



[www.forgetit-project.eu](http://www.forgetit-project.eu)

## ForgetIT

Concise Preservation by Combining Managed Forgetting  
and Contextualized Remembering

**Grant Agreement No. 600826**

### **Deliverable D9.1**

<b>Work-package</b>	WP9: Personal Preservation & WP10: Organizational Preservation
<b>Deliverable</b>	D9.1: Application Use Cases & Requirements Document
<b>Deliverable Leader</b>	Heiko Maus, DFKI
<b>Quality Assessor</b>	Mark Greenwood, USFD
<b>Estimation of PM spent</b>	8
<b>Dissemination level</b>	PU
<b>Delivery date in Annex I</b>	M6 (July 2013)
<b>Actual delivery date</b>	August 13th, 2013
<b>Revisions</b>	2
<b>Status</b>	Final
<b>Keywords:</b>	Personal Preservation, Organizational Preservation, Scenario, Requirements

**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
DFKI	Heiko Maus
dkg	Olivier Dobberkau
UEDIN	Maria Wolters
LUH	Claudia Niederée



# Contents

<b>List of Authors</b>	<b>3</b>
<b>Contents</b>	<b>5</b>
<b>Executive Summary</b>	<b>7</b>
<b>1 Introduction</b>	<b>8</b>
<b>2 Personal Preservation</b>	<b>9</b>
2.1 Current State of Personal Preservation and Main Obstacles . . . . .	9
2.1.1 How does ForgetIT address this? . . . . .	11
2.1.2 Motivation for Personal Information Management using the Semantic Desktop Approach in ForgetIT . . . . .	11
2.1.3 How will the Semantic Desktop approach contribute to ForgetIT? .	13
2.2 Relevant Findings from Pilot Survey Results . . . . .	15
2.3 Requirements collected from ForgetIT for Characteristics of Application Scenarios . . . . .	19
2.4 Application Scenario Personal Proactive Diary . . . . .	22
2.4.1 Motivation for the Application Scenario . . . . .	22
2.4.2 The Personas: Meet the Stainers . . . . .	22
2.4.3 S.E – Content for the diary . . . . .	23
2.4.4 S.R – Reminiscence . . . . .	25
2.4.5 S.I – Interests . . . . .	26
2.4.6 S.N – Noise . . . . .	26
2.4.7 S.P – Accessing Preserved Material . . . . .	26
2.5 Requirements for Personal Preservation . . . . .	28
2.5.1 R.PIM – Personal Information Management . . . . .	28
2.5.2 R.F – Forgetting . . . . .	30
2.5.3 R.R – Remembering . . . . .	31

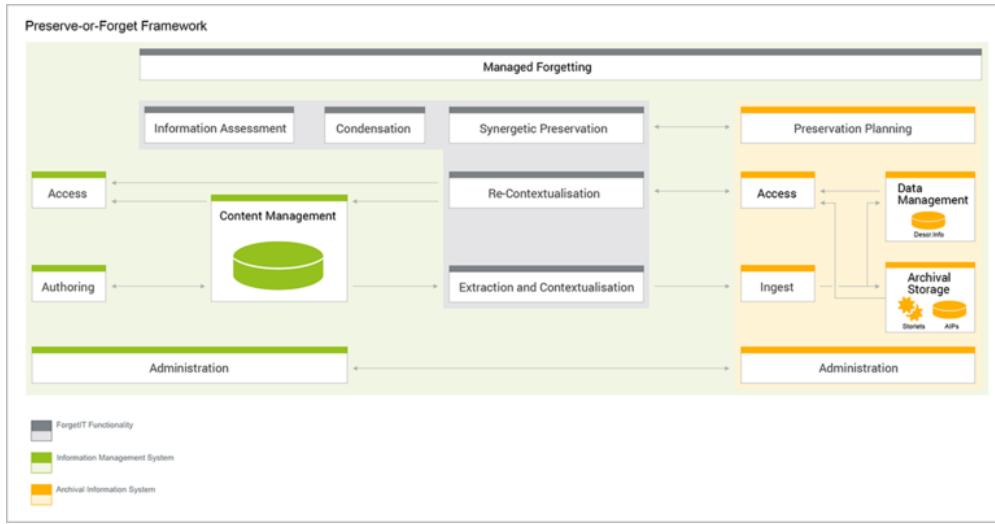
2.5.4 R.P – Preservation . . . . .	31
<b>3 Organizational Preservation</b>	<b>34</b>
3.1 Organisational Preservation in ForgetIT . . . . .	34
3.2 Conceptual Model for Requirements . . . . .	35
3.3 TYPO3 CMS . . . . .	36
3.4 Pilot Interviews . . . . .	37
3.5 Use Cases . . . . .	39
3.5.1 Use Case Actors . . . . .	39
3.5.2 Use Case: Synergetic preservation for TYPO3 . . . . .	40
3.5.3 Use Case: Preservation-aware Digital Asset Management . . . . .	42
3.6 Working with Users—Going Forward . . . . .	44
<b>4 Conclusions</b>	<b>46</b>
<b>Appendix A Survey Scales</b>	<b>47</b>
A.1 Items to Assess Storage, Use, and Management of Digital Photos . . . . .	47
A.2 Attitude to Technology . . . . .	48
<b>References</b>	<b>50</b>

## Executive summary

In this document, we summarise the results of initial work on design requirements for the ForgetIT Personal Preservation solution (WP9) and the Organisational Preservation solution (WP10). These requirements will guide the development of ForgetIT and finally show the benefits of the services in these scenarios.

The Personal Preservation scenario is based on a Personal Information Model (PIMO) which is represented through a flexible ontology that evolves with the user's needs. The core application of the PIMO will be personal information management in personal life situations such as managing personal photos, which will allow to provide the ForgetIT services as an integral part and thus reduce the effort for preservation. The requirements were informed by a top-down analysis and an online survey that was distributed to team members, friends, and colleagues in Sweden, Germany, the UK, and Italy.

The Organisational Preservation scenario focuses on managing and preserving the content (web pages) and assets (such as media files) of company web sites in a way that creates immediate benefit and fosters the use of preservation solutions. The Content Management System (CMS) we chose for our implementation is TYPO3, an open source system that is widely used in Germany. We have designed a framework for further work based on a Web Content Management Lifecycle model. Based on experiences with potential users, it was decided to, initially focus on two use cases, namely Synergetic Preservation for TYPO3 and preservation-aware digital asset management. Instead of a detailed list of requirements, we have opted for an action research model, where relevant changes are co-designed with the two companies who have agreed to trial ForgetIT solutions and evaluated in situ.



**Figure 1: ForgetIT architecture**

## 1 Introduction

Preservation and forgetting are the main goals in ForgetIT. To show the range of applicability of the methods and technologies to be developed as well as to derive requirements from real world scenarios, work packages 9 and 10 address two different areas: WP9 considers personal preservation, i.e., preservation for everyone at home, whereas WP10 considers preservation in organizations. This deliverable describes the envisioned application use cases for both work packages.

For realizing the application scenarios, we will use the Preserve-or-Forget framework as shown in Fig. 1. Both application scenarios address a specific information management system (in green). For personal preservation this will be an implementation of the Semantic Desktop paradigm for knowledge management that is underpinned by a Personal Information Model which represents relevant concepts from the user's mental model and relations between concepts (PIMO). For the organizational preservation, our implementation will be based on the TYPO3 Content Management System for managing web site content.

The deliverable will report on research that has informed initial application scenarios, and identify requirements for all components of the ForgetIT system, including managed forgetting (in grey) as well as the archival information system (in orange).

The deliverable is organized as follows: Section 2 focuses on the Personal Preservation application scenario whereas Section 3 considers the Organizational Preservation application scenario. A conclusion will indicate the next steps for both work packages.

## 2 Personal Preservation

In the following the personal preservation application scenario gets motivated and requirements for ForgetIT are derived.

### 2.1 Current State of Personal Preservation and Main Obstacles

When considering preservation for personal, non-professional, or home usage, we have to contend with a vast increase in ways to create digital artefacts (computers, smartphones, tablets, digital cameras, ...) as well as an ever increasing amount of storage for this digital material. These days, users' personal information space consists of a substantial number of information objects connected to the person's life such as wedding videos, travel pictures, or graduation keepsakes. It requires serious dedication and cognitive effort to organize all this data and keep it accessible as time passes.

Moreover, these digital artefacts often represent past moments but are not associated with a physical memento. Therefore, they form a valuable resource for the user and future generations. If the material is lost or corrupted due to improper conservation, it will be useless. Most users still use backups as their main form of preservation. John et al. in [1] surveyed 2600 academics and members of the digital public about their preservation strategies. 60% of the respondents relied on backups (p. 42). If data was lost, which happened to 30% of the participants, the most common reason for the loss (70% of all cases) was inability to find the files again (p. 23).

In addition, many people keep everything. Marshall [2] points out five main reasons:

- It is difficult to assess value in advance;
- Keeping everything aligns well with current practice;
- Deletion is itself a cognitively demanding exercise;
- People are rarely methodical about culling their files, so why even try; and
- A full chronological and contextual record is essential for using one's archives as a memory prosthesis.

There are preservation guidelines aimed at the general public that show how to go beyond backups. For example, the Library of Congress raises awareness of personal archiving solutions on their web site and provides practical information [3, 4]. The recommended steps<sup>1</sup> are:

1. Identify what you want to save,
2. decide what is most important to you,

---

<sup>1</sup>There are some specializations depending on the media, e.g., video or social media.

3. organize the content (descriptive filenames and folders),
4. save copies in different places,
5. manage your archive (including migration plans).

This recommendation leaves all steps to the user, i.e., what to save, how to organize, where to store (hard disk and online storage), and when to migrate. All this is a lot of effort for users, various decisions need to be made and it requires discipline in, e.g., maintaining and updating the archive. Creating a structure for preservation in particular is one of the major problems. Every person who adds material has to follow this structure every time further material is added, and people who want to search for files need to be aware of this structure. After a long period of time, someone else such as descendants need to be able to interpret the structure.

This cognitive up-front effort is one of the reasons why the cloud storage offered by DropBox, Microsoft SkyDrive, or Google Drive is not a preservation system in itself, but only a tool in a larger preservation strategy. Started as syncing, file sharing, and backup solutions, those services offer organisation methods such as file folders or (keyword-) tags, but do not comply with the OAIS standard [5]. Other services, such as Amazon Cloud, comply with OAIS, but do not support users before ingesting data into the store. Either way, users are left to their own devices for large parts of the preservation process. An overview of the preservation functionality of major cloud services can be found in [6].

Another issue is preserving social media content. Social media services are typically developed by engineers and interface designers. While service providers carefully store users' data for in-depth analytics, they often do not provide any support for dedicated archiving and preservation. Even when archivists are called in at a later stage, preservation is not optimal, since this is not part of the business model [7].

A recent survey [8] shows that social media users often post information about contemporary issues and lifestyle, with another cluster of data that includes travel, festivities/parties and funny events. Users tend not to share information about private topics, such as their relationships, or identifying information, such as birthdays or places of residence.

Many services have sprung up around curating social media content in a form that is easily accessible, and would lend itself well to further preservation. At the most basic level, Twitter allows users to download a compressed file that includes all of their tweets, which can then be viewed in a browser. The Storify service allows users to curate conversations on Twitter - tweets that are replies to each other are collated into a single web page and can be annotated further [9].

Considering the current state of personal preservation, the main obstacles we see so far are:

- Users are **not aware of personal preservation** of digital content. There is a huge gap between current practices, such as backup by copying material to a different hard disk, and a proper preservation strategy.

- When starting with personal preservation, the user faces **high up-front costs** in terms of time, effort, and resources, and there are **very few tools** to help users prepare material for preservation and interact with an archiving service.
- There is no **personal preservation service** for the majority of end users which supports the whole preservation process. Cloud storage alone is not preservation.
- The **vast increase in digital content with relevance to a person's life** poses challenges to personal information management as well as preservation.
- Designing and organising an archive so that its **structure can be understood in a century from now** is cognitively challenging for users.

### **2.1.1 How does ForgetIT address this?**

Within the ForgetIT framework, we will address these challenges as follows.

**Synergetic Preservation:** Most personal preservation should be automatic. After deciding on a preservation policy and the amount and cost of storage, the ForgetIT system then determines what to preserve and how to represent it in a way that will be accessible to future generations. User intervention will be minimal, and therefore the burden on the user will be low.

**Managed Forgetting:** As material accumulates over time, not everything can be preserved. As material becomes less relevant, it is gradually forgotten, using a process inspired by features of human forgetting.

**Contextualised Remembering:** Context helps find and access archived material; it is also crucial for interpreting the data that was archived.

In the next section, we will show how a Personal Information Management system can support all three functions based on the ForgetIT approach and therefore address the five obstacles to personal preservation we have identified in this section.

### **2.1.2 Motivation for Personal Information Management using the Semantic Desktop Approach in ForgetIT**

While some users are concerned about preservation, it is not part of most users' regular practice. Preservation requires manual effort and the users need to think about it to actually do it, it poses a cognitive burden on the users.

Therefore, the approach envisioned in ForgetIT for Personal Preservation is to embed it in the user's activities in the personal information space in order to collect material to be

preserved, evidence for preservation values, and triggers for preservation while keeping user involvement minimal.

But how can this be achieved? By concentrating on the Personal Information Management (PIM) of users, we can cover various life events together with associated digital material, usage of the digital material, and evidence for preservation values. For example, we can detect whether a file is only relevant for a certain time frame (such as time tables) or has emotional relevance (such as a picture showing the user's daughter). Furthermore, it is a chance to derive the user's mental model on the contents of the material, and thus, get a means to describe the preserved material from a user's point of view with less effort.<sup>2</sup>

By providing an ecosystem for PIM we can show that collecting material and deriving evidence for preservation is possible. Motivated by the research done in the Semantic Desktop field, by using the Semantic Desktop paradigm in ForgetIT we can

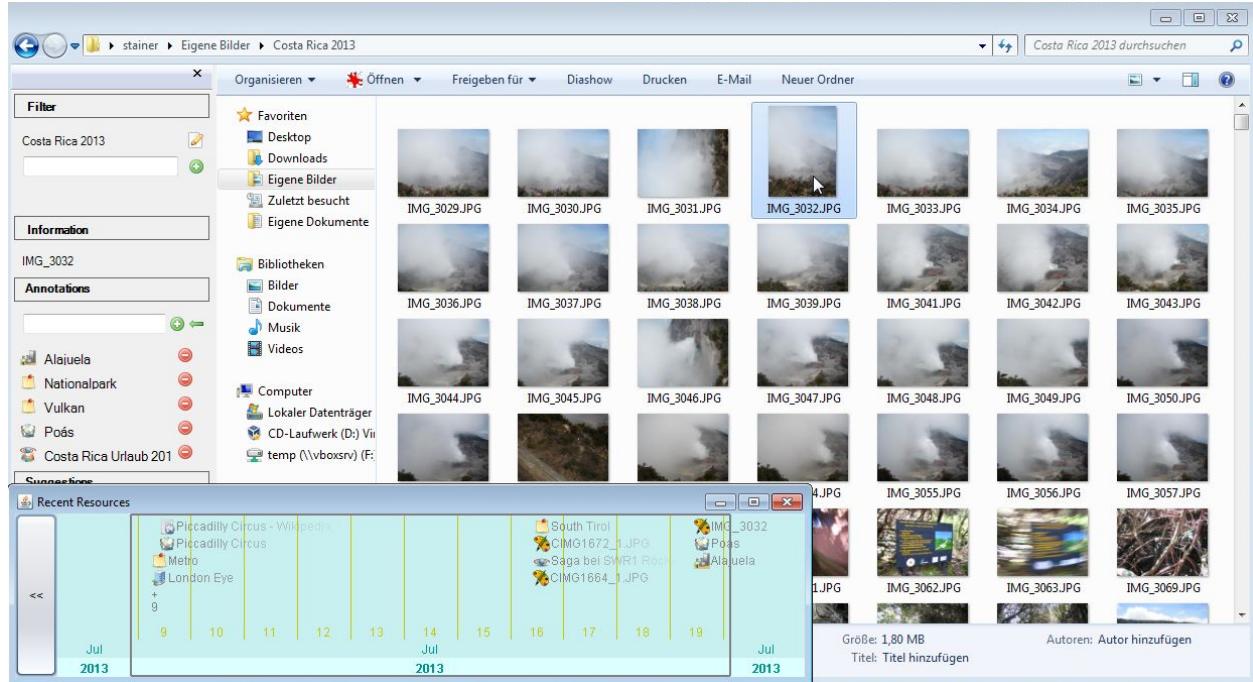
- use the **Personal Information Model (PIMO)** to represent a user's mental model over time. The PIMO [11] is a result of the EU IP Nepomuk (GA 027705) and provides a basic ontology of concepts that a person uses for their desktop and PIM. The PIMO is modelled as a semantic graph of interconnected concepts and information objects. Extensions adapt the ontology to specific domains or tasks.
- provide an **ecosystem of applications and plug-ins** which access the PIMO for vocabulary and knowledge representation. The ecosystem outlined in [12] will be the Semantic Desktop implementation in ForgetIT.
- provide means to **continuously update a user's PIMO** and adapt to new situations. We have shown that the PIMO can be used over time [13, 14]. The oldest PIMO still in use at the DFKI has been evolving steadily for more than 8 years.
- provide **context for information objects** such as files, webpages, or emails by using the PIMO. The PIMO provides the knowledge representation layer both for users and for semantic services [15].
- provide a means to understand – together with observing user actions and access, creation and deletion of information objects – the **context of the user** [16] and provide services such as context-aware task management [17].

In the ForgetIT deliverable D5.1, [18] Section 3.1.2, an initial vision for the integration of the Semantic Desktop in a smooth transition between preservation and active use is given together with the analogy to human forgetting and remembering.

Up until now, research focused on users in a professional setting, e.g., in PIM for research [13] or business administration [19]. This work indicates that people can build up and use rich PIMOs from and in their everyday work. Concepts range from professional topics

---

<sup>2</sup>In ForgetIT deliverable D6.1 [10] a world view for contextualization and de-contextualization is considered as an important ingredient for ensuring interpretation in long-term preservation.



**Figure 2: Example of a Semantic Desktop with the Stainer data set: Annotations of a holiday picture shown by a MS Windows File Explorer PIMO plug-in and a view on recent resources (such as events, topics, locations) from the PIMO in a timeline.**

such as projects, contacts, and meetings to personal concepts such as interests and friends.

We therefore believe that the PIMO has the potential to also support a focused personal use case. As a first step towards addressing non-professional use cases, Klinkigt et al. [20] have bootstrapped a vocabulary to describe photo collections to users.

Fig. 2 shows a first impression of the Stainer data set collected in WP9 with annotated holiday pictures embedded in the MS Windows File Explorer of the user “stainer”. This data set is intended to cover several aspects of the application scenario described in Section 2.4 with realistic data and a PIMO. The depicted file explorer shows pictures and their annotations of a holiday in Cost Rica.

### 2.1.3 How will the Semantic Desktop approach contribute to ForgetIT?

By using a Semantic Desktop approach in ForgetIT, we can support Preservation, Forgetting, and Remembering as follows:

**Preservation:** The Semantic Desktop ecosystem (applications, plug-ins, mobile apps) allows us to connect the PIMO to the user’s information objects through annotating photos and web pages, organizing documents and emails, and managing tasks and reminders.

Information objects are connected by reusing concepts such as contacts, which are part of the PIMO, for annotating pictures and writing emails. The resulting personal information space tightly links resources and concepts. Evidence for preservation values and context for preserving an information object can be derived from this information and formalised using the PIMO knowledge representation. Importantly, the continuously evolving PIMO not only covers information objects in current use but also objects which have already been stored in the archive for later use and are therefore no longer directly accessible for the user.

**Forgetting:** The data about information objects in the Semantic Desktop ecosystem that is held together by the PIMO provides evidence for preservation value, topical and long-term relevance. Observations in the PIMO are similar to files on the computer. For example, while topics of previous projects might still be relevant to the user, most of the associated resources, such as meetings, notes, and presentations, might no longer be of interest. The PIMO and its ecosystem therefore provide crucial input for the managed forgetting system.

**Remembering:** Just like the human brain, the PIMO is still capable to retrieve things which seem to be forgotten. Similar to humans, who can remember things or situations by starting with a cue and then follow associations, PIMO can provide paths through the semantic graph that start from a particular node. For example starting from a project (ForgetIT), we can follow a path to an associated event (the Kick-off Meeting in Hanover) to a photo (the group in front of the town hall) to a person (the professor from UEDIN). At each node along the path, the links from the node to other concepts provide the context required to remember. Thus, the PIMO contributes to contextualized remembering.

In WP9 we consider personal scenarios, the question arises what are the application scenario we want to address and which activities are in there. Therefore, the next section details the requirements collected from the ForgetIT project specification for such an application scenario, and Section 2.4 describes the application scenario.

The goal of this deliverable is to identify and describe a personal preservation scenario for ForgetIT. To get a clearer idea of aspects of such a scenario the next section addresses results and implications of a first survey conducted in ForgetIT on an envisioned personal use case.

## 2.2 Relevant Findings from Pilot Survey Results

In this section, we report relevant findings from the pilot survey on Managing Digital Photos which was conducted online between May 29, 2013 and June 29, 2013. The survey had three aims:

- Get an overview of common search and storage strategies for photos to complement the existing literature
- Develop a scale that allows us to detect user types with different attitudes to storing and managing photos
- Determine whether it makes sense to add validated scales for aspects related to storage and management of personal property, such as conscientiousness [21] and privacy [22, 23].

Here, we focus on the first aim; results on scale development will be reported later in Deliverable D2.1 [24].

### **Design:**

For the user type scale, we selected a total of 30 items based on a content analysis of research on managing digital photo collections [25, 26, 27] that covered attitudes to storing and managing digital photos as well as common practices that can make searching easier, such as tagging. The full list of items is given in Appendix A.1. Each item was worded as a statement; respondents then had to state whether it was very accurate, moderately accurate, neither accurate nor inaccurate, moderately inaccurate, or very inaccurate.

Common search and storage strategies were probed using free text entry fields. We also probed typical reasons for taking photos and searching for photos by asking respondents to rank a list of categories.

### **Method:**

The survey was conducted entirely online using LimeSurvey. The survey questions were published in German, English, Swedish, and Italian. The original version of the survey was written in English and translated into the other three languages by native speakers from the team. Links to language-specific versions were distributed through the ForgetIT team, and via social media. Since the survey was mainly about personal attitudes and practices, ForgetIT team members were encouraged to fill in the survey themselves in addition to forwarding it on to friends and colleagues.<sup>3</sup>.

---

<sup>3</sup>In all descriptive statistics that follow, percentages are rounded to the nearest percentage point, which is best practice when the number of respondents is around 100

**Table 1: Demographics of Survey Sample**

	Language				Total
	German	English	Italian	Swedish	
completed	19	37	16	83	155
% of attempted	56%	54%	76%	60%	59%
% female	10%	59%	0%	52%	43%
Age (years)					
median age	37	31	37.5	49	39
IQR age	28–41.5	25–37	27.75–43.75	37–57.5	32–51.5
age range	21–70	19–72	23–57	17–67	17–72
education (years)	18 ± 4	18.5 ± 4	18 ± 3	18 ± 3	18 ± 3.4
Occupation					
employed	63%	57%	94%	94%	
student	26%	35%	0%	4%	
other	11%	8%	6%	3%	

## Results:

261 users started the survey. 31 people (12%) only called up the start page. 75 participants (29%) filled in at least part of the survey, and 155 (59%) completed it. Table 1 presents the demographics of the completers by language. Due to the large differences in the demographics of the samples, we did not perform any language-specific analyses.

This section will focus on aspects that are relevant to the personal preservation scenario, such as current storage and preservation practices; a more detailed analysis of the results will be presented in Deliverable D2.2.

In their day-to-day lives, most of our respondents (86%, n=133) use a smartphone, and one in four (26%, n=41) use a basic mobile phone. Both types of phone are able to function as cameras, but 8% (n=13) never use that function. Typically, respondents take photos using their phones at least once per week (62%, n=96). Although 86% (n=134) of our respondents own a digital camera, this seems to be reserved for special occasions—74% only use their camera monthly or rarely.

Respondents were highly computer literate, scoring an average of 1.6 ( $\pm 0.8$ ) on our attitude to technology scale, which ranged from -3 for a highly negative attitude to technology to +3 for a very positive attitude. The scale is documented in Appendix A.2.

To help us categorise the content that needs to be stored, managed, and made searchable, respondents were asked to rank photograph content types, reasons for taking photos, and photo search criteria according to their importance. For this analysis, we focussed on the top three options listed, because feedback indicated that the sequence of options after this cut-off point may have been more or less random.

The most common reason for taking pictures was capturing a moment that was important to the photographer—82% named this in their top three reasons for taking photos.

Respondents also often took photos to document events, holidays, achievements or milestones (74%). About half of our sample took photos to share them with family and friends (55%) or to capture the essence of a place or a person (54%).

Respondents tended to take pictures of people they were close to (79%), people they were with at the moment (54%), landscapes (40%), milestones (37%), and scenes in a city (25%).

When searching for photos, 68% mainly looked for people, while 45% wanted to find photos that looked particularly good, and 41% searched for photos that evoked emotions in them. 39% looked for places, and 38% for informal events such as parties or holidays. Only 1 in 5 looked for milestones (21%) or photos they wanted to share with others (19%).

We asked respondents where they stored photos that were around one year old (current photos) and photos that were five years old (older photos). Current photos are commonly stored on a currently used computer (88%) or on CDROM/DVD (62%). A third also kept current photos on their camera (32%). Older photos were preserved mostly on CDROM (68%) and a computer that was currently in use (59%). Only one in four stored photos (both older and current) in printed format. 29% used web- or cloud-based services to store older photos; for current photos, the percentage increased to 39%. 5% used social media such as Facebook. As for photos that are 20 years old or older, these are almost always preserved as prints, either loose or in an album. A few respondents had digitised old photos and two mentioned that they intended to store them in the cloud.

72% (n=112) of respondents shared photos that they had put online or found on the Internet. Of these 112 people, 90% shared photos via email. Facebook is the main social media platform for sharing photos in our sample (83%, versus 21% sharing on Twitter and 6% sharing on LinkedIn). Dedicated photo sites such as Picasa (17%) and Flickr (20%) were used as often as Twitter. Half of our respondents distributed photos they had found online via MMS (48%) or by printing them off (50%).

The average score on the privacy scale [23], which ranged from 1 for not concerned at all to 5 for highly concerned, was 3.2 ( $\pm 0.8$ ), right in the middle. Although 64% (n=99) hardly ever or never read the privacy policies of web sites, two out of three respondents want to know when somebody uploads a photo where they are clearly identifiable onto the Internet (64% agreed or strongly agreed, n=100), 68% support the right to know what data a web site collects about them, and 68% (n=106) support the right to have personal information about themselves deleted.

Looking at the user type scale, we see high overall agreement on five of the 30 items. Respondents rarely tagged people automatically (12%)<sup>4</sup>, used keywords to tag their pictures (12%), or added titles (14%). Only 10% deleted most of their photos. 79% wanted to control who could see their photos.

Users liked to browse old pictures (73.5%). 68% kept copies of pictures before editing

---

<sup>4</sup>Percentages are based on the number of respondents who said that the corresponding statement was moderately or very accurate.

them, and 65% were worried about losing photos that were important to them. 63% used photos to reminisce about people. Only 19% regularly used geolocation data, and 46% knew where to find photos that they were interested in.

### **Discussion:**

It turns out that quite a few of the reasons why people take photos and want to retrieve photos are difficult to quantify using traditional image analysis methods. Most respondents take photos to capture important moments, but what is important depends on the person. Aesthetic judgements are also highly personal. Such information needs to be captured using a network, because what makes a photo memorable is its ability to evoke other memories, i.e. its connection to people and events in autobiographical memory.

Within the Personal Preservation work-package of ForgetIT, we need to find ways of eliciting information about this network that take into account the content of a photo and the extent to which it evokes emotions. Since our users rarely tag or annotate their photos, we need to find ways of inferring these connections that require minimal user input. This could be addressed by algorithms for semantic multimedia analysis in ForgetIT WP4.

Preservation values should be calculated not just based on the object itself, but also based on its links to other objects. Therefore, preservation values would need to be automatically updated as the content of the archive changes.

The preservation practices that emerge in the survey are problematic. Many people use DVDs and CDROMS, which can decay within a couple of years or decades, depending on their quality. There is also evidence that some people systematically store their photo collections on their current computer, and migrate collections as they change computers [1]. While some respondents see print as the ultimate archival medium, others are busy digitising old printed photos to help preserve them. Within ForgetIT, we need to define best practices for personal preservation that suit various budgets of time and money.

## 2.3 Requirements collected from ForgetIT for Characteristics of Application Scenarios

In order to identify an application scenario suitable for showing the benefits of the ForgetIT approach for personal preservation, we started by collecting requirements that allow each work package in ForgetIT the opportunity to address relevant issues in the personal preservation scenario.

Using these requirements, we identified new aspects that should be covered and investigated the suitability of scenarios already under consideration such as trip journal or reminiscence with photo collections.

In the following pages, we list the collected requirements and indicate how they are motivated. Each requirement is listed under the aspect for which it is most relevant. Section 2.4 presents the scenario that was selected based on these requirements.

For referencing purposes, the scenario requirements are named according to the following schema **SR.<Aspect> . <Number>**.

### **SR.PP – Requirements for “Personal” aspects**

These requirements address aspects that are specific to personal preservation.

**SR.PP.1** The scenario is relevant for people who are not specialists in IT, information management, archiving, or digital preservation.

**SR.PP.2** Focus on aspects of a person's private life, although some business-related data can be included as well.

**SR.PP.3** Scenario contains elements personal information management, e.g., calendar, contacts, related documents, emails

**SR.PP.4** Desktop use is part of the scenario (T9.4 in the Description of Work)

**SR.PP.5** Mobile use is part of the scenario (T9.5 in the Description of Work)

**SR.PP.6** Diversity of material and information sources

**SR.PP.7** Large number of resources involved

**SR.PP.8** Context matters; context switches occur

**SR.PP.9** Mental model: classifying things according to this model is helpful (tagging/annotation) and things can be reused in other life situations

**SR.PP.10** An immediate benefit for the user is visible when using ForgetIT (not only in the long-term)

**SR.PP.11** Project-like phases are covered in the scenario (e.g., preparation, execution, wrap-up)

**SR.PP.12** Social media integration make sense (sharing, tweets)

### **SR.P – Requirements related to Preservation**

**SR.P.1** Users will want to retrieve resources some time after ingestion

**SR.P.2** At least some of the resources require long-term preservation

**SR.P.3** Retrieving from archive through contextual links makes sense

**SR.P.4** Resources require different ingestion methods

**SR.P.5** There should be a need to change the service provider

### **SR.F – Requirements related to Forgetting**

**SR.F.1** Semi-automated/managed forgetting is acceptable

**SR.F.2** Relevance of and interest in stored resources typically decreases

**SR.F.3** Resources vary in importance over time

**SR.F.4** Resources can have redundancy

**SR.F.5** Condensing is possible (also landmarks/representatives are possible)

**SR.F.6** Distinction between memory buoyancy and preservation value: case allows to distinguish between things that get currently out of focus (memory buoyancy) and the general long-term value of things (preservation value)

**SR.F.7** Use case allows a rich set of forgetting options (summary, duplicate detection, information condensation, deletion, ...)

### **SR.R – Requirements related to Remembering**

**SR.R.1** Remembering occurs

**SR.R.2** The scenario can be described by a rich set of concepts from the user's mental model, i.e., various concepts can be used to describe a scenario such as persons, topics, locations, events.

**SR.R.3** There should be a reason for bringing archived things back to active context

**SR.R.4** There should be cases where re-contextualization is a challenge (major change of context)

**SR.R.5** Data needs to be remembered after time frames that are longer than a decade

## **SR.G – Requirements from the ForgetIT Project**

**SR.G.1** The combination of forgetting, remembering, and preservation results in a distinct advantage over existing applications

**SR.G.2** Scenario contains text (for WP6)

**SR.G.3** Scenario contains images (for WP4)

**SR.G.4** Data sets combine text and images with possible references to each other

**SR.G.5** Requirements towards the experimental settings (from WP2)

**SR.G.ES.1** Laboratory study where users work with a pre-defined information space/data set is possible

**SR.G.ES.2** Longitudinal experiment that requires people to use a prototype instance of ForgetIT is possible

**SR.G.ES.3** Revisiting the data after several months or years makes sense

**SR.G.ES.4** Individual data collection sessions can be condensed into an hour

## 2.4 Application Scenario Personal Proactive Diary

### 2.4.1 Motivation for the Application Scenario

Given the extensive list of requirements detailed in the previous section, we decided on a main scenario that could accommodate a variety of sub-scenarios.

The main scenario is documenting the life of a person through a personal proactive digital diary. Here, we use the term diary as a metaphor for an application that supports users in selected parts of their lives, and thus, together with the PIMO is able to provide a digital diary. This diary covers various situations (such as events or travel), includes various types of material such as photos or documents, and provides services that range from daily interaction, when data is proactively processed for later preservation, to long-term reminiscence.

This approach has the following benefit:

- people know what diaries are from their personal lives,
- it is easy to explain the main ForgetIT concepts using the diary metaphor,
- a diary is personal which stresses the privacy aspect, i.e., it is not intended as a facebook timeline,
- diaries have a long-term perspective which stresses the ForgetIT view,
- diaries cover a wide range of actions and life situations, thus a chance to cover relevant material to be preserved,
- various mobile apps are already available where aspects of a diary are provided<sup>5</sup> and which show the interest of users in such services, and finally,
- social media activities such as tweeting or sharing photos from a life situation fit nicely into this concept.

The scenario and its sub-scenarios are defined using a set of personas, the Stainer Family. In the following, we introduce the Stainers and list the sub-scenarios we have designed. For referencing purposes, the sub-scenarios are named according to the following scheme **S.<Aspect> . <Number>**.

### 2.4.2 The Personas: Meet the Stainers

**Peter Stainer** likes travelling, taking photos, and his hobby is whisky. He works at a consulting company. The family has a computer and a tablet. Peter manages the family's files on the computer and has a separate hard disk for backups. Peter is married to **Jane**.

---

<sup>5</sup>For instance apps such as Momento, PhotoCal, or Day One in Apple's App Store.

Jane likes music, theatre, and comedy. Peter and Jane have smartphones. Jane has a grandmother in her Eighties. Jane is in her late thirties, Peter is in his mid thirties. They have two **children, Sandra and Tim**, both teenagers.

### 2.4.3 S.E – Content for the diary

These sub-scenarios are designed around situations that

- cover a wide range of situations as well as organizing them,
- provide means for collecting data
- motivate the relevance of the material,
- allow to distribute situations over a longer time period,
- add diversity in the material collection,
- allow to add more events to populate the data set(s) if required in ForgetIT.

There are two main types of situations, trips and events.

#### S.E.1 Trips

Trips cover a variety of user actions and materials, including preparation, travel, and post-processing of files, web pages, emails, photos, and notes. Trip data will be created and processed both on the desktop and on the go.

**S.E.1.1 Holiday trips with family** *The Stainers plan to spend their holiday in X. Therefore, Peter and Jane invest some time in preparing the trip including searching the web for hotels, sightseeing, tips, and arrange the material for the travel such as notes and a bookmark collection. On the travel the material is accessed, photos are taken, impressions are noted. At home, the Stainers want to prepare the photo collection as a means to remember this nice holiday trip.*

The data set will be populated using anonymised photos, e.g., from trips by members of the research team or by sources from the web. A set of trip photos, such as a holiday trip to Costa Rica documented using anonymised private photos (see also Fig. 2), has already been made available to the consortium; more are to follow.

**S.E.1.2 Business trips alone** *Peter Stainer's work allows him to travel to various cities. If possible, he takes some time to explore his destinations and take photos.*

Most of the cities covered here will be locations of ForgetIT meeting locations. We will merge photos taken by the research team with photos from photo sharing sites that are licensed for research under Creative Commons. The first trip in this data set involves travel to Edinburgh, the site of the second ForgetIT meeting.

### S.E.2 Preparing and taking part in events

This sub-scenario covers different life situations of taking part in or preparing an event such as an anniversary, which provides some motivation for personal information management and yields a rich set of material with varying relevance over time.

**S.E.3.1 Anniversary** *Jane is turning 40. The family plans a party. Various things need to be discussed, investigated, decided, and prepared. An invitation has to be written. Jane wants to keep track of tasks and decisions related to the party. She uses notes, calendar reminders, and task lists. She also wants to keep relevant photos and files in order to relive the memory of the event. She would like to access some material related to planning on her mobile (for example, venue details). After the party, she would like to keep selected photos on the mobile to share with friends.*

**S.E.3.2 Taking part in an event** These can be one-off events, such as the 2013 Black Sabbath reunion tour, or regular events, such as the Edinburgh Fringe arts festival. We will be able to collect public domain data such as tweets, photos, or blog posts from many people. These scenarios definitely include a mobile component, because many relevant impressions are generated during the event. Resources which are involved are photos, people, stories, tasks, reminders, appointments, and notes.

Two examples:

**A concert** *Peter hears an announcement of his favourite band Saga to play at a location near his hometown (Rheinland-Pfalz-Tag 2013, Pirmasens). He plans to visit this. He does some research on material on how to get there and what is shown. While there, he checks the program, takes photos of the venue as well as the band (with varying quality) and processes the photos back home.*

**Taking part in the Fringe festival in Edinburgh** *Peter and Jane love going to Edinburgh. Jane regularly goes to the Fringe festival because it allows her to discover exciting new performers. Before her trip, Jane buys tickets for several shows. While she is at the Festival, she buys further tickets based on recommendations and reviews. All the shows Jane considered are documented in her Personal Information Manager. Back home, she looks at the back catalogue of her Fringe favourites and finds out more about them on the Web. Sometimes, this leads her to buy books or DVDs.*

UEDIN is well placed for field experiments featuring the Fringe Festival; a pilot study on documenting street performers is planned for 2013.

## 2.4.4 S.R – Reminiscence

Here, we cover two aspects of reminiscing, contributing to a data set that can be used for reminiscence as well as browsing data for reminiscence. In some cases, the material (photos, notes, etc.) are still available locally, in some cases, all or part of the material needs to be retrieved from the archive.

### S.R.1 Organizing the family photos

*Peter's responsibility is to organize the photos of his family. He does this occasionally. However, it is cumbersome to select the best photos out of a collection of snapshots which contains also mistakes, redundant ones, photos of bad quality, or even boring ones.*

This scenario addresses the challenge of organizing photo collections which is usually a task isolated from the actual event because photos are imported from cameras days, weeks, or months after the event. We also want to encourage those users which store their photos on camera only to use the ForgetIT services. Hence, we have to offer some benefits including the PIMO for organizing photos instead of using file folders only.

### S.R.2 Archiving family history

*On her grandmother's birthday, Jane sees several old photos of her family. Jane would like to make sure that other members of the family can use these photos for reminiscence as well. She organises the photos and gets information about their context from her grandmother and other relatives. She writes some of this information up as stories. Jane also scans additional material such as postcards that are relevant to these stories. All of this material is stored in the PIMO.*

In this scenario we introduce an act of personal preservation known from “analogue” times: preserving family history. Again, this covers several topics of ForgetIT and covers a concern that was mentioned often in our survey, the preservation of photographic documents of family history.

### S.R.3 Telling the children about their grandmother

*The Stainers are home from the funeral of Jane's grandmother. Jane and the children browse photos of her grandmother and reminisce about her, telling stories about her life, about the events when the photos were taken (such as Jane's 40th birthday party), and the places where Jane's grandmother lived.*

This is the complementary scenario to preserving the material in the previous section. Here, we assume that some material is retrieved from the archive, while other material, such as photos from Jane's party, are still local. The PIMO provides important context, in particular for older photos.

## 2.4.5 S.I – Interests

This sub-scenario covers special interests such as Japanese art and hobbies such as weight lifting. People will usually research their area of interest, organise material, and revisit it from time to time. Special interests can inform trips and plans and be the subject of reminiscence.

### S.I.1 Hobby

*After a trip to Scotland, Peter found a new hobby: Whisky. He is fascinated by the taste, the tradition, the different styles and flavours, and the distilleries that sit in the landscape. Therefore, he started to collect whisky himself and began to learn more about whisky. He occasionally searches for material about it. His plan for next year is to make a trip to Scotland to visit several distilleries.*

## 2.4.6 S.N – Noise

This sub-scenario provides material which serves as noise in the overall scenario: Activities of daily living that are of short-time interest which produce material that clutters the personal information management tool. This is required for including normal user behavior, having material of varying relevancy ranging from completely irrelevant to only temporally relevant and likely to be sorted out (i.e., being noise in the data set).

### S.N.1 Buying a new gadget

*The Stainers want to purchase a new smartphone for Jane. They collect information from various web pages, user reviews, photos, and prices. After the decision is made, the material related to the purchase is no longer relevant, except for the receipt, which Jane files in their personal finance system.*

### S.N.2 Assortment of material from personal life

Further ingredients of the file collections of a user are an assortment of material from the personal life such as invoices, school stuff, insurances.

## 2.4.7 S.P – Accessing Preserved Material

This sub-scenario takes care on the actual idea of the preservation approach: Accessing the preserved material in different life situations. This will show the benefits of the synergistic preservation. Accessing the material is close to the reminiscence sub-scenario

in Section 2.4.4, but here we focus on actually retrieving material from the archive on purpose.

### S.P.1 Disaster with family PC

*The Stainers were burgled. All PCs, tablets, smartphones, and hard disks were stolen. Fortunately, the family archive is hosted at an external site. With one click, Peter reinstalls a copy of his PIMO, the archive contents, all relevant interfaces to software such as email and note taking, and recovers his archiving policies. He decides to revise them to ensure that new content is ingested more frequently into the archive.*

In this scenario, we want to focus on retrieving material from the archive after a disaster, highlighting the difference between ForgetIT and backup-specific solutions such as Apple Time Machine. This scenario is also intended to foster project-internal discussions about specifications for such a professional preservation service.

### S.P.2 The Stainers want to change the archive service provider

*After the bankruptcy of their existing provider, Peter and Jane move the core archive to a competitor that also provides OAIS-based services. The new provider offers a special package that includes bespoke archive migration services.*

In this sub-scenario, a change of service provider for the archive is addressed to ensure long-term availability of the archive and access to it. A further aspect here is, that there might not a PIMO application be available anymore. In that case, the metadata from the PIMO describing resources should be readable, e.g., via Semantic Web standards such as RDF<sup>6</sup>, which gives future application developer the possibility still be able to interpret the metadata from the PIMO.

### S.P.3 Peter's children remember Peter after his death

*After Peter died aged 85, his children talk about life with their father. They start to dig around in the annotated photo collection that mainly reflects of the highlights of each period of Peter's life. They add their own photos of Peter's final months in hospice. Sandra recalls how she danced with her father at her mother's 40th birthday party. She cannot find a photo of their dance in the highlights, so she searches the archive and finds the photos she was looking for.*

The focus here is on what happens to the archive after death. Although close to the Reminiscence scenario, the focus here is on the questions of what happens to an archive after death.

---

<sup>6</sup>Resource Description Framework, a W3C standard model for data interchange in the Semantic Web; <http://www.w3.org/RDF/>

## 2.5 Requirements for Personal Preservation

This section lists the requirements derived from the scenario in the previous section. The requirements are organized along the three main areas of ForgetIT, namely Preservation, Forgetting, Remembering, as well as the personal case of Personal Information Management (PIM).

For referencing purposes, the requirements are named according to the following scheme **R.<Aspect> . <Number>**.

### 2.5.1 R.PIM – Personal Information Management

Most people don't like tagging and organising their data. This is similar to knowledge acquisition problems in knowledge management. Therefore, using the Semantic Desktop approach, we embed (largely automatic) knowledge management technology into people's daily work flow (e.g. [28, 17]). In ForgetIT, the Semantic Desktop is used for PIM. In the following we list the requirements from this personal information management perspective.

#### **R.PIM.1** Engage the user in order to get the resources, their context and usage

- For any decisions about archiving, forgetting, and preserving, we need the user to interact with the system. The PIM is presented as a diary that automatically integrates data from different relevant apps such as calendars, email, photo management software, and note management software.

#### **R.PIM.2** Proactively support users in organizing data about their daily life

- The Semantic Desktop paradigm has great potential for ForgetIT (see also Section 2.1.2). In this project, we will integrate a Personal Information Model (PIMO) and create proactive services within the PIMO ecosystem that support this functionality.
- The PIMO allows users to describe data with their own concepts and vocabulary. To make users' lives easier, world knowledge about, e.g., locations, monuments, or events should also be incorporated into the PIMO as and when applicable.

#### **R.PIM.3** Support users with low computer literacy

- The personal digital archive should be open to everybody who generates digital material. Therefore, interfaces need to be highly intuitive and attractive without compromising the underlying functionality.

#### **R.PIM.4** Concepts and resources should be interchangeable between apps

- Interchangeable here means that resources can be used in different applications, a photo from the photo collection app can be added to a task, the task

can be used to annotate a file or email, etc. This ensures that effort spent by the user in one application, e.g., annotating photos, can be used in other applications, thus ensuring that the PIMO grows with the user.

**R.PIM.5** Cope with different computing devices

- Users use mobile devices, smartphones, and tablets as well as desktop computers. Manual syncing requires considerable effort. Therefore, the system should have a cloud component that allows automatic syncing and access across several devices.

**R.PIM.6** Provide a personal cloud infrastructure,

- Service providers could offer customers cloud storage linked to an archive that uses ForgetIT technology. The PIMO server is well equipped for such a service, because it has a dedicated Application Programming Interface which would enable third parties to use the PIMO in their own applications.
- Another solution could be a personal cloud on a server at home. This server should require almost no configuration. This would address privacy concerns about existing cloud services. The main ForgetIT modules (and potentially also the encrypted archive itself) could still reside on a secure server outside the home.<sup>7</sup>

**R.PIM.7** Mobile production and access of information objects shall be part of the preservation strategy

- Mobile devices need to be part of the PIMO ecosystem and they should be covered by the preservation strategy defined for the overall system.
- A potential proactive PIM service would be to offer mobile access to information objects that are likely to be used on the go, such as travel documents to ensure access when offline in a foreign country. We will use PIM information sources such as calendar, tasks, annotations, the sensors of the mobile device, and the forgetting and remembering methods in ForgetIT in order to determine which information objects need to be on the mobile device, and which can be kept in the cloud or on the archive. Objects on the mobile device that have not been used or relevant for a period of time can be automatically pushed off the device onto the cloud.

**R.PIM.8** Accommodate changes in the mental model of the user over time

- With the personal diary application scenario, we address a long-term perspective of involved resources and the mental model. Thus, we need to cope with changes in the PIMO over time, if, e.g., material from the archive is brought back, there might be the need of re-contextualization to ensure the interpretability.

---

<sup>7</sup>For example, OwnCloud (<http://owncloud.org>) offers Personal Information Management solutions that can be installed at home or Networked Attached Storage (NAS) devices which support services such as remote access photo collections and bit torrent clients for downloading files.

**R.PIM.9** Preserve social media as and when needed

- Many social media services, including Facebook and Twitter offer APIs and export services that could be used by a dedicated ForgetIT application to preserve relevant content. WP3 has a dedicated use case that covers the analysis of timelines.

**2.5.2 R.F – Forgetting**

Over time, the Semantic Desktop, the PIMO, and the archive itself risk becoming cluttered with material. The following requirements cover forgetting both for the PIMO/Semantic Desktop and for the archive.

**R.F.1** A managed forgetting process is required for the PIMO and its resources

- Concepts in the PIMO may become outdated, irrelevant, or less important over time. Some concepts will remain relevant, but only in certain contexts, for example when a user revisits work on an old project. A managed forgetting approach modelled on human forgetting appears promising: discard irrelevant detail quickly, make valuable but currently not relevant material less accessible (which prevents information overload), and use associations to recall this material if necessary.

**R.F.2** Ensure users support decisions about forgetting and deletion

- ForgetIT needs to address concerns about accidental deletion of important material. Forgetting policies should allow for different degrees of deletion (e.g., deleted from local phone memory, moving from cloud to archive, making content progressively less accessible in the archive, etc.). We also need to consider to what extent users will want to be consulted, in particular when content needs to be deleted from the archive due to space reasons—or even when moving content off the local hard drive / device memory.

**R.F.3** Develop different forgetting policies for concepts of the mental model (the PIMO) and archival resources

- For data that is indexed by concepts, we can envision different forgetting methods such as information condensation, lighthouse concepts, proxies, and annotate data with relevance, usefulness, and appropriate preservation values.
- For the PIMO, we may want to hide concepts from the main concept tree or archive sub-concept trees that are no longer in use, but may still be needed to interpret associated data that has been stored in the archive.

### 2.5.3 R.R – Remembering

The following requirements can be identified for helping users remember, i.e., find and retrieve data.

#### R.R.1 Support users in remembering

- Users can only recall information they have provided. Within ForgetIT, we support the recall of information that was integrated in the Personal Information Management system. Our key sub-scenarios are recall of information that is relevant in the short term (e.g., for organising a birthday party) and for reminiscence (e.g., talking about the party years later).
- To make this possible for users, we have to provide user interfaces to foster remembering if required, but keep data from view if not needed. Memory buoyancy is a key ForgetIT concept here—objects with a high memory buoyancy will be more likely to be recalled first, and can serve as anchors for retrieving other objects with lower buoyancy.

#### R.R.2 Consider life events as reminiscence triggers

- The survey data has shown that life events and milestones such as marriage, birthdays, graduations, or child birth are suitable templates for knowledge acquisition and readily lend themselves to further annotation of people, places, and connections.

#### R.R.3 Enable users to explore their personal information space as they wish

- We don't want to simply present users with a set of file folders where they have to identify what is relevant on their own. Instead, we exploit the potential of the PIMO to allow users to search their information efficiently and to leave space for browsing and serendipitous finds. For instance, choosing a particular ancestor as a starting point, users could delve further into their life story, or search for information about other people in that generation. Furthermore, additional information from the resource itself (e.g., if it is a document or a picture) could be derived as context such as entities mentioned in a text. This can be provided by the analysis algorithms used for contextualized remembering from WP6.

### 2.5.4 R.P – Preservation

We propose to implement synergistic preservation by embedding it into the PIMO ecosystem. Considering preservation from the user's point of view, we can see three aspects: First, preserve the material, second, access if required, and third, the legacy of the user's archive. Regarding these aspects, we can derive the following requirements from the application scenario:

## Preserve

### R.P.1 Preserve material to be retrieved even decades later

- From the application scenario we see the benefits of preserving material. However, preserving only the material is not enough, we also have to preserve the context of the material. This context can be derived from the mental model, hence, in our approach from the PIMO supported by the methods for evolution-aware contextualization.

### R.P.2 Preservation should be transparent

- To foster the adaptation of preservation in everyday life, preservation needs to be transparent (most of the time) and as automatic as possible. As long as users are responsible to trigger preservation, it won't be done, or there will be danger that things get lost, etc.
- We need to design preservation policies that cover space consumption, deletion behavior, privacy, and other relevant aspects. We also need to allow users to manually trigger archiving if they decide that a certain set of information is no longer relevant.
- Transitions between the live system and the archive should be as seamless as possible.

### R.P.3 Provide information on what is preserved and allow deletion

- Users should be able to see what is preserved and be able to delete information objects in the archive. Users must have the full access and control to their archive.

### R.P.4 The archive is not a backup

- ForgetIT is not a classical backup, but an intelligent archive where data is organised according to PIMO concepts.

### R.P.5 The archive must be private and secure.

- Since personal data will be archived, we need to ensure high levels of privacy and confidentiality, if users are to entrust their private data to ForgetIT. Therefore, ForgetIT needs to consider means to ensure privacy of the archived data.

## Access

### R.P.6 Allow to access archived material at any time

- This will be a major challenge. The PIMO has functionality to embed accessibility in its data infrastructure, such as landmarks and proxies.

### R.P.7 Allow to restore archived material on the computer

- In addition to accessing the material such as viewing a photo, it should be possible for the user to restore archived material on a computer, bringing it back into active use. This could be necessary in order to restore a local copy of the archive, or to process the archived material further.

**R.P.8** Preserved material should be contextualised so that it can be interpreted after and kept interpretable for long-term preservation

- As the diary scenario indicates, material from the archive can be embedded in a context when it is archived. When it is accessed, the user should be able to explore the context in order to supplement missing information. For example, words that have changed their meaning can be annotated with an explanation in present day terms, and people visible on photos can be annotated with links to their names and basic biographical and genealogical facts.

## Digital Legacy

**R.P.1** The personal archive as legacy

- We need to consider what happens with the archive if the administrator dies or loses the mental capacity to make informed decisions about it. Similar questions arise when it comes to social media accounts, web sites, and blogs, which may also contain highly personal data. Drawing on existing work on designing technology for the end of life [29], we will suggest policies for handing over the archive, controlling access, and deletion. This is especially addressed in Deliverable D11.4 [30] where the requirements for a personal preservation service are investigated.

**R.P.2** Allow change of service provider

- It should be possible to move the archive to a new service provider while requiring minimal intervention from the user

## 3 Organizational Preservation

### 3.1 Organisational Preservation in ForgetIT

Many commercial organisations do not appear to deduce a need for preservation from their daily business operations. The ForgetIT project aims to identify a path to meaningful preservation in organisations by developing solutions that benefit organisations in the short term (while solving long-term problems), by delivering guidance on implementing these solutions, and by demonstrating the importance of preserving digital content. In doing this the ForgetIT project focusses on digital content related to the Web sites of organizations (Intranet and Internet). The technical basis will be the TYPO3 CMS.

These days, web sites are a cornerstone of the branding, reputation, and customer relations of a company. Therefore, companies need to extend existing records management strategies for keeping relevant documents, emails, and computer files to encompass their web presence. Legal and regulatory requirements that force companies to preserve documents also apply to the respective web content.

However, Web sites are very difficult to preserve because they consist of many different types of content that are linked by a complex structure and often feature executable code. In addition, digital content evolves rapidly, and web sites are often seen as ephemeral, not worth preserving.

The web site as a communications and revenue channel for commercial organisations used to be the responsibility of the IT department. These days, marketing and communications tend to be responsible for it, often in close collaboration with one or more external service providers. In some cases, the web presence is managed by a dedicated business unit reporting directly to the executive levels.

Preservation and archiving on the other hand are mostly managed by records managers, file clerks, and archivists. These employees tend to work with the legal department and core business departments. The web presence is often not covered by records management policies and procedures.

Most preservation happens through backups, since that is what most users are familiar with. However, a backup is not an archive. Versioning systems are a more sophisticated way of storing old versions of digital artefacts (documents, web pages, photos, etc.). They store not only the artefact itself, but also about the time it was created, used, and withdrawn from use. Indeed, as we will see later, version management provides the starting point for one of the two key use cases that emerged in our user research.

Web sites that are highly integrated in the business processes of the organisation are more likely to profit from the technical framework provided by ForgetIT. As part of our preservation and digital asset management framework, we aim to develop tools that give administrators and editors an overview of the preservable content of the web site, determine preservation triggers and preservation values, and visualise the connections be-

tween digital content.

We contend that introducing a well-managed and well-supported digital asset management and preservation framework, companies can also gain a better understanding of their web site. Analyses that are required to determine memory buoyancy and preservation value can help companies understand the value of their web presence and their digital assets, improve information architecture and content and therefore lead to a better user experience for visitors. It also can foster the re-use of existing digital assets. The ForgetIT framework will also emphasise the strategic value of the web site and promote further integration of the web presence into the company's core processes and strategy.

### 3.2 Conceptual Model for Requirements

Our conceptual approach for integrating the ForgetIT framework functions and features within the content management systems used by organisations is based on the Web Content Lifecycle.

This lifecycle is the multi-disciplinary and often complex process that web content undergoes as it is managed through various publishing stages. It consists of several stages. Each stage is associated with functionality such as records management, digital asset management, collaboration, and version control, that needs to be supported by various technologies and processes.

Within ForgetIT, we propose the following six-stage process:

- 1. Collect** Find relevant data online, in local repositories such as databases, and in the archive. This can include media files and other assets. This step corresponds to the Discovering and Using / Reusing steps and the Selecting process of the Digital Content Life Cycle (DCLC) model described in D5.1 [18], Section 1.1.2.
- 2. Create** Create content that integrates the data and enrich it with semantics. This corresponds to the Create stage of the DCLC.
- 3. Process** Prepare content for distribution, including editing and translation. At this stage, processes associated with the Describing stage of the DCLC happen. It is important that any tags and descriptors used here interface well with preservation.
- 4. Publish** Approve and publish the content on the site, making it accessible via the web.
- 5. Analyse** Monitor the behaviour of visitors to the web site, assess the value of content and assets, and check whether content is up-to-date.
- 6. Archive** Unpublish content and assets that have been updated or removed, move it to the archive and initiate the forgetting process as appropriate. This corresponds to the Preserving process of the DCLC.

The Managing stage within the DCLC encompasses our stages Publish and Analyse. Managing also involves starting the process of creating, updating, or preserving content.

### 3.3 TYPO3 CMS

The solutions we plan to develop in ForgetIT will be based on the TYPO3 Content Management System<sup>8</sup>. The TYPO3 CMS is an open source CMS that was first released in 2000 and has since gone from strength to strength.

According to CMScrawler, TYPO3 is the fifth most popular CMS world wide with a total of 273,923 installations<sup>9</sup>. Among German-speaking web sites, TYPO3 CMS has a market share of 21.3%<sup>10</sup>, which makes it the second most popular content management system after Wordpress. It is used for 18.43% of all web sites under the .de domain that use a CMS<sup>11</sup>. Working with TYPO3 has a key advantage for ForgetIT. Since TYPO3 is an open source system, any extensions and improvements created during ForgetIT can be disseminated quickly to its large international user base, creating long-lasting impact. Project partner DKD is very well networked in the TYPO3 community. We plan to actively work with the TYPO3 developer community, thus engaging with a key stakeholder in Web CMS, the developers themselves, throughout the project.

Content for a TYPO3 CMS can be created in several different ways:

- Content is created outside of TYPO3 CMS with word processing software or other appropriate pieces of software and copied and pasted onto the corresponding page.
- Content is created within the content management system and set to a hidden state before publishing. For publishing, a link needs to be sent to the appropriate authoriser, who needs to release the content for publication.
- Content is created with the help of a predefined workflow where the content is kept in a workspace that gets published once it is approved.

Unfortunately, the existing workspace mechanism of TYPO3 CMS has a highly complex user interface and is not sufficiently stable for enterprise use.

Another problem is the handling of digital assets such as media files that are used in the web site together with content pages. TYPO3 CMS currently does not fully support digital asset management, but this has been acknowledged by the working group of active contributors to the TYPO3 CMS software and will be addressed in upcoming releases.

---

<sup>8</sup><http://typo3.org>

<sup>9</sup><http://www.cmscrawler.com/tool/>, last checked July 26, 2013

<sup>10</sup>[http://w3techs.com/technologies/segmentation/cl-de-/content\\_management](http://w3techs.com/technologies/segmentation/cl-de-/content_management), last checked July 22, 2013

<sup>11</sup><http://www.cmscrawler.com/tool/TYPO3>, last checked July 22, 2013

### 3.4 Pilot Interviews

In order to develop this use case, a number of organizations were contacted with a view to testing ForgetIT prototypes. Two organisations have agreed to serve as test sites, Spielwarenmesse (Nuremberg, Germany), and ACO Holding, Rendsburg, Germany. To develop the use cases and establish potential applications for ForgetIT technology in each company, a number of on site pilot interviews with relevant stakeholders, who were mostly from Marketing, were conducted.

#### The Companies:

**Spielwarenmesse** is a SME that organises fairs for toy producers in Nuremberg, Germany, and satellite fairs in India, Russia, and the Middle East. They also support toy companies that wish to present their products at other fairs. They have representatives in over 60 countries who maintain country-specific sections of the Spielwarenmesse web site. The company's web presence, [www.spielwarenmesse.de](http://www.spielwarenmesse.de), is closely linked with traditional press marketing activities, active social media presence, and an industry news service. They also offer extensive online services to exhibitors and visitors including an online shop.

There are several major update and revision cycles which are linked to the yearly fairs and the Toy Award, and several minor update cycles which are mainly linked to press releases and the industry news service. Spielwarenmesse has a company archive which is managed by a professional archivist; the web presence is also ingested into this archive.

**ACO Group** is a family-owned company with 29 production sites and 40 distribution partners throughout the world. ACO is organised as a holding, and we are working with the main office in Rendsburg that maintains the main company web site, [www.aco.com](http://www.aco.com), and provides web solutions for ACO partners if they wish. There is no central online shop, therefore, the web sites mainly showcase ACO itself and the products and activities of each partner.

Both companies use the TYPO3 CMS for their main web presence, both have web sites with multilingual content, and both focus on the business-to-business market and do not target consumers. However, they fundamentally differ in their organisational structure and in the markets they work in. Whereas ACO produces and distributes goods, Spielwarenmesse is a service provider and organises events. As ACO and Spielwarenmesse are so different, working with both will ensure that the solutions developed within ForgetIT are not targeted narrowly at a single type of company and that processes that are developed within ForgetIT will work in different organisations and management cultures.

**Interview Method:**

At each company, we talked to several people who are responsible for the web presence. Both companies presented their history, their structure, and the web sites that they maintain. In the interviews, the following aspects were covered:

**Stakeholder map:** Who are the stakeholders in the web presence, and how do they interact with it? How do stakeholders communicate?

**Workflows:** The path of information from origin (database, press release, etc.) to the web site; where are problems and sticking points?

**Genres:** Which genres of documents [31] are produced in which contexts?

**Preservation:** Which preservation and archiving strategies are already in place, and how are they integrated into the workflows?

As these were initial interviews, no audio recordings were made, but there are extensive hand-written notes, and discussions were also documented on flip charts, post it notes, cards, and photos.

## 3.5 Use Cases

The following use cases were derived from application scenarios and problems that emerged in both interviews. They are presented in a format based on [32]. First, we define the human and non-human actors, then, for each use case, we summarise the task, outline the workflow, and list the requirements that result from the workflow.

### 3.5.1 Use Case Actors

There are eight main types of actors that play a role in the use cases we have identified. Each actor corresponds to a role; often, a single person will have several roles. For example, a manager is a Visitor when they refer to the web site for information they need when interacting with customers, but a Manager when they tell the Editor to start a major revision of the web site as part of a new product cycle. In small companies, the Editor is often also the main Web Admin, and create material for the web site (Contributor).

**Visitor:** Everybody who visits the web site.

**Manager:** Everybody who can give the go-ahead for site-wide changes as part of major cycles.

**Information Architect:** Everybody who plans the structure of the site and have an overview of all parts that are affected by regular changes.

**Editor:** Everybody who needs to check and, if necessary, modify content before going live.

**Web Admin:** Everybody who administers the web site.

**Contributor:** Everybody who contributes content and assets for publishing on the web, or who administers a piece of software that provides material for automatic updates.

**Archivist:** Everybody who is responsible for archiving and preservation aspects of the web site.

**Legal:** Everybody who checks legal aspects of the web site and who may need to access web site archives when needed to process formal complaints and legal action.

Non-human actors are pieces of software and software systems that execute actions which are relevant to the preservation cycle.

**AIS:** Storage subsystem that persists numerous preservation objects (Archival Information System).

**T-CM Service:** Today's content management system for live content.

**F-CM Service:** Future content management system for live content.

**Ingest Service:** Service that ingests content into the archive (collection, packaging).

**Retrieve Service:** Search service that retrieves content from the archive.

**Generation App:** Application used to generate content for the web site, e.g., a word processor, picture processing software, a database.

**Re-contextualization Service:** Service used to process content retrieved from the archive and to prepare it for re-use in the current context.

**Scheduler:** Schedule of dates and events that trigger a systematic review of assets and/or a web site update and service to manage the schedules and the triggering.

### 3.5.2 Use Case: Synergetic preservation for TYPO3

Either on a regular basis, or before major changes or updates to a web site or based on a strategy based on preservation values, the whole site, sections of the site that change or assets that have been used in the creation of the Web site will be archived for archival purposes. Users can access the archived site later to research company history, obtain information for legal use, retrieve assets for reuse, and inform the creation of new documents (e.g., by using old documents as a template). The main idea here is to make the preservation an integral and easy to use part of the overall workflow enabling Synergetic Preservation for TYPO3 and smooth transition to and from the archive.

#### Initial idea for the Workflow:

1. The archiving is triggered either by the Scheduler (regular basis, preservation strategy, change-triggered) or by a Manager based on input from the Scheduler and/or based on relevant developments in the company. An example is a advertising campaign has run its course, and the tie-in web site that was created is switched off.
2. The parts of the web site that need to be archived and degree of detail of the archive are specified in T-CM Service based on preservation values and the preservation strategy. Preservation values are computed based on requirements and the perceived importance of the documents for the company. Degrees of detail may include multiple snapshots of the look and feel of the web site, link structure, executable code, and HTML as well as the digital assets used as part of the Web site.
3. Ingest Service accesses T-CM Service and collects and ingests the specified content in the required level of detail and saves further information about document formats and browsers used to generate the look and feel snapshots. As a possible extension in the second phase of the project further information from the context of the archived content can be collected and also archived (context information) for ensuring that the content can still be interpreted correctly, when retrieved at a later

point in time, e.g. photo x was used in the advertising campaign y at time t and in combination with photos a, b and c. This is linked to the idea of contextualized remembering in ForgetIT. It also creates links between the new material and relevant material that has already been archived to facilitate contextualised remembering. The current preservation values are also stored in the archive and can be updated during archival time to e.g. reflect changes in priority.

4. The AIS stores and manages the archived assets. It also takes care of transformation and checking operations that become necessary over time. Converters might have to be provided to the AIS archive for preserved material, if new formats are introduced in the T-CM.
5. After some time has passed, Marketing has decided to run a new advertising campaign similar to the old one. Using Retrieve Service, relevant digital assets are retrieved from the AIS. Assets, documents, look and feel, and the information architecture are recovered, prepared for re-use by the Re-contextualization service, and transferred to the F-CM Service for modification.

#### **Initial List of Requirements for Synergetic Preservation (SP):**

- R.SP.1** The AIS needs to be integrated into overall digital asset management with smooth transitions.
- R.SP.2** A user interface for specifying assets to be archived as well as services for suggesting assets to be archived.
- R.SP.3** Levels of archiving need to be defined (screen shots only, screen shots and information architecture, browsers to capture, etc.).
- R.SP.4** A archiving and preservation policy needs to be defined.
- R.SP.5** Strategies for computing the preservation value have to be defined.
- R.SP.6** the Scheduler has to be implemented for organizing the preservation process.
- R.SP.7** Formats for exchanging content and context between active system and the AIS have to be defined (CMIS is investigated as a candidate here).
- R.SP.8** A system for automatically generating snapshots needs to be implemented. Each snapshot is a preservation object that is tagged with the time and date when it was taken and the browser and operating system used.
- R.SP.9** The special case where the objects are not removed from the live T-CM System after ingestion into the archive has to be supported.
- R.SP.10** An Ingest Service needs to be provided, which can interact with TYPO3 needs to be provided

**R.SP.11** A Retrieve service needs to be provided, which supports searching and retrieving of digital assets from the archive; (see also next use case).

**R.SP.12** A Re-contextualization service needs to be provided.

### 3.5.3 Use Case: Preservation-aware Digital Asset Management

In Web content management a wide variety of assets are created over time, such as text blocks, images, logos and PDF files. Over time, assets are updated, new assets are created and assets become out of date after a while. Without a proper and intelligent digital asset management strategy and support, which also embraces preservation, T-CM Service will quickly become bloated with files representing all types of assets (including files that are no longer linked from the main web site). Assets that are still of value might get lost or just cannot be found anymore. The idea here is to have a digital asset management support that helps in better structuring the existing assets, assess the current value of assets e.g. via memory buoyancy (taking into consideration forgetting), supports the search of assets based on context and current value and fosters the re-use of assets as well as the preservation of important assets. The asset management support will also embrace ideas of version management. This service relies on the core ideas and services of the ForgetIT project, while also creating direct benefits to the users: de-cluttered asset collections and increased re-use of assets.

#### Initial idea for the Workflow:

1. Digital assets management covers assets as well as web pages. It also covers active use as well as preservation. There are a variety of initial situations:
  - (a) The Web Editor adds a new asset or page to the web site. A trigger for information value review is automatically assigned to the asset, and it is timestamped with the date at which it was added to the live version of the T-CM System. Information value assessment computes preservation value (PV) and memory buoyancy (MB) for the asset. PV refers to the long term value of the asset for the company; MB reflects its current importance. (in the initial phase of the project we might focus on one of the two values). Information value review re-computes the values for information value assessment based on new situations.
  - (b) When a new asset replaces an old asset, the previous asset is added to the information value assessment queue.
  - (c) When a web page is modified, the previous version of the page is added to the information value assessment queue, and triggers for information value assessment may be adjusted. Modifications are timestamped.

Triggers for information value review can be events such as a new product line, a date, a time span (e.g., a year), or a reaching of a memory buoyancy threshold.

2. When the Scheduler detects a trigger, it creates a list of digital assets that this trigger applies to and adds them to the information value review queue. For the assets in the queue information value assessment (PV and MB) are re-computed.
3. According to the new values of MB and PV for individual assets the system might suggest actions to the Web Editor such as archiving an asset or removing it from the T-CM. The Web Editor checks the suggested actions such as the implications of removing the asset from T-CM Service, and selects the objects to be ingested into the archive. The Web Editor may also check the level of preservation detail required and may adjust the preservation values. Much of this process is automatic, but the Editor, in consultation with the Archivist and the Legal team, has the final say.
4. Furthermore the MB values are also used in ranking assets, when they are returned as search results in the digital asset management. This enables the user on assets that are still of high importance (de-cluttering).
5. Similar to the previous use cases the Ingest Service ingests the selected assets and creates links to existing preservation objects to facilitate contextualised remembering. Preservation objects that were replaced or that are old versions of updated web pages are linked to their replacement in the T-CM System through a unique common identifier. All ingested objects are time stamped with the date and time of their ingestion.
6. After some time has passed, a Contributor needs to get an existing asset, such as a photo of a prizewinner, for a new document. The Retrieve Service is used to find the object independent if it is still in current asset management or already in the archive. PV and MB values as well as context information can be used for ranking. In case the asset is already in the archive it is retrieved from there, and provided in an adequate format for further use (exploiting the re-contextualisation service, if it comes from the archive).

### **Initial List of Requirements Preservation-aware Digital Asset Management (DAM):**

**R.DAM.1** A strategy for associating assets and web pages with triggers, a set of defaults, and an intuitive user interface that makes setting triggers easy.

**R.DAM.2** An implementation of triggers for information value review that encompasses automatically detected events, such as edits, manually set events, such as “general web site update for new product list” and specific dates/times.

**R.DAM.3** A information value review queue that stores links to old assets and versions of web pages.

**R.DAM.4** Methods for computing MB and PV values.

**R.DAM.5** An option for reviewing objects in the information value review queue that allows users to select standard maintenance options, such as “globally replace Version 2.3 of Asset A with Version 2.4 of this Asset”.

**R.DAM.6** An option for revising preservation actions suggested by the system; this includes a report on each object in the preservation review queue that shows where the object is referenced in the live T-CM System. Special actions are required if assets are still linked to any live content.

**R.DAM.7** An option for reviewing preservation values - some assets may need to be preserved particularly carefully because they might be needed to resolve complaints (see eDiscovery Use Case in [32]).

**R.DAM.8** A Retrieve service, which can jointly search assets in active asset management and in the archive (extension of R.SP.11). Ranking should take into account MB, PV and context information.

**R.DAM.9** The Ingest Service needs to classify the object using ontologies and concepts that have already been implemented in the archive and add links to all relevant preservation objects in the archive (contextualization).

**R.DAM.10** A Scheduler that works in the background, moves objects to the information value review queue.

**R.DAM.11** Links between objects that are in the live T-CM System and their earlier versions in the archive.

### 3.6 Working with Users—Going Forward

We plan to use an Action Research approach when implementing and evaluating the Organisational Preservation solutions for ForgetIT. Action Research has been used in business and health care research for planning, implementing, and evaluating new systems [33, 34]. It consists of several iterations of a basic four-step cycle:

**Identify an Issue:** What problems have arisen that can be addressed through an integrated preservation approach such as ForgetIT? Work with companies who have offered to test ForgetIT (use case hosts) in order to get enough data to get a deep understanding of the underlying issues. Collect baseline qualitative and quantitative data that will be used in the subsequent evaluation steps.

**Design a Solution:** Based on the data collected in the first step, analyse the situation and design possible solutions together with the use case hosts. If the proposed changes are extensive, identify a suitably small step that can be implemented and evaluated in a couple of months. Design tests to guide the implementation.

**Implement the Solution:** Extend the TYPO3 CMS with the functionality required to provide the solution. For example, while functionality has been defined for the TYPO3 CMS to support versioning [35], this functionality may not be sufficiently stable for use in enterprise contexts.

**Evaluate:** Implement the new functionality at the use case host site and collect qualitative and quantitative data through logs and interviews. Analyse whether there was an improvement in quantitative measures, and whether users felt that there was an improvement in their user experience.

We expect that this cycle will be iterated 3–4 times throughout the project. In the pilot interviews, we have partially completed the first step of the first cycle, and this document discusses proposed solutions for the design problem. We have also identified web sites and parts of web sites that would be appropriate test beds for development at each host site. The next step will be a co-design session (Step 2), followed by the Implementation/Evaluation part of the first cycle. Updates on this research will be provided in the relevant deliverables and the quarterly progress reports.

## 4 Conclusions

In this document, we introduced the Personal Preservation and Organisational Preservation use cases for ForgetIT.

For Personal Preservation work package of ForgetIT (WP9), we defined a set of requirements for the development of synergetic preservation with PIMO support. These requirements have been designed to showcase the main aspects of the ForgetIT model and to fit with a realistic scenario, a personal proactive diary. The requirements will be used to plan further research, determine evaluation strategies, and guide the implementation. We expect that as the project unfolds, some requirements will be prioritised over others. In Deliverables D9.2-D9.5, we will report on changes to the list of requirements, including added requirements that emerged from the development work, revisions to existing requirements, and requirements that had to be dropped for reasons of time and scope.

For the Organisational Preservation work package (WP10), we identified two main use cases, synergetic preservation for TYPO3 content enabling the preservation of assets and Web sites, and preservation aware digital asset management building upon version management of updates to web site content and assets and managed forgetting functionality. We provided a formal specification of both use cases and the resulting requirements. Within WP10, we will use an action research paradigm that involves several cycles of identifying problems, designing a solution, implementing the solution, and evaluating whether the solution works. The use cases specified here represent the initial identification of problems. The use cases may change and become more specific as we proceed to actual implementations. In D10.1, we will supplement these use cases with user stories and personas, and report on further iterations of the action research process in D10.2-D10.4.

## Appendix

### A Survey Scales

The following sections list the scales used for the pilot survey on Managing Digital Photos (see Section 2.2).

#### A.1 Items to Assess Storage, Use, and Management of Digital Photos

1. I delete most of the digital photos I take.
2. I keep all my digital photos meticulously organised.
3. I rely on geolocation data to annotate digital photos with locations.
4. I like browsing digital photos.
5. When I want to find a digital photo, I immediately know where to look.
6. I enjoy putting together themed collections of digital photos.
7. I find organising digital photos tedious.
8. I use digital photos to reminisce about people.
9. I worry about losing digital photos that are important to me.
10. My digital photo collection is a mess.
11. I put a lot of effort into taking good digital photos.
12. I only keep high quality photos.
13. I like editing digital photos.
14. Most of my digital photos are snapshots.
15. I want to control who can see my digital photos.
16. I like browsing digital photos in which friends have been tagged.
17. I don't want other people to tag me in digital photos.
18. When I edit a digital photo, I always keep a copy of the original.
19. Digital photos can take me right back to the time and place where they were taken.

20. I often get sidetracked when looking for a particular digital photo.
21. I like it when people comment on digital photos that I have taken.
22. I like to look at photos that were taken a couple of years ago.
23. I like to browse through old photo albums.
24. My collection of printed photos is a mess.
25. I print digital photos that are important to me.
26. I use automated tagging to tag photos of people.
27. I mainly keep photos that I associate with a strong emotion.
28. I tag all of my photos with keywords.
29. I add titles to all my photos.
30. I never add descriptive text to my photos.

## A.2 Attitude to Technology

1. I enjoy using technology.
2. Being able to use technology is important to me.
3. I keep up with the latest gadgets.
4. Computers are beneficial because they save people time.
5. The internet is beneficial because it is a great source of information.
6. I avoid using technology whenever possible
7. I enjoy reading about technology.
8. I enjoy installing new apps and software.
9. Email is an easy way to communicate.
10. My friends often ask me technology-related questions.
11. I do not read web logs.
12. I often visit web sites.
13. I avoid having to learn how to use new gadgets.
14. I enjoy learning to use new software and applications.

15. When using a computer, I often lose data.
16. When I use technology, I am afraid that I will damage it.
17. I must have a reference manual or online help to use technology.
18. Overall, I feel that I don't know how to use a computer.
19. Overall, I don't feel comfortable with social media such as Twitter or Facebook.

## References

- [1] Jeremy Leighton John, Ian Rowlands, Peter Williams, and Katrina Dean. Digital lives: Personal digital archives for the 21st century – an initial synthesis. research paper Beta Version 0.2, British Library, march 2010. <http://britishlibrary.typepad.co.uk/files/digital-lives-synthesis02-1.pdf> last accessed July 22, 2013.
- [2] Catherine C. Marshall. *I, Digital: Personal Collections in the Digital Era*, chapter Challenges and Opportunities for Personal Digital Archiving, pages 90–114. Society of American Archivists, 2011.
- [3] Library of Congress. Personal archiving: Preserving your digital memories. <http://www.digitalpreservation.gov/personalarchiving>, last accessed: July 21, 2013.
- [4] Library of Congress. Preserving digital memories. In *7th International Conference on Preservation of Digital Objects (iPRES 2010)*, 2010. [http://www.digitalpreservation.gov/personalarchiving/documents/NDIIP\\_PA\\_poster.pdf](http://www.digitalpreservation.gov/personalarchiving/documents/NDIIP_PA_poster.pdf), last accessed: July 21, 2013.
- [5] Space Communications CCSDS Secretariat and Navigation Office. *Reference Model for an Open Archival Information System (OAIS)*. Space Operations Mission Directorate, 2009.
- [6] POWRR. Tool-grid, 2013. <http://digitalpowrr.niu.edu/tool-grid/>, last accessed: July 21, 2013.
- [7] L Nathan and E Shaffer. Preserving social media: Opening a multi-disciplinary dialogue. In *Proceedings of the The Memory of the World in the Digital Age: Digitization and Preservation. Vancouver, BC, 26-28 September 2012*. UNESCO, 2013.
- [8] Miriam Julius. Themen in Sozialen Netzwerken (topics in social networks). blog, march 2013. <http://multimedia-memories.tumblr.com/post/45185122672/fragestellung-diese-infografiken-basieren-auf>, last accessed July 18, 2013.
- [9] James Cohen and Paul Mihailidis. Storify and news curation: Teaching and learning about digital storytelling. In *Second Annual Social Media Technology Conference & Workshop*, volume 1, pages 27–31, 2012.
- [10] Andrea Ceroni, Mark A. Greenwood, Nattiya Kanhabua, Vasileios Mezaris, Claudia Niederee, Jörgen Nilssonrgen Nilsson, and Olga Papadopoulou. D6.1: State of the art and approach for contextualization. Deliverable D6.1, ForgetIT consortium, August 2013.

- [11] Leo Sauermann, Ludger van Elst, and Andreas Dengel. PIMO – A Framework for Representing Personal Information Models. In T. Pellegrini and S. Schaffert, editors, *I-SEMANTICS Conference 5-7 September 2007, Graz, Austria*, J.UCS, pages 270–277. Know-Center, Austria, 2007.
- [12] Heiko Maus, Sven Schwarz, and Andreas Dengel. Weaving personal knowledge spaces into office applications. In Madjid Fathi, editor, *Integration of Practice-Oriented Knowledge Technology: Trends and Prospectives*, pages 71–82. Springer, 2013.
- [13] Leo Sauermann and Dominik Heim. Evaluating long-term use of the gnosis semantic desktop for pim. In *The Semantic Web. 7th International Semantic Web Conference (ISWC-2008), 7th International Semantic Web Conference, October 26-30, Karlsruhe, Germany*, volume 5318 of *LNCS*, pages 467–482. Springer, 10 2008.
- [14] Leo Sauermann. *The Gnosis Semantic Desktop approach to Personal Information Management*. PhD thesis, Fachbereich Informatik der Universität Kaiserslautern, 06 2009.
- [15] Leo Sauermann, Andreas Dengel, Ludger van Elst, Andreas Lauer, Heiko Maus, and Sven Schwarz. Personalization in the epos project. In *Proceedings of the Semantic Web Personalization Workshop at the ESWC 2006*, 2006.
- [16] Sven Schwarz. *Context-Awareness and Context-Sensitive Interfaces for Knowledge Work Support*. Phd thesis, University of Kaiserslautern, Fachbereich Informatik, March 2010.
- [17] Heiko Maus, Sven Schwarz, Jan Haas, and Andreas Dengel. CONTASK: Context-Sensitive Task Assistance in the Semantic Desktop. In Joaquim Filipe and José Cordeiro, editors, *Enterprise Information Systems. 12th International Conference, ICEIS 2010, Funchal-Madeira, Portugal, June 8-12, 2010, Revised Selected Papers*, volume 73 of *Lecture Notes in Business Information Processing*, pages 177–192. Springer, March 2011.
- [18] Jörgen Nilsson, Parvaneh Afrasiabi Rad, Tero Päivärinta, and et al. D5.1: Foundations of synergetic preservation. Deliverable D5.1, ForgetIT consortium, August 2013.
- [19] Constanza Lampasona, Oleg Rostanin, and Heiko Maus. Seamless integration of order processing in ms outlook using smartoffice: an empirical evaluation. In *ESEM '12 Proceedings of the ACM-IEEE international symposium on Empirical software engineering and measurement*. ACM Press, 2012.
- [20] Martin Klinkigt, Koichi Kise, Heiko Maus, and Andreas Dengel. Semantic retrieval of images by learning from wikipedia. In Andreas König, Andreas Dengel, Knut Hinkelmann, Koichi Kise, Robert J. Howlett, and Lakhmi C. Jain, editors, *Knowledge-Based and Intelligent Information and Engineering Systems*, volume 6884 of *Lecture Notes in Computer Science*, pages 212–221. Springer, 2011.

- [21] T Buchanan, J A Johnson, and L R Goldberg. Implementing a five-factor personality inventory for use on the Internet. *European Journal of Psychological Assessment*, 21:115–127, 2005.
- [22] Joseph Turow, Jennifer King, Chris Jay Hoofnagle, Amy Bleakley, and Michael Hennessy. Americans Reject Tailored Advertising and Three Activities that Enable It. *SSRN Electronic Journal*, September 2009.
- [23] Tom Buchanan, Carina Paine, Adam N Joinson, and Ulf-Dietrich Reips. Development of measures of online privacy concern and protection for use on the internet. *Journal of the American Society for Information Science and Technology*, 58(2):157–165, 2007.
- [24] ForgetIT team. D2.1: Evaluation plan. Deliverable, ForgetIT consortium, due M9 2013. t.b. published.
- [25] Morgan Ames, Dean Eckles, Mor Naaman, Mirjana Spasojevic, and Nancy House. Requirements for mobile photoware. *Personal and Ubiquitous Computing*, 14(2):95–109, June 2009.
- [26] David Kirk, Abigail Sellen, Carsten Rother, and Ken Wood. Understanding photowork. In *Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI '06*, page 761, New York, New York, USA, April 2006. ACM Press.
- [27] Kerry Rodden and Kenneth R. Wood. How do people manage their digital photographs? In *Proceedings of the conference on Human factors in computing systems - CHI '03*, page 409, New York, New York, USA, April 2003. ACM Press.
- [28] Uwe Riss, Alan Rickayzen, Heiko Maus, and Wil van der Aalst. Challenges for Business Process and Task Management. *Journal of Universal Knowledge Management. Special Issue on Knowledge Infrastructures for the Support of Knowledge Intensive Business Processes*, 0(2):77–100, 2005.
- [29] Michael Massimi, Wendy Moncur, William Odom, Richard Banks, and David Kirk. Memento mori: technology design for the end of life. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '12, pages 2759–2762, New York, NY, USA, 2012. ACM.
- [30] ForgetIT team. D11.4 personal preservation as a service: Opportunities, challenges and exploitation plans. Deliverable, ForgetIT consortium, due M30 2013. t.b. published.
- [31] Takeshi Yoshioka, George Herman, JoAnne Yates, and Wanda Orlikowski. Genre taxonomy: A knowledge repository of communicative actions. *ACM Transactions on Information Systems*, 19(4):431–456, October 2001.
- [32] Long Term Retention Technical Workgroup. Self-contained information retention format (sirf) use cases and functional requirements. Technical report, Storage Network Industry Association, September 2010.

- [33] Davydd J. Greenwood and Morten Levin. *Introduction to Action Research*. SAGE, Thousand Oaks, CA, 1998.
- [34] David E. Avison, Francis Lau, Michael D. Myers, and Peter Axel Nielsen. Action research. *Commun. ACM*, 42(1):94–97, January 1999.
- [35] *Inside Typo3 CMS 4.2.0: Versioning and Workspaces*, 2013. <http://docs.typo3.org/typo3cms/InsideTypo3Reference/CoreArchitecture/VersioningAndWorkspaces/Index.html>, last accessed July 22, 2013.