job in the digital marketing department. You never met your predecessor, but you have access to press releases and digital assets that were archived and managed using TYPO3, extended with ForgetIT functionality.

Tasks in this scenario include:

- finding suitable image assets for a new press release
- tracing the evolution of the context used in annotating assets and press releases
- Business intelligence—what were successful marketing campaigns, which of the digital assets are valuable?

Participants will have some experience with marketing and/or web content management, but will not necessarily be familiar with TYPO3.

D Outcomes and Success Indicators

This section summarises key project outcomes, the associated success indicators, and the deliverables where those success indicators will be documented. Some key outcomes will be cross-referenced in several deliverables. Where outcomes are associated with

several deliverables from the same work packages, deliverables will show progress in reaching the desired outcomes.

WP2

Outcome	Indicator	Del.
1) Principles of human and organisational preservation and forgetting that can be implemented in digital systems	conceptual principles	D2.2
2) User evaluation framework	number of users; analysis of results	D2.3/D2.4
3) Human information preservation and forgetting when interacting with digital systems	revised conceptual models	D2.4
4) Conceptual framework for organisational/societal forgetting	in-depth case studies and validated conceptual framework	D2.4

WP3

Outcome	Indicator	Del.
1) Conceptual process for managed forgetting incorporating WP2	conceptual process	D3.1-D3.3
	complementing human memory and meeting human expectations	D3.1-D3.3
2) Information assessment framework	evaluation of effectiveness in summative evaluation and cross-sectional studies	D3.2-D3.4, D2.4, D9.4, D10.3
	method for selecting features for information assessment	D3.2-D3.4
3) Forgetting strategies and policy framework	ability to express constraints and strategies required by both use cases	D3.3-D3.4, D9.4 and D10.3
4) Implemented managed forgetting process	implementation and integration of concepts	D3.3-D3.4, D9.4, D10.3
	evaluation of user satisfaction	D3.3-D3.4, D2.4, D9.4, and D10.3

WP4

Outcome	Indicator	Del.
1) Textual Similarity and Redundancy	ability to achieve deep understanding	D4.2-D4.3
2) Visual Quality, Similarity, Redundancy	ability to detect undesirable artifacts	D4.2-D4.3
	similarity assessment	D4.2-D4.3
	number of information dimensions during clustering	D4.2-D4.3
3) Semantic multimedia analysis	ability to detect concepts	D4.2-D4.3
	ability to detect complex events	D4.3-D4.4
4) Information condensation and consolidation	ability to summarise documents and multiple documents	D4.2-D4.4
	ability to combine all of above	D4.2-D4.4
5) Evaluation of information condensation and consolidation	Internal evaluations	D4.2-D4.4
	External benchmarking activities	D4.3-D.4.4

WP5

Outcome	Indicator	
1) Workflow for transition between active system and archive	seamless transition based on forgetting methods	D5.3/D5.4
2) Pre-ingest function for SIPs, integrating WP6 work	degree of automation of SIP generation	D5.3/D5.4
	degree of automation of SIP quality control	D5.3/D5.4
	ease of integration	D5.3/D5.4; D9.3/9.4; D10.2/D10.3
3) Re-contextualisation for DIP	degree of automation	D5.3/D5.4
	number of types of content supported	D5.3/D5.4
	flexibility with respect to context	D5.3/D5.4
4) Support for communicating major changes in active system to preservation system	types of change that can be communicated	D5.3/D5.4

WP6

Outcome	Indicator	
1) Conceptualisation of contextualisation	Framework for context modelling	D6.1-D6.3
2) Contextualisation tools	context identification modules	
	component delivery	D6.2-D6.4
	integration into WP9 and WP10	D9.2-D9.4, D10.2, D10.3
3) Techniques for dealing with information evolution	types of evolution supported	
	degree of context change that can be supported	D6.3-D6.4
4) Evaluation	Internal evaluations	D6.3-D6.4

WP7

Outcome	Indicator	Del.
1) Architecture and open interface for computational storage	Interfaces specification	D7.3/D7.4
2) Storlet Engine in OpenStack	working prototype Validation of the advantages of the Storlet Engine in reducing bandwidth, improving security and reducing cost	D7./-D7.4 D7.3/D7.4
3) Dissemination within storage and preservation communities	involvement in storage standardisation efforts	D7.3/D7.4

WP8

Outcome	Indicator	Del.
1) Preserve and forget reference model	reference models adoption in work packages adequate format for content and meta data	D8.2/D8.5 D3.3/D4.3/D5.3 D6.3/D7.3/D9.3 D10.2 D8.2/D8.5
2) preserve and forget framework	interfaces and protocols from technical WPs effectiveness of integration approach and tasks available facilities for managing software development	D8.3/D8.4/D8.6 D8.3/D8.4/D8.6 D8.3

WP9

Outcome	Indicator	Del.
1) Semantic Desktop personal information preservation strategy	documented use cases	D9.5
	documentation of preservation strategy	D9.5
	integration of Semantic Desktop into preservation	D9.3, D9.4
2) Prototypes	preserving personal desktop, used in summative cross-sectional evaluation	D9.3, D9.4
	preserving mobile information assistant, used in summative longitudinal evaluation	D9.4
3) Evaluated best practice of personal preservation and knowledge modelling	User satisfaction in summative evaluation	D9.4
	Workload assessment and efficiency measures in summative evaluations	D9.4

WP10

Outcome	Indicator	
1) Documented organisational archive and presentation needs	persona models available	D10.1
	user stories available	D10.1
2) TYPO3 pilot application	list of features implemented in pilot	D10.2/D10.3
	list of features from other work packages integrated	D10.2/D10.3
3) Open source TYPO3 extension	uptake by TYPO3 community	D10.3
	summative evaluation results	D10.3
4) evaluation documentation	completed design	D10.2
	number of users	D10.3
	insights into applying ForgetIT to organisational preservation	D10.3

References

- [1] ISO Ergonomic Requirements for Office Work With Visual Display Terminals (VDTs), Part 11: Guidance on Usability, 1998.
- [2] M. Ames, D. Eckles, M. Naaman, M. Spasojevic, and N. House. Requirements for mobile photoware. *Personal and Ubiquitous Computing*, 14(2):95–109, June 2009.
- [3] L. J. Bannon and K. Kuuti. Shifting Perspectives on Organizational Memory and Active Remembering. In *Proceedings of the 29th Annual Hawaii International Conference on System Sciences*, pages 156–167, 1996.
- [4] J. Brooke. SUS: A Quick and Dirty Usability Scale. In P. Jordan, B. Thomas, B. Weerdmeester, and I. McClelland, editors, *Usability Evaluation in Industry*. Taylor and Francis, London, UK, 1996.
- [5] Canadian Council of Archives. Newsletter 68, 2011.
- [6] D. J. Caron and R. Brown. The Documentary Moment in the Digital Age : Establishing New Value Propositions for Public Memory *. *Archivaria*, 71(Spring):1–20, 2011.

- [7] N. Conci, F. D. Natale, and V. Mezaris. Synchronization of multi-user event media (sem) at mediaeval 2014: Task description, datasets, and evaluation. In *Proceedings of MediaEval 2014 Workshop Barcelona, Spain*, 2014.
- [8] T. Cook. Evidence, memory, identity, and community: four shifting archival paradigms. *Archival Science*, 13(2-3):95–120, June 2012.
- [9] S. Della Sala, editor. *Forgetting*. Psychology Press, Hove, UK, 2010.
- [10] M. Easterby-Smith and M. A. Lyles. In Praise of Organizational Forgetting. *Journal of Management Inquiry*, 20(3):311–316, May 2011.
- [11] H. Ebbinghaus. *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie*. Duncker & Humblot, Leipzig, Germany, 1885.
- [12] S. G. Hart and L. E. Staveland. *Human Mental Workload*, volume 52 of *Advances in Psychology*. Elsevier, 1988.
- [13] N. Kanhabua, C. Niederée, and W. Siberski. Towards concise preservation by managed forgetting: Research issues and case study. In *Proceedings of the 10th International Conference on Preservation of Digital Objects*, iPres '2013, 2013.
- [14] T. Kindberg, M. Spasojevic, R. Fleck, and A. Sellen. How and Why People Use Camera Phones. Technical report, 2004.
- [15] S. E. Lindley, M. Glancy, R. Harper, D. Randall, and N. Smyth. “Oh and how things just don’t change, the more things stay the same”: Reflections on SenseCam images 18 months after capture. *International Journal of Human-Computer Studies*, 69(5):311–323, May 2011.
- [16] H. Maus, S. Schwarz, and A. Dengel. Weaving Personal Knowledge Spaces into Office Applications. In M. Fathi, editor, *Integration of Practice-Oriented Knowledge Technology: Trends and Prospectives*, Berlin / Heidelberg, 2013. Springer.
- [17] W. Moncur and A. Waller. Digital inheritance. In *RCUK Digital Futures*, Nottingham, UK, 2010.
- [18] W. J. Orlikowski and J. Yates. Genre Repertoire: The Structuring of Communicative Practices in Organizations. *Administrative Science Quarterly*, 39:541–574, 1994.
- [19] P. Over, G. Awad, M. Michel, J. Fiscus, G. Sanders, W. Kraaij, A. F. Smeaton, and G. Quéenot. Trecvid 2013 – an overview of the goals, tasks, data, evaluation mechanisms and metrics. In *Proceedings of TRECVID 2013*. NIST, USA, 2013.
- [20] M. K. Sein, O. Henfridsson, S. Purao, M. Rossi, and R. Lindgren. Action Design Research. *MIS Quarterly*, 35(1):37–56, 2011.
- [21] A. J. Sellen, A. Fogg, M. Aitken, S. Hodges, C. Rother, and K. Wood. Do life-logging technologies support memory for the past? In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07*, page 81, New York, New York, USA, Apr. 2007. ACM Press.

- [22] T. Tullis and B. Albert. *Measuring the User Experience*. Morgan Kaufmann, Burlington, MA, 2008.
- [23] T. Yoshioka, G. Herman, J. Yates, and W. Orlikowski. Genre taxonomy: A knowledge repository of communicative actions. *ACM Transactions on Information Systems*, 19(4):431–456, Oct. 2001.