www.forgetit-project.eu

ForgetIT

Concise Preservation by Combining Managed Forgetting and Contextualization Remembering

Grant Agreement No. 600826

Deliverable D5.1

Work-package	WP5: Joint Information and Preservation Management
Deliverable	D5.1: Foundations of Synergetic Preservation
Deliverable Leader	Jörgen Nilsson
Quality Assessor	Simona Cohen
Estimation of PM spent	4
Dissemination level	PU
Delivery date in Annex I	2013-07-31
Actual delivery date	2013-07-26
Revisions	4
Status	Final
Keywords:	Digital Preservation Life-cycles, workflow, content management

Disclaimer

List of Authors

Partner Acronym	Authors
LTU	**Tero Päivärinta , Jörgen Nilsson, Parvaneh Afrasiabi Rad**
DFKI	**Heiko Maus**
Dkd	**Olivier Dobberkau**

Table of Contents

Executive summary

This report covers a systematic study of existing research literature on smooth bi-directional preservation transition between producers system (active system) and a Synergetic Preservation system. This transition is meant to be as transparent as possible to the user, and also be able to fetch material from the preservation system in a smooth way. The report also includes a study on gaps that could be identified from existing research and what this integration would need to cover, and how it conceptually could work in Content Management Systems (use case: Typo3) and Personal Information Management (use case: Semantic Desktop / PIMO).

The report groups activities into three high-level workflows; Preservation planning and administration, Preservation, and Access and retrieval. In these workflows, the *contextual preservation administration,* which link the preservation and storage services to particular information systems and maintains the contextualized preservation and access workflows, interfaces and ontologies, is something that needs further development during the project, since it includes many of the project specific peculiarities. This also goes for the other components of the middleware, which also are contextually aware, and part of the "forgetting" approach.

Both workshop results and the literature review shows that integration of these workflows with the producing workflows should be done in a way that makes it low-effort, generic, and transparent – which goes well together with what the project set out to do. Most likely some sort of automatic appraisal needs to take place already in the producing systems, and the challenge here would be to make it non-intrusive but effective. The ideas and results from this report will serve as a starting point for design and integration that will take place in WP5 of the ForgetIT project, in particular for T5.2 and T5.3 as well as D5.2. The ideas in this report are however not strict rules to be followed, and may be altered or improved as work continues in above mentioned tasks. Any such actions will be reflected in D5.2 or subsequent reports.

1 Introduction

This section broadly introduces the scope of the report and also relates it to other tasks and deliverables in the project. The aim and process are also described here.

The study of Long-Term Digital Preservation has provided an engaging journey for scholars and organizations that attempt to build a better understanding of requirements of digital preservation systems [1], planning for and strategy of digital preservation [2, 3, 4], and the process through which an information object is preserved [5]. Digital preservation has emerged to be a prominent construct in research for Digital Libraries [6], and a new one in studies of engineering [7]. Moreover, digital preservation has been associated with revenue and profit at the organizational level of strategic planning [4, 6]. Although researchers has made great theoretical progress, considering the diversity of outcomes digital preservation is associated with, empirical and technical research on digital preservation lags in two ways. First, there is a lack of research in how to integrate or communicate between the systems that produce information (production systems) and the preservation systems. Second, most efforts have been put into handling professional information (e.g. business records) and there have been few attempts aiming specifically at preservation of information from the personal sphere (family photos and such) where the incentives might be less formal. The ForgetIT project aims to tackle both of these issues.

As mentioned earlier, much work in digital preservation has revolved around organisations that already preserve information in a professional manner (e.g., archives and libraries) while the ForgetIT project on the other hand target; business with less formal needs and requirements for preservation; individuals, and in particular in their private role. These circumstances, together with the increasing amount of digital information created on a daily basis, led to the idea of learning from human memory and human forgetting in assisting the preservation system (i.e. the ForgetIT system) with making decisions, or at least suggestions, on what to preserve.

In order for the system to be able to assist the users, there is a need for a close cooperation between the ForgetIT system and the production system (as well as the Archival Information System). This work package (WP5) deals with this integration. The integration is labelled as "smooth bi-directional transition" in the project, which means that there is a need for the transfer of information going both ways, and that it should be done in a non-invasive way – all for the purpose of enabling Synergetic Preservation.

In order to handle these transitions smoothly and seamlessly, this work package also need to tackle such issues as quality of information package, file format identification, support for automated creation of (preservation) metadata and, last but not least, a communication model between active systems and preservation system. First of all though, we need to look on the integration of workflows.

1.1 Life cycles and Workflows

This section briefly describes some preservation oriented life-cycles and workflows. Most are on high level and not as detailed as an "actual" workflow would need to be.

Most preservation oriented life cycles and workflows, for good reasons, usually start with something being delivered or ingested to them. In this case we are interested in both the workflows within the delivering organisation (pre-ingest, and reuse) and what takes place within the preservation organisation/system. This section contains a short overview of some preservation related workflows and lifecycles as an introduction to some related concepts.

1.1.1 Digital Curation Centre – Curation Lifecycle Model

The Curation Lifecycle Model (Figure 1) from the Digital Curation Centre [8] shows typical steps in a preservation lifecycle (in the outer circle). An object is created or received, then evaluated if it

should be preserved, ingested into a preservation system/organisation, some preservation action might be undertaken to better preserve the object, the object is then stored eventually used/reused and perhaps transformed which might yield a new object. The *preservation planning* and *community watch* activities are continuously on-going as shown by the full circles towards the middle of the figure. What is shown here as well, although conceptually, is the difference between *storing* and *preserving* where we see that the first is just one part of the latter.

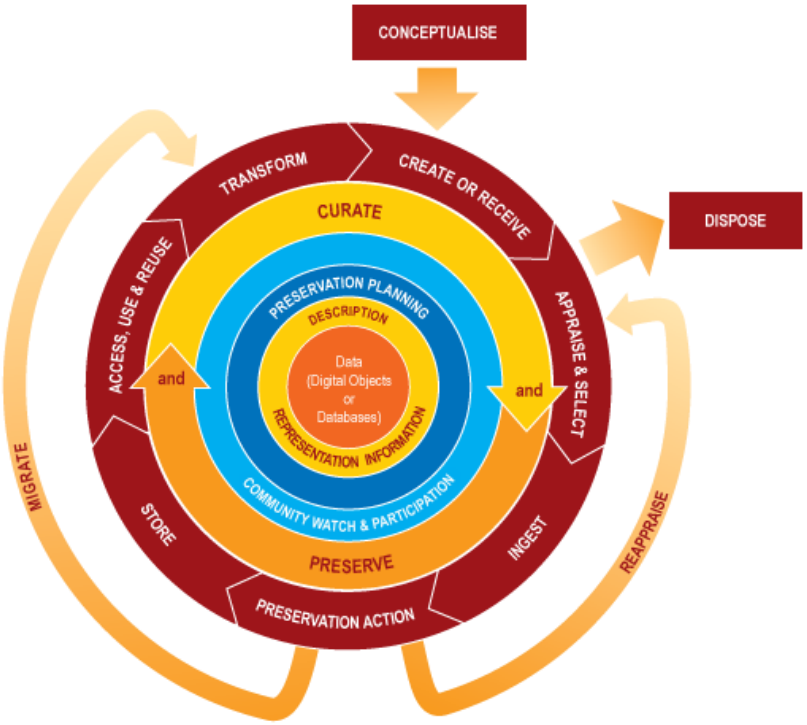


Figure 1: Key elements of the DCC Curation Lifecycle Model [8]

1.1.2 DigitalNZ – Make it Digital

The Digital Content Life Cycle (Figure 2) from DigitalNZ [9] certainly have similarities to the one from DCC [8] and in this case mostly serves to reinforce the typical steps that are involved in a preservation life cycle. What is shown here though, is a more use-oriented life cycle, where preservation is a task on the side. This would be typical of e.g. a library organisation that is focused on use and that does not have any particular preservation obligations (as opposed to e.g. a national library or a research library).

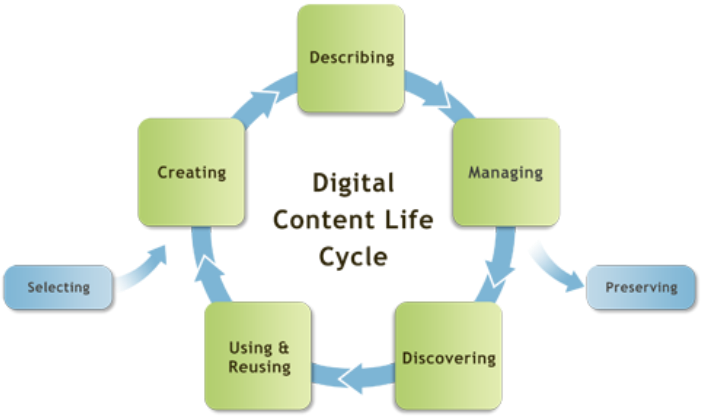


Figure 2: The Digital Content Life Cycle [9]

Ideas from such types of life-cycles might be useful when looking at more use oriented perspectives, but need to be combined with more explicit preservation life-cycles to serve their purpose in this project.

1.1.3 Caspar Workflow

The Caspar Workflow (Figure 3) from the Caspar Project is certainly more of a workflow than the two lifecycles mentioned above. We recommend that you look this up through the reference[1], since there is an animated version of it that shows the steps in the order they (should) take place.

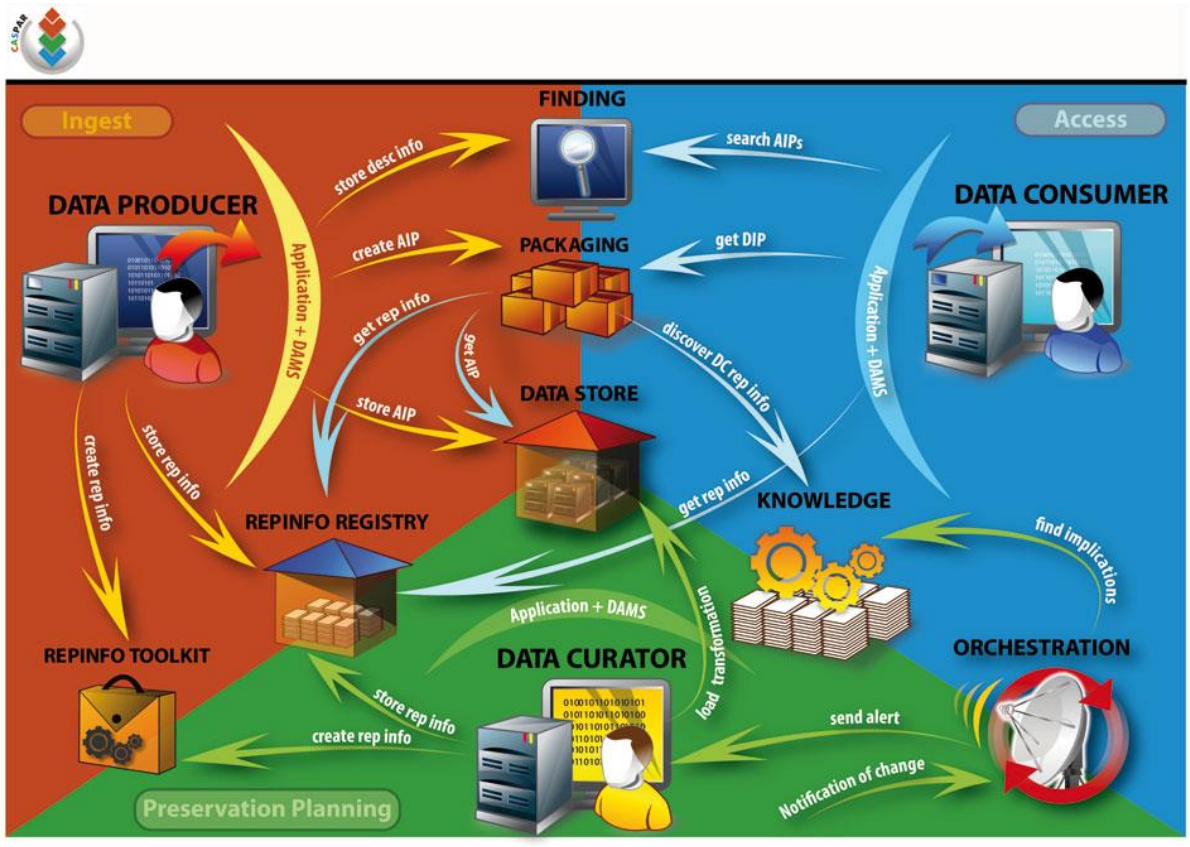


Figure 3: CASPAR Workflow[1]

Here we can see that the *Data Producer* need to create representation information (e.g. descriptions about how to interpret certain data both semantically and technically), and descriptive information which would be used later on for finding. We also see that the *Data Curator* create some representation information (e.g. about general things, such as file format specifications). This is something that is worth to consider in general, but in this project (ForgetIT) we should remember that the users might not be that interested to actively create (representation and descriptive) information.

1.1.4 Portico

The Portico "Preservation Step-by-Step" (Figure 4) is interesting since it clearly shows that you should start with *Preservation Planning* [10]. Preservation planning here includes e.g. file format and package analysis as well as a policy based preservation plan including for example migration into more suitable formats. This is normally a good idea, and might be something that we consider that the system should do (to its best abilities) automatically and proactively without the user's active participation. The following steps are also typical for a preservation system where we see: *Receipt & inventory management* handling a typical pre-ingest phase where material is transferred to the Portico system; *Processing & archival deposit* handles the actual ingestion to the system; *Monitoring & management* encompasses the daily operation of the archive, keeping content secure and taking actions to ensure future accessibility; *Content delivery* handles access to the archive.

[1] http://www.casparpreserves.eu/other-caspar-products/other-caspar-products/caspar_workflow.jpg

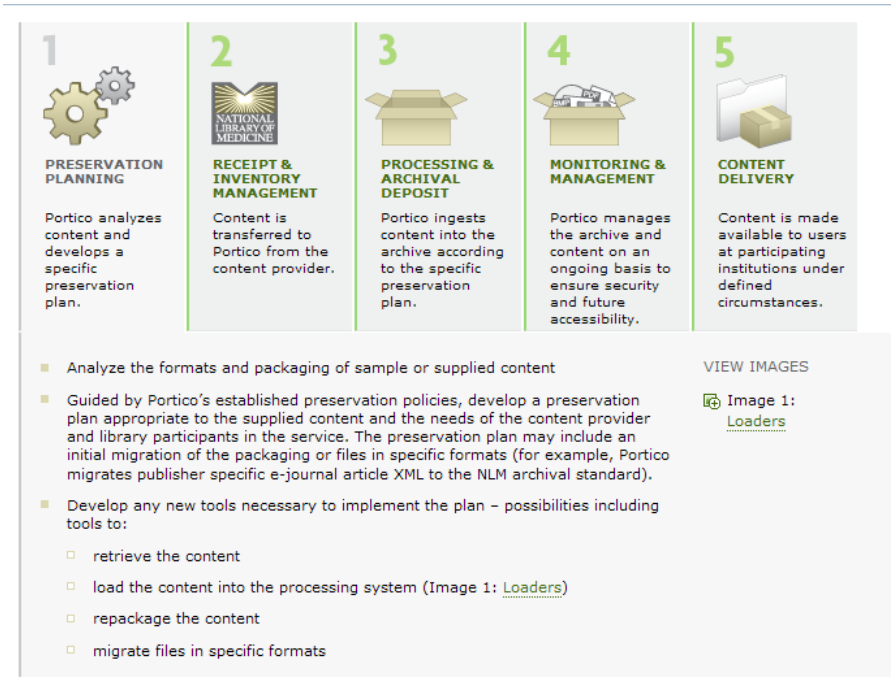


Figure 4: Portico - Preservation Step-by-Step [10]

1.1.5 Information Continuum Model

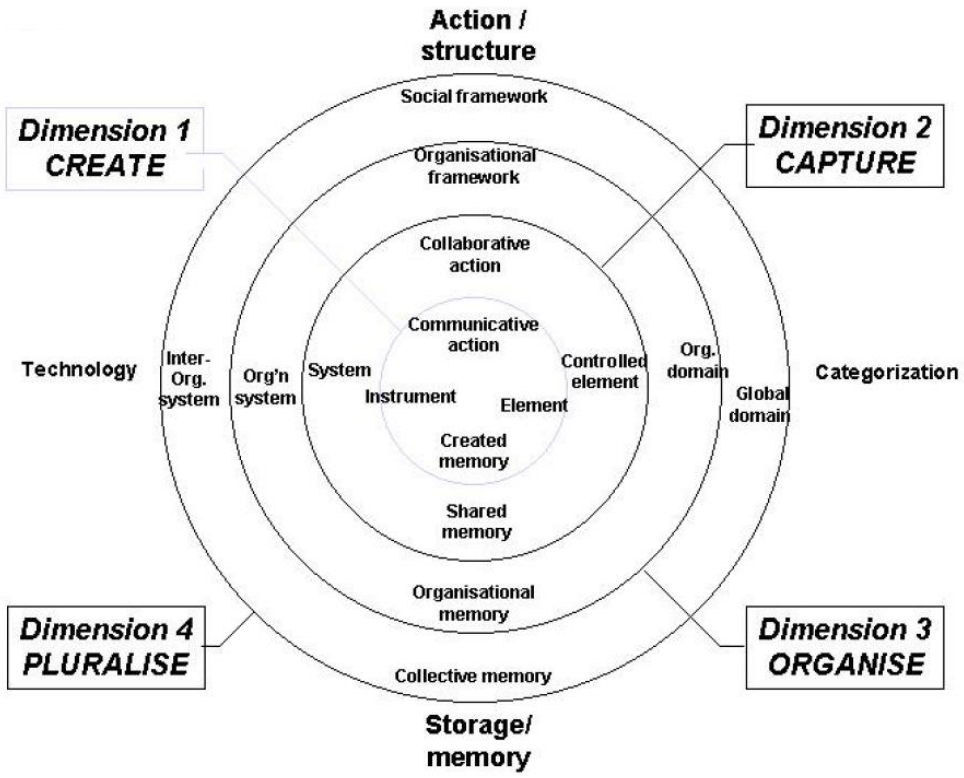


Figure 5: Information Continuum Model [11, p 130]

The Information Continuum Model (Figure 5), conceived at Monash University [11], is a way of representing the idea that information lives on, and that it has no clear boundaries in space and time between its active use (in records management) and its archived state. The work originated

as the Records Continuum Model, and subsequently took some other forms, of which Information Continuum Model is one of them. The dimensions themselves represent different stages of "being" and the axis's different concepts that are relevant for the particular model. In the case of Information Continuum these are *Action/structure, Categorization*, *Storage/memory*, and *Technology*. In the different fields that are formed we see, for example, what kind of technology that is involved in a particular dimension. Clearly this model takes some getting used to, and we recommend reading more in the referred work, or other work on the model(s). What we can bring with us without much effort, is the notion of time-space continuum and the idea of constantly becoming which goes rather well together with the idea in ForgetIT about information transitioning between different stages of awareness (buoyancy).

Overall, these models and workflow show steps that are relevant from a preservation perspective and some of them also involve a dialogue with the creator and the future user. There still exists a need to make these preservation oriented concepts enter the work flows of the (information) producing organisations and/or systems to facilitate better preservation possibilities. This is where this report aims to help.

1.2 Aim

The purpose of this report is to review the available literature and organize the current body of research in order to take advantage of what is so far learned and where future research could/should be directed. Part of the report also explores a conceptual way to integrate productions systems with preservation systems in a way that is transparent and seamless and supports Synergetic Preservation and the Foundations for Managed Forgetting.

1.3 Process

We conducted a literature review with the following question in mind: Which preservation tasks and information management workflows have been reported which relate to the “seamless interaction” between active information systems and digital preservation systems?

We investigated literature and articles in areas such as preservation workflow, digital preservation workflow, and content management, in order to find discussions about integration of content management systems with digital preservation systems.

In addition, we reviewed the recent European research projects on digital preservation [1] with regard to their experiences and suggested solutions, which could be used for seamless interaction and integration between active information systems and archival services. While the major proportion of “preservation workflow” –articles was focusing on the digital preservation workflows *inside* an OAIS-compliant archival system, we found altogether 32 articles (among a few hundreds of resulting hits with the above-mentioned keywords), which focused on interaction between archival services and separate production systems. The literature review can be found in chapter 2

We also informed ourselves about the situation in the application cases that we have in the project by conducting two discussion oriented workshops, one with each case owner. The statements from the workshops are taken at face value, and not scrutinised and analysed as such. The results from the workshops are summarised in chapter 3.

The results from the workshops are compared to what is said in the literature review and a suggestion for how to move on in the project is presented in chapter 4.

2 Literature Review / State-of-the-Art

This section covers the literature review described in 1.3 and concludes with reflection on what this entails for the ForgetIT project.

2.1 Review: Seamless interaction between digital preservation and information systems

Need for digital preservation has emerged lately in many areas. The traditional areas for digital preservation have been libraries, archival / curation institutions, and certain scientific institutions. Previous research on digital preservation has largely focused on the issues and problems in these special areas [1]. However, during the latest years, new research and educational institutions [5, 12], e-health, e-government, e-commerce [1] and other business areas, such as design and engineering [7], have started to increasingly recognize the need for systematic long-term preservation of their data and content resources.

While the amount of content to be preserved is exploding in the digital era, the diversity of data resources and methods to store and organize them poses a challenging problem as well [13]. Digital libraries and archives in several specified areas may involve complex digital objects, relational databases [14], even whole workflows and their business semantics [2, 3], which use advanced data models and require rich models for organizing bibliographic, contextual, semantic and technical metadata. Often, these metadata models depend on the very type of the technical implementation of the content objects in question. Hence, current business information systems and their repositories rarely, if ever, provide methods for **automated** preservation workflows or support for preservation metadata. This makes digital preservation, still, a manual and labour-intensive task [13].

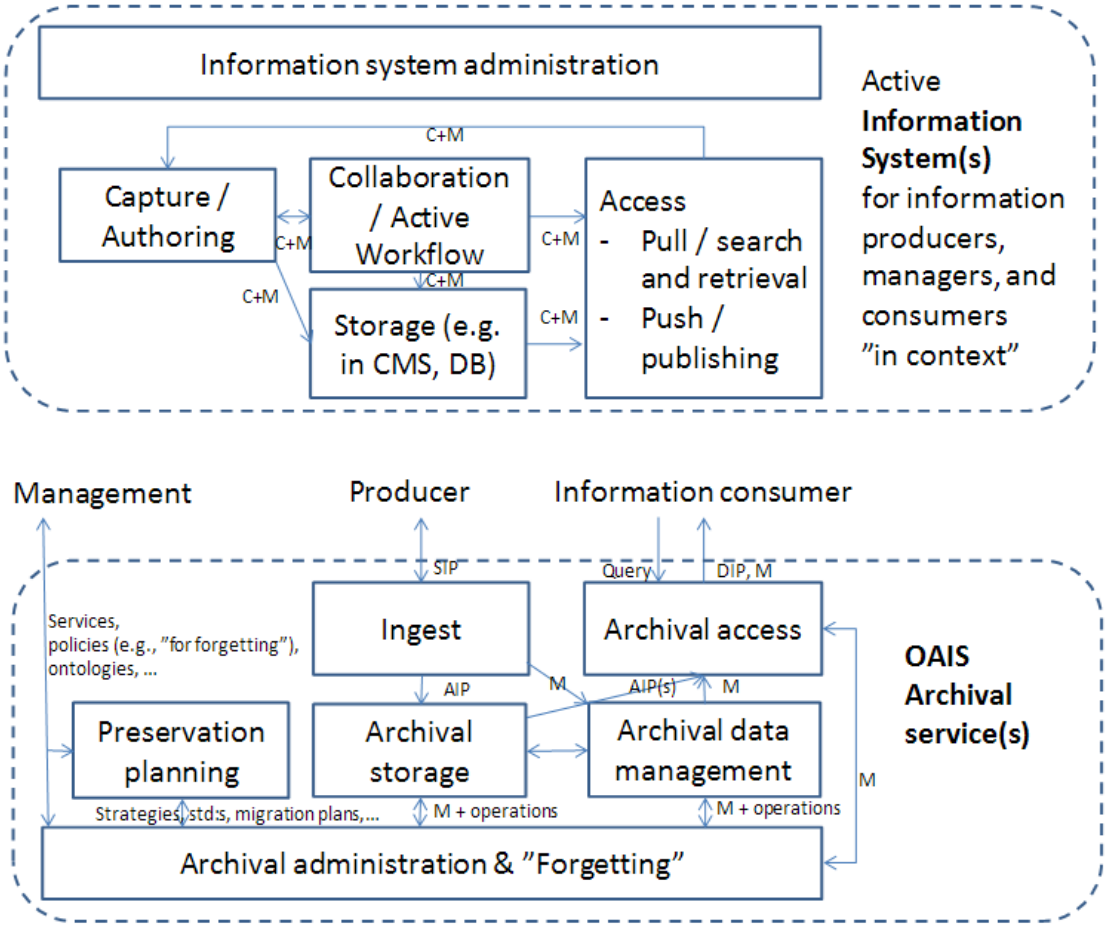
However, the penetration of digital preservation practices in industry is still low and organizations are not willing to use a lot of resources on the preservation activities [4]. Long-term preservation processes should thus be as effortless, transparent, and automated as possible, despite of dealing with challenging data formats [7], because the manual content preservation of the growing digital volumes will become unsustainable [6]. This requires increased integration of digital records management and archival systems and services with “external systems”. Modern preservation actions should thus be based on (automated) business rules, not on human actions, and this causes a need for the “external” information systems to evolve as well, to meet the requirements of automated submission management for digital preservation [6]. Moreover, digital preservation services are becoming cloud-based, which poses the customers of preservation with new set of challenges and decisions with regard to acquiring and implementing the desired “service levels” for preservation [15].

In the field of information systems, especially with regard to enterprise content management (ECM) systems, preservation and records management are from early on mentioned as relevant functional areas of interest [16]. However, preservation issues have rarely, if ever, been demonstrated or evaluated as parts of the reported cases [17] or analysis methods [18]. In the recent reviews on the field of ECM, no automated or deeply integrated digital preservation workflows have been presented in relation to the ECM solutions [17, 19]. Even the content management interoperability standards, such as CMIS [20], focus more on the active storage and immediate management and sharing of digital content. Hence, the issue of “effortless, transparent, and automated” integration between content production and management systems and the archival services is still in its infancy. However, several of the recent research efforts on digital preservation have addressed **parts** of such workflows on a couple of specific content domains [1], which warrants a more detailed review on this issue.

Many, if not most, of the literature found still focused on plainly mentioning the problem of seamless interaction instead of suggesting solutions. However, we attempt to summarize the main points of the “state-of-the-art” with regard to the related research below.

2.2 OAIS-compliant archival services in relation to enterprise information/content management systems

The OAIS-standard defines the internal functionality to preserve digital information which is submitted to the archival system. The OAIS-standard does not define how the management, producer and consumer stakeholders, who are supposed to interact with the OAIS system, are technically supposed to interact with the service. However, in the era of ever-increasing amounts of digital content and data, the most of the actions taken by management, producers and information consumers would most likely be through other, active information systems – such as enterprise content management systems



(

Figure 6).

ECM systems have varying methods to capture metadata and content objects into the **active** storage systems (see the upper half of

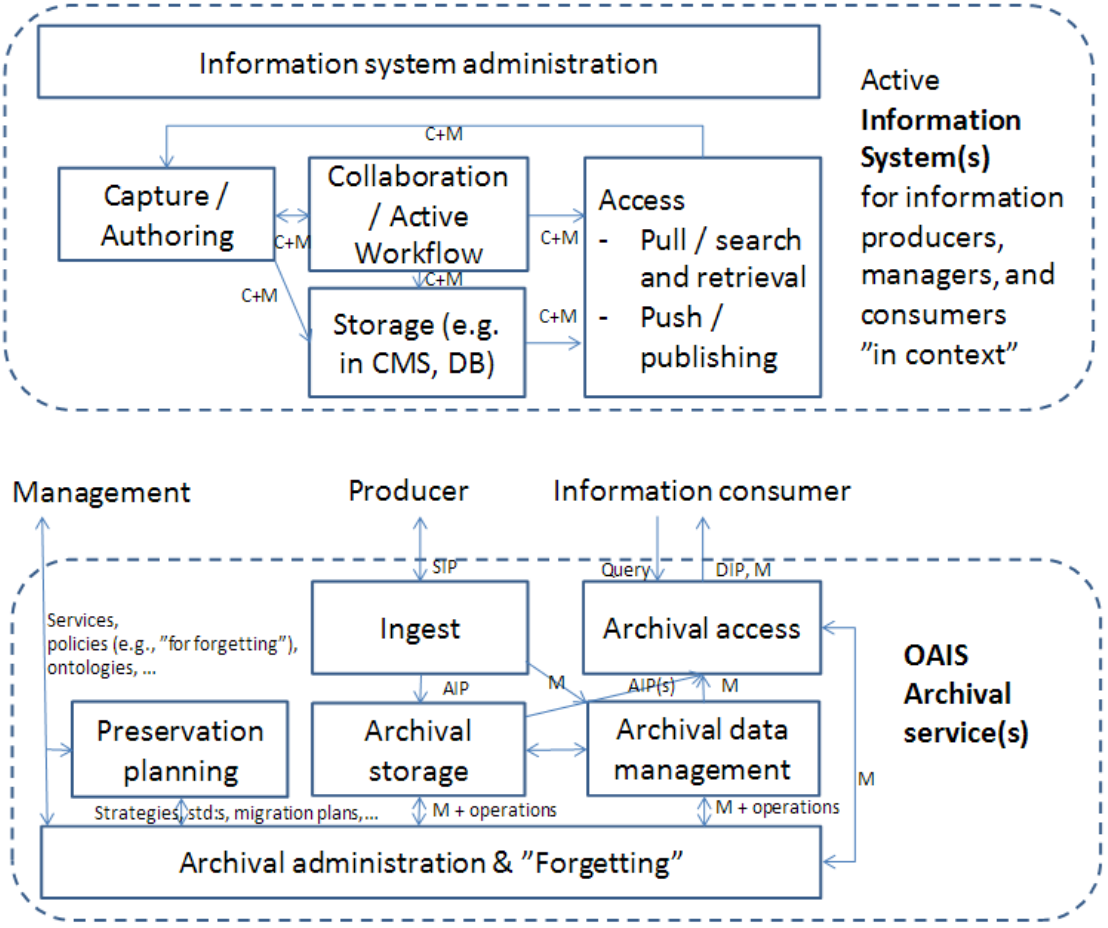


Figure 6; cf. [17]), they also mention "record management" and "preservation" as areas of ECM [16, 17], but the literature has so far discussed or evaluated few, if any, solutions for those functions in connection to ECM systems. For example, the CMIS standard on CMS interoperability does not discuss about preservation or records management –related functionality in detail [20]. As well, the CLIF Project [21] has recently determined requirements for moving content repositories across different systems – however, the CLIF project talks little about really long-term preservation. While ECM gets ever more common also in smaller organizations, some recent cases have recognized maximally 5-year periods as the longest needs to preserve their digital content beyond the active systems, e.g. based on product guarantees [19]. Hence, the focus of ECM storage function has mainly resided within the existing and active ECM systems, and less on the longer-term digital preservation integration.

All in all, the review of recent literature on both ECM / information systems and OAIS-compliant preservation and records management reveal that their integrations, especially holistic and

seamless integrations, remain to be clarified in detail.

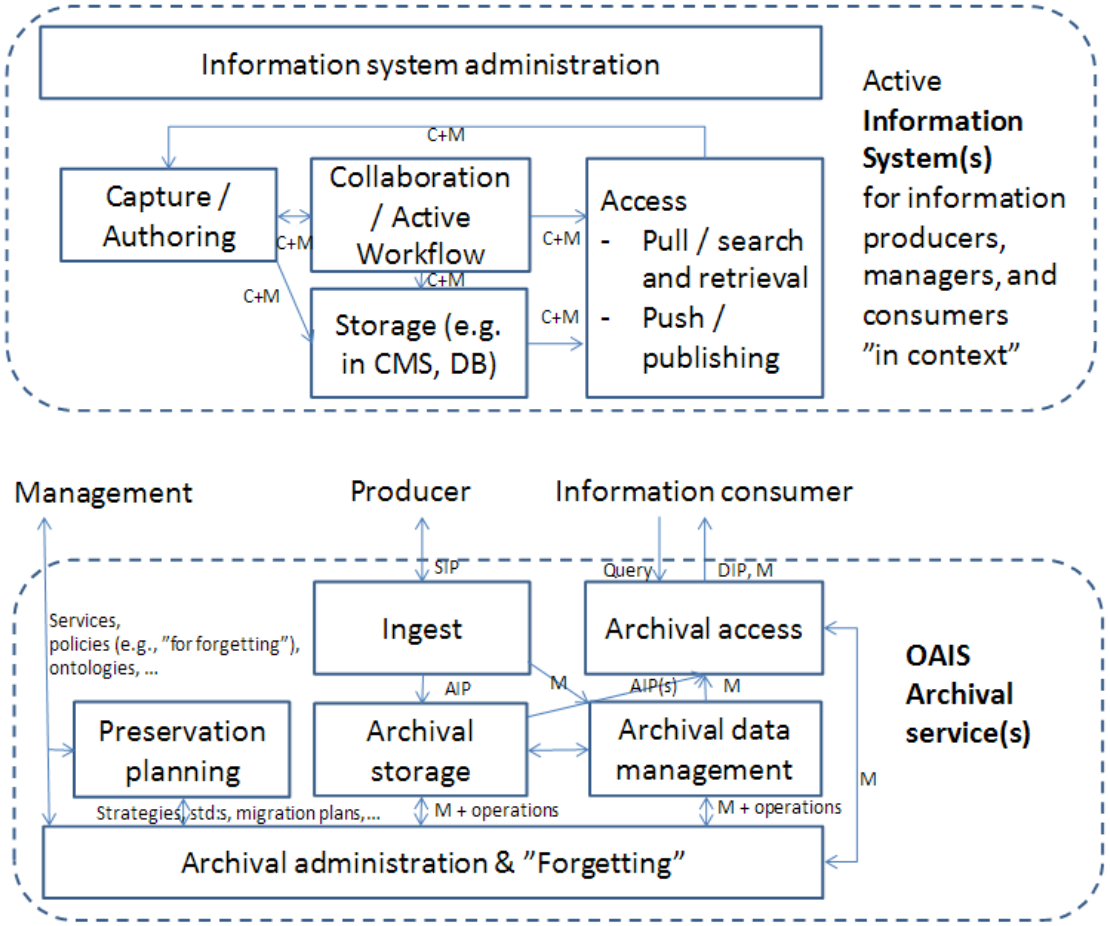


Figure 6 illustrates this "gap", in which the active information systems need still to be more explicitly aligned with the management, producer and information consumer roles in order to interact smoothly with OAIS-compliant archival services.

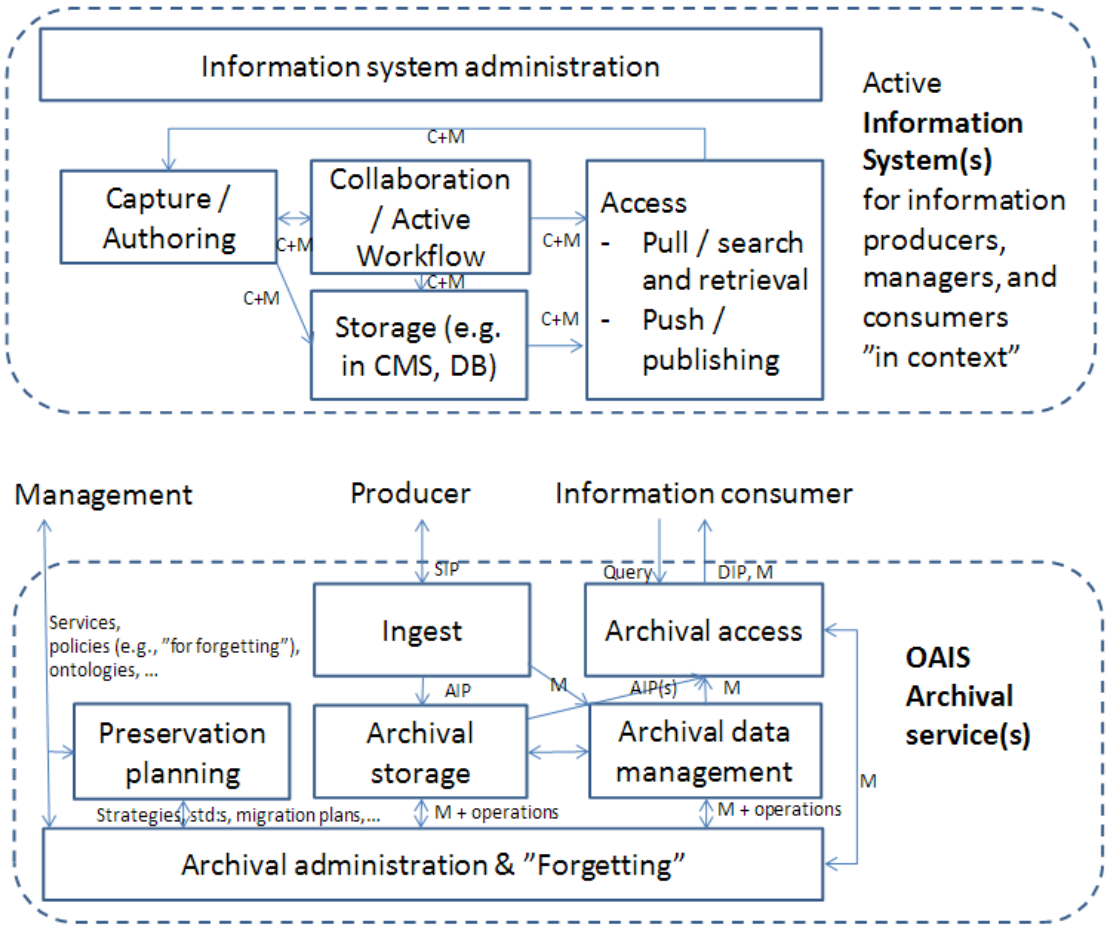


Figure 6: A gap between enterprise information/content management and OAIS archival services, as observed based on the literature review. (Legend: C = content, M = metadata, SIP = submission information package, AIP = archival information package, DIP = dissemination information package)

Especially, Korb & Strodl [22] have argued for the following gaps to exist between contemporary enterprise content management (ECM) systems and archival services:

- ECM provides no preservation planning and preservation control functionalities; those are to be taken care by the Preservation administration of OAIS Archival services [22]. However, it remains unclear how to implement these issues integrally into ECM systems.

- In the records management part of a typical ECM, the migration has been largely the migration of data from one storage medium to another, not so much about migrating file-formats when the old ones have become obsolete due to a change in an ECM system [22]. Hence, there is a gap between the processes of how to administer information systems and how to plan for and to administer the archival services connecting these two together.

- ECM-capture collects information produced by the organization. Content capture gathers also (in a few systems also automatically) metadata about content ownership, access rights, and other organizational issues related to the context and content lifecycle. However, the capture / edit also need to provide information which is to be preserved (so

that the ingest-function can prepare for that). So far, this functionality is largely lacking from the active information systems [22].

- ECM-systems are mostly integrated to the organizational IT-infrastructure, while preservation services are often external from them. Administrations of organizational information systems and external preservation services thus often lack alignment [22].

- While ECM-capture has collected some contextual metadata, the ingest function needs to add to the preservation-related metadata (file formats, representation, and other preservation metadata) [22]. This should be based on preservation planning and administration (and forgetting) policies.

- The preservation system needs to store descriptive metadata separately from the actual content. In the ECM systems, these are often tightly coupled [22].

However, despite of the general-level lack of integration between active systems and archival services, our review also revealed that a few potential workflow elements in a few content domains have started to emerge due to research in the fields of content management and digital preservation. In the following, we will look at the hitherto documented related research on content workflow integration which may happen in between the active information systems (such as ECMS) and OAIS-compliant digital preservation and archival services.

2.3 An integrated model for smooth interaction workflows between ECM and OAIS-compliant archival services: need for contextualization middleware

To organize our literature review, we follow the idea introduced by the Protage [23] and SHAMAN projects [e.g. 24], which have reported conceptual frameworks and prototypes for middleware to be placed between (cloud-based) preservation services and the information producers, managers and consumers. However, whereas the SHAMAN project itself discusses little about integration of the middleware to the active systems, such as ECM systems, we suggest further three logical elements of that middleware, which are to be related to three abstract-level workflow types that need to be implemented between maximally automated and smoothly interacting information systems and preservation services.

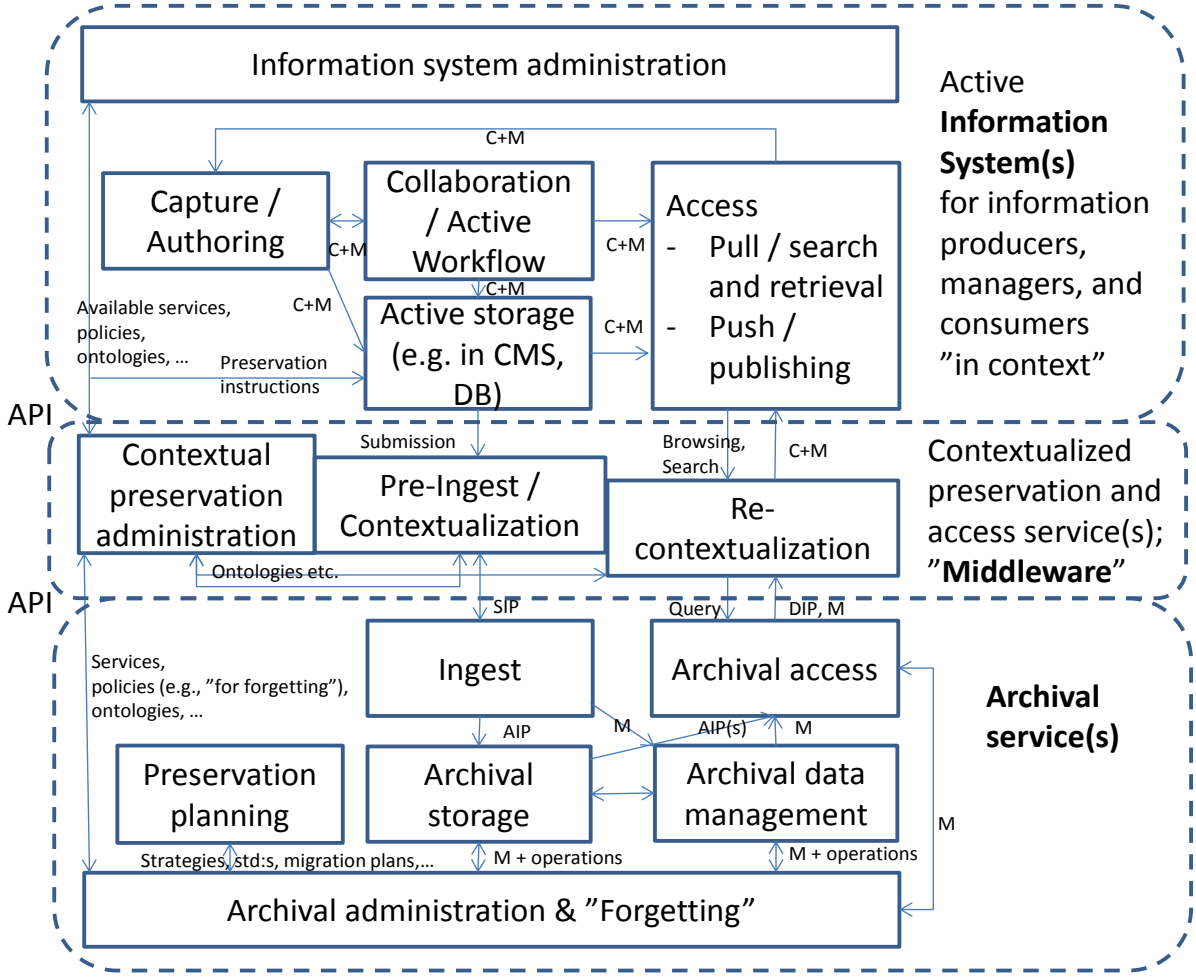


Figure 7: A “unified view” on the “smoothly integrated” active information system, middleware for contextualized preservation and access services, and archival services. (Legend: C = content, M = metadata, SIP = submission information package, AIP = archival information package, DIP = dissemination information package)

In line with the SHAMAN project, we recognize the need for the **preservation** (pre-ingest) **workflows** and **retrieval workflows** [24]. The SHAMAN middleware [24] recognizes also a need for additional support services (such as natural language processing and data mining) to support these two workflow types. However, in total we would also like to introduce a third type of workflows, namely **contextual preservation planning and administration** workflows, which need to be prescribed and implemented in order to configure the maximally automated preservation and retrieval workflows. In the middleware layer, which is located between an active information system and selected archival/preservation service(s), this involves at least three logical architectural components under which the actual functions involved in these workflows are taking place: *contextual preservation administration, pre-ingest/contextualization,* and *re-contextualization* (see the middleware layer in Figure 7). These parts will be discussed in more detail in what follows.

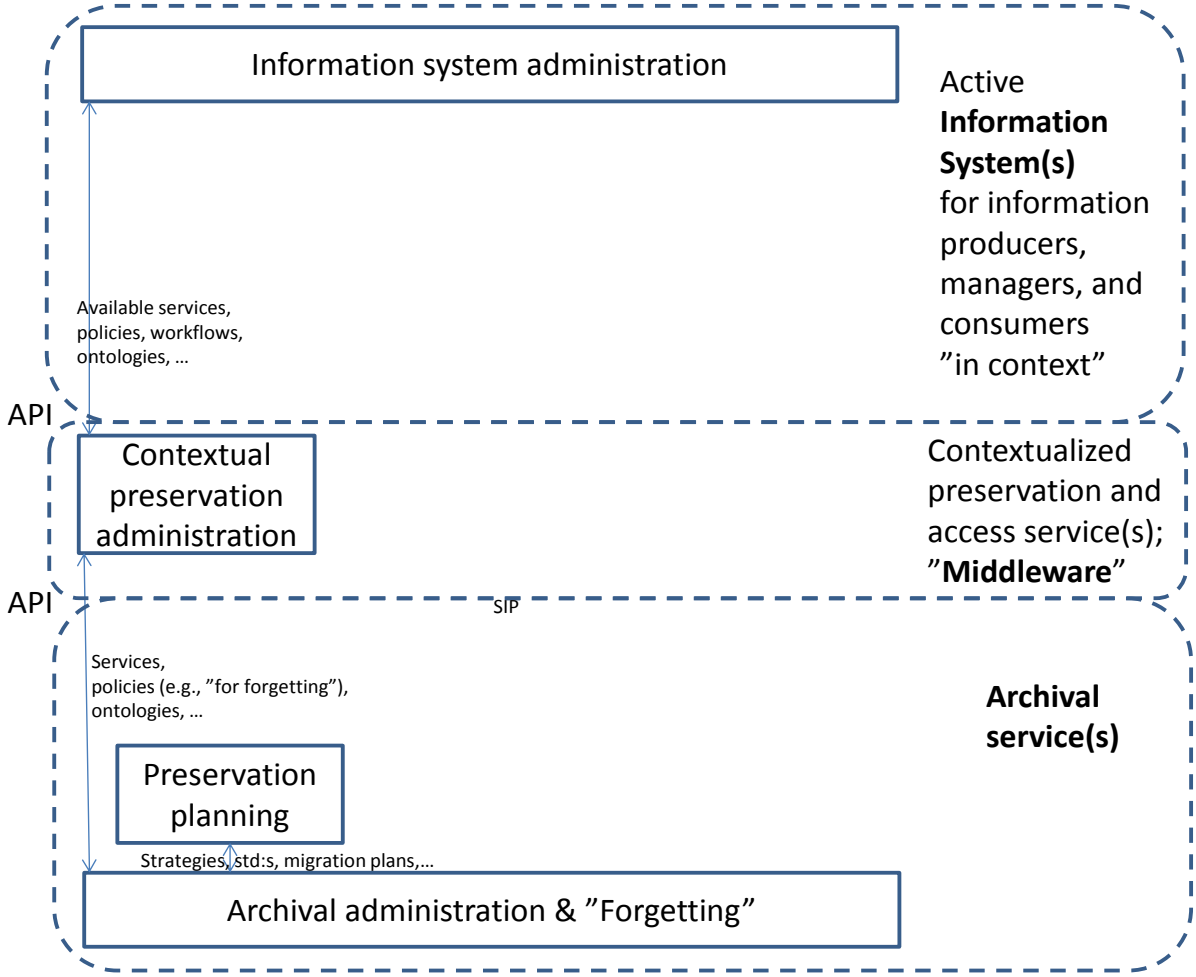


Figure 8: Abstract elements of the contextual preservation planning and administration workflow

As Korb and Strodl [22] have addressed in general, not much research and development has taken place to fully integrate the preservation planning and administration to the information systems in their context. For example, the OAIS-model and archival services do not necessarily include mechanisms to involve content retention rules which are set in the contextual content management systems. On the other hand, CMSs may lack mechanisms to integrate preservation planning rules and policies that have been defined in the OAIS. In order to make such interaction to happen, especially in the modern environment where one organization could use even more than one (cloud-based) archival service, a middleware element for contextual preservation administration is necessary.

Although this kind of workflows at the preservation planning, administration and implementation of the policies have not been comprehensively discussed in the literature, we could recognize a few elements in the previous research which are worth considering while designing for more detailed workflows for particular solution types (Figure 8). The SHAMAN project has developed “Preservation Management & Planning Interface” [25] to manage creation of preservation information under the “pre-ingest” phase. The PLANETS infrastructure provides a web portal framework that integrates a set of end user applications with a number of data repositories and a federation of grid/web and other services, such as data/metadata management, preservation, information and workflow execution [26]. Tarrant et al. [27] discuss about possibilities to connect the preservation planning element of the PLANETS project, Plato, with digital repository interfaces. Laleci et al. [28] introduce a semantic backend for content management systems, which could be useful for creation of relevant ontology mapping definitions between content repositories (including ontologies for preservation).

At the conceptual level, Nguyen and Lake [15] propose a set of differentiated service levels for defining utilization of cloud-based archival services. For example, the administrators and preservation planners should be able to decide whether to ingest with “transfer only”, “format identification” or even with “metadata extraction”. Archival storage could be chosen to be implemented with delayed access & near-line storage or rapid access & high-performance storage. Content servers could be implemented to function either just-in-time or to be always active. The decision of the preferred service level furthermore will have impact on the actual workflows, tasks, and the experienced smoothness and seamlessness of the preservation and access workflows.

However, these recent advancements remain at the level of prototypes which are so far (reportedly) created in research projects for research purposes. Moreover, it remains unclear, from the academic documents of the reported services, how the distribution and execution of the administration and preservation planning workflow tasks would take place in these solutions. Hence, the ForgetIT project needs to evaluate the technical environments of the SHAMAN and PLANETS administration frameworks further in more detail, to adopt the available readily-made components, and to define the full integration automated preservation and access workflows of this area with regard to our active information system use cases in more detail by itself.

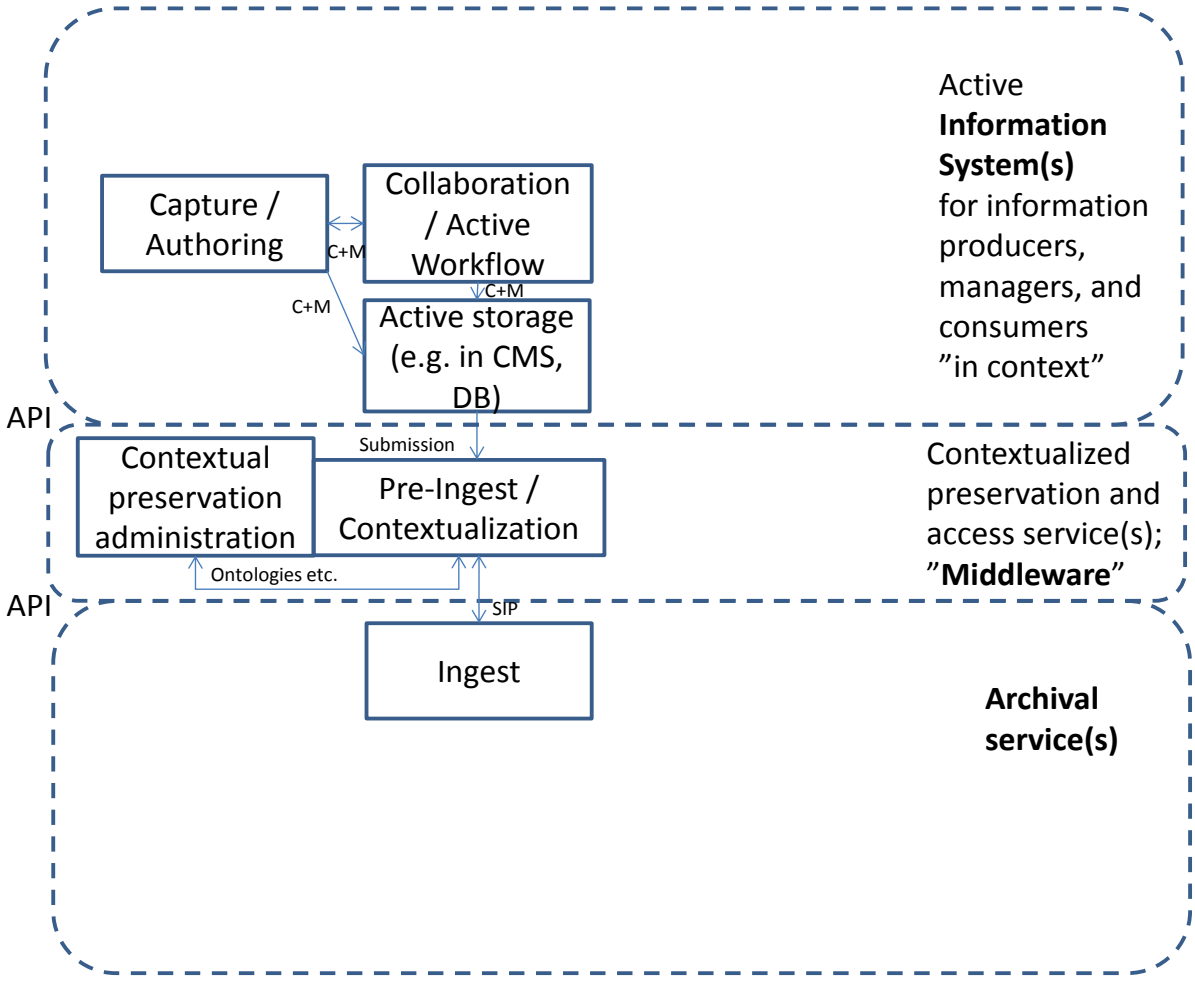


Figure 9: Abstract elements of the preservation (pre-ingest and contextualization) workflows

A few initial and partial solutions for preservation workflows to cover pre-ingest and contextualization tasks have also been suggested, while we need to keep in mind that many, if not the most, content management systems do not really provide support for producing preservation

description information and implementing longer-term preservation appraisal in the active production/storage systems [22]. The content management interoperability standard [20] provides data models for storing and sharing information objects services for interoperable CMS repositories. This could perhaps be regarded as a potential “first step” through which to enhance also interoperability of the CMS system with the pre-ingest and contextualization middleware (Figure 9).

From early on a few targeted, automated techniques to extract preservation metadata from particular types of content (e.g. journal articles; [29]) have been developed. The PAIMAS standard [30] was created already a decade ago to recommend practices for a producer-archive interface communications. Although the reported applications of the PAIMAS standard have been few, the CAST project [31] adhered to PAIMAS recommendations in their implementation of a collaborative archiving services testbed for web content.

Recently, the PLANETS project has provided a set of pre-ingest and preservation workflow support techniques, to integrate particular end user applications with data repositories. PLANETS also provides testbeds to test pre-ingest processes, and provides tools for data/metadata management, preservation, information, and workflow execution. (such as Digital Object abstractions, Technical Registry, Workflow application processing interface, Data Registry and Workflow execution engines) [26].

Alongside, the SHAMAN project has also created its initial workflow support and tools from creation of objects and context data, through “assembly” of objects with descriptive, context, and preservation metadata to forming submission information packages [25].

In this workflow, decisions whether to allow human intervention to the appraisal, pre-ingest and contextualization services need to be made. For example, Heutelbeck et al. [7] highlight the definite need for automatic extraction of metadata and validation of manually entered metadata in the field of engineering and design. However, some researchers, such as the Prometheus project for the needs of libraries [32], have simultaneously continued to develop tools which ease human interaction with archival services during the pre-ingest and ingest workflows.

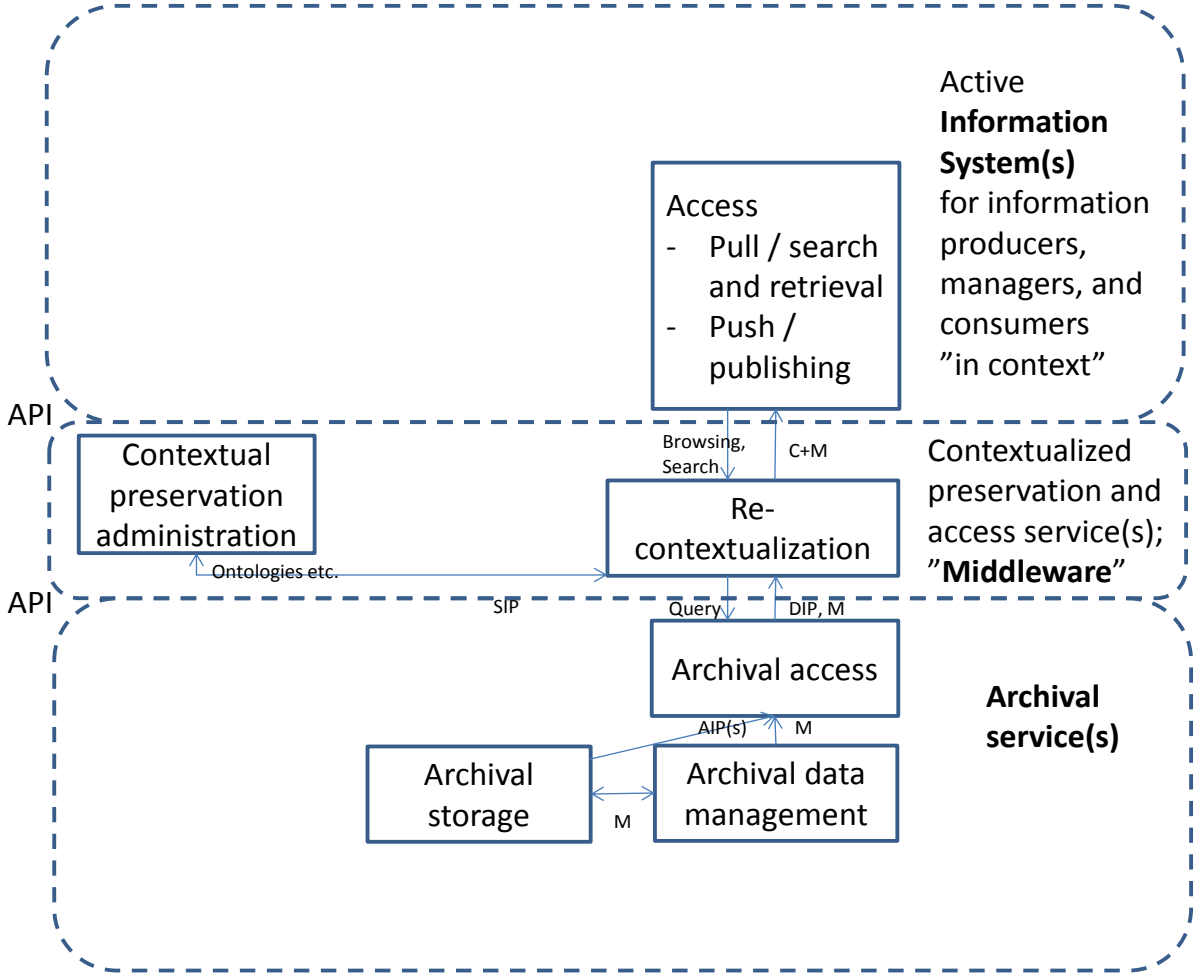


Figure 10: Abstract elements of the retrieval and access workflows

Although content management systems mostly do not include functionality to process dissemination packages [22], several prototypical solutions for retrieval and access workflows have been suggested as well. If the middleware could process the dissemination information packages to the CMIS-compliant format, any content management system could then regard the archival storage as one “CMS repository” to which it could access by its standard getRepository –services [20]. The advances of “semantic backends” for content management systems could also be useful here for creating mappings between the query ontologies and preserved ontologies [28]. After the early focus on preservation and preservation planning workflows [33], the SHAMAN and PLANETS projects have as well provided their own toolsets, such as PLANETS Digital Object manager, Data Registry, graphical user interfaces and pluggable access components (such as xml – PMH/ORE, SOAP-Fedora, and html-WEB) [26] and SHAMAN’s “Multivalent browser” [25] to access and retrieve preserved information resources. SHAMAN also provides tools for “adoption” of objects, including their descriptive metadata, context metadata, and preservation metadata through a graphical user interfaces so that the objects can be processed for reuse [25].

In the field of preserving research experiments, Page et al. [3] have even reported about their advancements on how to retrieve and re-contextualize preserved workflows.

Again, the selected “service level” for discovery and access services to archives will influence the actual workflow design and implementation. For example Nguyen and Lake [15] suggest that the discovery service can be based on either metadata or full content searches; and access services may involve possibilities to access through a passive viewer, interactive viewer or full-scale content mining facilities. Hence, the access and retrieval workflow needs truly to interact with the contextual preservation planning specifications (Figure 10).

2.3.1 Conclusions

- The need for digital preservation has been lately identified in several business areas and public organizations beyond the traditional fields (museums, libraries, and research organizations) where the preservation research started a bit more than a decade ago. Organizations in the fields of e-health, e-government, engineering and design, and industry in general have started to recognize their digital preservation challenges. However, the field in general is still in its infancy and the preservation practices and systems have not been widely adopted (e.g. [4]).

- The need for seamless integration of preservation to varying kinds of information systems has been recognized and a few authors have mentioned this issue. Especially, the ever-exploding amounts of digital content require maximal automation of the preservation workflows in business environments. However, the solutions to automatize preservation processes and integrated access to preserved information from information systems are in their infancy. At best, the research has provided some early demonstrations and prototypes for the purpose.

- Organizational information systems, especially in the field of enterprise content management, have recognized the functionalities for records management and digital preservation for a long time, but research and software development on these fields have not yet focused on how to integrate really long-term digital preservation seamlessly to the ECM systems. Rather, focus has been on the active content storage and interoperability between active content management systems through standardization efforts (e.g. CMIS), but not so much on interoperability between content management systems and long-term preservation systems.

- For the purposes of the ForgetIT-project, we suggest that we need to distinguish between the following functional areas and to have a closer look at their interoperations to concretize the “seamless interaction” between active information systems and digital preservation

    1. The **active information systems**, such as an ECMS, a database management system, or a knowledge management system in use at the organization are the sources of data and content to be preserved. More than one active system may need to preserve content in an organization. However, the following functional areas need to be clarified further and implemented for the seamless integration:

        - Information Systems Administration needs to define and implement the content appraisal, selection and preservation policies and (to the extent possible) automated workflows in the context of the information system in question. This functional element needs also to be able to interact with Contextual Preservation Administration in order to contribute to preservation ontologies and other elements needed in the preservation contextualization and re-contextualization tasks. At least, this function needs to understand the content interchange and preservation interface standards which the information system in question needs to be integrated. The administration function also needs to coordinate and implement the control functions which

guarantee that the content creation/capture function produces the necessary contextual metadata for the preservation purposes.

- The Active Storage function needs to be able to interact (automatically, according to pre-defined preservation workflows) with middleware (or preservation system(s)) to which it is supposed to submit content and extract the necessary metadata packages which are to be preserved.

- The Access function in the information system(s), which are meant to utilize the preserved content in the selected preservation system(s), needs to be able to interact with the re-contextualization services in order to pull content from them for (re-)use or publication; also so that the inquiries and uses of the preserved content leaves a trace in the preservation system. As well, it needs to be able to utilize metadata provided by the re-contextualization services for discovery, browsing and retrieval of the preserved content.

2. Contextualized **preservation and access middleware** forms a bridge between particular information systems and one-to-many preservation and storage services. One organization can, on the one hand, use many preservation and storage services (e.g. in the cloud), while, on the other hand, a particular preservation and storage service can accept submissions and access requests from several information systems. To manage these workflows, a middleware is thus required with the following functional areas.

   - **Contextual preservation administration** is needed to link the preservation and storage services to particular information systems and to maintain the contextualized preservation and access workflows, interfaces and ontologies.

   - **Pre-ingest and contextualization functionality** is required for finalizing a submission information package for the ingest function of the particular preservation and storage service.

   - **Re-contextualization functionality** is required for linking the selected preservation storages to the selected active access and retrieval solutions, by using adequate ontologies and standards. For example, the re-contextualization interface could “process” the dissemination information package to conform to the CMIS standards where the receiving CMS could see the preservation and storage system in question as a “standard repository” to be shared with the active content repositories. As well, the re-contextualization middleware might provide adequate metadata and ontology for the active system to enhance discovery and browsing of the preserved resources.

3. In an **OAIS-compliant archival** service, the functions of preservation planning and administration, ingest, and archival access need to interoperate with the contextualized preservation and access middleware.

- In total, the “seamless interaction and integration” needs to define three more detailed workflows and their relations, which engage the above-mentioned functional areas of the

active information system administration, contextual preservation middleware, and the archival service.

1. The **preservation planning and administration workflow** results, on the one hand, in the implemented preservation policies and configured workflows in the information system to capture preservation and other relevant contextual metadata and policies already in connection of content capture, workflows, and active storage. On the other hand, it will result in the implementation of the preservation middleware interactions with the information system and the adequate preservation plan in the archival service.

2. The **contextualization and pre-ingest preservation workflow** from active storage and appraisal through contextualization and pre-ingest actions to ingest resulting in submission information packages should be automatized through the extent possible through pre-defined workflows reacting to events and utilizing the pre-defined interfaces between the information system, middleware, and the archival service(s).

3. The **access and retrieval workflow** from the active access through re-contextualization services to the selected archival access functions is required for seamless discovery, browsing, and re-use of the preserved resources. It may involve automated publishing and discovery requests, as well as human-computer interaction with the adequate access tools.

- The above-mentioned workflows depend a lot on the actual domain of information to be preserved and accessed and the related metadata standards and models. Hence, the subsequent development may require several application processing interfaces between the middleware and the selected information systems in question. ForgetIT-project will focus on the TYPO3 content management system (and the CMIS repository interface standard) and the challenges provided by the DFKI personal knowledge repository, semantic desktop and PIMO, and the related preservation needs.

Closer evaluation of the recognized related research prototypes and results has not yet been possible within this initial state-of-the-art review of the literature. However, we recommend that the next version of the ForgetIT workflow design first decides on the adequate “service levels” for the preservation administration, preservation and retrieval workflows. After that, a closer mapping and more detailed task list under each of the above-mentioned abstract workflow components needs to be created and the available solutions (such as CMIS functions, PLANETS- tools and SHAMAN-tools) are mapped and evaluated in light of a more detailed model. However, our expectation is that the contextual preservation planning element of the middleware still needs to be developed largely in the project – especially if and when the “forgetting” functionalities, which form the core of the ForgetIT-project, are to be added to the overall workflow picture.

3 Workshops

This section describes the workshops and the outcome of them.

The idea with the workshops was to gather ideas and opinions on what would be suitable ways to integrate the active systems with the preservation system. The opinions in this case were gathered from the two use case owners (DFKI and dkd), in separate workshop sessions. In one of the sessions, the integration workpackage leader (EURIX) also participated in a mostly passive role, since the discussion of course was interesting for them as well.

3.1 Personal Information Management case

DFKI have worked on the Semantic Desktop and the related PIMO (Personal Information Model)[2] for several years. This system, and model, holds (personal) information that is semantically linked to make discovery easier. The information types could be of several different kinds, but in a professional case typical information would be in the form of e-mail, text documents, images and maybe also audio files. They could contain invitations, travel plans, meeting notes, business records, customer data, and so on.

3.1.1 Scenario

In the example we consider a transition between desktop, server, and archive.

After a project finishes, the results, material, and working details decrease in importance. However, importance decreases differently depending on the type of the resource: The user accesses project results such as deliverables, related work, presentations, etc. after the project from time to time because of reuse, consultation, presentations, etc., however, in decreasing frequency. Also new material comes up which summarizes the project details for use in new situations (e.g., slides that present the project and its results to a new audience) for which the need for the original material decreases again. Concerning working materials such as versions of deliverables, fragments, notes, tasks dealing with them, are decreasing faster in importance, if not already unimportant as soon as the results (e.g., deliverable) were produced. Concerning the user, there could be a clean-up which removes all those working material and clutter from the desktop and out of immediate view, i.e., in order not to occasionally stumbling into it, but which keeps it on access somehow, because you never know what you might need (e.g., want to reuse). Again over time, all those details decrease more in importance which could mean even to delete those details. Expectation of the user would be not to care about it but to retrieve it in the end.

Files: If the project is finished a cleaning of the individual files could start: variants of files can be moved to the archive, only most important files stay on access on the server such as deliverables, project overview and result slides. Over time these files can also be moved to archive and removed from the server (decay; reduced likelihood of required access to those files).

Model: Condensation takes place in the model, i.e., working traces are reduced to “proxies”, e.g., files, notes, tasks used to work on a deliverable are removed from immediate view of the user (i.e., if the user browses the project, only the deliverable proxy will be seen), if details are required the user can access the details. This is done over time, i.e., after a period of time, further condensation takes place and also the deliverables are merged in the project proxy, and so on.

Currently we will keep the statements for a proxy still in the model, to allow to answer queries, which are expected to be known to the user if the user tries to remember (hard). However, over time the PIMO model will be compacted and such details will be cleaned-up in the model. The removed statements of the model are then only available from the archive. The user might ask a

[2] http://www.semanticdesktop.org/ontologies/pimo/

question to such a proxy where the details are required for the user. In that case, the detailed statements would be retrieved from the archive and fed back into the PIMO. (We are not yet sure how to do this technically, i.e., to introduce a separate and temporary db table, which can be deleted afterwards. During this, it is possible that some statements are taken over to the main PIMO (reuse)).

Seamless transition back from the archive in this respect would be, that the user is able to find the proxies in the model and if requested is able to retrieve the archived material. There might be several steps: first retrieve the archives part of the model and then – again on request – retrieve resources such as files from the archive. Thus, the model serves as a preview, the user could choose what to retrieve and access from the structure.

3.1.2 Main Ideas for Integration

From the perspective of PIMO and Semantic Desktop, the main idea is that the preservation should be something that is fairly transparent or non-intrusive to the user of the PIMO. This might demand a fair amount of integration in the background, where for example CMIS can be used for exchanging objects between the PIMO and the ForgetIT system. Worth noting here is that we here talk about a ForgetIT system that acts as a middleware between a production system (active system) and the preservation system (an OAIS).

Table 1: Seamless transition over the whole range of stages in ForgetIT

	Mobile	**Desktop**	**PIMO server**	**Archive**
Recency	Immediate	Up to medium-term	Medium-term	Long-term / backup
User Expectation	• Immediate relevant material is available • Recency in model as well as important concepts (hotness function) • Only most recent versions of resources required • Constant swapping/refreshing (in WLAN) to keep recency is OK	• Quick availability of recent & work related material • Broader range of resources: work relevant are available on immediate access (e.g., think of offline requirements); outdated ones are available on request • Clutter of not-relevant stuff is gone, i.e., details can be removed from desktop • "relieve the burden to see everything"	• Provides access to all material • Condense and compact to get a sharper view • Old versions of files can be deleted & moved to backup/archive	• All versions are available upon request • However, over years, detailed material would be condensed, i.e., keep the deliverable but not the versions of the file, not necessary any more (e.g., after 1 year)
Analogy to human memory	working memory & short-term memory	working memory to medium-term memory. Currently activated episodic and semantic memory	working memory & long-term memory. Episodic and semantic memory	Long-term memory and life-time storage

DFKI have an idea about what the *seamless transition* would entail in the PIMO context [Table 1]

The idea is that the PIMO/Semantic Desktop should be able to provide some preservation related metadata that suggests e.g. if an object *must* be preserved or if there is a known retention period. Otherwise the actual "forgetting" should be taken care of by the ForgetIT system, or in some cases even in the OAIS/storage.

One benefit of the above is that relatively small changes are required in the existing production system, and most of the decisions are made by the ForgetIT middleware. Some prompts may be made back to the user in certain circumstances set out in e.g. a profile for the user.

The above idea of course require quite a lot from the middleware, but if we in that way can focus the implementation of (automated) decision making into a universal component it seems as a good use of resources. This also means that the middleware can be enhanced to support several transmission standards for exchanging objects between the production system and the preservation system.

The two-step model suggested in the scenario description fits very well with the OAIS access, where the user usually first queries the system for a list of relevant "hits" on a specific query and then chooses more specifically what objects they want to retrieve in full.

3.2 Content Management System case

In the ForgetIT project, we are using TYPO3[3] as base for our content management system case. Dkd serves as the TYPO3 experts in the project and are the case owners of the CMS case.

3.2.1 TYPO3 context

People mainly use TYPO3 as a web based content management system, creating a web site both for external and internal usage. TYPO3 has versioning, but it is handled from a more "engineering" kind of way instead of something (end) user friendly. This versioning system also includes an audit trail that shows what has been done and by whom. Worth noting is that comments made on articles (pages) in TYPO3 also are stored in the system and therefore also could be preserved, other on their own, or with a relation to the original post.

A TYPO3 installation can host several websites, e.g. for subdivisions of a company, or different national branches. Very seldom, one installation hosts different companies or organisations since the backend is sharing the TYPO3 database. In the cases with several subdivisions on the same server, the preservation of them would most likely need to be dealt with separately. Here it is worth mentioning that most of the customers (to dkd) idea of preservation is "backup" at best.

For transition (to the preservation / ForgetIT system) it will most likely be the administrator of the TYPO3 site that sets up such an integration or workflow. The editors (i.e. the ones creating the content) will probably neither want to be bothered with that, nor be capable of doing that (e.g. due to lack of rights). In general, the users probably do not want to make any proactive decisions about preservation, but could/should be prompted when something is about to happen in the ForgetIT system.

Regarding workflows, most customers do not use any advanced workflows in Typo3. There is an "advanced workflow" which whenever a page is created or edited it is held in a separate workspace. Otherwise there is no particular workflow in that sense. There is however an important thing to consider, which could be seen as part of a workflow. The system need to be able to deal with access restrictions, that could be either individually based or group based.

[3] http://typo3.org/

3.2.2 Main Ideas for Integration

At the moment dkd is working on how to send and receive information to and from TYPO3, with the main candidate being CMIS. Dkd is also working on visualization of what is within TYPO3. This is done in order to get a good grasp of what type of content that are maintained in TYPO3, which would help in managing content including making preservation decisions. TYPO3 in itself is not arranging content according to any standard, but has its own way of doing it which is "page centred" with hierarchical organisation of content. Everything happens on a page, which can have modules and elements in it. The same content element could however be used on several pages (e.g. a logotype, or a page footer).

Some kind of preservation or ForgetIT module could help Typo3 administrators (and users) to view what content that has been changed recently, or not changed at all. This module could then also serve as a point of access for the material that is brought back from the preservation system. Regarding what should be sent to the preservation system, the first idea is that we should send everything we create to the ForgetIT system, since the users most likely would not bother doing any selection or preservation decisions up front. In other words, if it is published; send it to the ForgetIT framework.

The audit trail that is already in place in Typo3 could certainly help in analysing content and how often it has changed, including the scope of the changes. Small changes might perhaps be merged in the ForgetIT system, instead of keeping each version separately.

An idea is that when an editor starts working on a new page (or editing an existing one) the system could give hints on that similar content already exists in the "archive" e.g. if the new page is linking to the same resource (web page, pdf document …) as an existing one.

Overall, the "seamlessness" in access and retrieval from the ForgetIT system should preferably have:

- Ability to browse through the content with regard to some aspects of the metadata that you actually have (e.g. author, last time changed)
- Dashboard showing information on different aspects of preservation
- Indexing preserved record in the active system
- Being able to define which preserved records that automatically should be updated to the active system, which mainly should be in the case of using the ForgetIT framework for sharing content between several CMSs.

The actual information sent to the ForgetIT system could be configured by e.g. typo script, to decide for example that images should not be sent to preservation.

3.3 Summary of Workshops

In general, similarities between the cases came up and the idea about smooth transition very much translated into that the user should not need to be bothered that much up front with preservation decisions. Instead the suggestions went towards prompting the user when e.g. something is about to be removed from the preservation system. So, instead of having a proactive preservation planning, the idea is that the ForgetIT system would be handling information objects, and when it decides to do something (drastic) with it, it will alert the creator (or whoever is responsible) thereby promoting something that could be labeled as "reactive preservation planning".

One important thing to consider is the access restrictions that might be in place on certain digital objects. These restrictions need to be respected by the ForgetIT system. This also should include typical privacy concern.

4 Recommendation – Project Approach

This section describes and summarizes the outcome from the workshops and the literature review. It also outlines design principles for the integration.

4.1 Analysis

Based on both what came up in the literature review, as well as in the workshops, we can see that the workshop ideas concur with e.g. Heutelbeck et al. [7] stating that the processes should be effortless and automated. In order to get non-professional users involved in actually preserving their digital information, and helping them in contextualizing the information as well as removing some of the surplus, the system needs to be non-intrusive. The system should not pester the users with constant questions on if something should be preserved. The cases also show that information systems need to evolve to better suit the needs for preservation, as stated by Stewart [6]. Both cases raise some concern about preserving privacy and access control to the preserved information, which should be taken into consideration although the project would not be making extensive research or development on something extra regarding those aspects.

4.2 Conceptual Integration

We recommend that the next version of the ForgetIT workflow design first decides on the adequate "service levels" for the preservation administration, ingest and retrieval workflows. After that, a closer mapping and more detailed task list under each of the above-mentioned abstract workflow components needs to be created and the available solutions (such as CMIS functions, PLANETS-tools and SHAMAN-tools) are mapped and evaluated in light of a more detailed model. In this way, the (producing) information systems might not need to evolve so much that it would be overwhelming, but could instead make use of community best practices or standards relevant for the application area, which the ForgetIT middleware then would (need to) support. So as next steps we should:

- **Investigate suitability of CMIS for interaction between producing systems and ForgetIT middleware:**

  Tentative results show that it might be useful for exchanging the information packages, but that much of the "forgetting" might need to be dealt with on the side. Not necessarily a problem, since this would mean that we could have one way to handle forgetting related communication with *all* systems, and then handle the information packages according to relevant standards or best practices in each area of application.

- **Define preservation planning and administration workflow:**

  This would include deciding on appropriate service levels [15] for the different workflows (Administration, Ingest, Retrieval). After that a closer mapping and more detailed task list of the mentioned abstract workflow components (for more details, see 2.3) would be needed. This could be done using e.g. the ECM-blueprinting framework described by vom Brocke et al. [12]. An evaluation of existing artefacts from e.g. PLANETS, SHAMAN and DuraSpace should be undertaken in light of the more detailed models coming out of this process.

- **Consider security/integrity/privacy:**

  The project should not conduct any new research or extensive development into the area of information security or access control, but we should at least consider having access control implemented in order to keep information objects private over time.

Although we would be evaluating existing tools and workflows for their suitability, our expectation is that the contextual preservation planning element of the middleware still needs to be developed largely in the project – especially if and when the "forgetting" functionalities, which form the core of the ForgetIT-project, are to be added to the overall workflow picture. One thing that still is undecided is how much of the "intelligence" that should reside in the active systems. One idea would be to send everything that is created in the systems to the ForgetIT middleware for

preservation decisions to be made there. This would however entail a lot of communication and might be undesirable in the long run. But the more decision we require the active (producer) systems to make, the more we raise the level of tight integration which also is undesirable.

What is reported and suggested here, will guide the initiation of coming tasks in WP5 of the ForgetIT project. The ideas and suggestions from this report will be put to test in both practice and theory in the coming tasks and deliverables of WP5, and further development of these ideas are to be expected.

References

[1] S. Strodl, P. Petrov, and A. Rauber, “Research on Digital Preservation within Projects Co-Funded by the European Union in the ICT Programme.” Vienna University of Technology, May-2011.

[2] R. Mayer, S. Proell, and A. Rauber, “On the Applicability of Workflow Management Systems for the Preservation of Business Processes,” in *Proceedings of the 9th International Conference on Preservation of Digital Objects (IPRES 2012)*, Toronto, 2012, pp. 58–65.

[3] K. Page and et al, “From workflows to Research Objects: an architecture for preserving the semantics of science,” in *Proceedings of the 2nd International Workshop on Linked Science*, 2012.

[4] S. Ross, “Digital Preservation, Archival Science and Methodological Foundations for Digital Libraries,” *New Review of Information Networking*, vol. 17, no. 1, pp. 43–68, 2012.

[5] S. Hitchkock, “Setting institutional repositories on the path to digital preservation.” JISC Keepit project, 28-Jun-2011.

[6] C. Stewart, “Preservation and Access in an Age of E-Science and Electronic Records: Sharing the Problem and Discovering Common Solutions,” *Journal of Library Administration*, vol. 52, no. 3–4, pp. 265–278, 2012.

[7] D. Heutelbeck, J. Brunsmann, W. Wilkes, and A. Hunsdörfer, “Motivations and Challenges for Digital Preservation in Design and Engineering,” in *InDP’09*, Austin, TX, 2009.

[8] Digital Curation Centre, “Digital Curation Lifecycle Model,” [Online]. Available: http://www.dcc.ac.uk/resources/curation-lifecycle-model. [Accessed 11 06 2013].

[9] DigitalNZ, “Make it Digital - Getting started with digitisation,” [Online]. Available: http://www.digitalnz.org/make-it-digital/getting-started-with-digitisation. [Accessed 11 06 2013].

[10] Portico, “Preservation Step-by-Step,” [Online]. Available: http://www.portico.org/digital-preservation/services/preservation-approach/preservation-step-by-step#step1. [Accessed 11 06 2013].

[11] F. Upward, "Modelling the continuum as paradigm shift in recordkeeping and archiving processes, and beyond - a personal reflection" *Records Management Journal*, vol 10, nr 3, pp. 115-139, 2000

[12] M. Wolski, N. Simons, and J. Richardson, “ECMs and Institutional Repositories. The Case for a Unified Enterprise Approach to Content Management,” presented at the THETA: The Higher Education Technology Agenda 2013, Hobart, Tasmania, 2013.

[13] R. Schmidt, R. King, A. Jackson, C. Wilson, F. Steeg, and P. Melms, “A Framework for Distributed Preservation Workflows,” *International Journal of Digital Curation*, vol. 5, no. 1, pp. 205–217, Jun. 2010.

[14] R. A. P. Freitas and J. C. Ramalho, “Relational databases digital preservation.” University of Porto, 2009.

[15] Q. L. Nguyen and A. Lake, “Content Server System Architecture for Providing Differentiated Levels of Service in a Digital Preservation Cloud,” in *2011 IEEE International Conference on Cloud Computing (CLOUD)*, 2011, pp. 557–564.

[16] P. Tyrväinen, T. Päivärinta, A. Salminen, and J. Iivari, “Characterizing the evolving research on enterprise content management,” *Eur J Inf Syst*, vol. 15, no. 6, pp. 627–634, 2006.

[17] K. R. Grahlmann, R. W. Helms, C. Hilhorst, S. Brinkkemper, and S. van Amerongen, “Reviewing Enterprise Content Management: a functional framework,” *Eur J Inf Syst*, vol. 21, no. 3, pp. 268–286, May 2012.

[18] J. vom Brocke, A. Simons, and A. Cleven, “Towards a business process-oriented approach to enterprise content management: the ECM-blueprinting framework,” *Inf Syst E-Bus Manage*, vol. 9, no. 4, pp. 475–496, Dec. 2011.

[19] A. Haug, “The implementation of enterprise content management systems in SMEs,” *Journal of Enterprise Information Management*, vol. 25, no. 4, pp. 349–372, Jul. 2012.

[20] D. Choy, F. Müller, and R. McVeigh, “Content Management Interoperability Services (CMIS) Version 1.1.” OASIS Content Management Interoperability Services (CMIS) TC, Sep-2011.

[21] S. Waddington, R. Green, and C. Awre, “CLIF: Moving repositories upstream in the content lifecycle,” *Journal of Digital Information*, vol. 13, no. 1, 2012.

[22] J. Korb and S. Strodl, “Digital preservation for enterprise content: a gap-analysis between ECM and OAIS,” presented at the iPres 2010, Wien, 2010.

[23] C. Saul and F. Klett, “Conceptual Framework for the Use of the Service-oriented Architecture-Approach in the Digital Preservation,” presented at the iPres 2008, 2008, pp. 229–234.

[24] P. Wittek and S. Darányi, “Digital Preservation in Grids and Clouds: A Middleware Approach,” *J Grid Computing*, vol. 10, no. 1, pp. 133–149, Mar. 2012.

[25] SHAMAN Project, “Report on demonstration and evaluation activity in the domain of ‘Memory institutions’.”SHAMAN, 15-Sep-2011.

[26] R. Schmidt, A. Lindley, R. King, A. Jackson, C. Wilson, and F. Steeg, “The Planets IF: a framework for integrated access to preservation tools,” in *Proceedings of the 1st International Digital Preservation Interoperability Framework Symposium*, New York, NY, USA, 2010, pp. 10:1–10:8.

[27] D. Tarrant, S. Hitchcock, L. Carr, H. Kulovits, and A. Rauber, “Connecting preservation planning and Plato with digital repository interfaces,” presented at the 7th International Conference on Preservation of Digital Objects (iPRES2010), 2010.

[28] G. B. Laleci, G. Aluc, A. Dogac, A. Sinaci, O. Kilic, and F. Tuncer, “A semantic backend for content management systems,” *Knowledge-Based Systems*, vol. 23, no. 8, pp. 832–843, Dec. 2010.

[29] S. Mao, J. W. Kim, and G. R. Thoma, “A dynamic feature generation system for automated metadata extraction in preservation of digital materials,” in *First International Workshop on Document Image Analysis for Libraries, 2004. Proceedings*, 2004, pp. 225–232.

[30] Consultative Committee for Space Data Systems, *Recommendation for Space Data System Practices. Recommended Practice for a Producer-Archive Interface Methodology Abstract Standard*. Washington DC / St. Hubert, Canada: CCSDS Secretariat / Magenta Book, 2004.

[31] I. Andersson, L. Lindbäck, G. Lindqvist, J. Nilsson, and M. Runardotter, “Web Archiving Using the Collaborative Archiving Services Testbed,” in *e-2011 Conference Proceedings*, 2011.

[32] N. del Pozo, D. Elford, and D. Pearson, “Prometheus: managing the Ingest of Media Carriers,” *National Library of Australia Staff Papers*, Apr. 2009.

[33] A. Farquhar and H. Hockx-Yu, “Planets: integrated services for digital preservation,” *Serials: The Journal for the Serials Community*, vol. 21, no. 2, pp. 140–145, Jan. 2008.