Thèse présentée à

Université Lumière Lyon 2

Ecole Doctorale Informatique et Mathématiques (ED InfoMaths, EDA 512)

Laboratoire Décision et Information pour les Systèmes de Production (DISP)

Par

TENG, Fei

Pour l'obtention du grade de

Docteur en **Informatique**

Méthodologie et Architecture de Préservation des Connaissances à Long Terme

- Approche de préservation dynamique et architecture multi-couches

Methodology and Architecture for Products Long Term Knowledge Preservation

- A dynamic preservation approach and a multi-layer architecture

---

Directeur de thèse: **Abdelaziz BOURAS**

Co-encadrant: **Néjib MOALLA**

Devant un jury composé de :

M. Alain BERNARD	Ecole Centrale de Nantes	Rapporteur
M. András GÁBOR	Corvinus University of Budapest, Hungary	Rapporteur
M. Bruno BACHIMONT	Université de Technologie de Compiègne	Examinateur
M. Xavier BOUCHER	Ecole des Mines de Saint-Etienne	Examinateur
M. Abdelaziz BOURAS	Université Lumière Lyon 2	Directeur de thèse
M. Néjib MOALLA	Université Lumière Lyon 2	Co-encadrant

Dédicace

In memory of XU Jiucheng (1983-2012),

classmate, colleague, roommate,

dream chaser,

and best friend

that I should have got to know since a damn long time ago

À la mémoire de XU Jiucheng (1983-2012),

camarade de classe, collègue, colocataire,

chasseur de rêve,

et le meilleur ami

que j'aurais appris à connaître depuis un moment sacrément longtemps

Remerciements

Ces travaux de thèse ont été réalisés conjointement au sein de l'équipe l'Université Lumière Lyon 2, du laboratoire Décision et Information pour les Systèmes de Production (DISP).

Je remercie Abdelaziz Bouras, Professeurs à l'Université Lumière Lyon2, et Néjib Moalla, Maître de conférences à l'Université Lumière Lyon2, d'avoir accepté de codiriger cette thèse.

Tout deux ont su me faire partager leur enthousiasme et leurs connaissances. Ainsi que le soutien, les conseils et le temps qu'ils m'ont consacrés m'ont permis de mener à bien ce travail.

Je remercie Madame et Messieurs les membres du jury qui ont supervisés mon travail et participés à ma soutenance.

Je remercie tout particulièrement Alain Bernard et András Gábor, Professeurs à l'Ecole Centrale de Nantes et Corvinus University of Budapest (Hongrie), d'avoir accepté de rapporter cette thèse et du temps consacré à sa lecture.

J'adresse également mes remerciements à Bruno Bachimont et Xavier Boucher, Professeurs à l'Université de Technologie de Compiègne et l'Ecole des Mines de Saint-Etienne, pour m'avoir fait l'honneur de participer au jury de thèse en cotutelle.

Je tiens à remercier aussi les personnes que j'ai côtoyées à l'IUT Lumière et dans le laboratoire DISP, pour leur soutien, aides et conseils. Un grand merci enfin à ma famille pour leurs encouragements, leur soutien et les sacrifices qu'ils ont du faire pour moi tout le long de cette thèse. Sans eux, ce travail n'aurait jamais pu voir le jour.

Résumé

La préservation des informations numériques est l'un des objectifs de la création et la gestion d'un système d'information dans une entreprise. Cette préservation des informations et des connaissances devient cruciale en terme de sécurité et au niveau des services liés aux produits. Elle s'appuiera principalement sur des fondements scientifiques issus des travaux de recherche autour de la représentation et du partage des connaissances. Notre problématique de recherche concerne plus spécifiquement la Préservation des Information et Connaissances à Long Terme (PCLT ou LTKR en anglais pour Long Term Knowledge Retention) en garantissant la traçabilité, la réutilisation et la sécurité de l'information numérique. L'objectif est de proposer une méthodologie et une architecture pour l'archivage des connaissances liées aux cycles de vie des produits.

Nous avons structuré notre travail de recherche en quatre parties génériques. La première partie concerne l'identification des besoins de la préservation à vision préservation à long terme, la deuxième s'intéresse aux aspects Gestion des Connaissances (GC ou KM en anglais pour Knowledge Management), la troisième partie présente la préservation numérique et enfin la quatrième partie se concentre sur une nouvelle proposition. Dans le contexte des travaux de préservation numériques, nous avons identifié les aspects qui affectent la préservation préoccupations, d'après les recherches de préservation à long terme précédente.

Notre recherche est liée à la préservation numérique, ce qui dans notre travail développé, la connaissance, qui est destiné à être conservé, est sous forme numérique. Un centre de notre recherche est faite sur des méthodologies et des technologies de préservation des connaissances en fonction des perspectives de long terme, à commencer par un état de l'art des approches et des exigences de la préservation des connaissances et la long terme. Préservation des connaissances est dans le contexte de gestion des connaissances. La gestion des connaissances ne permet pas toujours l'accent ou de préciser les détails du processus et l'utilisation des technologies de référentiel de connaissances. Afin d'établir les architectures et systèmes pour la vision à long terme, la gestion des connaissances et les technologies de conservation numérique ont été étudiées et analysés. Par conséquent, nous organisons notre travail en cinq chapitres, en ce qui concerne les quatre parties que nous avons discuté avant. La démarche de recherche de notre travail et les relations entre les chapitres de la thèse est illustré à la figure ci-dessous :

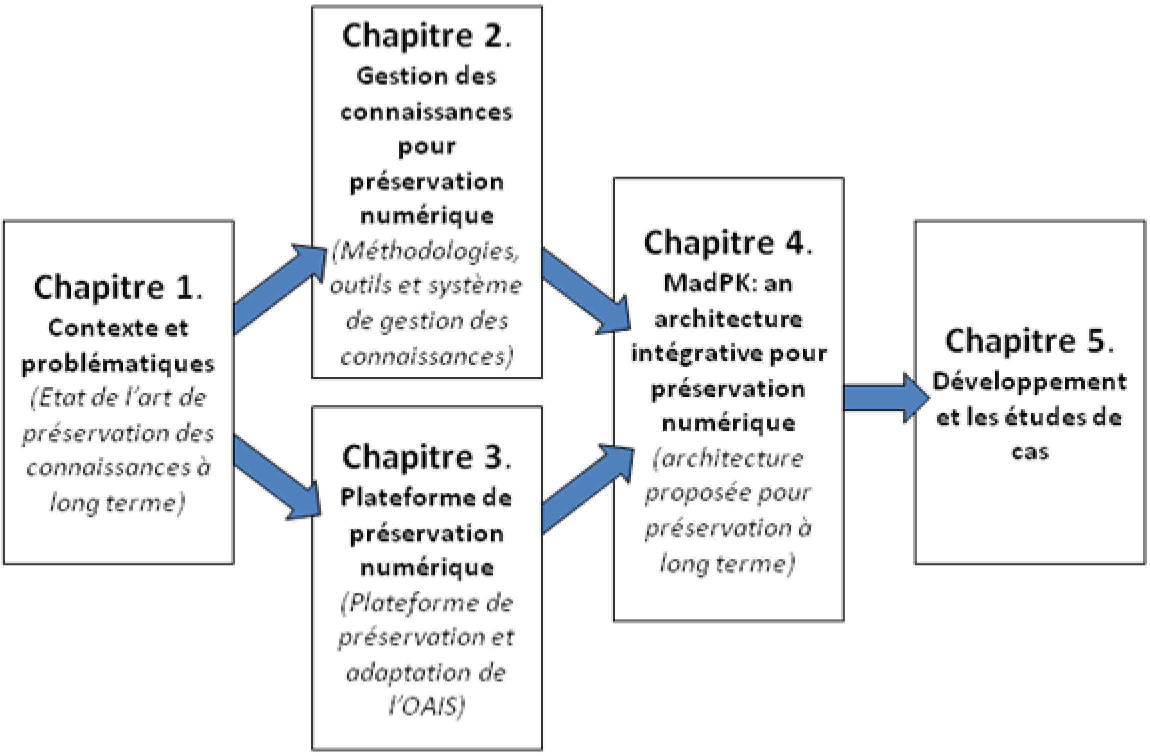

Figure R.1: Présentation des chapitres de la thèse

Chapitre 1. Contexte et problématiques

Dans le contexte des travaux de l'Europe depuis 2005 et en prenant en considération les projets existant, dont les objectifs sont à préservation numérique à long terme, nous avons synthétisé les besoins communes de la préservation des connaissances à long terme et trouver des "besoins manquantes" sur les recherches existantes. Notre objectif du travail de recherche est de proposer une méthodologie et de l'architecture, afin de surmonter le fossé entre les besoins, que nous avons découvert, et les approches existantes. Nous avons identifié les aspects qui préservation numérique préoccupations, d'après les recherches de projets préservation numérique précédente et de conférences.

- Certaines exigences de préservation des connaissances à long terme sont soutenues par les technologies et outils existants et les fonctionnalités. Cependant, il ya encore des écarts entre les exigences et les technologies et outils existants ;
- Nous avons constaté que les exigences de fonctionnalités existantes ou technologies sont l'interopérabilité, les caractéristiques dynamiques, l'agilité de la portée à long terme et des normes pour le système d'archivage.

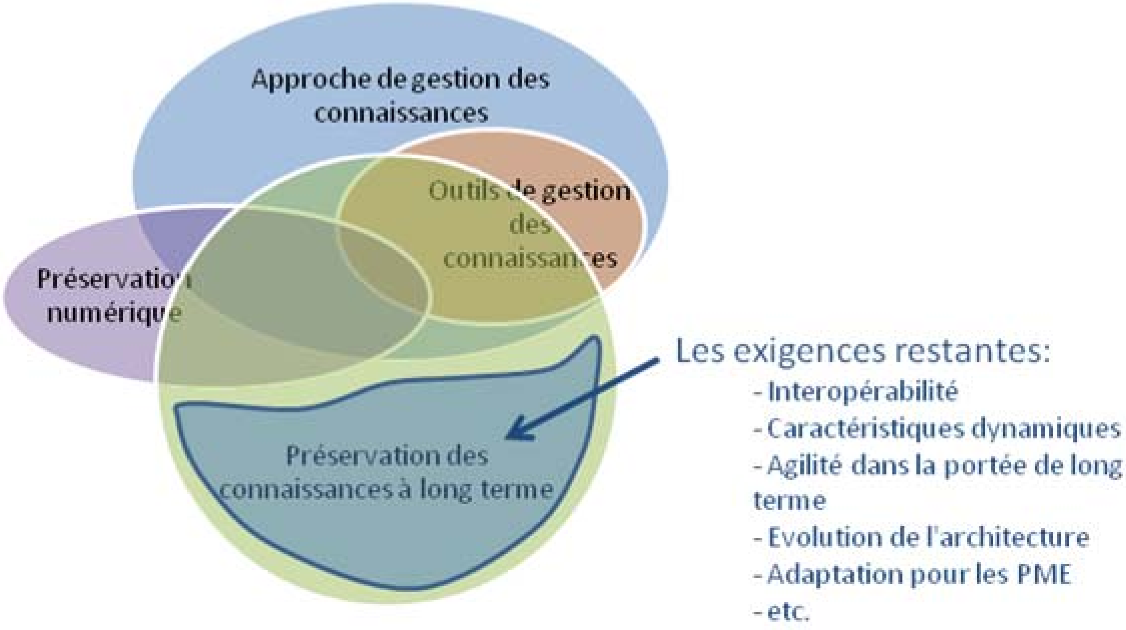

Figure R.2: Les exigences restantes pour préservation des connaissances à long terme

Dans cette thèse, notre proposition est de développer des modèles dynamiques de la conservation numérique. La condition idéale de conservation dynamique, c'est que le processus de préservation, la plateforme de conservation, et le contenu seront tous préservés être dynamique et toujours à jour. Nos missions de cette thèse sont :

- Tout d'abord, nous devons développer une approche de gestion des connaissances extensible pour la conceptualisation des connaissances et la création de connaissances;
- Deuxièmement, nous devons établir une plateforme de la conservation numérique dans la perspective de la préservation des connaissances à long terme. Nous allons adapter le modèle de référence OAIS (Open Archival Information System) en bénéficiant de ses fonctions pour la préservation des connaissances à long terme, grâce à des recherches antérieures de conservation numérique.
- Ensuite, nous devons construire une architecture pour connecter la gestion des connaissances et les approches de la préservation numérique de façon dynamique. Cette architecture sera d'assurer l'interopérabilité des processus d'affaires et des technologies des deux approches.
- Certaines technologies et applications de gestion (e.g. les processus d'affaires, les applications d'architecture orientée services, etc.) ont fourni des fonctionnalités pour les systèmes dynamiques et en développement interopérabilités élevés. Ainsi, dans nos

propositions, nous allons utiliser ces technologies et applications existants pour atteindre les caractéristiques dynamiques de notre conception.

Dans ce chapitre, nous avons présenté la recherche préliminaire et l'analyse du contexte de la recherche à long terme la préservation des connaissances. Technologies et fonctionnalités fournis par les plates-formes existantes de préservation numérique ont été étudiées. Toutes ces fonctionnalités ont été adaptées à partir de plateformes existantes. Néanmoins, il ya quelques références préservation à long terme et des perspectives (i.e. "les aspects manquants"), qui ne sont pas tout à fait atteint par des fonctionnalités existantes. Grâce à la synthèse de l'état de l'art dans le contexte de la préservation à long terme (i.e. gestion des connaissances, la conservation numérique, la préservation à long terme), nous avons identifié les aspects manquants pour la conception d'une architecture à long terme la préservation des connaissances.

Les états dynamiques théorie de préservation que si toutes les connaissances conservées dans la plateforme de la conservation numérique est toujours à jour en fonction des situations stratégiques actuelles et technique des sources de connaissances, la préservation sera toujours récupérable et réutilisable dans le long terme. Afin de rendre notre conservation numérique proche de la situation idéale de conservation dynamique, nous avons besoin de développer une architecture pour diriger la prise de décision stratégique, planification et les processus fonctionnels.

Chapitre 2. Gestion des connaissances pour préservation numérique

Les fonctionnalités existantes de gestion des connaissances sont identifiées par l’analyse des méthodologies et des outils de gestion des connaissances :

- Nous avons synthétisé les 9 approches générales de l’exécution gestion des connaissances, par l’analyse de 16 méthodologies de gestion des connaissances, qui sont utilisées dans différent domaines
- Nous avons classifié les outils de gestion des connaissances en 4 classes métiers et 16 sous-classes fonctionnelles, par l’analyse de 78 outils de gestion des connaissances, qui sont développés par 33 entreprises de différents logiciels
- En fait, les 4 classes métiers sont définies par le cycle de gestion des connaissances, qui est au cœur de gestion des connaissances. Ainsi, la taxonomie fournit également un lien. Ce lien est des aspects fonctionnels de gestion des connaissances aux les outils de gestion des connaissances

Afin de construire connexion entre gestion des connaissances et de préservation numérique, nous proposons une méthodologie gestion des connaissances étendue de la méthodologie CommonKADS. Les approches de la méthodologie sont divisées en 4 modules séquentiels:

- Contexte: Il s'agit de la phase d'identification des connaissances. Dans cette phase, nous devons saisir la structure organisationnelle et la stratégie de la conservation numérique dans l'entreprise. Et dans le même temps, nous identifions les processus d'affaires et connaissance des produits ;
- Concept: Il s'agit de la phase d'intégration des connaissances et transfert des connaissances. Dans cette phase, les connaissances sont manipulées selon les modèles de connaissances, que nous définissons. Les modèles des connaissances sont définis dans la perspective de préservation à long terme ;
- Design: Dans cette phase, nous synthétisons les résultats des phases précédentes, et nous concevons l'architecture de préservation à long terme du point de vue de métier. A l'intérieur de cette architecture, nous devons identifier et concevoir les fonctionnalités et les sous-fonctionnalités dans chaque processus métiers;
- Implémentation: Dans cette phase, nous mettons en œuvre les modèles précédents.

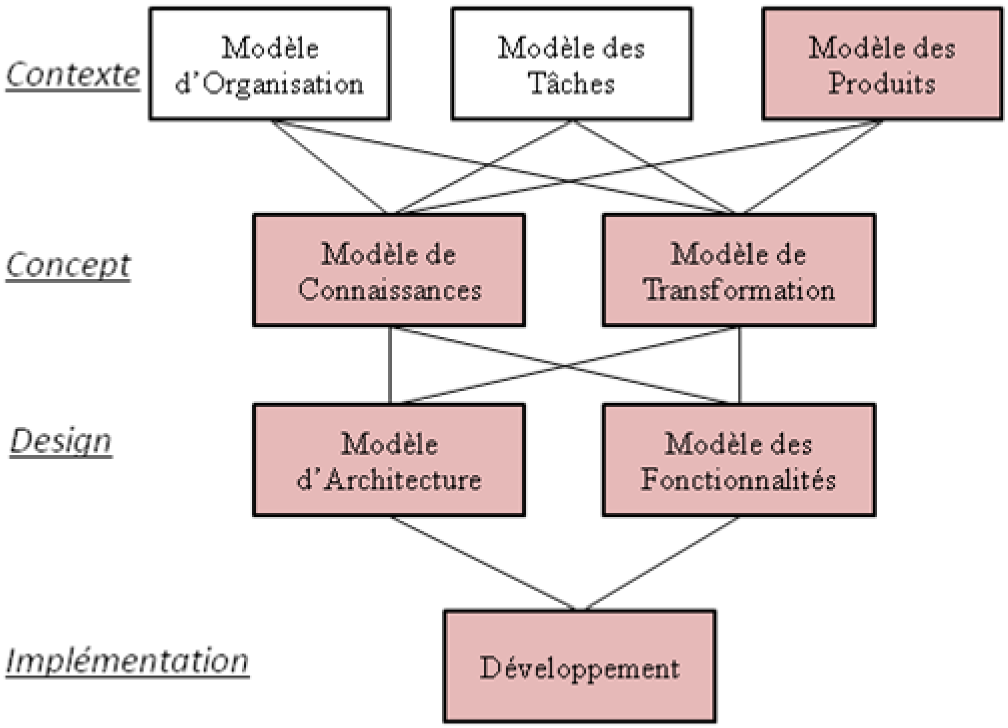

Figure R.3: Méthodologie proposée de gestion des connaissance pour préservation à long terme

Chapitre 3. Plateforme de préservation numérique

Nous avons étudié les plates-formes libres de la préservation numérique, et grâce à l'analyse fonctionnelle sur les plates-formes existantes :

- Nous avons découvert qu'il y a plusieurs solutions/logiciels de préservation (i.e. DSpace, Fedora repository, EPrints, etc) qui supportent de la préservation à long terme
- Nous avons étudié et testé les plates-formes, et de cette façon nous avons constaté que le flux de travail d'archivage est toujours processus statique dans ces plates-formes. Il s'agit d'une barrière de mise en place d'un système dynamique de préservation
- Nous avons étendu le modèle de référence OAIS (Open Archival Information System) pour installer un système d'archivage.

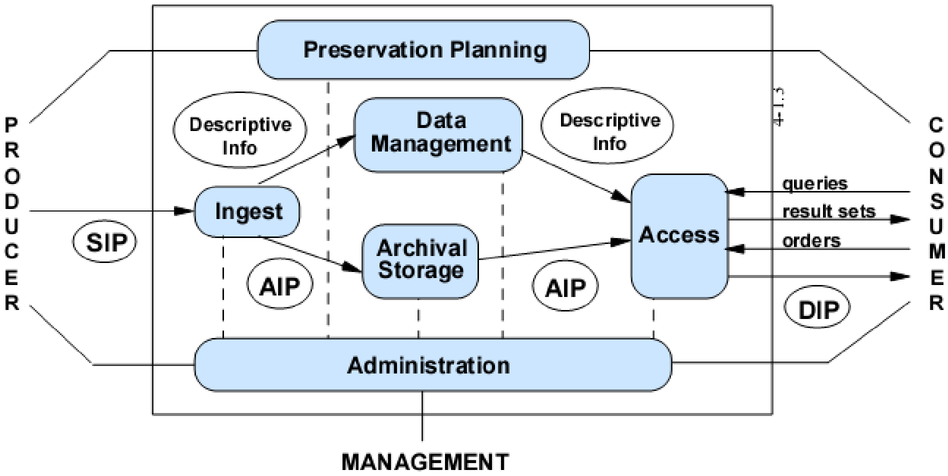

Figure R.4: Entités fonctionnelles d'OAIS [CCSDS 650.0-B-1; ISO 14721:2003]

Selon les résultats de plates-formes existantes du référentiel, nous sommes réorganisés les six grandes entités fonctionnelles de l'OAIS, et nous avons ajouté ou modifié certaines fonctions selon les fonctions ou d'outils existants. Dans notre méthodologie CommonKADS étendue, le modèle de connaissances est la structure de l'information qui est pour l'archivage. Ainsi, les associés modèles de connaissance avec les fonctions ou les sous-fonctions de la plateforme de la conservation numérique. Les fonctions de la plateforme de la conservation numérique sont liées à la préservation numérique à long terme, et nous avons également étendu les fonctions de la norme OAIS:

- Stockage d'archives: « Stockage d'archives » fournit des services et fonctions pour le stockage, la maintenance et la récupération des AIPs (Archival Information Packages).

Nous avons à traiter de la obsolescence des connaissances dans le long terme, donc en fonction de stockage d'archives, nous avons ajouté une sous-fonction appelé « terminer données », ce qui met fin à la préservation de certaines données en fonction de la direction de l'entité d'Administration. Terminer données ne sont pas fournis dans le modèle de référence OAIS originale, mais nous avons besoin de mettre fin à certaines informations obsolètes dans la conservation numérique dans le but de réduire l'utilisation des ressources et de fournir de meilleurs environnements pour l'information qui a toujours eu des valeurs pour les utilisateurs finaux.

- Gestion des données: « gestion des données » pourrait aussi être considérée comme un système de gestion de base de données, qui remplit et maintient l'information descriptive et des données administratives. L'Information de description identifie le dossier d'information préservée. Comme les IPs (Information Packages) sont encapsulés, de l'information descriptive est la seule information disponible pour la récupération des IPs. En outre, la gestion des données génère des rapports d'enregistrer les événements dans la plateforme de la préservation numérique. Puis cette entité filtre les IPs, qui ne sont pas utilisés dans le terme désigné de long, selon les journaux et les rapports. En ce qui concerne la résiliation des données, nous avons également ajouté une sour-fonction appelé « filtrer les données » en être dépassées le modèle OAIS référence d'origine.
- Administration: « administration » fournit des services pour le fonctionnement global de la plateforme de la conservation numérique. Il vérifie les IPs selon le modèle de connaissances prédéfini. Le processus de vérification est nécessaire chaque fois qu'il ya des transferts d'IP. L'administration permet de configurer les politiques d'archivage et les questions techniques de la plateforme de la conservation numérique. Une autre fonction importante de l'administration, c'est qu'il contrôle l'autorisation d'accès aux données et l'actualisation des données. Lorsque la gestion des données essaie d'envoyer des requêtes pour effectuer l'évaluation des connaissances dans une approche de gestion des connaissances, les requêtes d'envoi doit être autorisée par l'entité d'administration.
- Planification de la pérennisation: « planification de la pérennisation » développe la stratégie de préservation de base et le modèle de connaissances pour la plateforme de la conservation numérique. En outre, il attrape les signaux de variation critique de l'approche KM et en avise le processus de mise à jour des connaissances.

Chapitre 4. MadPK: an architecture intégrative pour préservation numérique

Afin de résoudre l'écart entre le système dynamique espéré et le flux d'archivage statique existant, nous proposons une architecture dynamique multi-couches appelé MadPK (Multi-layer Architecture for Dynamic Preservation of Knowledge). Bien que le flux de travail d'archivage soit statique, les différentes fonctionnalités du système d'archivage réaliseront différents services pour soutenir la préservation des connaissances. La plateforme d'archivage pourra réagir rapidement aux changements du système d'information dans l'entreprise.

Les travaux ont été réalisés dans les différentes couches de notre architecture proposée.

- Couche d'entreprise: Nous avons réalisé analyse des processus PLM, en utilisant la méthodologie de l'ingénierie des connaissances (i.e. modèle de produit, modèle de processus, modèle de organisation).
- Couche d'archivage électronique : cette couche est la plateforme d'archivage des connaissances (une instance d'OAIS). Nous avons ajouté quelques fonctionnalités supplémentaires que OAIS sur la plateforme de conservation numérique, basée sur la fonction, nous avons trouvé de la plateforme préservation numériques existants et les exigences de préservation à long terme.
- Couche de médiation : Nous créons les services qui sont correspondants à les fonctionnalités critiques (i.e. identification, acquisition, encapsulation, archivage, transfert, recherche des connaissances) de la gestion des connaissances. Les composants correspondants seront fournis par la plateforme d'archivage et de la technologie extérieure. Nous avons nommer cette couche comme EKB (Enterprise Knowledge Bus).

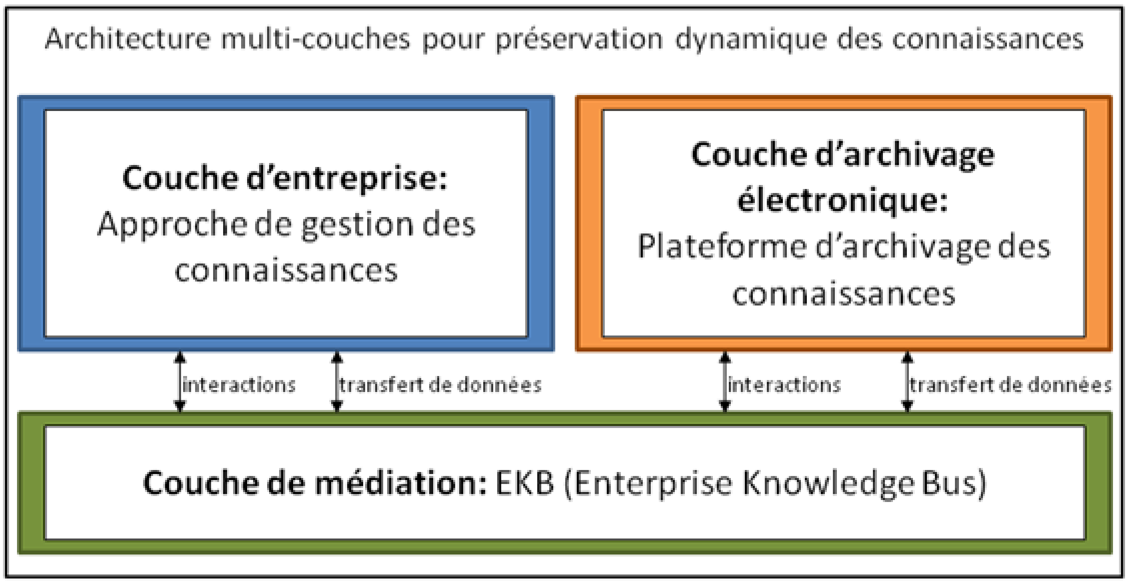

Figure R.5: Architecture de MadPK

Les flèches de deux sens entre l'état des couches qu'il existe des interactions amont et en aval et l'échange de données entre la couche. En général, MadPK fonctionne comme:

- La plateforme de la conservation numérique est stable dans une certaine mesure, et de la stabilité est une des clés à long terme et la précision des données de sécurité. C'est la raison pour laquelle nous séparer le référentiel de connaissances de la production de connaissances (couche d'entreprise, des approches de gestion des connaissances).
- La raison pour laquelle la plateforme de la conservation numérique peut être stable dans une certaine mesure, c'est que son entrée et de sortie sont des connaissances toujours emballé dans le modèle formel, même si les connaissances et les données de la source des connaissances peut être différent dans les modèles et dans des formats. Nous développons toutes les activités de transfert de modèle à l'intérieur de la couche de médiation (EKB).
- Dans la couche d'entreprise, les changements des systèmes d'information conduire à la modification des approches de gestion des connaissances, et par conséquent le changement de modèle de connaissance de sortie, qui est l'entrée de EKB. Le modèle de connaissance de sortie de l'approche de gestion des connaissances peut être différent en fonction de la complexité des systèmes d'information dans l'entreprise. Cependant, chaque fois une partie ou l'ensemble des modèles de ces connaissances changement, des règles de connaissance modèle de conversion sont mis à niveau correspondante dans la couche de médiation (EKB).

Chapitre 5. Développement et les études de cas

Nous mettons en œuvre des modèles en utilisant Oracle SOA et BPM suites 11g. A la fin de cette thèse, nous utilisons plusieurs scénarios pour montrer les processus globaux de la conservation à long terme numérique, et en même temps, prouver l'efficacité de notre conception et des propositions.

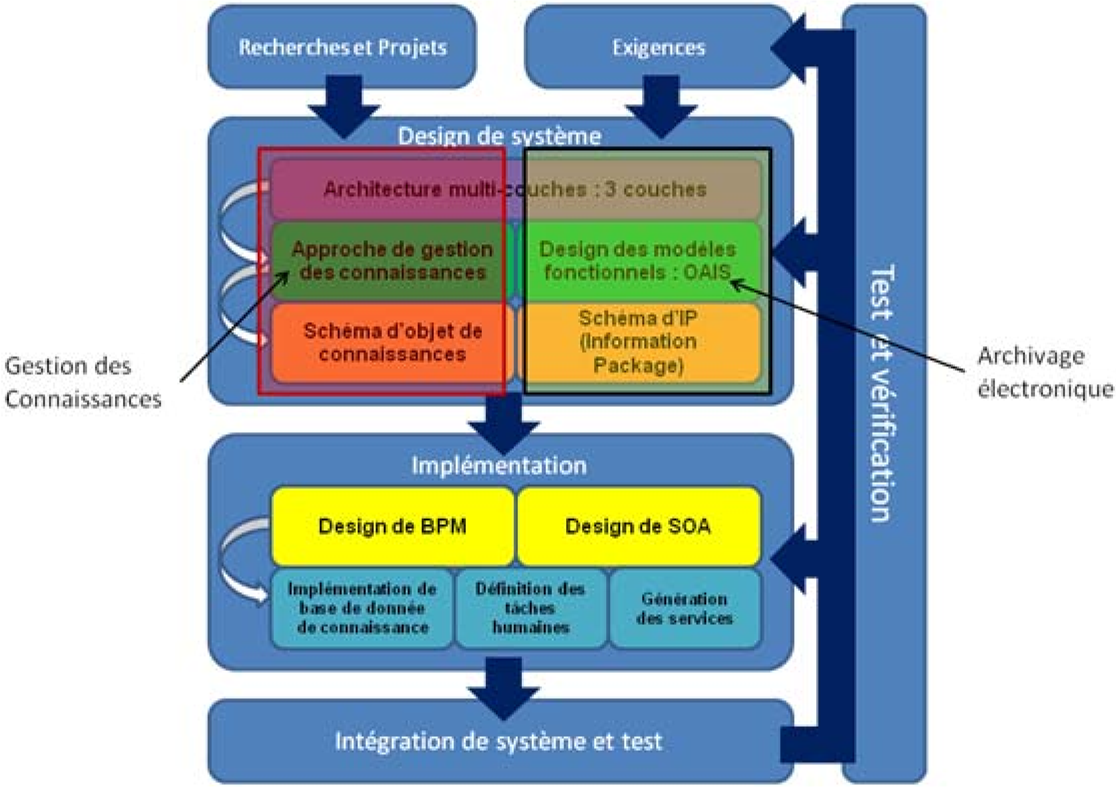

Figure R.6: Processus de conception basé sur un modèle et des modèles désignés pour MadPK

Nous utilisons la modélisation des processus d'affaires dans le but de valider les modèles prédéfinis (qui est, les modèles structurels, fonctionnels et de données) et la mise en œuvre de gestion des connaissances et des approches d'archivage électronique. Dans les modèles des processus d'affaires, les fonctions qui incluent des fonctionnalités dynamiques sont définis et déployés, grâce au principe SOA et des composants, qui collaborent avec l'approche BPM

Mots-clés : préservation à long terme ; gestion des connaissances ; archivage électronique ; OAIS ; MadPK

Abstract

Preserving digital information is one of the objectives of the creation and management of information systems in a company. The long term preservation of information and knowledge becomes crucial in terms of safety and availability. Therefore, our research targets the Long Term Knowledge Retention (LTKR) and preservation in terms of traceability, reusability and security of digital information; and aims at proposing a methodology and an architecture for this purpose.

Regarding the objective of this thesis, we have structured our research into four generic parts:

Firstly, through analysis of existing works and projects, whose objective is long-term digital preservation, we have synthesized long term preservation common requirements and found out the gaps between existing requirements and the new requirements we have identified.

Secondly, existing methodologies and tools for Knowledge Management (KM) are identified. We have extended the CommonKADS methodology in order to build connections between KM and digital preservation. The knowledge objects we have produced in a KM approach are thus better suited as inputs for the digital preservation platform.

Then, we have studied the functional features of the existing digital preservation systems, and some main features to support a long term preservation approach. We have extended the Open Archival Information System (OAIS) reference model to establish a preservation platform, by adding specific features in order to fulfill the remaining requirements in the long term preservation area.

Finally, we have proposed a dynamic preservation method and an architecture for long term preservation, adapting Service Oriented Application (SOA) and Business Process Management (BPM) concepts. The proposed architecture provides dynamic features and enables us to have better interoperability between the KM approach and digital preservation approach/platform. .

Key words: long term preservation ; knowledge management ; digital preservation ; OAIS ; MadPK

TABLE OF CONTENTS

List of Figures

List of Tables

Abbreviation

AIP	Archival Information Package
API	Application Programming Interface
BOM	Bill Of Material
BPM	Business Process Management
BPMN	Business Process Model and Notation
CRC	Cyclic Redundancy Check
DBMS	Database Management System
DIP	Dissemination Information Package
EKB	Enterprise Knowledge Bus
ESB	Enterprise Service Bus
HTTP	Hypertext Transfer Protocol
IP	Information Package
IPR	Intellectual Property Right
ISO	International Organization for Standardization
JSP	JavaServer Pages
KE	Knowledge Engineering
KM	Knowledge Management
KMS	Knowledge Management System
MadPK	Multi-layer Architecture for Dynamic Preservation of Knowledge
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OAIS	Open Archival Information System
OM	Organization Model
PDI	Preservation Description Information

PLM	Product Lifecycle Management
PM	Product Model
PPO	Product-Process-Organization
SIP	Submission Information Package
SOA	Service-Oriented Architecture
SOAP	Simple Object Access Protocol
SRB	Storage Resource Broker
TM	Task Model
UML	Unified Modeling Language
XML	Extensible Markup Language

Introduction

Digital information has played important roles in enterprises for the last decades. From the first day when digital information was produced by human beings, the digital preservation has been “forced” into a critical position. The aircraft A319 of Airbus was launched in June 1993, and delivered in April 1996. After more than 15 years, A319 is still used at airline companies. However, in the meanwhile, technologies have developed rapidly. In the next 10 years time, the type of hardware or software that will be developed and popularized is unknown right now, whereas we do know products such as A380 or even A319 will still be in-service. That is why “in industries there exist extremely high requirements to the processes for administration, archiving and reuse of product defining data” [LOTAR 03]. In spite of the application of traditional document engineering methods, the long term digital preservation issues have been mostly neglected in traditional standard information lifecycle implementations. The challenges of long term digital preservation include legal, policy, organizational, managerial, educational, and technical aspects. The long term retention of digital information is a work in progress and there are various issues that need to be addressed. Long term digital preservation aims at creating technological solutions and innovative methods for keeping digital resources available and useable over time. Our research on long term digital preservation consequently focuses on long term preservation of knowledge, or in other words, long term knowledge retention (LTKR).

Our research is related to digital preservation, thus in our developed work, the knowledge, which is intended to be preserved, is in digital form. A focus of our research is made on “knowledge retention” methodologies and technologies according to “long term” perspectives, starting with a state of the art of “knowledge retention” and “long term” approaches and requirements. “Knowledge retention” is in the context of “knowledge management”. Knowledge management does not always focus or specify the detailed process and utilization of knowledge repository technologies. In order to establish architectures and systems for the long term vision, both knowledge management and digital preservation technologies have been studied and analyzed. Therefore, we have organized our work into five chapters:

Chapter 1 discusses the context and research problems of the long term knowledge preservation research. In this chapter, we review the state of the art of relative methodologies and technologies, which will benefit long term preservation. The long term changes exist in companies and industries, and the changes imply constraints on data and knowledge preservation. The constraints produce challenges when launching knowledge preservation projects in a production context. In this chapter, we identify the long term changes in industries, and then we track the challenges for long term knowledge preservation. At the same time, we discover the various projects regarding digital preservation in Europe. According to the results and contributions of these completed or ongoing projects, we identify the improvable aspects in long term knowledge preservation, and we state the position and research goals of our research work among all these research works.

Chapter 2 proposes a KM approach as a part of the long term preservation. We have studied multiple KM methodologies (i.e. MASK, CommonKADS, etc.), which would contribute to our research work, so as to propose an appropriate methodological approach. According to the preliminary research work on knowledge management and digital preservation, we extend the CommonKADS [Orsvarn et al. 95] methodology as we have stated in our proposal on the KM part. At the same time, we have done analysis on software tools, which support knowledge management, in order to identify the existing functionalities and technologies for establishing a knowledge management system.

Chapter 3 proposes digital a preservation approach and a platform for long term preservation. We design models by adapting Open Archival Information System [CCSDS 650.0-B-1; ISO 14721:2003] reference model and web service concepts. We experiment on open source platforms to acquire functionalities from existing technologies in order to support our research work. Based on the long term preservation requirements identified in **Chapter 1**, we establish an OAIS based digital preservation platform.

Chapter 4 constructs a multi-layer architecture to connect the KM approach to the digital preservation platform, which is named MadPK (Multi-layer Architecture for Dynamic Preservation of Knowledge). The architecture enhances the interoperability of business processes of KM and digital preservation. The multi-layer architecture also meets some long term preservation requirements, which we have discovered.

Chapter 5 consists in the development of the previous designs, as well as the integration of functionalities and models. We implement the models by using Oracle SOA and BPM suites 11g. At the end of this thesis, we use several scenarios to show the overall processes of long term digital preservation, and at the same time, prove the efficiency of our design and proposals.

The research approach of our work and the relations of the chapters in the thesis is shown in Figure 0.1.

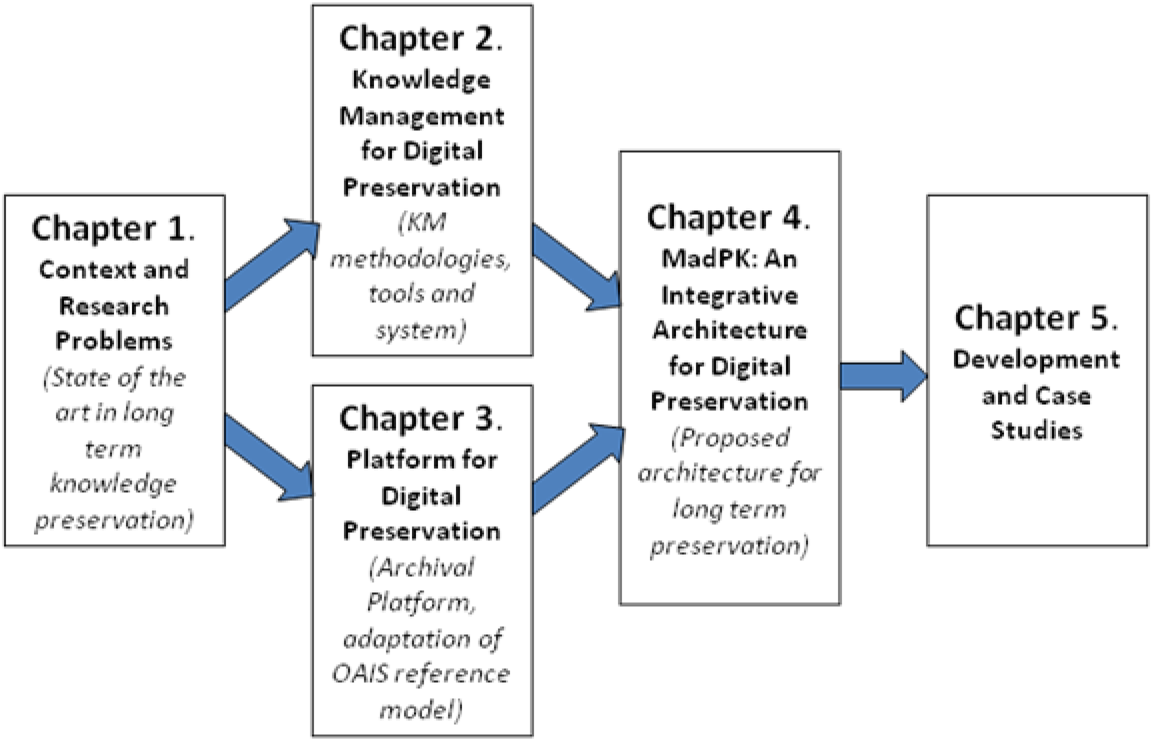

Figure 0.1. Presentation of the chapters in this thesis

Chapter 1. Context and Research Problems

1.1. Introduction

In the context of product lifecycle management (PLM), the rapid increase of product data and more complex data type requires an improved approach for digital information preservation. In this chapter, we discuss the challenges of long term knowledge retention in a product lifecycle, and we have analyzed the existing researches and projects, which relate to digital preservation. From the existing researches and projects, we identify the requirements of long term digital preservation and clarify the objectives in our research work.

1.2. Needs for Product Knowledge Retention

Ancient manuscripts based on solid materials, or even paper-based information could last longer (thousands or hundreds of years). In contemporary business environments, data origin is electronic, and depends on elaborate hardware and software systems with defined data and information models. However, the technologies are upgraded regularly. Preservation of digital objects is a multidimensional conceptual and technological challenge [Blažič et al. 07]. Digital information also plays a significant role in a production context, and the digital preservation is one main objective of construction and maintenance of a production information system. However, regarding the product lifecycle management (PLM) aspect, the product lifecycles are often far longer than the expected lifetime of a manufacturing software application used to interpret the data. In spite of the application of traditional document engineering methods, the long term digital preservation of product lifecycle data issues have been mostly neglected.

A product's lifecycle describes various phases from the design until the dissemination of the product. Generally, we consider that the whole lifecycle consists in three generic phases:

- Beginning-Of-Life (BOL): the design and manufacturing processes of products;
- Middle-Of-Life (MOL): the delivery, maintenance and service processes of products;

- End-Of-Life (EOL): the obsolescing, recycling and business process reengineering processes of products.

We notice from the description above that the products go through different phases, and thus are manipulated in different information systems in each phase. The variety of information systems and the rapid development in information technologies leads to the complexity of digital product data and data correlations. In order to enhance the traceability and reusability of the preserved product data, comprehensive metadata are required. From the name "metadata", we could notice that it refers to "data about data". Because of the rapid increase of digital data, in the modern repository and library field, metadata describe the digital data by using specific metadata standards or schemas, in order to help users understand or retrieve digital data. Our long term digital preservation aims to archive digital data as well as metadata. If metadata provide comprehensive information on descriptions and contexts about the digital data, we consider the preservation of data and metadata as the preservation of knowledge. Knowledge has a larger scope than data, and is considered as data plus more information (metadata), which is related to the context. Knowledge will provide more value to people not only on utilizing the digital facts (i.e. digital data), but also on learning and reproducing more data and knowledge. In our research work, when mentioning long term knowledge retention, we particularly refer to the preservation of digital objects and their comprehensive metadata over time. Regarding digital product data, it may require a specific knowledge management methodology for handling the knowledge.

In the following sections, we introduce the long term changes and challenges for digital preservation, and these descriptions explain why we need to develop "knowledge" retention approach rather than "data" retention.

1.2.1. Long Term Changes

In the environment of industrial products many objects can change in the long term and hence demand the setting up of an archive which will guarantee the digitally encoded information [Giaretta 09]:

- Hardware and software changes: as we have discussed earlier, hardware and software will be changed, upgraded or replaced over time. The use of many digital objects relies on specific software and hardware. Experience shows that while it may be possible to keep hardware and software available for some time after it has become obsolete, it is not very convenient in the long run. However, there are several projects and proposals which aim to emulate hardware systems and hence run software systems.

- Knowledge Base changes: a knowledge base is a special database that is used in knowledge management. The knowledge base of organization and individuals changes over time. Thus some archives, which are considered as perfectly understandable now, may not be recognized or reused in the future.
- Environment changes: these include changes to licenses or copyrights and changes of organizations, affecting the usability of digital objects. External information, ranging from name resolvers such as the DNS (Domain Name System) to DTDs (Document Type Definition) and Schema, which are essential to the use and understandability, may also become unavailable.
- Archival deployment status changes: the deployment of repositories requires continuous maintenance. And if supporting key elements are missing over time (e.g. cease of funding, lost of chain of evidence, etc.), the preserved knowledge will not be retrieved or be reused.

Of course, among the changes we have illustrated above, there are possibilities that accidents or incidents may prevent long term preservation. Although we cannot always avoid accidents or incidents, we will always add principles to reduce the chance of their occurrence. Beside these kinds of changes, there are also a few unavoidable changes linked to long term preservation. These certain changes are the threats which would be resolved.

1.2.2. Challenges of Product Knowledge Long Term Digital Preservation

The companies are concerned by the knowledge they are archiving as well as the data they are producing. For enterprises and industries, the big challenges for archiving are: 1. the variety of engineering data types and 2. the complexity of the relationships between the information units comprising these data types. The reality is that an archive must capture all the data required to completely define the product. And in some instances, information of processes also has to be captured. For each type of data, or each type of relations between the information units, specific methods of archiving have to be used. As there is an evolution in the data formats, types and volumes which always increase, data and their relations during the engineering process are more and more complex.

Moreover, for the companies and industries, production systems can be considered as dynamic systems within the products' lifecycles. Here a "dynamic system" means that during the whole Product Lifecycle, the system process and product data are both assumed to be available, even if some of the data have been produced decades before, by using technologies, hardware or software that have become obsolete. Therefore, the extensibility and reusability of the digital models and systems have been required by industries for a long while. Technologies, hardware and software may be obsolete, yet digital models and systems should be able to be extended and reused after a longer time. However, in real life, there is always a dilemma between the innovation of the digital technologies and the digital preservation. New technologies may always surprise enterprises and industries thanks to unpredictable fabulous features, which are different from the original digital models and systems.

Furthermore, the accuracy of digital data after a long while and changes in archival technologies and media is one other challenge [Lubell et al. 09]. Extending the digital models and systems will guarantee the availability of the preserved data for a long time. However, information may lose its accuracy during the long archiving and extension process. The lack of accuracy of digital information may imply companies and industries to social and economic consequences. Beside the data themselves, the metadata to describe or to locate the data are required to be semantically rich, for technical as well as organizational purpose. The evolutions of the organizations and people represent regular improvements for enterprises, thus different people will be in the same position over long time. However, during the movements of human resources, some digital information may be misunderstood or unrecognized overtime, because of accidents or incidents. Therefore, semantically rich metadata are required for long term digital preservation.

For synthesizing the descriptions above, we illustrate the following challenges of long term preservation perspectives, and the aspects that we should focus on in our research work:

- Knowledge identification and conceptualization
- Complexity of digital formats, types and relationships
- Extensibility and reusability of digital preservation models
- Knowledge lifecycle assessment
- Traceability and reusability of digital information

- Interoperability of digital preservation platforms

1.2.3. Synthesis of Long Term Preservation Requirements

We have to figure out what functions or features are really needed in Long Term Preservation, according to the previous discussions. In other words, we have to identify the long term preservation requirements. We have identified some aspects that concern long term preservation, from the results of previous LTKR projects and research works [NIST 07] to which our group has participated. These aspects, as well as the OAIS reference model together, are the comprehensive requirements for any long term preservation project. The referenced long term preservation requirements are shown in Table 1.1. The requirements are divided into 14 general aspects:

- Administration: management concerns of long term preservation project and long term preservation organizations.
- Policy Issues: planning and policies, which concern digital preservation.
- Legal IPR: legal and intellectual property right (IPR) on long term preservation methodologies, technologies, and standards.
- Standards and Architectures: from this aspect, we notice that researchers have already considered OAIS as a fundamental helpful standard and reference model for long term preservation project. And this is also the reason why we adapt OAIS in our research work.
- Digital Formats: the issues of the digital object formats, as well as the technologies of dealing with digital formats (e.g. migration, merging, transferring, etc.)
- Data Relationship: rational structure of preserved data, in the perspective of traceability and reusability of the data in repositories.
- Storage: distributed archival systems have been recommended in storage process, which would be more accessible in the way of distributed archival systems with the help of web services. However, for implementing a long term digital preservation platform, the distributed systems are not always required.
- Design information: this aspect is particularly crucial when working with information systems for production. The design and manufacturing history and descriptive

information should be captured and stored for the purpose of redesigning or designing a new product in the future.

- Granularity: our research work based on information systems through product lifecycles aimed at the production industry leads us to think that the capturing of information should be performed during PLM stages.
- Searching metadata: this aspect concerns the retrieving issues. Comprehensive metadata is one key for long term preservation. And at the same time, methodologies and technologies such as ontologies and data mining will offer many opportunities in metadata searching.
- Validation and Fidelity: this aspect targets the stability and accuracy of preserved knowledge, information and data.
- End Users: the end users of a long term preservation architecture or software tools are different in their professional needs, thus in the design of standards for long term preservation it is worth adding end user perspectives and options, in order to enhance the end users' interests levels of prototypes in long term preservation.
- Case study: there should be case studies or testing procedures in research projects.
- Vendor Support: the features provided by a methodology, an architecture or a standard should display enough reasons for increasing the incentives of software vendors. In other words, the methodology and architecture for long term preservation prove more convincing when applied than in theory.

No.	**Aspect**	**Sub-No.**	**Sub-aspect**
1	Administration	1.1	Archival policies
		1.2	Disaster & risk management and security
		1.3	Data Reliability and persistence
		1.4	Funding level and human resources
		1.5	Education and training
		1.6	Project management
		1.7	Organizational issues
		1.8	Access right
		1.9	Configuration management
		1.10	Revised/deleted archives
2	Policy Issues	2.1	Integration of preservation with business workflow
		2.2	Cultural issues
		2.3	Archival management from organization structure view
		2.4	Sustainability for preservation
		2.5	Digital archival systems for designated user community
3	Legal IPR	3.1	Open standards for archival systems
		3.2	IPR issues
		3.3	Legal issues

No.	Aspect	Sub-No.	Sub-aspect
4	Standards and Architectures	4.1	Extensions to OAIS for multimedia documents and environments
		4.2	AIP interoperability
		4.3	Classification of AIP
		4.4	Reference architecture based on OAIS
		4.5	Standards landscape for archival
		4.6	Identification and use of established standards and best practices
5	Digital Formats	5.1	Preservation of format registry
		5.2	Merging and differentation standards
		5.3	Detection of formats obsolescence
		5.4	Format migration
		5.5	Flexible handle of formats
		5.6	Workflows
		5.7	Software version
6	Data Relationship	6.1	Traceability
		6.2	Information models for archival systems
7	Storage	7.1	Storage media
		7.2	Software/hardware integration
		7.3	Distributed archival systems
8	Design information	8.1	Design history capture
		8.2	Standards for extraction of design and manufacturing rules from knowledge based support tools
9	Granularity	9.1	Information capture through PLM stages
		9.2	Level/detail of information
		9.3	Information extraction by reconstruction
		9.4	Information refinement
10	Searching metadata	10.1	Metadata standards
		10.2	Ontologies for application domains
		10.3	File structure mechanisms
		10.4	Research in search technologies
11	Validation and Fidelity	11.1	Testbed
		11.2	Fidelity of data and data operation
		11.3	Tools for metadata extraction and validation
		11.4	Provenance
12	End Users	12.1	End user perspective
		12.2	Reliance on middleware application support
		12.3	Usecases
13	Case study	13.1	Archival engineering body of knowledge
		13.2	Tools repository
		13.3	Analysis of existing archival systems
		13.4	Engineering data corpus
14	Vendor Support	14.1	Incentives for software vendor to implement standards
		14.2	Reconciling multiple data formats
		14.3	Vendor support for post delivery maintenance

Table 1.1: Referenced requirements of Long Term Knowledge Retention [NIST 07]

This conclusion includes all aspects that affect knowledge management and knowledge retention in the long term. We use some of these requirements as a base for our long term preservation architecture and platform.

1.3. Related Works

1.3.1. Digital Information Preservation

The archiving of engineering and manufacturing information has been carried out for a long time in the paper based world. Technologies such as microfiche, while preserving the loyalty of paper, also provided a means to reduce the space required for storing information. Generally, manufacturing organizations created departments to maintain company archives able to adapt the available technologies. The advent of computing brought about new means for creating and storing the information, and generated new demands for archiving.

Data has always been a critical asset to manufacturing. Increasingly, data is in digital form with no corresponding analog equivalent [Glandney 07]. For example, 3D digital models have become the preferred method for specifying designs in the transportation sector. Digital information plays a significant role in production context, because the digital preservation becomes one main objective of the construction and the maintenance of a production information system. However, regarding the product lifecycle management (PLM) aspect, the product lifecycles are often far longer (e.g., aircraft – fifty years) than the expected lifetime of a manufacturing software application used to interpret the data (approximately three years).

Sometimes people have difficulties in distinguishing between Knowledge Management, data mining and digital preservation. In fact, there are obvious correlations between these three activities but also clear boundaries.

Knowledge Management (KM) is the approach, as well as the system, to gather, manage, use, analyze, share, and discover knowledge in a designated organization in order to maximize the performance of the functions within the organization [Chen 01]. There is no standard definition of what knowledge should consist of, but it is generally agreed that data, information, and knowledge have different levels. Data are mostly constructed, packaged,, in an electronic form, and is stored in database management systems. Information is normally unstructured, and sometimes in textual form. Knowledge is more comprehensive than the previous ones, and is needed to support decision making or business processes

Data mining is normally utilized during the knowledge or data discovering process. Data mining is one of the most important sub-functions in KM. Data mining aims to analyze a set of given data or information, in order to identify the most useful patterns for the users[Fayyad et al. 96]. The techniques(e.g. decision trees Bayesian models, associate rule mining, artificial neural networks, genetic algorithms, etc.) are often used to discover patterns or knowledge for the end users, while the end users have not known these patterns or knowledge yet [Dunham 02; Chen and Chau 04].

Digital preservation has correlations with KM, but develops in a totally different context from data mining. Digital Preservation – the process of keeping electronic material accessible and usable for a certain period of time – has turned into one of the most critical challenges within the digital library community [Strodl et al. 07]. In our research work, we focus more on digital preservation, than on knowledge discovering and data mining. Our objective is to hold the preserved knowledge for a longer time, in order to make sure that knowledge users in the future will benefit from the knowledge which is created now.

1.3.2. Research Works and Initiatives on Digital Preservation

Recognizing the importance of these electronic records for its mission of preserving "essential evidence," some research works and projects in Information and Communication Technology (ICT) launched a major new initiative, the Electronic Records Archives (ERA) initiative, in 1998. The Consultative Committee for Space Data Systems (CCSDS) recommendation established a common framework of terms and concepts which comprises an Open Archival Information System (OAIS) and was later adapted as ISO (International Organization for Standardization) 14721:2003. Various other efforts have been explored to address the needs for long term knowledge retention in specific areas such as manufacturing, health care, life sciences, legal, and military applications [Glandney 07].

The importance of digital preservation is clearly emphasized by various efforts. In Europe, there are more than fifteen projects, in the perspective of long term digital preservation, funded by the European Union (EU) since the year 2006, through its Framework Programmes. As an example, the 6th Framework Programme (FP6) had been completed. The FP6 projects include CASPAR, DPE and PLANETS.

Among these early completed projects, CASPAR provides a completed integrated architecture for digital preservation. **CASPAR** (Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval) project [CASPAR 06] was launched in April 2006 and has already been completed. CASPAR has addressed the growing challenge facing society of a deluge of intrinsically fragile digital information, upon which it is increasingly dependent. CASPAR is performed by building a pioneering framework to support the end-to-end preservation "lifecycle" for scientific, artistic and cultural information, and it is based on emerging existing standards. CASPAR tries to build up a common preservation framework for heterogeneous data. CASPAR brings together a consortium covering important digital holdings, with the appropriate extensive scientific, cultural and creative expertise, together with commercial partners, and world leaders in the field of information preservation. Figure 1.1 shows the layered architecture of the CASPAR project. This layered architecture provides the functionalities to ingest, to preserve and to reuse the digital objects while following the information workflow of the OAIS reference model. CASPAR combines the knowledge management and digital preservation approaches, but focuses on cultural, artistic and scientific contexts. Our research work targets the product data and knowledge, thus the knowledge management in our research requires a specific methodology and approach. However, the layered architecture has inspired our design in the integrated architecture for long term knowledge retention.

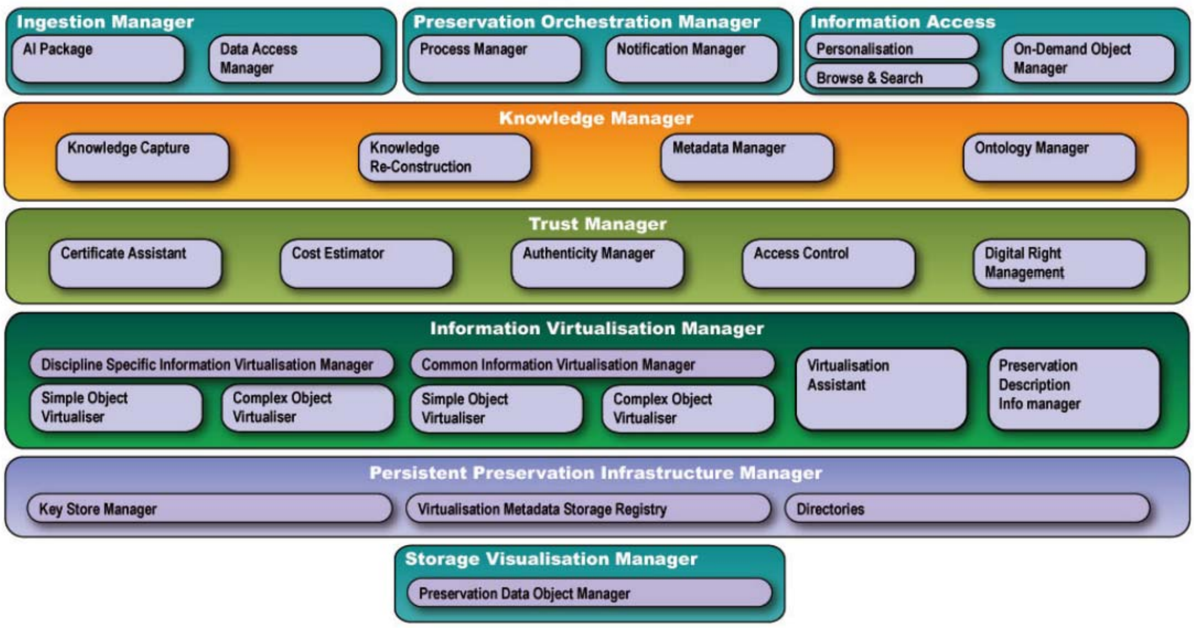

Figure 1.1: CASPAR integrated architecture [CASPAR 06]

Besides the architecture for digital preservation, projects such as PLANETS collect the experience of digital preservation researches and form a network for further researches. **PLANETS** (Preservation and Long-term Access through Networked Services) project [PLANETS 06] was launched in June 2006. It is a four-year program which addresses core digital preservation challenges. The strong PLANETS consortium brings together expertise across Europe from national libraries and archives, as well as leading research universities and technology companies, and it is coordinated by the British Library. The PLANETS project ended on 31 May 2010. PLANETS results will be maintained and developed by a follow-on organization called the Open Planets Foundation (OPF) [Open Planets Foundation].

OPF is a not-for-profit company, registered in the UK. OPF has been established to provide practical solutions and expertise in digital preservation, building on the research and development outputs of the Planets project. OPF's mission is to ensure that its members around the world are able to meet their digital preservation challenges with a solution that is widely adopted and actively being practiced by national heritage organizations and beyond. OPF believes that establishing digital preservation practice requires an open community that actively shares the best practice and is able to apply group learning. OPF founders foresee that making tools available under an open source license where and when possible will stimulate the adoption of the digital preservation practice. By communicating with the researchers in the digital preservation domain, we acquire the latest requirements for long term knowledge preservation, which will be stated in the following sections.

The EU ICT Programme, which is planned to be completed in the year 2013, is the 7th Framework Progamme (FP7). The first work programme under FP7 (2007-2008) [Europea Commission: CORDIS/FP7] consists of proposal of new digital preservation approaches. In these approaches, the advanced ICTs have the capacities, which include acting on huge volumes of dynamic web content, keeping integrity, authenticating and accessing over long term, and tracking of contexts, etc. The first FP7 projects include PROTAGE, SHAMAN, LiWA, PrestoPRIME and KEEP.

The new projects in EU ICT programme provide some new ideas on digital preservation. For example, PROTAGE brings agents in the digital preservation environment, in order to simplify the usage of preservation for end users. **PROTAGE** (Preservation organizations using tools in agent environments) [PROTAGE 07] was launched in November 2007. The mission of the PROTAGE project is to investigate and initiate complementary new approaches to digital preservation. The goal of PROTAGE is to make long-term digital preservation easy enough for users to be able to help preserving their own content, while reducing the cost and increasing the capacity of memory institutions to preserve digital information. The targeted end users of the PROTAGE project are digital curators and content creators, including individuals creating and managing their own digital information. The developed solutions will be flexible and extensible, so that they can be utilized by archives, libraries, museums, private and public sector organizations as well as individuals. Figure 1.2 shows the agents network in PROTAGE. We notice that this project identifies the agents in digital preservation approaches, to ensure that the preservation process will last in the long term. We also consider the aspects of separating the whole preservation approach and constructing services, in order to build a more dynamic system. However, in PROTAGE, the researchers focus more on the digital preservation architecture and environment than on the creation of digital objects (knowledge objects). In our research work, we must consider the product knowledge creation and knowledge management process according to our research context.

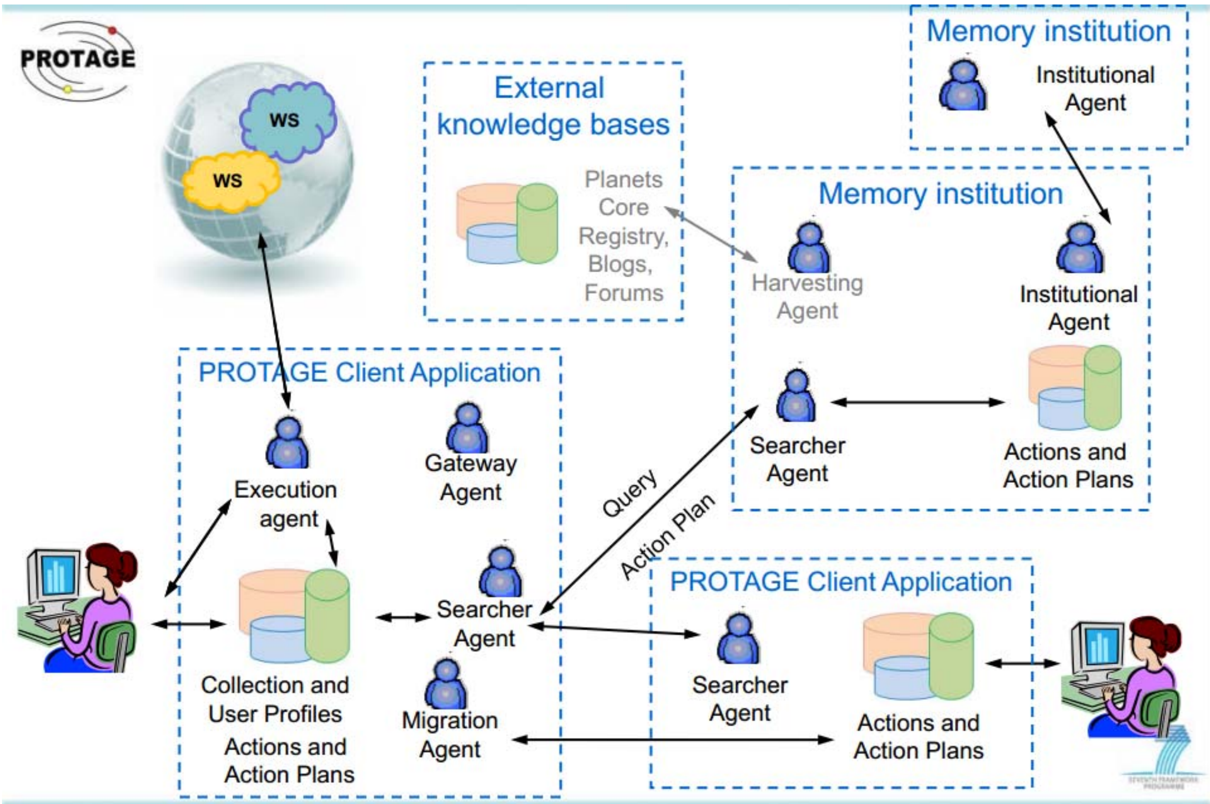

Figure 1.2: Agents network of PROTAGE [PROTAGE 07]

Other projects, such as **LiWA** (Living web archives) [LiWA 08] and **PrestoPRIME** [PrestoPRIME 09], focus on the preservation of the web or online information. They provide frameworks on web information archives and transfer. The thoughts of transferring data and information through the web is not our primary goal in this work, still the online tools they have developed will be helpful if we try to establish a distributed system in a production context which requires online communications.

KEEP (Keeping emulation environments portable) [KEEP 09] was launched in February 2009. KEEP will develop emulation services (KEEP Emulation Services) to enable accurate rendering of both static and dynamic digital objects: text, sound, and image files; multimedia documents, websites, databases, videogames etc. The general architecture of its Transfer Tool Framework (TTF) is shown in Figure 1.3.

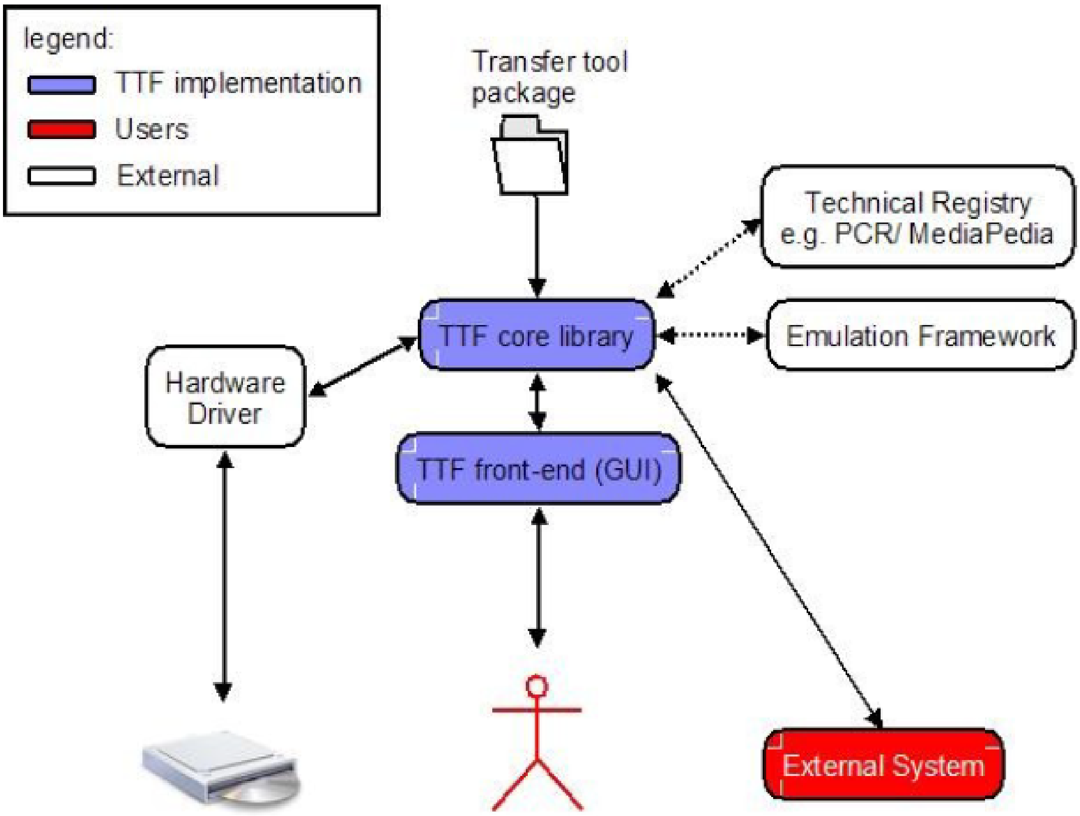

Figure 1.3: TTF system overview [KEEP 09]

The overall aim of the project is to facilitate universal access to our cultural heritage by developing flexible tools for accessing and storing a wide range of digital objects. The architecture is composed by two main parts: core library and GUI. The core library interacts with both machine interface and end user interface (i.e. TTF front-end, GUI). Although in Figure 1.3 we see the TTF is just core and interfaces, in KEEP the code organization is also layered (Figure 1.4).

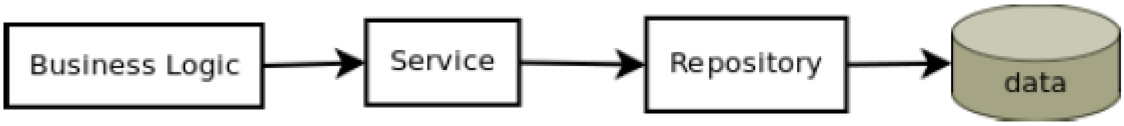

Figure 1.4: Layered code architecture of TTF [KEEP 09]

The implementation of KEEP is thus in a multi-layer architecture: from the top level business logic to low level repository implementation; service layer is needed to encapsulate the low level access methods into the top level.

Through this analysis, the researchers have achieved a design, in order to create the emulation services, repository media and transfer tools which are meant to be flexible and user friendly.

In the 2009-2010 Work Programme of FP7, the target outcomes for digital preservation were [European Commission: CORDIS/FP7]:

- The scalability of preservation systems and services. The systems and services are for preservation of digital objects, preservation workflows for different types of digital objects;
- More scenarios for preservation. The scenarios will be advanced, and thanks to these scenarios, the models and tools can be developed for overcoming the current preservation challenges, which are caused by current models and tools.

Furthermore, the FP7 also defines the priorities for 2011-2012, and the target outcomes are:

- Technologies and methods which will be more reliable and more secure
- Systems which are used for Intelligent Preservation Management
- Interdisciplinary research networks
- Schemas for implementing the outcomes of digital preservation research

Therefore, in 2011, other projects were launched in the context of ICT and digital preservation:

- Scalable solutions for digital preservation (**SCAPE**: SCAlable Preservation Environments), which will create automated quality assured workflow and support preservation planning based on the organizational policy.
- Intelligent digital curation and preservation systems (**ARCOMEM**, **TIMBUS**, **ENSURE**): leveraging the Wisdom of the Crowds (social Web) for content appraisal, selection and preservation; preserving access to services and software that support

business processes; commercially relevant data; issues specific to industry and services (IPR, privacy, legal compliance, use of existing IT tools and infrastructures).

- Preserving complex objects (**Blog Forever**, **Wf4Ever**): weblogs; continuously evolving challenges; scientific workflows (i.e. not only results but also discovery processes).
- Network of Excellence (**APARSEN**: Alliance Permanent Access to the Records of Science in Europe Network): strengthening and extending collaboration amongst major European stakeholders in digital data and digital preservation (focus on science records); creating a virtual digital preservation research centre.

Besides the EU-funded projects, there are other projects address the issue of long term digital preservation in the context of product lifecycle management: LOTAR (LOng Term ARchiving) project, KIM (Knowledge and Information Management Through Life) project, DCC (Digital Curation Centre) project, MIMER project by EPM Technology and ITI TranscenData by International TechneGroup Inc. (ITI), etc. Among them, the LOTAR project has made successful achievements since it was launched in 2002.

The objective of **LOTAR** International is to develop a process for the long-term archiving (LTA) of digital data, and this process should be auditable (on archiving 3D CAD and PDM data). The standards in industries will be used in LOTAR. The results are based on the ISO 14721 and OAIS Reference Model. The documents for this standard are published as the EN9300 series. LOTAR International is a leading project in the aerospace area and collaborates with ASD-STAN, AIA, PDES Inc. and the ProSTEP iViP Association [JSA 10]. Figure 1.5 shows the architecture of LOTAR by adapting the OAIS reference model.

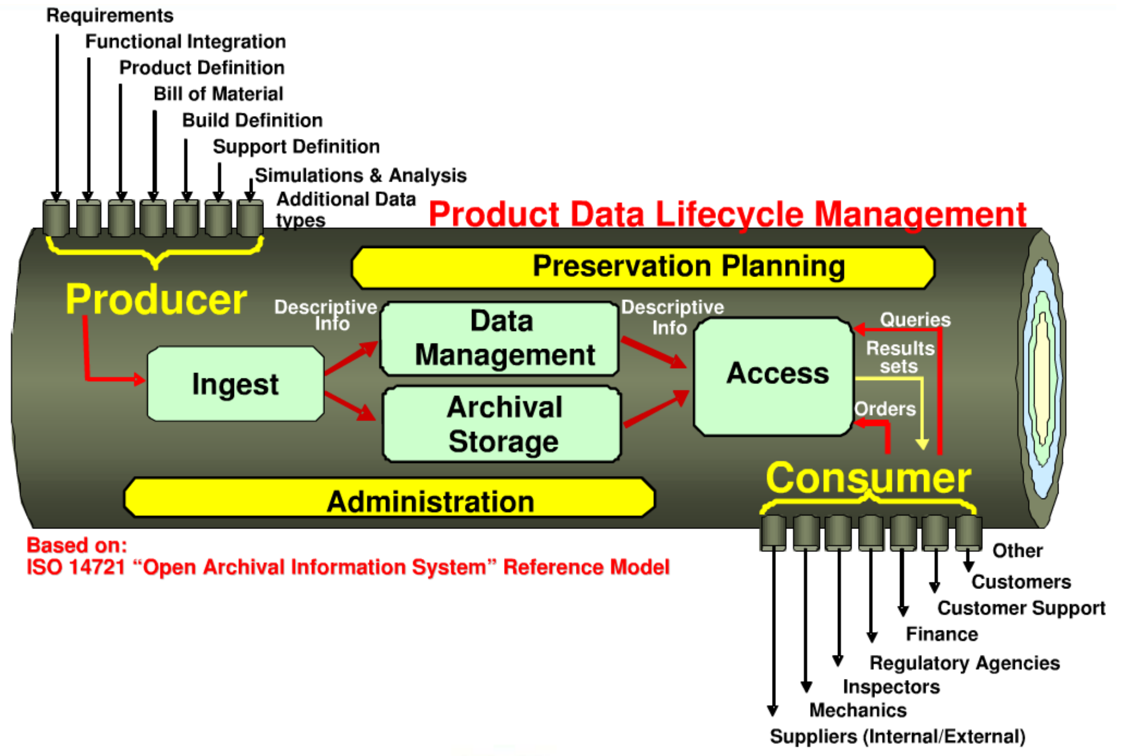

Figure 1.5: Adaptation of OAIS in LOTAR project [LOTAR 03]

The prominent research issues of the recent projects are shown in Table 1.2. We have noticed that the existing research projects on digital preservation focus more on preservation methods and models than the data context. However, as we have argued, knowledge retention should be considered rather than data preservation, in order to overcome the long term changes and issues. Therefore in our research work, we try to discover a fair methodology and architecture to combine knowledge conceptualization and digital preservation.

Issue / Project	Scalability	Methods & Models	Integrity	Complex Objects	Context	Automation
PROTAGE						✓
SHAMAN		✓				
LiWA		✓	✓	✓	✓	✓
PrestoPRIME		✓		✓		
KEEP				✓		
SCAPE	✓	✓				
ARCOMEM		✓			✓	✓
TIMBUS		✓	✓			

Issue / Project	Scalability	Methods & Models	Integrity	Complex Objects	Context	Automation
ENSURE	✓	✓				
BlogForever	✓			✓		
Wf4Ever		✓	✓			
APARSEN		✓	✓			
LOTAR		✓		✓	✓	

Table 1.2: Prominent research issues of ICT projects

There exist several technologies for long term preservation. These technologies could be easy to understand but in the meanwhile may not be always successfully archived, which is mainly accounted for the multi-formity of digital data. However, these technologies are truly the basis and guidance for our research [Strodl et al. 07].

- Preserving technology: collections of obsolete hardware and operating systems being maintained. However, to adapt this technology for long term, the increasingly large quantity of hardware and software is also a major problem.
- Emulation: the development of emulator programs that can mimic the behavior of obsolete hardware and operating systems. Emulation denotes the duplication of the functionality of systems, be it software, hardware parts, or legacy computer systems as a whole, which are needed to display, access, or edit a specific document. In the preservation context, this most often means emulating a certain (version of) a software system needed to access a file in an outdated version or format.
- Migration: the periodic transfer of information from one generation of computer technology to a subsequent one. Migration is the method of repeated conversion of files or objects. A file is converted to either a more current version of its own file format, or to another, which is easier to handle and access. A good example of migration to an easier preservation format is the recently adopted PDF/A standard. It implements a subset of the PDF standard and is especially well-suited for long-time preservation due to its omitting of, for instance, embedded scripts. Other examples would be the conversion from Microsoft Word to RTF and vice versa. In our work, the migration seems quite efficient in dynamically keeping the format of preserved knowledge up to date. Surely, the comprehensive migration planning is required.

- Encapsulation: based on the idea that preserved objects should - to some extent - be self-describing. The content is encapsulated with all of the information required for it to be deciphered and understood. (Concept of OAIS) Encapsulation needs more content than data. For self-describing, descriptive metadata is necessary and critical.

Of course, it does not mean that, we will easily solve the long term preservation problems by just integrating one or more of these technologies. However, we consider these technologies as the basics for establishing a comprehensive platform for long term preservation. The prominent technologies used in recent digital preservation projects are shown in Table 1.3. In fact, these technologies are not always necessary in digital preservation projects. This is the reason why some projects (e.g. SHAMAN, etc.) aim to develop scalable environment for digital preservation, instead of proposing standards for digital migration or encapsulation.

Issue / **Project**	**Preserving Technology**	**Emulation**	**Migration**	**Encapsulation**
PROTAGE				✓
SHAMAN	✓			
LiWA	✓			
PrestoPRIME			✓	
KEEP		✓		
LOTAR			✓	✓

Table 1.3: Prominent technologies in recent digital preservation projects

1.4. Discussion

Some long term preservation requirements are supported by existing technologies/tools and functionalities. However, gaps between the requirements and the existing technologies/tools still remain.

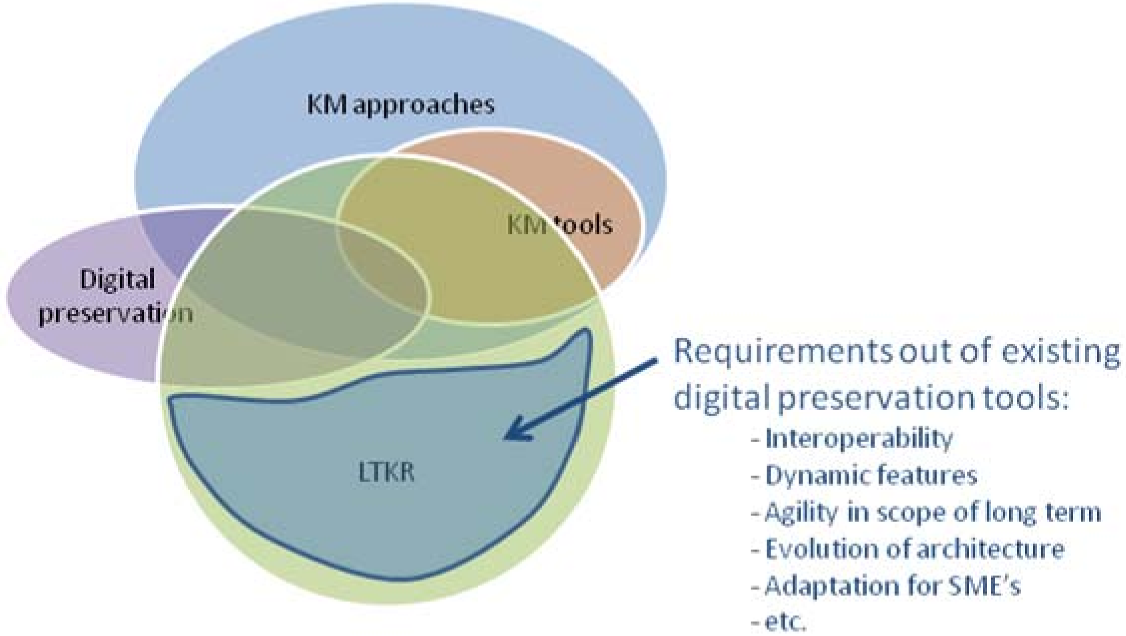

Figure 1.6: Non-reached long term preservation requirements based on existing technologies

We call these aspects "missing aspects", but in fact they are not all missing in existing knowledge management projects or systems. On the contrary, nearly all the aspects are somehow mentioned in various existing KM projects, even if not fully implemented. Indeed, in our perspective of long term preservation, these aspects are either not fully implemented or not implemented at all (some functionalities, which have the same name, can in fact be really different). The "missing aspects" mainly concern the interoperability of systems, the dynamic features of archival systems and the agility of digital archival systems in the scope of long term preservation. We have discovered the general missing aspects as follows with brief explanations:

- Cultural issues: this culture issue affects KM a lot, and it has been considered by many KM projects. However, it is maybe the most difficult issue to solve. If all the knowledge-based systems obey the same rule or standard, knowledge could be captured, managed and transferred quite efficiently. We cannot just argue it is totally impossible, still, by now, we have tried to overcome the cultural issue by adapting some technologies or methodologies (e.g. ontologies, etc.).
- Archival management from an organization structure view: this requests the correlations between knowledge repositories and the organization which is the source

- Sustainability for preservation: in other words, this is the objective of long term preservation.
- Detection of formats obsolescence: in long term preservation, if obsolescent information cannot be detected, of course it will not be removed from repositories. And more and more resources are taken by unnecessary archives. This situation not only wastes digital preservation resources, but also brings more critical problems in knowledge retrieving.
- Information models for archival systems: the OAIS reference model has proposed Information Package as information models in digital preservation platforms. However, through the preservation process, digital information goes from a source information system to a digital preservation platform, and during the transition information may appear in different forms or structures. Therefore, intermediate information models should also be concerned.
- Information captured through PLM stages: this aspect is shown here because in our research work, we focus on knowledge retention in the field of product lifecycle management (PLM). The information capturing process though PLM stages requires the dynamic connections between the digital preservation platform and the information systems of PLM.
- Information extraction by reconstruction: this is also concerns the information model aspect we have discussed above.
- Information refinement: information refinement in long term also requires dynamic connections between source information systems and digital preservation platform. Moreover, it depends on the information model transfer rules or techniques (i.e. refining information in the same information model, transferring knowledge from one model to a new better model, etc.)
- Analysis of existing archival systems: we have already done this work of evaluation of existing digital preservation platforms and identified the functionalities, which may be utilized in long term preservation.

- Incentives for software vendors to implement standards: the unique features, which would attract software vendors to adapt one methodology/technology/standard, should be highlighted in long term preservation project proposals.
- Reconciling multiple data formats: the various data formats (e.g. in engineering fields, 3D, CAD data, etc.) should be recognized and associated by supporting software for reuse. And sometimes people also tend to adapt common standards (e.g. STEP, etc.) for data migration.

The referenced "missing aspects" for long term preservation, along with the KM methodologies and functionalities we have analyzed, are the preliminary fundamental research for our research work. They constitute the basis to establish a functionality-based architecture for long term preservation of knowledge.

1.4.1. Information System Engineering and Long Term Preservation Requirements

Modern information systems solve many problems: data, information and knowledge-based problems, although in the past most of the information systems were developed focusing on data; in other words, they were data-oriented systems. The primary purposes of these data-oriented information systems are to preserve, manage, retrieve and display data. The application domains of the information systems are organizations or industries which need to deal with huge numbers of data, such as banks, airlines or governments. From the 1980s, more requirements on the capabilities of information systems have been raised and theories have been developed. Information systems have expanded from data-oriented computing to analytical computing. The developed information systems provide people with analytical support such as planning, decision-making, simulations, etc.

Information system engineering has consequently developed in the last decades. System engineering covers the whole lifecycle of the systems: including system requirements, system functional design, system development, system testing and system evaluation. Regarding the development of information systems, many methodologies have appeared, as well as various models and applications to support the system engineering. With the help of these new advantages, information system engineering provides more capabilities in each phase.

Regarding long term preservation, some methods and applications focused on information system engineering will help to achieve some of the long term preservation requirements. The interoperability and dynamic system long term preservation requirements will be supported by technologies and applications in Business Process Management (BPM) and Service-Oriented Architecture (SOA) or distributed systems.

According to the study of Kohlbacher [Kohlbacher 09], BPM focused on information system engineering has helped organizations or enterprises to shorten development time and enhance customer satisfaction. As BPM system and BPM suite have been released in the technical field, BPM is now seen as a critical component of Operational Intelligence ways of delivering real-time, actionable information. Normally, the BPM suite emerges from the thinking of SOA (e.g. Oracle BPM/SOA suite). The BPM/SOA suite delivers dynamic features, which ensures the sub-systems and sub-functions' dynamic connections and interactions. The business process design transfers the system requirements into systems' "to-be" processes. The process flow and data identification share common agreements and allow the authorizations and authentications in enterprises.

According to archiving professionals of the digital preservation projects we have spoken to, there are existing technologies in system engineering for long term preservation. These technologies seem easy to understand but may not always be successfully deployed. This is mainly due to the multi-formity of digital data. These technologies are still important today and our research is based on emulation, migration and encapsulation. The critical problem of long term digital preservation is linked to the fact that we cannot determine which technologies are the most promising for the long term. One reason of this uncertainty is that we cannot foresee whether one of these technologies is going to prevail or not. At the same time, in long term perspective, an ever growing stock of digital documents which comes from different sources is stored in very different formats and requires to be preserved for an indefinite period of time. Therefore, the long term digital preservation problems cannot be solved by simply integrating one or more digital preservation technologies. Thus, a comprehensive platform for long term preservation is really needed.

From the previous research work and projects, we have synthesized the requirements and "missing aspects" of long term preservation. We have chosen the method of requirements-driven information system engineering [Castro 02] in our research work. In the requirements-driven engineering, the critical phases are:

- Early requirements, which means the understanding of the problems in an organization or an enterprise;

- Late requirements, which signifies the system-to-be is described in its operational environment, including some necessary functionalities and features;
- Architectural design, that is to say the global architecture of the system. The architecture is defined regarding sub-systems, interchange of data, control workflows and other dependencies.;
- Detailed design, which represents the detailed design in each part of the global architecture.

The long term preservation requirements, which we have synthesized and proposed, span the early requirements and late requirements. Because some of the requirements (e.g. detection of format obsolescence, etc.) means the analytical results of existing systems of long term preservation, these requirements will be considered in the late requirements phase. In our research work, we have followed this sequence of information system engineering. We have identified the long term preservation requirements and detailed requirements in each sub-section of long term preservation (i.e. KM and digital preservation). Then according to our research approach, we need to propose an architecture for long term preservation of knowledge regarding KM and digital preservation, as well as the system of KM and digital preservation. At last, we go deep in the sub-systems or sub-functions of our long term preservation architecture, which benefit from the existing features through existing technologies or applications (e.g. BPM/SOA suite, etc.).

1.4.2. Dynamic Preservation Process

In this thesis, our proposal is to develop dynamic models of digital preservation. The ideal condition of dynamic preservation is that the preservation process, the preservation platform, and the preserved content will all be dynamic and always up-to-date. Dynamic preservation concerns several aspects with dynamic characteristics and concepts. These concepts are intended to be generic ones so as to be used as a reference for any specific long term digital preservation project. Benefiting from the models, the preservation process, preservation platform, and preserved content will all be dynamic and always up-to-date thus corresponding to the information systems. Dynamic preservation concerns several aspects, including dynamic characteristics and dynamic perspectives.

In the dynamic preservation, one dynamic part is the preserved content; the knowledge (i.e. digital data content and metadata). Data refers to a collection of facts usually collected as the result of experiences, observation or an experiment, or other processes within a computer system or a set of premises; while “knowledge” contains more information than data with a comprehensive explanation of the data context. However, the information object (i.e. data content), contextual metadata or any other preserved metadata may change over long term: data content may be migrated to another form because of the development of information technology; and metadata may be restructured or augmented because of organizational changes. Our proposal is not to develop a formal knowledge model standard, but to get advantages from the existing standards (e.g. STEP standard for production data), and dynamically update knowledge models corresponding to changes of the standards and technologies. That is why in this level, we need to study and adapt KM approaches, in order to produce knowledge in a dynamic way.

Besides, the connection between the preservation system and the other systems (e.g. Knowledge Management System or simple information systems in enterprises, where KM approaches are performed) must be established and maintained in the long term. We should establish links and connections between preservation system and the other information systems, as well maintain the links with the domain experts in companies. In other words, the activities of preservation and knowledge creation ought to be dynamically connected, to make sure that changes or improvements of one or more activities will not affect the other activities.

Another major part of dynamic preservation is the interoperability of the digital preservation platform and the knowledge source (i.e. information system in enterprise, organization, etc., from which the preserved knowledge is extracted). The dynamic preservation system should always monitor tools, technologies and strategies in enterprises, so as to ensure that the preserved knowledge will be updated when critical changes are detected. This part requires both an analysis and decision making technology and tools (e.g. Multi-criteria decision analysis), in order to identify the critical changes that could affect the long term preservation. The term “critical changes” should be predefined in companies, as critical changes may include:

- New types of data or information to go into operations
- New types of processes in designated organizations
- New versions of information systems introduced in designated organizations
- New versions of software used by designated organizations

Thus in the KM process of a designated organization, there must be such a process as a change notification. When some critical change signals are received by a digital preservation platform, the digital preservation platform itself or its operator must decide whether to reconfigure the knowledge model or not, reconstruct preserved knowledge or not, migrate preserved data or not, in order to adapt to the critical changes. In other words, we need a complete digital preservation platform, which ensures the interoperability with the information systems in enterprise or the other designated organizations.

1.4.3. Proposed Research Issues

To sum up the discussions above, our research work includes a KM approach and a digital preservation approach. The KM approach identifies enterprise knowledge and produces digital knowledge objects, which represents the digital objects in digital preservation. The digital preservation platform performs the whole digital preservation process. In knowledge reuse process, the digital preservation platform also produces digital knowledge objects for companies. These two parts of long term knowledge preservation are both necessary and important. However, the KM environments are complex and various in enterprises, and compared to them, the digital preservation environment is more consistent and lasts longer in the long term. Therefore, in order to achieve dynamic preservation to connect the two major parts together, we need to develop an architecture for long term knowledge preservation. This architecture should allow both parts to perform smoothly in their own environments, and also allow them to communicate and establish available and dynamic connections. Therefore, in our research work, we have identified some missions, which will enable us to achieve dynamic preservation. These missions are based on the related works on KM and digital preservation:

- Firstly, we have to develop an extensible KM approach for knowledge conceptualization and knowledge creation;
- Secondly, we have to establish a digital preservation platform in the perspective of long term knowledge preservation. We will adapt the OAIS reference model by benefiting from its features for long term knowledge preservation, thanks to the previous digital preservation researches.

- Then we have to construct an architecture to connect the KM and digital preservation approaches in a dynamic way. This architecture will ensure the interoperability of business processes and technologies of both approaches.
- Some technologies and applications (e.g. Business Process Management, Service-Oriented Architecture applications, etc.) have provided features for dynamical systems and developing high interoperabilities. Thus in our proposals, we will use these existing technologies and applications to achieve the dynamic features in our design.

From the previous discussion, one can notice that there have been many researches related the long term preservation issues. The timeline of the digital preservation researches and our research work is shown in Figure 1.7. Our research work is half way through the long term digital preservation researches. The outcomes of these researches have not been completed but could still be considered in our research work. And the results of our research work will contribute to the future continuous long term digital preservation researches.

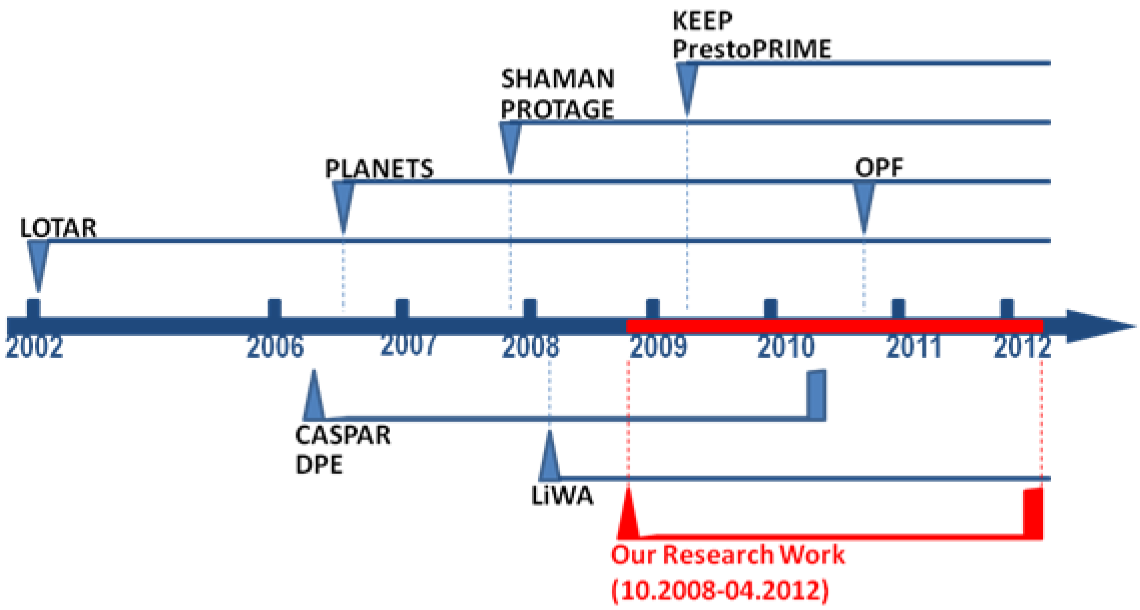

Figure 1.7: Timeline of digital preservation researches and our research work

In order to deal with the long term preservation problems in the context of product lifecycle management, we have to face some challenges. In order to achieve our goal of developing long term knowledge preservation for digital product data, we have focused on these three levels in order to develop a prototype:

- A methodology for long term knowledge preservation: our research work will be developed along this proposed methodological approach. The methodology could mean an adaptation of existing methodologies. As there is no specific methodology for long term knowledge preservation, we will study the existing major KM methodologies, which do not particularly focus on knowledge retention but comprehensively describe the process of KM. The methodology consists of a KM approach (i.e. knowledge conceptualization) and the digital preservation approach (i.e. knowledge retention).
- A model-based architecture for long term knowledge preservation: not only do we need the methodological approach, but also the structural reference to deploy a long term preservation project. Thus we propose architecture for long term knowledge preservation. The reason why we have chosen to construct a "model-based" architecture is that model-based development enhances agility. And in the long term, technologies will improve and certainly change, but models will be reused by associating different sorts of technologies. As our architecture is composed and described by models, the changes in certain technologies will not change our architecture, and this is also taken into account in a long term preservation perspective.
- Integrated systems for long term knowledge preservation: since the perspectives of our work consist of aspects and technologies in different fields (i.e. knowledge management, digital preservation, etc.), we are required to improve the interoperability of enterprise knowledge management solutions and digital preservation technologies. And in order to achieve "long term" preservation, agile technologies (e.g. web service, etc.) will also be integrated in our research work. So, if the first and second objectives are respectively tacked from the methodological and the structural point of view, the third objective of the thesis is considered from the technological point of view.

1.5. Conclusion

In this chapter, we have presented the preliminary research and analysis of the context of the long term knowledge preservation research. Technologies and functionalities provided by existing digital preservation platforms have been studied. All these functionalities have been adapted from existing platforms. Nevertheless, there are some long term preservation references and perspectives (i.e. "missing aspects"), which are not quite achieved by existing functionalities. Through the synthesis of the state of the art in the long term preservation context (i.e. KM, digital preservation, long term preservation), we have identified the missing aspects for the design of a long term knowledge preservation architecture.

The dynamic preservation theory states that if all the preserved knowledge in digital preservation platform is always up-to-date according to the current strategic and technical situations of the knowledge sources, the preservation will always be traceable and reusable in the long run. In order to make our digital preservation closer to the ideal situation of dynamic preservation, we need to develop an architecture for directing the strategic decision making, planning and functional processes.

Chapter 2. Knowledge Management for Digital Preservation

2.1. Introduction

The major threats of long term preservation are the changes over the long term. Generally, the changes include not only the preserved knowledge or data themselves, but also the Knowledge Base inside people's mind. Either change will reduce the traceability and reusability of appropriate knowledge. As we have discussed in the previous chapter, in order to achieve long term knowledge preservation, we need to develop a KM approach as well as a digital preservation approach. In this chapter, we develop a KM methodology for providing knowledge objects for digital preservation. The KM approach needs the strategic decision making, planning and functional processes for knowledge management and knowledge retention. Based on the preliminary research on KM methodologies, we choose to extend from an existing KM methodology (i.e. CommonKADS) in our research, in order to adapt the Product-Process-Organization (PPO) design principle in knowledge management approach for products.

2.2. Knowledge Management Methodology

As we have already discussed in the previous chapter, "knowledge retention" is in the "knowledge management" context. And we have to study KM, in order to develop approaches for "knowledge retention". In this section we have studied KM methodology, tools and systems. The KM methodologies will help to develop methodological approaches and models for KM and thus knowledge retention. The KM tools provide functionalities and features to achieve the knowledge retention approaches and models. The KM system is a composition of KM technologies and tools, in order to support KM. In our research work, we introduce the "system" concept, in order to establish a platform for the perspective of long term digital preservation.

2.2.1. Knowledge and Knowledge Management

The formal definition of "knowledge" is defined by Webster's dictionary and implies that knowledge extends beyond information. It gives the following description:

Knowledge - noun. 1. applies to facts or ideas acquired by study, investigation, observation, or experience 2. rich in the knowledge of human nature 3. Learning applies to knowledge acquired especially through formal, often advanced, schooling 4. a book that demonstrates vast learning.

Data, *information* and *knowledge* are three often encountered words that are close together. However, data refers to a collection of facts usually collected as the result of experience, observation or experiment, or processes within a computer system, or a set of premises. Information is the basis of knowledge, might be directly associated with the facts of the real world. And knowledge often depends on the context, thus one person's knowledge could be another person's information.

Knowledge Management (KM) is the practice of selectively applying knowledge from previous experiences of decision-making to current and future decision-making activities with the express purpose of improving organizational effectiveness [Zhang 04]. For enterprises and industries, an archive of knowledge must capture all of the data required to completely define the product, and in some instances, processes. Thus KM is not just a technology, but it is about people, processes and practice.

Our research concerning knowledge management represents initial work for long term digital preservation. According to the study of Dave [Dave 07], the discipline of KM has changed very quickly over the last decade, and KM can generally fall into two generations. The first generation of KM is from 1995 to 2005, and the second generation is from 2005 to today. Besides the strategic differences between the two generations of KM (from central large repositories to shared personal repositories), the content formats of the first generation of KM are mostly categories of texts organized by subjects (taxonomy), while the content formats of the second generation of KM are graphics and multimedia organized by applications (ontology). Therefore, in modern KM discipline, managing knowledge activities are performed in a collaborative environment [Sureephong 09]. The complexity of producing, keeping and reusing knowledge from graphic and multimedia formats implies that digital information preservation approaches and technologies play an important role in modern KM projects.

2.2.2. Knowledge Management Methodologies

In order to achieve dynamic preservation and keep the preserved knowledge and data always up-to-date, we need to develop the strategic decision making, planning and functional processes for knowledge management and knowledge retention. Therefore, we need to adapt the KM approach and develop a methodology.

Ritendra Banerjee in his article [Banerjee 05] claims that the KM Cycle is described as

- Knowledge Generation
- Knowledge Codification
- Knowledge Retrieval
- Knowledge Transfer
- Knowledge Purging

Organizations need to determine a formal process in order to organize knowledge: identify, capture, store, and retrieve critical knowledge. Thus, organizations need KM to help them identify what they know, what they need to know, and how to effectively use what they know [Jennex 08]. According to the generalization of KM methodologies, the following KM steps will be used in our research:

- Researching and aligning with Corporate Strategy
- Identifying business process and knowledge
- Capturing knowledge
- Communicating and organizing knowledge
- Creating a knowledge-sharing culture
- Benchmarking
- Improving the process continuously

Although via the definition of operational steps of KM methodology we get references for the deployment of long term knowledge preservation, we still need recommendation design models for the process of organizing knowledge to reach the enterprise needs of preservation, as well as maintenance and transaction of knowledge. In other words, we have to develop knowledge engineering methodology to carry out long term knowledge preservation.

Knowledge engineering refers to the designing, developing and maintaining of Knowledge Based Systems (KBS) in the knowledge management project. It has a great deal in common with software engineering [Sureephong 09], and is related to many computer science domains such as artificial intelligence, databases, data mining, expert systems and decision support systems. Our methodology of long term knowledge preservation has established a dynamic strategy and preservation system to keep the stored knowledge (Product-Process-Organization: PPO model of PLM in enterprise [Noël 08]) always up-to-date; thus, to overcome the long term changes. Through the research on knowledge management and digital preservation, we notice that there is gap between existing digital preservation technologies or tools and our goal of establishing dynamic preservation.

A KM methodology is proposed and developed, and the utilizations of this methodology are:

- Drawing a whole guideline for development of KM in long term knowledge preservation projects;
- Identifying a hierarchical structure of KM;
- Identifying a sequential deployment process for KM;
- Designing data and process models for each level and each step.

One of our objectives, which are discussed in **Chapter 1**, is to propose a methodology, which is based on KM methodology, in order to carry out our research work. We propose a worksheet formula, in order to analyze and evaluate the existing KM methodologies. Although the KM methodologies are performed in different procedures and focus on different dimensions, we intend to use this worksheet formula to collect the information that we are interested in for our research work. We are going to separate the KM methodologies into different functional modules, and a functional module can by a process, a sub-process, or a simple activity. The functional modules are integrated and interacted in one KM methodology. By filling the worksheet formula (Table 2.1) that we have defined, we will have a comprehensive understanding of these KM methodologies in the viewpoint of functionalities. Because we use model-based development process to carry out our research work, the results of this functional module analysis of KM methodologies will help us to initialize the required functional modules of our own KM methodology, as well as to choose a KM methodology as our initial reference. The initial referential KM methodology will be combined with the extra functional modules. The worksheet formula we use to summarize the KM methodologies is shown in Table 2.1. The purpose of each column in the worksheet formula is presented, too.

Number	Processes /Activities	How to /Phases	Outputs	Why/Objectives	Tool(s)	Deployment
Step 1						
Step 2						
……						

Table 2.1: Worksheet formula for analyzing KM methodologies

- Number: The number to identify each process in the methodology.
- Processes/Activities: The general processes or activities, which are proposed by the methodology.
- How to/Phases: The detailed description of one process/activity. All we consider this part as decomposition of one process/activity. In most of the KM methodologies, the processes and sub-processes are well illustrated.
- Outputs: The outputs of each phase. The output could be specific documents or signals or triggers for other activities or phases.
- Why/Objectives: The objective of one phase, and also the reason why we should add this phase in the methodology.
- Tool(s): The supporting tool(s) or software that would will achieve the methodology.
- Deployment: The deployment of the KM phase or the supporting tool(s).

Of course not all the KM methodologies are described in a sequential manner, but we have utilized this formula to synthesize them, in order to identify the common functionalities as well as special features in KM (**Annex 1**). At the end of this synthesis, we are going to propose a methodology. Table 2.2 shows one example of one of the methodologies we have analyzed – CommonKADS methodology.

Number	Processes /Activities	How to /Phases	Outputs	Why /Objectives	Tool(s)	Deployment
Step 1	Context level	Coping and feasibility study: analysis	OM: 5 worksheets	Identify problem/opportunity areas and potential solutions; Put them into a wider organizational perspective.	Organization Model	-
		Coping and feasibility study: synthesis	OM: 5 worksheets	Decide about economic, technical and project feasibility; Select the most promising focus area and target solution.	Organization Model	-
		Impact and improvement study: analysis	TM: 2 worksheets; AM: 1 worksheet	study interrelationships between the task, agents involved, and use of knowledge for successful performance; what improvements may be achieved here	Task Model, Agent Model	-
		Impact and improvement study: synthesis	TM: 2 worksheets; AM: 1 worksheet	Decide about organizational measures and task changes; Ensure organizational acceptance and integration of a knowledge system solution	Task Model, Agent Model	-
		-	Summary: 1 worksheet	-	-	-
Step 2	Concept level	Knowledge identification	Knowledge Model	survey the knowledge items; prepare them for specification	-	-
		Knowledge specification		complete specification of knowledge except for contents of domain models	-	-
		Knowledge refinement		Validate knowledge model; Fill contents of knowledge bases	structured walk-troughs; software tools for checking the syntax and find missing parts; paper-based simulation; prototype system	-
		-	CM: 2 worksheets	specifies knowledge/information transfer procedures; top-level control over task execution; additional communication tasks	-	-

Number	Processes /Activities	How to /Phases	Outputs	Why /Objectives	Tool(s)	Deployment
Step 3	Artefact level	-	specification of a software architecture; design of the application within this architecture	Specify the architecture of implementation of the KM project	-	-

Table 2.2: Analysis result – CommonKADS methodology

From the analysis results of the CommonKADS methodology, we notice that there are three major functional modules, which is separated by different levels of viewpoints. In Context level and Concept level, there are decompositions and work processes, and each phase produces one or several worksheets of knowledge. On the contrary, the Artefact level of CommonKADS is not described in detail, and depends on the implementation methods and tools that the specific users of CommonKADS want to use. The tools that are utilized in each phase are predefined models (i.e. Organization Model, Task Model, etc.). The deployment of each phase is not specified in CommonKADS methodology, and some of the activities are performed without specific tools. Thus parts of the worksheet formula are blank. The detailed explanations of each KM methodology are shown in **Annex1**, and here we just discuss the results of our functional analysis of the KM methodologies. We illustrate the KM methodologies that we use in our analysis.

- Accelerated Knowledge Management (AKM) [Balafas et al. 03]: aims to develop KM methodology for long term perspectives. This methodology has been carried out at the Danwood Group in Lincoln and in collaboration with the Department of Computer Science, Loughborough University.
- Distributed Knowledge Management (DKM) [Cuel 03; Schwotzer et al. 04]: tries to manage knowledge in an autonomous way and introduces a concept of Knowledge Node (KN), which could represent a potential knowledge model for long term knowledge preservation.
- MASK [Ermine et al. 96; Barthelme et al. 98; Benmahamed et al. 05]: provides comprehensive approaches of knowledge capitalization: knowledge analysis and modeling. And the efficiency of the MASK methodology is proved by many KM projects.
- MaKE [Sharp et al. 03]: concerns the Information System (IS) development with KM approaches.
- SAKE [Ntioudis et al. 07-4, Ntioudis et al. 07-6]: describes the whole KM/KE approaches from analysis to deployment in the public administration environment. The approaches it describes and the connections between each step in KM/KE are helpful in our research work, which aims to develop an architecture and methodology for long term knowledge preservation.
- SMARTVision [Bubenstein-Montano et al. 01]: presents a more micro-view of a specific KM methodology, based on the existing KM methodologies. Different from

some existing KM methodologies, which are more strategic, SMARTVision provides a detailed description of each step of implementation of KM in designated organizations. The research on SMARTVision is helpful for implementing KM methodology in our research work, and contributes to discovering more threats and challenges in deploying a KM methodology into applications.

- KM-Beat-It [Bures 05]: is quite similar to the common KM methodologies. However, in the development of KM-Beat-It, some strengths and weaknesses of KM implementation are stated as a basis for this methodology. Thus KM-Beat-It really considers the issues of KM implementation, which are also our concerns in long term knowledge preservation projects.
- Common Knowledge Acquisition and Design System (CommonKADS) [Orsvarn et al. 95]: offers a structured approach to break down and structure knowledge engineering process. CommonKADS provides model-sets for creating requirements specifications for knowledge systems.

The results of the KM methodologies analysis lead to several sequential general functional aspects that KM should consider. They are:

- Strategy alignment: KM strategy must be predefined when starting a KM project in designated organizations. This strategy will lead to the following steps or functions in one KM project.
- Knowledge identification: to identify existing data, information and knowledge from reliable sources within or across organizations.
- Knowledge acquisition: based on the knowledge identification and knowledge source identification, the procedures/functions/technologies for knowledge acquisition will be determined and performed.
- Knowledge modeling: a knowledge model is an interpretable model of knowledge or standard specifications about a kind of knowledge.
- Knowledge adaptation: after knowledge has been extracted and encapsulated thanks to the knowledge models, knowledge should be stored in repositories, as the security of the preserved knowledge must be ensured.
- Knowledge transfer: the process and tools of knowledge transfer should be identified. The tools for transferring knowledge could be in the form of a static network approved by the designated organization, or web services, etc.

- Knowledge evaluation: the dynamic analysis and identification of knowledge.
- Knowledge Revise: the dynamic maintenance of knowledge.

These aspects appear in nearly all the KM methodologies. Through our analysis results, these aspects together constitute the basics of the KM architecture. The identification of these fundamental aspects constitutes the initial work for our research. If certain tools provided the functionalities that achieve these aspects, it would be quite possible to use the very tools to support a KM project. And moreover, if these tools, which are with KM functionalities, collaborate, we would establish an integrated KM system.

2.2.3. Knowledge Management Tools

We try to find out if the fundamental aspects of KM are supported by existing tools, thus we have performed on the analysis in existing commercial software tools, which concern data/information/knowledge management. From the research and analysis of [Banerjee 04] the KM tools are divided into 4 business classes and 16 functional sub-classes:

<table>
<tr><th>Technology</th><th colspan="2">Tool Type / Functional Feature</th></tr>
<tr><td rowspan="6">Knowledge Ceneration</td><td rowspan="3">Meeting Support Tools</td><td>Structuring</td></tr>
<tr><td>Visualizing</td></tr>
<tr><td>Polling</td></tr>
<tr><td colspan="2">Group Decision Support Software</td></tr>
<tr><td colspan="2">Data-mining</td></tr>
<tr><td colspan="2">Online Analytical Processing (OLAP)</td></tr>
<tr><td rowspan="4">Knowledge Codification</td><td colspan="2">Knowledge Repositories</td></tr>
<tr><td colspan="2">Document Management Systems</td></tr>
<tr><td colspan="2">Text-mining</td></tr>
<tr><td colspan="2">Taxonomy Generators</td></tr>
<tr><td rowspan="3">Knowledge Retrieval</td><td colspan="2">Retrieval Systems</td></tr>
<tr><td colspan="2">Search Machines</td></tr>
<tr><td colspan="2">Navigators</td></tr>
<tr><td rowspan="3">Knowledge Transfer</td><td colspan="2">Online Collaboration</td></tr>
<tr><td colspan="2">Online Coordination</td></tr>
<tr><td colspan="2">Training Tools</td></tr>
</table>

Table 2.3: Classification of KM tools [Banerjee 04]

In order to prove this taxonomy, we have carried out a functional analysis of 78 KM software tools, which are developed by 33 different software companies. The result of the analysis shows that the taxonomy of KM tools is fair (Figure 2.1, **Annex 2**).

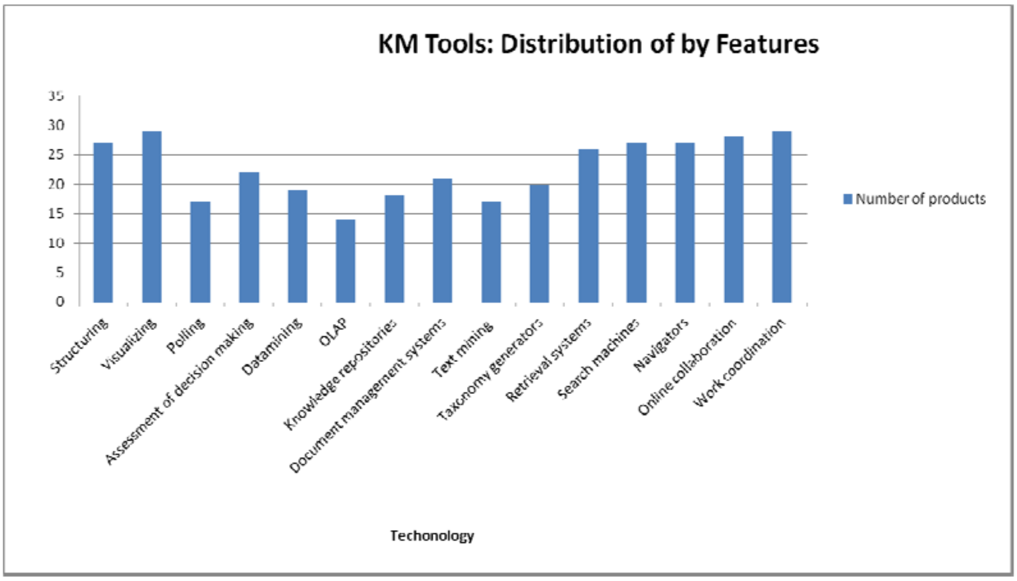

Figure 2.1: Distribution of KM tools by functional taxonomy of Banerjee

This taxonomy is fair and it seems that, on the market, the distribution of each kind of commercial tools is similar. However, there are no KM tools that cover all or at least most of these features (Figure 2.2). So If we try to establish an integrated KM system, we need to get benefits of all sorts of features from different tools.

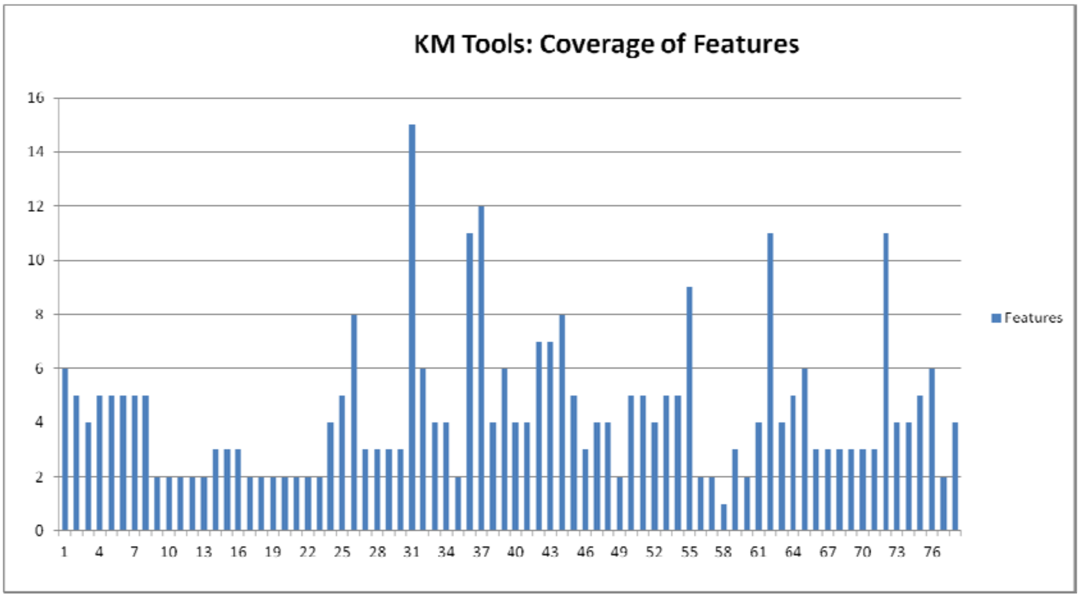

Figure 2.2: Coverage of features by KM tools

2.2.4. Knowledge Management System

Knowledge management systems (KMS) have been defined as "an emerging line of systems which target professional and managerial activities by focusing on creating, gathering, organizing and disseminating an organization's 'knowledge' as opposed to 'information' or 'data' [Alavi et al. 99]. We consider that such a system is an enhanced digital information preservation system, which integrates technologies (e.g. artificial intelligent technologies [Becerra-Fernandez 00], etc.) to support modern knowledge management.

The KMS could be utilized in different ways as in different organizations. However, through the research and observation of Becerra-Fernandez [Becerra-Fernandez 99], KMS fall into three categories:

- Educational KMS: this type of KMS is usually an educational tool for training in organizations. However, the educational training is also one way of eliciting and inheriting tacit knowledge.
- Problem-solving KMS: this type of KMS is implemented with intellectual technologies, and aims to capture knowledge for reusing and solving problems. The problems concerning knowledge could be new problems, which need solutions derived from existing knowledge; and the problems could also be old problems, which need recurring to the existing solutions. In any condition, problem-solving KMS' perspective is problem-oriented.
- Knowledge repositories: this type of KMS is the widest used in organizations. Although there are sub-categories under knowledge repositories, we just consider it as a preservation system, which keeps knowledge in certain designated forms.

We introduce the "system" concept, in order to establish a platform for the perspective of long term digital preservation. In other words, we are going to develop a digital preservation system, which implemented as repository as well as a problem-solving KMS. A digital preservation system refers to a system in which objects are stored for preservation archiving. Thus, we have to analyze the existing functionalities and technologies of digital preservation system/KMS, and to integrate the functionalities and technologies as one system in the perspective of long term knowledge preservation. The detailed architecture of our research work will be discussed in **Chapter 4**.

2.3. Proposal of KM Approach

CommonKADS stands for Common Knowledge Acquisition and Design System which is a present version of KADS. CommonKADS is one of the KM methodologies that we have analyzed in the section "**2.2.2. Knowledge Management Methodologies**". It enables a top-down approach and provides model sets at each level. The reason why we have chosen CommonKADS as our KE approach for long term knowledge preservation is that we intend to get benefits of its abundant models. We are going to develop an approach and an architecture in the perspective of model reusability in the long term. The top-down structure and model-sets composition of CommonKADS are fit for our needs. Therefore, we choose CommonKADS from other KM methodologies, because we will benefit from its model sets and KADS templates to develop a model-based architecture.

The method has been developed since 1984 through two major CEC ESPRIT (Commission of the European Communities, European Strategic Programme for Research and development in Information Technologies) funded research projects. The methodology aims to support structured knowledge engineering. It indicates the opportunities and bottlenecks in the organizations, distributes and applies their knowledge resources, and so gives tools for corporate knowledge management. It also provides the methods for performing a detailed analysis of knowledge-intensive tasks and processes. CommonKADS supports the development of knowledge systems that support selected parts of the business process [Schreiber et al. 99].

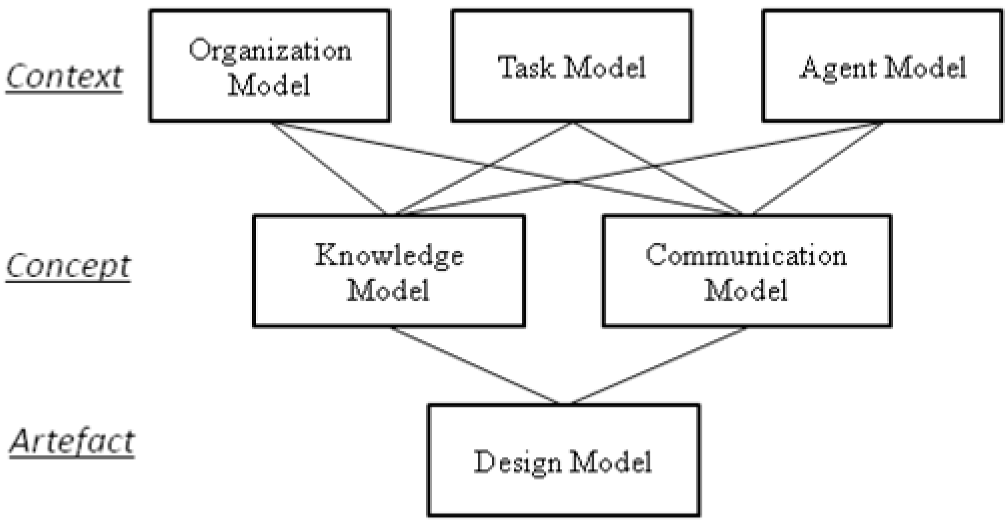

Figure 2.3: CommonKADS model set

The three levels of CommonKADS methodology are:

- Context Level analyzes the organizational environment and the corresponding critical success factors for a knowledge system. In other words, the model and functionalities in this level interact with the knowledge source (i.e. data, information in designated organization).
- Concept Level yields the conceptual description of problem-solving functions and data that were handled and delivered by a knowledge system. The models and functionalities in this level manipulate the knowledge, which is acquired from knowledge source.
- Artifact level integrates the above levels together in the design model in order to construct the requirements specification for the knowledge system. In original CommonKADS methodology, this level is not specifically defined.

CommonKADS is a complete methodological framework for the development of a Knowledge Based System (KBS) [Schreiber et al. 99]. It supports most aspects of a KBS development project, such as:

- Project management
- Organizational analysis (including problem/opportunity identification)
- Knowledge acquisition (including initial project scoping)
- Knowledge analysis and modeling
- Capture of user requirements
- Analysis of system integration issues
- Knowledge system design

We have proposed the KM methodology [Teng et al. 10], which is an extension of CommonKADS. We have introduced the Product-Process-Organization (PPO) concept in our proposed KM methodology, in order to acquire production related knowledge. In general, comparing to the three levels of the original CommonKADS methodology, the approaches of the methodology are divided into 4 sequential modules (levels):

- Context Level: this is the knowledge identification phase. In this phase, we must capture a corporate organization structure and strategy of digital preservation in enterprise. Then simultaneously, we identify the business processes and product knowledge. Here the “product” not only refers to the product in the manufacturing industry, but any kind of product, which uses an information system and digital descriptions. We have introduced PPO concept in our KM methodology, thus in Context Level, we have to modify the original models into Organization Model, Task Model and Product Model, in order to construct the models of knowledge acquisition dedicated to product and production related information. Therefore, we added a Product Model. The original Agent Model is canceled in our proposal because we don’t require too much information of the operators in production workshop.

- Concept Level: this is the knowledge integration phase. In this phase, knowledge is manipulated according to the knowledge models and templates we have defined. The models and templates are defined in the perspective of long term preservation. The knowledge captured in Context Level will have to be formalized in order to archive. As our perspective of this research work is long term preservation, we have to consider the digital preservation issue even we are developing the KM methodology. In Concept level, we redesign the Knowledge Model to fit for digital preservation. We also propose a Transformation Model, which is used for constructing and transferring knowledge between different models. The original Communication Model is canceled in our proposal because we have already canceled the Agent Model and Communication Model describes the communication plan between agents.
- Design Level and Implementation Level: In Design Level, we have to synthesize the results of the previous phases, and we design the architecture of long term knowledge preservation from the business point of view. Inside of this architecture, we have to identify and design functionalities and sub-functionalities for each business process. In Implementation Level, we implement the previous designs. The reason we add design and implementation model is that the original CommonKADS methodology provides little instruction on its Artefact Level, but in our research work, we need to propose an architecture and functional designs for long term preservation. Therefore, we replaced the original Artefact Level by Design Level and Implementation Level, in order to provide more detailed development instruction for deploying the KM methodology.

The general view of our extended CommonKADS methodology is shown in Figure 2.4. We have marked the models that are modified or added.

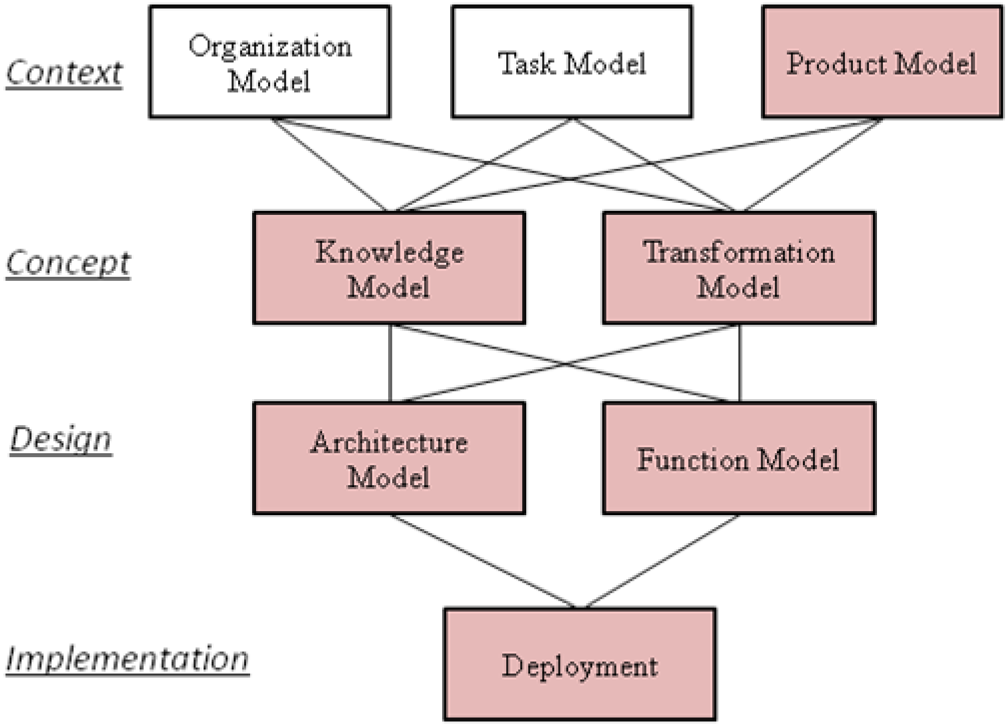

Figure 2.4: Proposed KM methodology for long term knowledge preservation [Teng et al. 10]

From Figure 2.4, we notice that, compared to the original CommonKADS, some models are replaced by new models. Even for the models that are not replaced, the detailed identification and design of models will be different from CommonKADS. The adaptation and extension of CommonKADS can be considered as:

- Adaptation of the general structure of CommonKADS, or in other words, the way that CommonKADS manages knowledge related projects.
- Adaptation of certain models (e.g. Organization Model), and extension of some models (e.g. Knowledge Model). The extension means that the model may be similar to the original model, but some features are changed by our needs.
- Introduction of new levels and models (e.g. Architecture Model).

In our research work, we work on the long term knowledge preservation with production information. This is also the reason why we have introduced the PPO (Product-Process-Organization) model in the Context Level of our proposal, although the "product" can be extended to other products rather than manufacturing products. However, in this thesis, we use production information systems for our research work. In this chapter, we introduce the utilization of each level and model of our proposed KM methodology. The detailed design of architecture and functionalities will be introduced in **Chapter 4**. And the implementation of our design will be introduced in **Chapter 5**.

2.3.1. Context Level Design

In production information systems, the management and maintenance of product structure is one of the most important functions in the whole system [Saaksvuori 08]. For example, modern PLM (Product Lifecycle Management) systems handle several product structures for the same product from different viewpoints (e.g. the product structure is different when considered from the engineering point of view than a manufacturing point of view). Thus, we propose to formalize the product related knowledge by mapping from data in production information systems to the PPO model. The reason why we have had this proposal is that data are not always managed by using PPO model in information systems in enterprise. In our proposed methodology, the modeling process is supported by the following models/documents:

- Organization Model (OM) is the scope and feasibility study which describes and analyzes a broader organizational environment: 5 worksheets. The Organization Model represents the organization environment for knowledge management in an enterprise;
- Task Model (TM) focuses on tasks (sub-processes, breakdown of business processes) of PLM and identifies the information systems that operate the tasks: 2 worksheets. The Task Model represents the real tasks performed in enterprise, not the KM tasks;
- Product Model (PM) collects all product related information: 1 document package, whose format depends on the product engineering data representation and exchanging implementation methods.

In a long term knowledge preservation context, some original CommonKADS models will not be promptly applied. And we must redesign the context models.

2.3.1.1. Organization Model

Organization Model supports the analysis of the major features of an organization, in order to discover problems and opportunities for the knowledge system, establish their feasibility, and assess the impacts on the organization of intended knowledge actions. In a long term knowledge preservation context, Organization Level is one significant aspect in PPO model design; therefore, we keep the Organization Model of CommonKADS in the model set.

The worksheets OM-1 to OM-5 are used for interviewing knowledge decision makers, who are responsible for a KM approach, in organizations, or analyze the information system organizational structure in the enterprise. Then, the outputs from the model are the list of the knowledge intensive processes and product knowledge assets which are related to each process. Finally, the feasibility of the knowledge management project was analyzed to see if the project was feasible in terms of business, technique, project and solution. It serves as a decision support for the study of business, technical and project feasibilities, in order to select the most promising focused area and targeted solution [Sureephong 09]. The five worksheets are show in Figure 2.5 (OM-1 to OM-5).

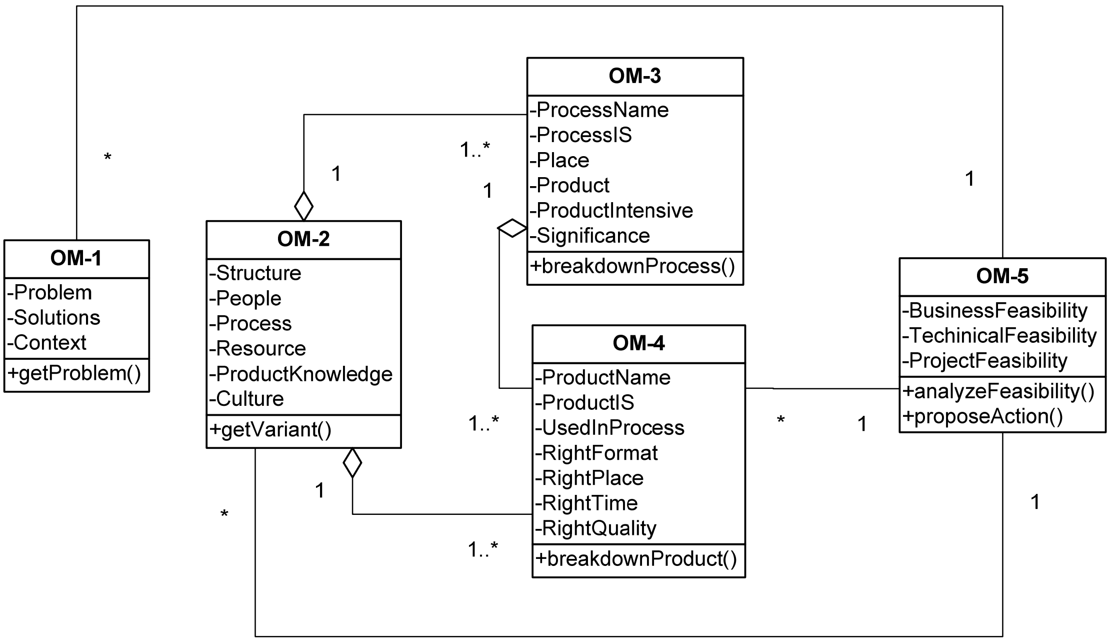

Figure 2.5: Organization Model worksheets

In general, the OM worksheets focus on the products, respecting the CommonKADS methodology.

- OM-1 analyzes the KM project problem (e.g. long term digital preservation) and organizational context (i.e. enterprise strategy, goals, missions, and important external factors, etc.), and lists suitable solutions and technologies that could be adapted.
- OM-2 represents a single problem solution of OM-1, and contains information regarding the organizational structure, business process, product and knowledge.

- OM-3 identifies the business process, and breaks down the process, which is concerned in OM-2, into tasks (sub-processes). Simultaneously, OM-3 identifies the corresponding products and information systems of the tasks.
- OM-4 identifies each product, introduced in OM-2 and OM-3. OM-4 specifies the corresponding information system as the source of product knowledge.
- OM-5 is a decision-making support document, which summarizes the worksheets from OM-1 to OM-4, and focuses on business, technical and project feasibilities. Then it proposes actions, risks and constraints of the KM project.

2.3.1.2. Task Model

For enterprises and industries, an archive must capture all the data required to completely define the product, and in some instances, processes [Lubell et al. 09]. Task Models are the relevant subparts of a business process. The Task Model analyzes the global task layout, its input and outputs, preconditions and performance criteria, as well as needed resources and competences skills. We reform the Task Model by reducing the knowledge item description and adding the modeling of executors of processes. The Task Model can be considered as a process model or sub-process model, corresponding to the PPO design.

The Task Model is a refinement of knowledge intensive tasks identified in the Organization Level. To investigate a task, three viewpoints are concerned in this model. The functional view divides a task into subtasks: input and output. The static information structure view is a description of the information content and structure of objects that are handled in the task. The control view (or dynamic view) provides understanding about triggering events, decision-making points, and other knowledge about the time aspects. The two worksheets are shown in Figure 2.6 (TM-1, TM-2).

- TM-1 aims at refining the task within the target process. The three views (i.e. functional view, static information structure view and control view) of tasks are addressed by this worksheet.
- TM-2 is a specification of the information system or sub-functions of an information system, where the target task performs. This worksheet, which concerns information system, is quite different from the original CommonKADS methodology, because we propose a methodology that is dedicated to production related digital preservation.

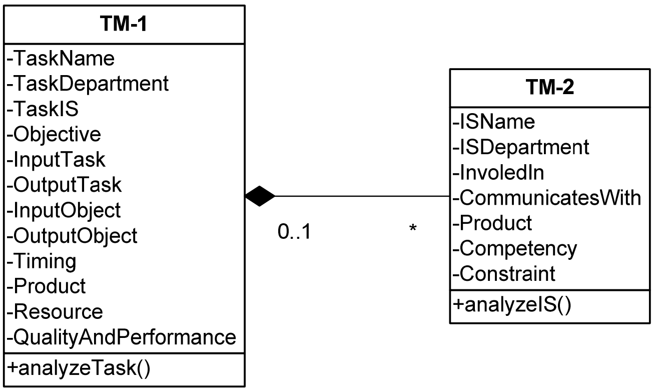

Figure 2.6: Task Model worksheets

2.3.1.3. Product Model

The PM-1 document is a specification of the product knowledge employed for a task, and possible bottlenecks and areas for improvement. In information systems of production, the Product Model is obvious and easy-to-captured. In fact in most information systems, there are already specific data structures for product, especially in PDM (product data management) or PLM systems. The composition of Product Models will cover the following aspects: component, function, behavior, structure, interface, specification and metadata, but will not be limited to these aspects. This worksheet structure depends on the product data structure in information systems.

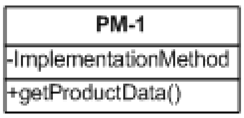

Figure 2.7: Product Model

Generally, the Context Level models collect information from information systems, which are the knowledge source and the results of the Context Level models (especially Organization Model) determine the feasibility of the long term knowledge preservation research. We have identified the correlations between the models in Context Level (Figure 2.8). Task Model is the decomposition of the OM-3, and Product Model is the decomposition of OM-4. Some parts of our proposed structure are from the original CommonKADS methodology (e.g. correlation between OM-1 and OM-5), but the detailed worksheet of each model and more other correlations between OM, TM and PM are proposed according to PPO design concept.

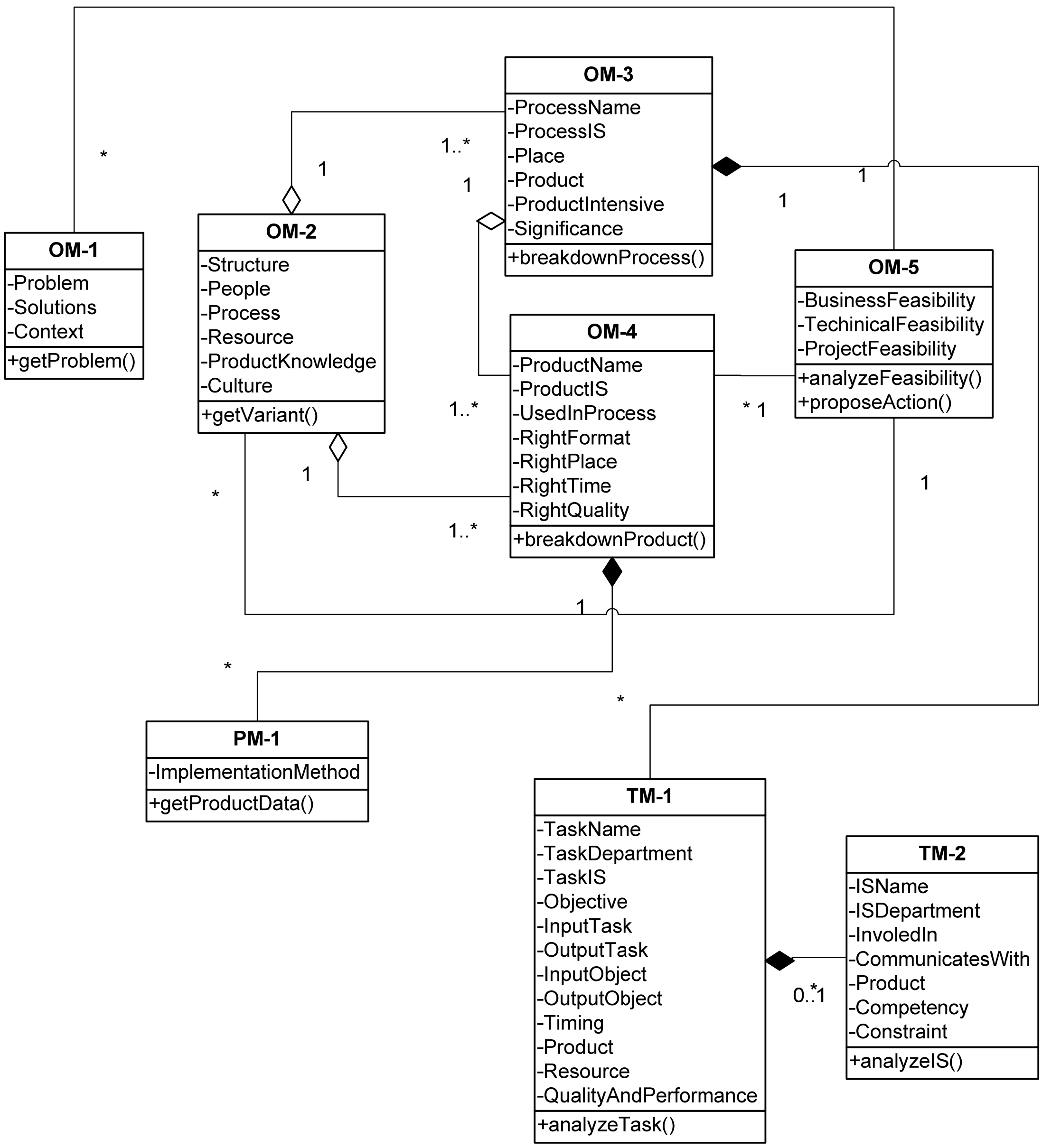

Figure 2.8: Context models: Organization Model worksheets, Task Model worksheets and Product Model

2.3.2. Concept Level Design

The worksheets in the Context Level act as checklist and information archive, and they should be used with flexibility. In the Concept Level we have to identify models, which are used for manipulating the information archive.

In order to establish the digital preservation platform, for knowledge archiving, we have adapted the OAIS reference model. Thus, in the digital preservation platform, knowledge appears as the form of Information Package (IP), which is the knowledge model provided in OAIS. Therefore, the extracted knowledge from information systems will finally be packaged in the Information Package form for long term knowledge preservation. In other words, the Knowledge Model will be constructed according to the Information Package structure.

2.3.2.1. Knowledge Model

The knowledge model is the formalization of the knowledge, which is acquired from Context Level. In other words, the knowledge model is a comprehensive knowledge object. Our proposed KM methodology targets long term digital preservation, thus the knowledge model is structured according to the digital preservation aspects. We have adapted the OAIS reference model to establish a digital preservation platform. Consequently, we adapt the knowledge object structure in OAIS, and the Knowledge Model in our methodology is an extension of Information Package (IP) model from OAIS. IP is a package containing data objects as well as comprehensive metadata to describe the data objects (Figure 2.9).

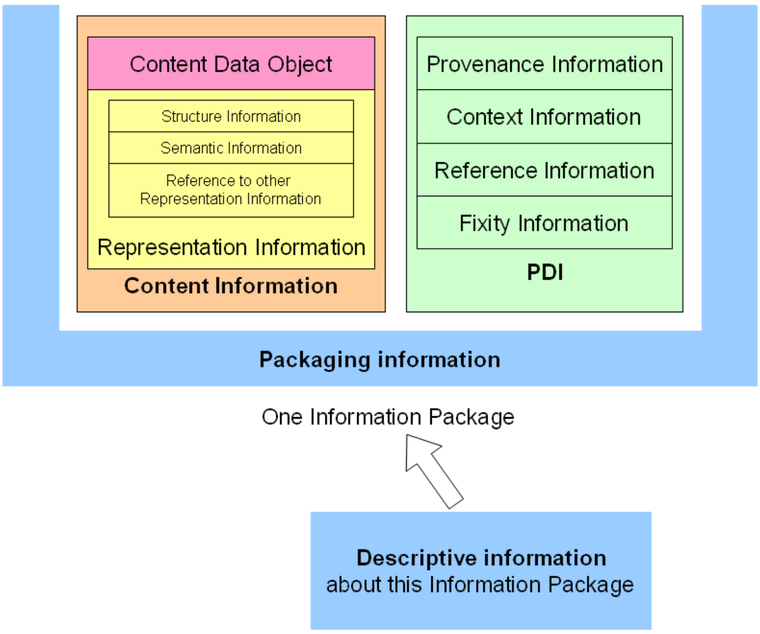

Figure 2.9: Information Package concepts and relationships [CCSDS 650.0-B-1; ISO 14721:2003]

The detailed explanations of the concepts of IP are illustrated as follows:

- Content Data Object: Physical Objects or Digital Objects.

- Representation Information: Information that makes the Content Data Object understandable to the Designated Community.
- Structure Information: Information to describe structure or data format.
- Semantic Information: Semantic description, varied and complex.
- Reference to the other Representation Information: When Representation Information itself is an Information Object, it needs other Representation Information to explain its own Digital Objects.
- Preservation Description Information (PDI): Information to preserve the Content Information, to ensure it is clearly identified, and to understand the environment in which the Content Information was created.
- Provenance Information: Source and history of Content Information.
- Context Information: Relationship of Content Information and other information outside the package.
- Reference Information: Identifiers or systems of identifiers to identify the Content Information uniquely.
- Fixity Information: Wrapper of protective shield to protect the Content Information from undocumented alteration.
- Package Information: Information which, either actually or logically, binds, identifies and relates the Content Information and PDI.
- Descriptive Information: Information to discover which package has the Content Information of interest.

The product data will be preserved as data objects, while the IP itself is task-oriented. In Descriptive Information of the IP, the usage of the IP (i.e. task objective of IP) is described based on the analysis in Task Model. Thus the end users of the preserved IP will locate the corresponding knowledge by their working requirements. For example, if one end user tries to find information about core design of a power transformer in the critical design phase, his/her task objective is “core design”. And by searching Descriptive Information with this task objective, all the corresponding IPs would be located. Figure 2.7 is just a logical view of the information package, and Figure 2.10 shows the class diagram of the knowledge model: Information Package.

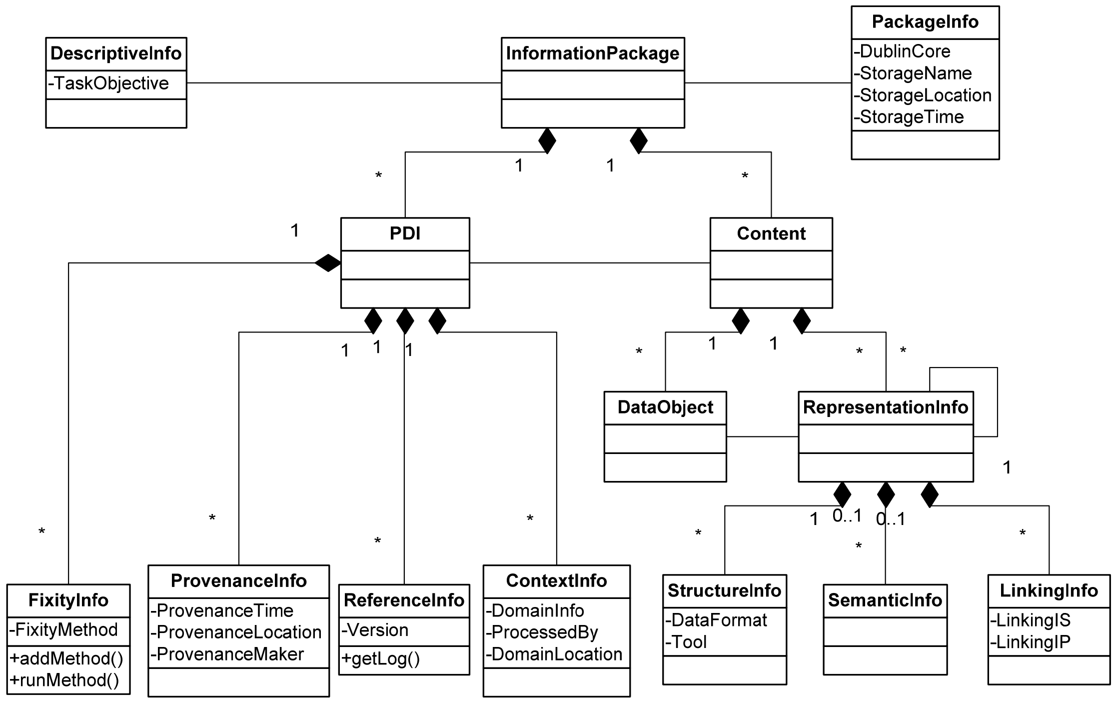

Figure 2.10: Knowledge Model: Information Package

2.3.2.2. Transformation Model

The Transformation Model in our KM methodology focuses on knowledge model-transfer, in other words, the knowledge mapping from one sort of model to another. The knowledge creation process in Context Level is in a top-down strategy. Thus, either Task Model or Product Model is identified in Organization Model. And in this manner, it is not too difficult to establish links to merge the Context models (Figure 2.6). Then the outcome of the Context Level models and worksheets will provide not only a clear idea on the target knowledge for long term preservation, but also a comprehensive knowledge of the organization, business process and product engineering. In other words, the Transformation Model includes the mapping from context model sets to formalized knowledge model. However, the knowledge transfer process is not always an integration process. For instance, in knowledge reuse process, specific information system may require just product knowledge, which needs to be separated from the content in Knowledge Model.

In digital preservation platforms, the worksheets are easy to convert from or to XML (Extensible Markup Language) format. We use XML as one example to explain how to identify the communication mapping.

The Product Model document may also be represented in XML. By using XML, we integrate the product knowledge with organization and process knowledge, in order to capture comprehensive metadata for long term preservation. According to the KM approach and worksheets structure, we have developed our XML Schema for generating XML files from worksheets. We provide these XML files as the knowledge source of construction of knowledge model. Figure 2.11 shows the mapping from Context models to knowledge model.

The knowledge model mapping (conversion) in Figure 2.11 is from the left side to the right side. According to this mapping, not all the worksheets are converted into XML. OM-1 and OM-5 are decision-making documents in our KM project, while other six documents are the source knowledge for integration. Thus, the source knowledge documents are converted into XML files and are put in XML database, waiting for more operations. Based on the IP structure, we query corresponding information and compose the information as the components in IP. All the components are generated into two XML files (i.e. Preservation Description Information and Representation Information), which is the metadata of the product data object. And at the same time, the product data object is also saved. The linking information of product data and metadata is described in Packing Information. Simultaneously, the Descriptive Information is generated, too.

Figure 2.11 just shows one example of knowledge model mapping. In fact, the knowledge model mapping (conversion) rules includes certain terms and scenarios that will decide when and how the knowledge model should convert from one to another.

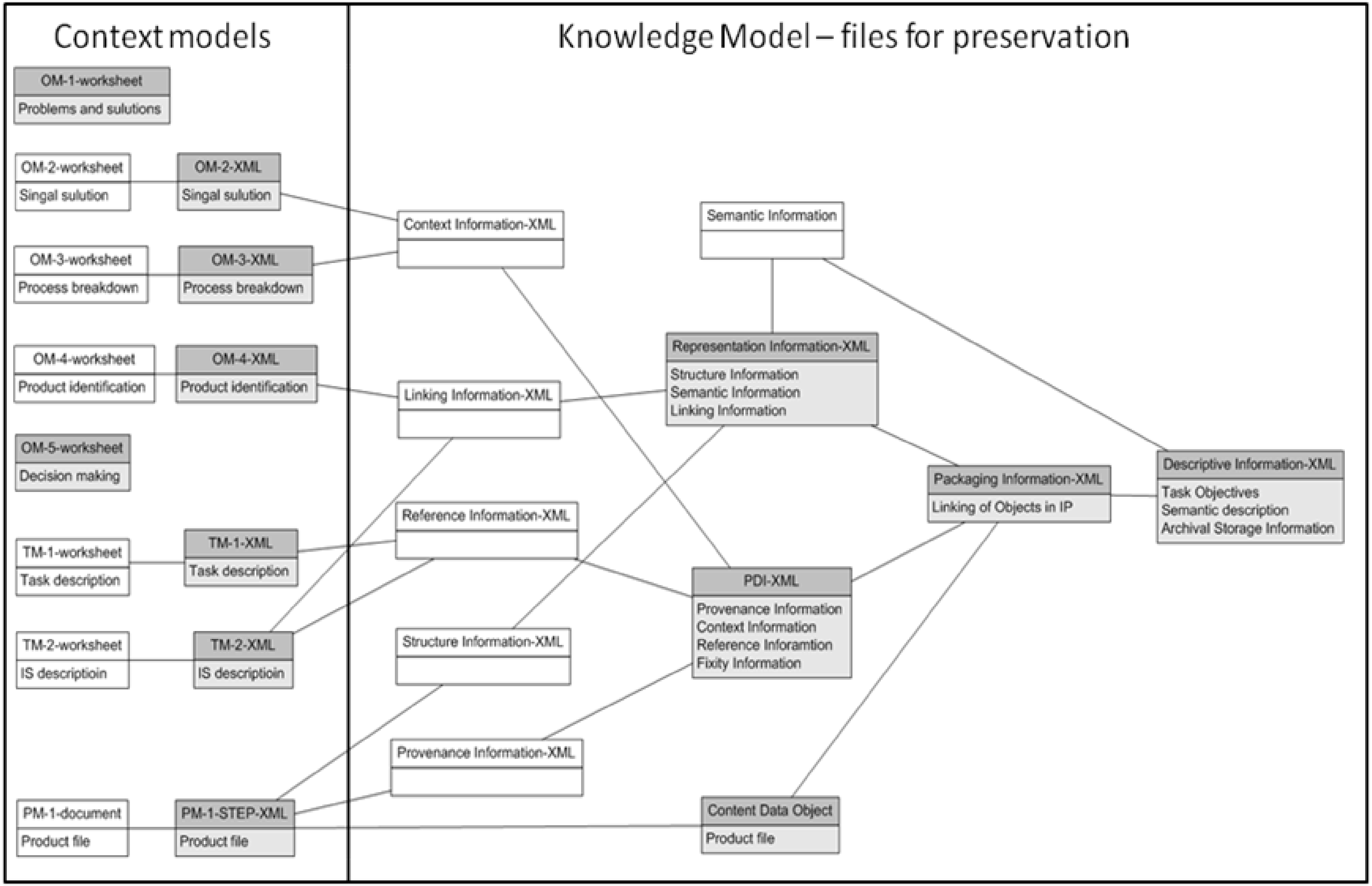

Figure 2.11: Transformation Model: mapping from Context models to Knowledge Model

2.3.3. Discussion of Design and Implementation Levels

It is argued that, the Artefact level (i.e., the design models) was lightly defined in the CommonKADS methodology. Hence, in our research work, software engineering concept is introduced to extend the CommonKADS and two levels (Design Level and Implementation Level) are proposed to replace the Artefact level.

In Design Level, Architecture Model should be developed. Not only the structure of long term knowledge preservation is established, but also specifications of required functionalities and services are produced in Architecture Model. Therefore, the Architecture Model and Function Model include not only the KM activities, but also the knowledge preservation activities. We will introduce the architecture and the required functional designs after the discussion of digital preservation platform (**Chapter 3**), where knowledge preservation is performed. The Design Level (i.e. Architecture Model and Functional Model) will be introduced in introduced in **Chapter 4**. We have noticed that in our research work, the extended CommonKADS methodology is the direction and blueprint for not just KM approach, but for the whole long term knowledge preservation approach. Nevertheless, during the discussion from **Chapter 2** to **Chapter 3**, we will see some general design of Function Model, and in **Chapter 4**, they are discussed in details.

The last level is Implementation Level is the deployment of KM architecture and models. The level translates the architecture and models into software programs. The Deployment Model depends on the models we have discussed above as well as the software platform. The research on digital preservation platform is one other major work of our research. Although the CommonKADS has this hierarchic structure, the Architecture Model or service model design depends on, to certain extent, the selection of digital preservation platform. The Implementation Level is the lowest in the hierarchic structure, but the Design Level needs the feedback from this level for architecture and service design. In fact the implementations of KM approach and digital preservation approach ought to be integrated, because they will not realize long term knowledge preservation by either one alone. Thus, we will introduce the implementation approaches in detail in **Chapter 5**.

2.3.4. Model Sets in KM Approach

We have already described the model sets of our CommonKADS methodology. These models are the basics of our research work. These predefined models will be used in each step of the long term preservation processes. In Section 2.2.2, we have already discussed the major steps of KM approach. Accordingly, in our research work, we have identified the functions in the processes of knowledge acquisition and knowledge reuse in enterprise. Table 2.4 shows the general description of the Function Model of the KM approach in our research work. In Table 2.4, we notice that the Context Level models are required to associate with each function or sub-function. In other words, the functional modules or KM approach depend on the Context Level models.

General Info.	Function	Sub-function	Associated Model
KM Approach	KM Planning	Develop KM Strategy	-
		Capture Critical Change	-
		Design Knowledge Model	Organization Model
			Task Model
			Product Model
	Knowledge Acquisition	Capture Organizational Info.	Organization Model
		Identify Business Process	Organization Model
			Task Model
		Idnetify Product Info.	Organization Model
			Product Model
	Knowledge Evaluation	Handle Query	Organization Model
			Task Model
			Product Model
		Determine Knowledge Obsolescence	-
	Knowledge Reuse	Create Query	Organization Model
			Task Model
			Product Model
		Distribute Knowledge	-

Table 2.4: Function Model of KM approach

The functions related to KM approach in enterprise:

- KM Planning: the KM planning is the decision making part of KM. According to different environments (e.g. different information systems used in different departments, etc.), different cultural issues would be captured, and consequently the KM objectives and KM methods would be determined (sub-function: Develop KM strategy). This function also designs the structure of the extracted knowledge from information systems and domain experts (sub-function: Design Knowledge Model) and identifies the changes (sub-function: Capture Critical Change) that would trigger the knowledge changes in digital preservation platform.

- Knowledge Acquisition: this function extracts data and knowledge from information systems or from domain experts, according to the knowledge model, which has been designed in KM planning function. And the packaged knowledge is transferred into the digital preservation platform.
- Knowledge Evaluation: the Enterprise Layer also receives queries when knowledge repository performs the knowledge obsolescence check. Knowledge evaluation function handles this kind of queries and makes decision whether the preserved knowledge should be kept or not, according to the query results. This function aims to reduce the redundancy of the preservation.
- Knowledge Reuse: this function plays as a contrary role as the Knowledge Acquisition. Firstly, this function sends knowledge reuse request to the knowledge repository. Then it gets the responses and results from the knowledge repository, whether appropriate knowledge is retrieved or not. If the appropriate knowledge is sent back, this function sends the knowledge back to information system or domain expert for reuse.

2.4. Conclusion

We have presented, in this chapter, the KM methodology for our research work.

We develop our KM methodology by extending the CommonKADS methodology. And our proposed methodology has a multiple levels structure and top-down approach, with several model sets. We have introduced each part of our proposed KM methodology level by level in this chapter. The methodology describes from the KM approaches in knowledge sources context, to knowledge modeling and communication activities among information systems. Although we have identified the functionalities of KM approach, the detailed functional design would be discussed in Chapter 5 after the presentation of the global architecture of our long term knowledge preservation research work.

Chapter 3. Platform for Digital Preservation

3.1. Introduction

Knowledge objects will be produced by going through the KM approach, which have been introduced in **Chapter 2**. However, our core problem is to preserve these knowledge objects in the long term. Therefore, we need to establish a digital preservation platform for knowledge retention. As we have already done the analysis on existing digital preservation platforms, we have identified existing features of these platforms. In this chapter, based on the existing repositories' functionalities and the OAIS reference model, we propose a design of an OAIS based digital preservation platform. In this chapter, we are introducing the functional design of our digital preservation platform, by adding features on the original OAIS functional models. The supplementary functions are the "missing" parts of the long term preservation requirement, as we have studied in **Chapter 1**. We intend to add functions and features to overcome the gaps between existing features and the long term preservation requirements.

3.2. Evaluation of Digital Preservation Platforms

There are a number of open source solutions that provide the capability to store, manage, re-use and curate digital materials. Digital preservation platform holds multiple functions which can be custom developed or extended.

In this research work, our objective is long term knowledge preservation, and we have to thoroughly analyze the knowledge retention tools and functionalities. Through the functional analysis on existing digital preservation platforms, we have studied several open source digital preservation platforms. From these platforms, we have chosen the most popular open source digital preservation platforms: DSpace, Fedora repository and EPrints, in our research. The three platforms are described briefly:

- DSpace is an open source repository solution that provides the tools for the management of digital assets. It supports the archiving of a wide variety of data including books, theses, 3D digital scanned objects, photographs, films, videos, research data sets and other form of contents. DSpace was developed as joint effort between Massachusetts Institute of Technology (MIT) Libraries and Hewlett-Packard (HP). It is freely available to all organizations under BSD (Berkeley Software Distribution) open source license. DSpace is written in Java and JSP (JavaServer Pages), using the Java Servlet API (application programming interface). It uses a relational database and supports the use of PostgreSQL and Oracle database. It makes its holdings available primarily via a web interface, but it also supports OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) v2.0, and is capable of exporting METS (Metadata Encoding and Transmission Standard) packages. DSpace supports multiple types of storage devices through a light weight storage API. The storage layer currently provides support for local file systems, Storage Resource Broker (SRB), Amazon S3 or Sun SAM/QFS.[Kesavan 09]
- Fedora (Flexible Extensible Digital Object Repository Architecture) repository is a modular architecture built on the principle that interoperability and extensibility is best achieved by integrating data, interfaces and mechanisms clearly defined as modules. Fedora repository provides a general purpose management layer of digital objects. Object management is based on content models that represent data objects or a collection of data objects. Fedora supports two types of access services: a management client for ingest, maintenance and export of objects or API hooks for customized web-based access services built on either HTTP (Hypertext Transfer Protocol) or SOAP (Simple Object Access Protocol). Fedora supports ingest and export of digital objects in a variety of XML formats. This makes it possible to have interchanges between Fedora repository and other XML-based applications, thus facilitating archiving tasks. Fedora Digital Object Model allows tight management of metadata and digital content, regardless of format. The platform is scalable and flexible allowing Fedora to associate objects with external or distributed repositories. Fedora's architecture is based on four APIs: manage, access, search and OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting). Various applications like Muradora, Vital etc., can be implemented as front end layer over Fedora. Fedora is the first open source repository designed to

work as part of an extensible framework of service components. This extensible framework also allows Fedora to support trusted and secure organizational repository needs. Fedora is developed by the Fedora Commons non-profit organization as a platform for providing sustainable technologies to create, manage, publish, share and preserve a digital content as a basis for intellectual, organizational, scientific and cultural heritage.

- EPrints is an open source software package for building open access repositories that are compliant with OAI-PMH. It shares many of the features commonly used in Document Management Systems, but is primarily used for institutional repositories and scientific journals. EPrints has been developed at the University of Southampton School of Electronics and Computer Science and is released under GPL license. EPrints is a web and command-line application based on LAMP architecture (mostly written using Perl). It has been successfully run under Linux, Solaris and MAC OS X[Kesavan 09].

EPrints puts a particular emphasis on OA content (preprints and postprints of institutional research output, plus theses), DSpace on digital curation in general, and Fedora repository describes itself as storage layer software requiring custom front-ends for any purpose. And in fact, some researchers have already set up the collaboration between DSpace and Fedora repository, so in our research work, we have carried out a number of work to collaborate the open sources solutions to get their features for our research work.

Here we present our key purposes for the evaluation of digital preservation platforms:

- To find out the design, architecture and the implementation details of short listed candidates.
- To understand the functionalities of the repositories according to the given criteria.
- To employ the members of the open source repository where needed.

In the purpose of evaluation, we noticed that the criteria will lead the evaluation. Many needs can be characterized by the functional requirements. However, others, which relate to extensibility of the software, sustainability of the user/developer community and usability, are equally important. Sometimes these needs may be conflicting to each other resulting in difficulties in finding out coherent criteria for the evaluation of the platform. Anyhow, we have identified the following important criteria:

- Ease of system deployment and testing;

- Versioning: versioning allows the repository to keep older versions of metadata and files;
- Scalability: in the current context the volume of the digital archives may be small, but the balance will change over time and archives are expected to receive huge quantities of digital materials in the near future. The volume of the metadata is also expected to increase over time. The repository should scale to manage millions of digital materials. Thus the repository should have capacity to manage large quantities of materials, and support mass inputs and outputs of materials when ingesting and exporting.
- Interoperability: the interoperability with other ISs;
- Security: privacy and other security issues will be retained for a high number of years for digital archives. Security is crucial in building the confidence of potential users;
- Archiving and Database management;
- Submission: the submission process and methods are the inputs of the digital preservation platform;
- System Configuration: the ease of system configuration will reduce the difficulties in adapting the open source platform, and vice versa;
- Working with the codes;
- Archival and administrative issues: they address the strategic information on the designated organization, which will adapt the digital preservation platform;
- Globalization: the usages in different environments and cultures;
- Searching and Browsing: the basic functionality of a repository for reusing information.
- Community and support: as the platforms are open sources, the developer community and technical support team will be quite helpful and necessary for the development of a dedicated digital preservation platform.

This evaluation methodology of the repositories primarily depends on the following methods;

- Deploying and testing the platform.
- Examining the release documentation of repositories platform.
- Examining the papers from the user community i.e. from research works, reports and papers as well as wikis, institutions and projects websites.
- Analyzing the earlier comparison of the repositories.

- Analyzing the mailing list of the repository communities.

We have performed taxonomy of these existing digital preservation platform functionalities, and from different points of view, so as to identify a classification of existing digital preservation functionalities [Timsina et al. 10]. Each selected criterion is given an importance rating when evaluation is performed on different repositories. The headline criteria are also broken into sub-criteria for the transparency of the heading and each one carries this importance rating. Since each sub-criterion is marked using these ranges, this rating can be display as:

- 0: worst or/and feature does not exist.
- 0.25: carries poor support or /and can be accomplished with significant labor.
- 0.5: reasonable but needs the adaptation to reach the desired condition or/and still need some efforts for completeness.
- 0.75: good but still needs some minimal efforts.
- 1: as desired or /and needs virtually no extra efforts.

We intend to use this evaluation approach to classify different repository platforms, in order to select an efficient digital preservation platform for the implementation of the digital preservation approach. Among the different preservation platforms used around the globe, we have selected the platforms matching the following criteria as our short list of cadidates:

- Open source
- Massively used
- Strong community and support
- Development history and future forecast of the platform

At the end of the evaluation (**Annex 3**), we have identified the positive and negative points of the three open source digital preservation platforms (Table 3.1):

Criteria	**Platforms**		
	DSpace	***Fedora***	***EPrints***
System Deployment	0.67	0.5	0.62
Versioning	0.125	0.75	1
Scalability	0.58	0.84	0.67
Interoperability	0.8	0.8	0.35
Security	0.92	0.83	0.75
Archiving and Database Management	0.59	0.84	0.17
Submission	0.7	0.5	0.5

Criteria	Platforms		
	DSpace	***Fedora***	***EPrints***
System Configuration	0.57	0.57	0.5
Working with the code	0.58	0.75	0.5
Archival and Administrative Concerns	0.45	0.75	0.2
Globalization	1	0.87	0.87
Searching and Browsing	0.75	0.87	0.75
Community and Support	0.88	0.81	0.56
As a total	0.71	0.75	0.58

Table 3.1: Evaluation of the open source digital preservation platforms

Through the evaluation results, we have learnt that the three open source platforms are focused on and clever in these various aspects:

- DSpace: it can be used within the a medium scale organization or institution. The overall glance of DSpace is good, but due to the scalability issue it cannot be recommended for the national platform where the number of objects may exceed millions. Furthermore, this scalability issue cannot be solved with simple efforts, and major works of programming and re-writing are required. Besides, another weakness is the versioning but it can be maintain by locally writing the module. The strength lies on the interoperability and the security of the platform. Indeed, it has inbuilt roles and an access policy. Basically, DSpace is the out of box platform which can be best suited for the medium and small institutions.
- Fedora: the key point is that it can be adapted to any kind of local platform. The platform can be configured to institution relevant workflows and branding. The security and the interoperation schema can be imposed as desired. More than a simple form of downloading, it also supports the additional operations inside the repository. Furthermore, it has huge scalability power. Finally Fedora is best suited to huge organizations that have a huge amount of complex data and are also able to provide enough value to overcome this issue.
- EPrints: it is best suited for the self–configuring institutions which want to build and host their own archiving platforms. They do not ensure a very tough security and interoperatibility schema, but on the overall rating they still hold a good position. The power of EPrints rests on the fact that it is a simple and out of box platform which needs very small technical and cost overhead. In other words, they provide an overall assistance for the institutions which cannot afford permanent technical staff. At last,

EPrints is most suited for small organizations which need overall common features of repositories but do not need a complex and high level inter-operation.

Unlike the commercial software tools, these open source tools enable us to access and have full operational experiences. As a result, in addition to the identification of usage of the three open source digital preservation platforms, the other outcome of the evaluation of digital preservation platforms is the classification of the functionalities that can be provided by such a platform (Table 3.2). These functionalities will be used in our following design of digital preservation platform.

No.	Function	Sub-No.	Sub-function
1	System Deployment	1.1	Software requirement
		1.2	Duplication
		1.3	Installation steps
2	Versioning	2.1	Notice of similarity of submission
		2.2	Keeping Versions
		2.3	Acquiring data and metadata from old version for new version
		2.4	Log of versioning
3	Scalability	3.1	Scale up
		3.2	Scale out
		3.3	Architecture
4	Interoperability	4.1	OAI-PMH
		4.2	SRW/SRU
		4.3	SOAP
		4.4	Bulk import and export
		4.5	Integration with other web pages
5	Security	5.1	Data transmission
		5.2	Server security
		5.3	Roles and authentication
6	Archiving and DataManagement	6.1	Archival media and database
		6.2	Storage hierarchy
		6.3	Backup and disaster recovery of archived content
7	Submission	7.1	User interface
		7.2	Authorization
		7.3	Individuation of user interface
		7.4	Submission report
		7.5	Workflow
8	System Configuration	8.1	Configuration of UI
		8.2	Configuration of system policies
		8.3	Configuring module of information package
		8.4	Configuring archival strategy
9	Working with Code	9.1	Writing plugins or other packages
		9.2	Alter the digital object type including metadata
		9.3	Documentation and understanding of code
10	Archival and Administrative Concerns	10.1	Complex inter object relationship
		10.2	Referenced metadata
		10.3	Support content model
		10.4	Realistic learning curve of system

No.	Function	Sub-No.	Sub-function
		10.5	Stability of monitoring of data and metadata
11	Globalization	11.1	Multi-language
		11.2	Unicode
12	Searching and Browsing	12.1	Searching engine
		12.2	Browser
13	Community and Support	13.1	Development community
		13.2	User community
		13.3	Supports for users

Table 3.2: Classification of digital preservation platform functionalities

3.3. Design of the Digital Preservation Platform

3.3.1. Open Archival Information System Reference Model

Since 2000, various research programs and workshops have been held all over the world. (e.g., National Digital Information Infrastructure and Preservation Program, LOTAR project, CASPAR digital preservation project, Atlantic LTKR workshop, Indo-US workshop on International Trends in Digital Preservation, etc.) These researches have been aiming to develop strategies as well as efficient methodologies and tools for digital preservation.

Due to these valuable researches, architectures, models and standards have been developed. The Open Archival Information System (OAIS) reference model has been proposed by researchers to provide a conceptual framework and a common vocabulary for digital preservation activities. OAIS is an ISO standard reference model which was developed by the NASA's (National Aeronautics and Space Administration) Consultative Committee for Space Data Systems (CCSDS),

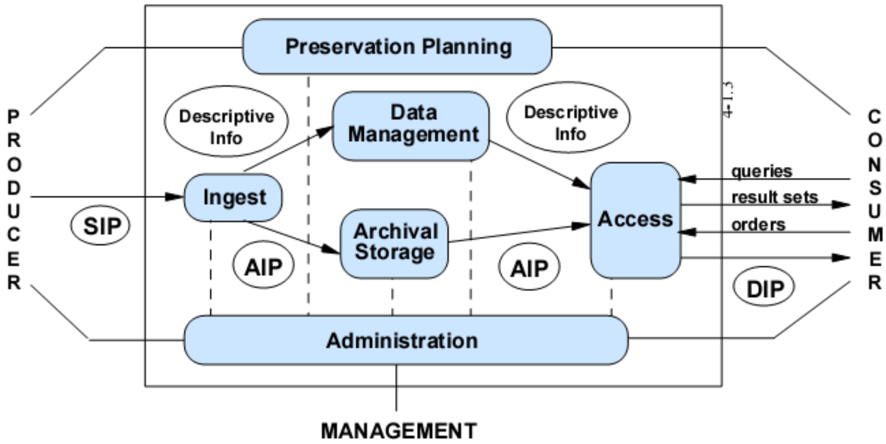

Figure 3.1: Open Archival Information System functional entities [CCSDS 650.0-B-1; ISO 14721:2003]

The entities provided by each entity in Figure 3.1 are briefly described as follows [CCSDS 650.0-B-1; ISO 14721:2003]:

- Ingest: This entity communicates with the submissions from outside of the OAIS. The services and functions provided by this entity will accept Submission Information Packages (SIPs) from Producers (or from internal elements under the control of the entity Administration). Then the submissions will be further examined and re-manipulated for preservation.
- Archival Storage: This entity aims to manage the repositories, which represents in other words, the Archival Information Packages (AIPs). The AIPs are stored and maintained in certain media.
- Data Management: This entity populates, maintains and accesses the Descriptive Information, which identifies the AIPs. The services and functions provided by this entity are similar to the functions in a DBMS (Database management system). Moreover, this entity deals with the administrative data of the OAIS, too.
- Administration: This entity is the "brain" of the whole archival system. The overall operations and communications of other entities are directed and authorized by this entity.
- Preservation Planning: This entity monitors the storage environment of the OAIS, and responds to the changes by providing recommendations for repositories. This entity's responsibility includes ensuring the accessibility of the stored information in the long

term, even when the original environment of the designated information has already changed or become obsolete.

- Access: This entity communicates with the designated users (Consumers) of the OAIS. A Consumer browses or searches through this entity, in order to determine the existence, the description and the locations of the stored information in the OAIS. And this entity allows the Consumer to demand for the information packages.

The OAIS reference model provides concepts concerning long term preservation, as well as functionalities that may support these concepts. And this is the reason why we have chosen it as our reference model for establishing a digital preservation platform, which is a major part in our proposed architecture. The detailed description and design based on OAIS will be presented in the following sections.

3.3.2. Function Models Design

The main features of the reference models of OAIS, in the long term preservation scope, focus on defining a stable environment for data storage, data management, data accessibility and data interpretation. In our previous section, we have already introduced the fact that there are six major functional entities composing an integrated OAIS. We have managed these six main functional entities for our digital preservation platform by adding or modifying the sub-functions under these entities (Table 3.3). According to the results of existing repository platforms, we have reorganized the six major functional entities of OAIS, and have added or modified some functions according to the existing functions or tools. In our extended CommonKADS methodology, the Knowledge Model is the structure of information that is for archival. Thus the Knowledge Model associates with the functions or sub-functions of the digital preservation platform (Table 3.3).

General Info.	Function	Sub-function	Associated Model
Digital Preservation Platform	Ingest	Validate Submission	Knowledge Model
		Generate AIP	Knowledge Model
		Generate Descriptive Information	Knowledge Model
	Archival Storage	Archive Data	-
		Maintain Storage	-
		Provide Data	-
		Terminate Data	-
	Data Management	Archive Data	-
		Generate Report	-
		Perform Query	-
		Filter Obsolescent Data	-

General Info.	Function	Sub-function	Associated Model
	Administration	Audit	Knowledge Model
		Configure Platform	-
		Configure Policy	-
		Authorize Data Evaluation	-
	Preservation Planning	Develop Preservation Strategy	-
		Capture Critical Change Signal	-
		Design Information Package Model	Knowledge Model
	Access	Handle Query	-
		Generate DIP	Knowledge Model
		Deliver Response	-

Table 3.3: Function Model of digital preservation platform

The functions in the digital preservation platform are related to digital preservation for the long term, and we have also extended the functions of OAIS. Our adaptation is different from other adaptations, because we consider KM approaches and functionalities in our design. We have also added sub-functions (i.e. Terminate Data, Filter Obsolescent Data, authorize Data Evaluation, Capture Critical Change Signal). However, we keep these six major functional entities:

- Ingest: Ingest function normally enables the services to accept Submission Information Packages (SIPs) from Producers (according to OAIS). We have identified the "Producers" as the information systems of production or the domain experts in KM approach; still they submit knowledge or data in their own knowledge model, which is in the form of Information Package. This function is almost identical as the original OAIS reference model.
- Archival Storage: Archival Storage provides the services and functions for storage, maintenance and retrieval of Archival Information Packages (AIPs). As we have discussed in previous section, we have to deal with the knowledge obsolescence issue in long term, thus in Archival Storage function, we have added a sub-function called Terminate Data, which terminates the preservation of certain data according to the direction from the entity Administration. Terminate Data is not provided in original OAIS reference model, but we do need to terminate certain obsolete information in digital preservation in order to reduce the usage of resources and provide better environments for the information that still has got values for end users.
- Data Management: Data Management could also be seen as a database management system, which populates and maintains Descriptive Information and administrative

data. The Descriptive Information identifies the preserved Information Package. As the Information Packages are encapsulated, the Descriptive Information is the only available information for retrieval of the Information Packages. Also, Data Management generates reports to log the events in digital preservation platform. Then this entity filters the Information Packages, which are not used in the designated long term, according to the logs and reports. Regarding to data termination, we also add a sub-function called Filter Obsolescent Data into the original OAIS reference model.

- Administration: Administration provides services for the overall operation of the digital preservation platform. It audits the Information Packages according to the predefined Knowledge Model. The audit process is required whenever there are transfers of Information Packages. Administration configures the archival policies and the technical issues of the digital preservation platform. Another important function of Administration is that it controls the authorization of data access and data updating. When Data Management tries to send queries to perform Knowledge Evaluation in a KM approach, sending queries must be authorized by the entity Administration.
- Preservation Planning: Preservation Planning develops the basic preservation strategy and Knowledge Model for the digital preservation platform. Besides, it catches the critical change signals from the KM approach and notifies the knowledge update process.
- Access: Access function provides services to support Knowledge Consumer, which is the role of "Consumer" in OAIS reference model. The basic services are required by Access: handle queries, generate DIPs, and deliver responses.

3.4. Conclusion

We have adapted the OAIS reference model in development of our digital preservation platform. Based on the functional entities and our analysis of existing functions of repositories, we identify the functionalities of digital preservation platform. According to these functions, we propose a digital preservation approach, which considers the KM approach in term of "Producer" and "Consumer". During the adaptation, we have also added some features (e.g. Terminate Data, Filter Obsolescent Data, authorize Data Evaluation, etc.) according to the long term preservation requirements. Moreover, we have separated the whole OAIS functional models into four parts, according to the different data objects.

Chapter 4. MadPK: An Integrative Architecture for Digital Preservation

4.1. Introduction

In this chapter, we construct an architecture to connect the KM approach and the digital preservation approach in a dynamical way. We propose a multi-layer architecture under the concept of web services. Thus, the interoperabilities of the business processes in various approaches are ensured. Besides, the interoperabilities of the knowledge objects (digital objects) are also ensured.

4.2. Proposed Architecture for Digital Preservation

Our research work includes the KM approach as well as the digital preservation approach. The KM approach identifies enterprise knowledge and produces digital knowledge objects. The digital preservation platform performs the whole digital preservation process. These two parts of long term knowledge preservation are both necessary and important. However, the KM environments are complex and diversified in enterprises, and compared to them, the digital preservation environment lasts longer than KM. Therefore, in order to achieve dynamic preservation to connect the two major parts, we need to develop an architecture for long term knowledge preservation. This architecture should allow the two parts to perform smoothly in their own environments, and allow the communication and connections between them to be accessible and dynamic. As we have introduced in **Chapter 2**, this architecture appear as the Architecture Model of our extended CommanKADS methodology.

4.2.1. Multi-layer Architecture for Dynamic Preservation of Knowledge – MadPK

The KM approach must selectively acquire comprehensive knowledge for long term preservation. On one hand, we need to adapt the ways of KM and develop a methodology, which is based on KM methodologies, for long term knowledge preservation. On the other hand, digital preservation needs a platform for the repository of knowledge (i.e. digital content). Digital preservation has correlations with KM, but presents a very different context compared to data mining. In order to achieve dynamic preservation and interoperability between knowledge sources and digital preservation platforms, we intend to construct a universal architecture for long term digital preservation.

There are plenty of projects in Europe which work on digital preservation. In the researches and projects of digital preservation, the KM approach is not usually integrated. Meanwhile, although the long term knowledge preservation research issues have been noticed by researchers after the year 2000, the existing researches and projects on long term preservation focus on specific field (mainly on production field). In their approaches, the confinement of "knowledge" narrows to product's engineering data (i.e. 3D, CAD data, etc.). The core work of these existing long term preservation researches and projects is the development of standards or approaches (e.g. STEP, OntoSTEP [Krima 09], etc.), which is not a fully KM process. The digital preservation platform is just one carrier or one box for the "knowledge". These researches and projects' critical work focus on knowledge identification, acquisition and encapsulation, and they integrate the digital preservation platform during the knowledge retention (engineering data retention) process.

In real life, there are different KM approaches in enterprises. Data acquisition in different information systems is also involved in different processes and adapted to different standards or principles. In some departments, there is even an integrated Knowledge Management System (KMS) dealing with the acquisition of knowledge. On the other hand, KM and digital preservation have different purposes. The data objects in KM (i.e. knowledge) are not always well structured for digital preservation, and vice versa. If we focus on just certain departments such as the previous long term preservation projects (e.g. manufacturing data), we have to re-define models, processes or systems in both KM and digital preservation. Nevertheless, this way narrows the utility of the long term preservation approach and lacks dynamic features.

Therefore, we propose a more adaptable architecture for long term preservation – Multi-layer Architecture for Dynamic Preservation of Knowledge (MadPK), which separates KM and digital preservation. This architecture will allow us to keep the existing various KM approaches or systems, while providing dynamically connected digital preservation platform. The MadPK architecture is shown in Figure 4.1. Our proposed architecture is different from any other existing digital preservation architecture. And we use this "separated-but-connected" structure of MadPK to benefit from existing technologies and models of both KM and digital preservation. The MadPK architecture is proposed for establishing the dynamic connections and communications. The layers in our MadPK are:

- Enterprise Layer: KM approaches among information systems and domain experts in a designated organization. This layer handles "knowledge source" (i.e. the information systems and people in enterprise) and produces knowledge from the enterprise business processes. There is more than one KM approach in this layer.
- Digital Preservation Layer: digital preservation platform. This layer captures, packages, transfers, stores knowledge and provides retrieving features for knowledge reuse. One of the responsibilities of the digital preservation platform is to ensure that the preserved information is in an independently understandable and reusable form to the end users in the long term. We adapt the OAIS reference model in our proposed preservation platform. Knowledge in this platform is formed on the basis of a formal Knowledge Model (i.e. Information Package of OAIS).
- Mediation Layer: integration methods and tools for digital preservation. This layer connects the other two layers, in a dynamic perspective. The data transferred between the two other layers are digital knowledge objects. So we name this layer as an enterprise knowledge bus (EKB), which enables the other two layers to call and acquire corresponding knowledge. Although this layer is named EKB, in the same way as the Enterprise Service Bus (ESB), more content and missions will be introduced in this layer (e.g. knowledge structure modeling, definition of knowledge transfer rules, storage and management of unfulfilled digital content for digital preservation, etc.). Therefore, our architecture does not only involve two layers (KM and digital preservation).The third layer (EKB) is the key to support the other layers' interoperabilities as well as enable them to adapt complete existing models with technologies and without worrying about the interoperability issues.

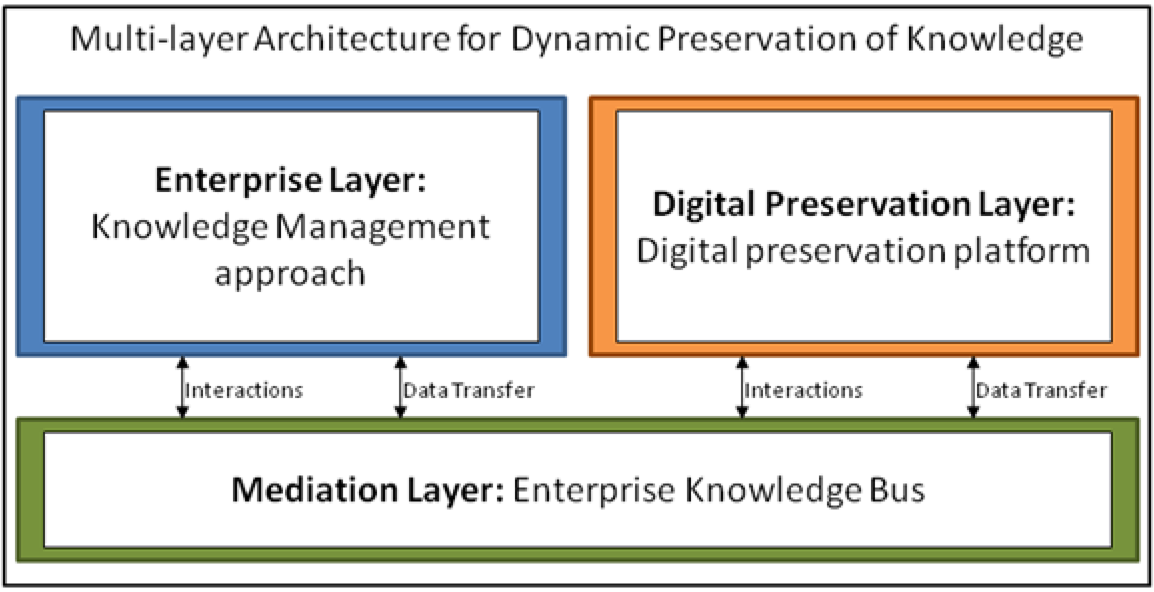

Figure 4.1: MadPK Architecture Model

In Figure 4.1, there are three layers that we have defined in our research work. The two-direction arrows between the layers state that there are forward and backward interactions and data exchanging among the layer. In general, MadPK works as:

- The digital preservation platform is stable to a certain extent, and stability is one key for long term data accuracy and security. This is the reason why we separate the knowledge repository from the knowledge production (Enterprise Layer, KM approaches).
- The reason why digital preservation platform can be stable to a certain extent is that its input and output knowledge are always packed in formal model, even if the knowledge and data from the knowledge source can be different in models and in formats. We develop all the model transfer activities inside Mediation Layer (EKB).
- In Enterprise Layer, changes of information systems lead to the change of KM approaches, and consequently the change of output knowledge model, which is the input of EKB. The output knowledge model of KM approach can be different depending on the complexity of the information systems in enterprise. However, whenever some or all of these knowledge models change, knowledge model conversion rules are upgraded correspondingly in the Mediation Layer (EKB).

4.2.2. Interoperability vs. Integration in MadPK

In our design of MadPK, we try to enhance the interoperability of the KM system and the digital preservation platform. Interoperability is the ability or the aptitude for two systems to understand each other and function together. The word “inter-operate” implies that one system performs an operation for another system [Chen 02, 04]. Generally, interoperability has the meaning of a coexistence, autonomy and federated environment, whereas integration refers to the concepts of coordination, coherence and uniformatisation. However, on the other hand, according to our proposed MadPK architecture, we have also developed an integrated platform (consisting of KM and digital preservation). From the Table 4.1, we notice that integration and interoperability represent different levels of coupling.

INTEGRATION	INTEROPERABILITY
Consistency of local objectives to global ones	No consistency between local and glaobal objectives
Tightly coupled Components interdependnent	Loosely coupled Components independent
Uniformation (languages, methods, tools, etc.)	Identity & diversity preserved
Intra enterprise Fusion, re-structuration, etc.	Inter enterprise (Virtual enterprise, etc.)

Table 4.1: Integration vs. Interoperability [Chen et al. 06]

Separated systems, which are interoperable with each other, will reduce the coupling of the information systems. However, whether this way is better than an integrated system in the long term is not certain, according to different situations in enterprises. Generally, lower coupling will ease the update of systems and be more dynamic. Nevertheless, in long term digital preservation, one integrated system for KM and digital preservation may ensure the data availability from a technical and legal point of view.

Therefore, we will not determine that the Enterprise Knowledge Bus carries out internal communications within a system, or external communications between systems. Instead, this communication will provide services of knowledge transfer between the KM approach and the digital preservation approach.

4.3. Roles and Requirements of MadPK

4.3.1. Roles in MadPK

As our research is under the context of KM, the roles, which interact in our work, are extensions of actors in the KM process. In [Rao 05], knowledge roles are suggested for the online communities, and we adapt this suggestion of roles to our research work. We have added one more role as "Knowledge Storage Manager" and redefine the other roles and to make them fit for our MadPK:

- Knowledge Consumer: the activities of this role are searching, browsing, accessing, applying, and learning knowledge. In our MadPK, this role is the domain experts and information systems in enterprise, who consume knowledge in the long term.
- Knowledge Creator: the activities of this role are to publish, improve, classify, and discuss knowledge. As we have introduced in **Chapter 3**, the domain experts and information systems in enterprise are also the knowledge source, from which the preserved knowledge is extracted. Analogously, in the OAIS reference model, the roles Producer and Consumer are represented by Knowledge Creator and Knowledge Consumer. Nevertheless, in MadPK, the two roles could be fulfilled by the same individuals, who are in different positions in the temporal dimension. For example, the designer of one product preserved the design data five years before, as the Knowledge Creator. After five years, the same designer (or other individuals but in the same position) retrieves the preserved product design knowledge for the maintenance of the product, or for designing a similar new product, in the same way as the Knowledge Consumer.
- Knowledge Editor: the activities of this role are interviewing experts, storytelling, and content management. In MadPK, this role collects information from Enterprise Layer, and submits the extracted knowledge into EKB. This role also edits the information in digital preservation platform (Digital Preservation Layer).
- Knowledge Expert: the activities of this role are validating, certifying, and legitimizing knowledge. In MadPK, this role works in all the three layers. The Knowledge Expert designs the knowledge models according to different environments

in Enterprise Layer and Digital Preservation Layer, and determines the knowledge transfer rules in EKB (Mediation Layer).

- Knowledge Broker: the activities of this role are locating experts and knowledge, identifying gaps, organizing the whole preservation architecture, filtering knowledge, and coordinating communities of practice (CoPs). In MadPK, this role plays in Mediation Layer, maintains the communications between the other two layers, and transfers knowledge into different forms.
- Knowledge Leader: the activities of this role are shaping KM agenda and aligning with enterprise business objectives. This role is the decision-maker of MadPK. Thus, this role makes knowledge preservation plans and strategies in each layer of the MadPK architecture.
- Knowledge Storage Manager: this role was not introduced in Rao's research, which mainly focused on KM. However, as we have stated that the MadPK is a combination of KM and digital preservation, the knowledge preservation approach and platform require a role to manage and operate. This role manages and maintains the archived data and metadata, performs queries, checks for knowledge obsolescence, and configures the digital preservation platform.

We have identified the roles that interact in our proposed MadPK architecture. Our purpose of the role identification is to determine the usage requirements of each layer of MadPK and consequently to propose the functional design and models. Table 4.2 shows the roles in each layer of MadPK. The definition of roles will lead to the interactivities between the roles, and consequently lead to the requirements of the MadPK.

Layer / Role	Enterprise Layer	Digital Preservation Layer	Mediation Layer
Knowledge Consumer	✓		
Knowledge Creator	✓		
Knowledge Editor	✓	✓	
Knowledge Expert	✓	✓	
Knowledge Broker			✓
Knowledge Leader	✓	✓	✓
Knowledge Storage Manager		✓	

Table 4.2: Roles in each layer of MadPK

4.3.2. Usage Requirements of MadPK

We introduce the use cases of the architecture by layers and by interactions between layers in MadPK. The usage requirements will illustrates the basic functional needs of the MadPK architecture. Moreover, the interactions between different roles will help to identify the business processes with the MadPK architecture.

4.3.2.1. Requirements of Enterprise Layer

Enterprise Layer supports the KM approach with information systems in designated organization. In our research work, we aim to achieve the knowledge preservation with the production information systems in industries or enterprises. This is why we named this layer as Enterprise Layer. This layer produces product, process, and organizational data. By adapting KM approach and Context Level models, knowledge will be generated and packaged for the following preservation. Figure 4.2 shows the use cases of this layer.

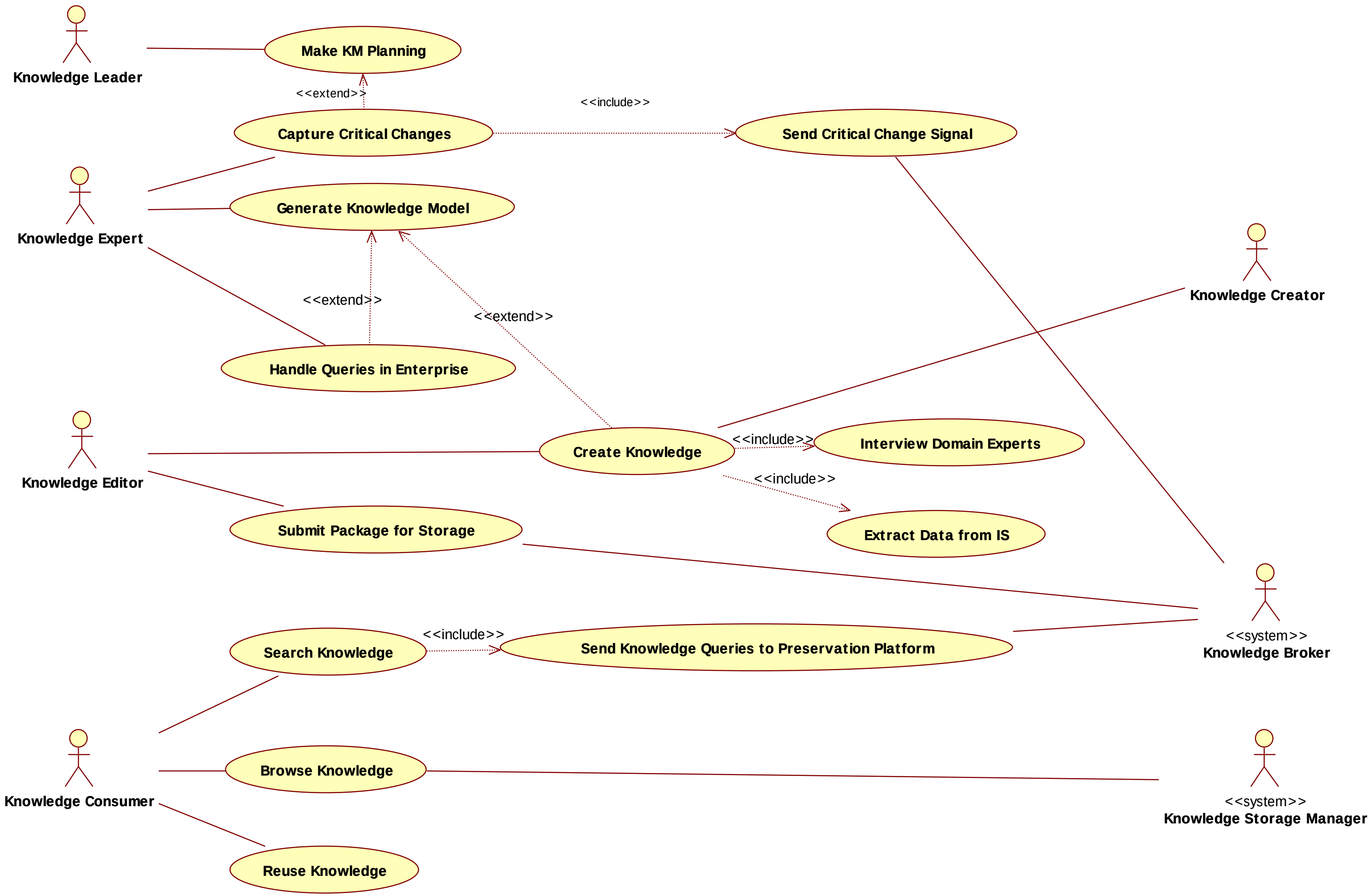

Figure 4.2: Use Case diagram of Enterprise Layer

The Knowledge Leader makes the KM planning, which determines KM strategies and identifies the critical changes in the long term. According to the KM planning, the Knowledge Expert generates the knowledge model, which is the structure of the exacted data. Then, the Knowledge Creator extracts knowledge from the domain experts or from information systems, based on the knowledge model. The knowledge packages will be sent to Mediation Layer, and the Knowledge Broker will operate these packages.

When the Knowledge Consumer would like to retrieve information from the preserved knowledge, queries need to be sent to the Knowledge Broker, too. By going through the communications with Digital Preservation Layer, the Knowledge Broker sends back the query responses and the knowledge packages, in order that the Knowledge Consumer can use the required information. Similarly, when critical changes (i.e. policies, resources, software changes, etc.) in Enterprise Layer are captured by the Knowledge Expert, signals (e.g. queries, notifications, etc.) also need to be sent to the Knowledge Broker, in order to inform and cooperate with the Digital Preservation Layer.

4.3.2.2. Requirements of Digital Preservation Layer

Digital Preservation Layer is the digital preservation platform. This layer captures, packages, transfers, stores knowledge and provides features such as search and browse for the knowledge reuse. The digital preservation platform in our research work is an instance of OAIS. The use cases of this layer are shown in Figure 4.3.

The Knowledge Leader also makes plans for digital preservation in this layer. As the knowledge is packaged according to the Information Package model of OAIS, the Knowledge Expert creates, audit and modify the Information Package structure and content in this layer. Normally, the Information Package structure is firmed. However, in certain circumstances, Information Package should be restructured according to critical changes in Enterprise Layer. Once the Information Package structure is settled, the Knowledge Storage Manager gets submission from the Knowledge Broker (from Mediation Layer). Then, the Knowledge Storage Manager validates submission, generates Information Packages, stores the packages in certain media (i.e. physical media such as hard disk, flash disk, optical disc, etc.) and updates database of the digital preservation platform. Moreover, according to the OAIS recommendation, the Knowledge Storage Manager handles the queries from the Knowledge Broker (originally from the Knowledge Consumer) and delivers responses. The dynamic feature of this layer is that the Knowledge Storage Manager updates the preserved knowledge according to the real time critical change signals from Enterprise Layer.

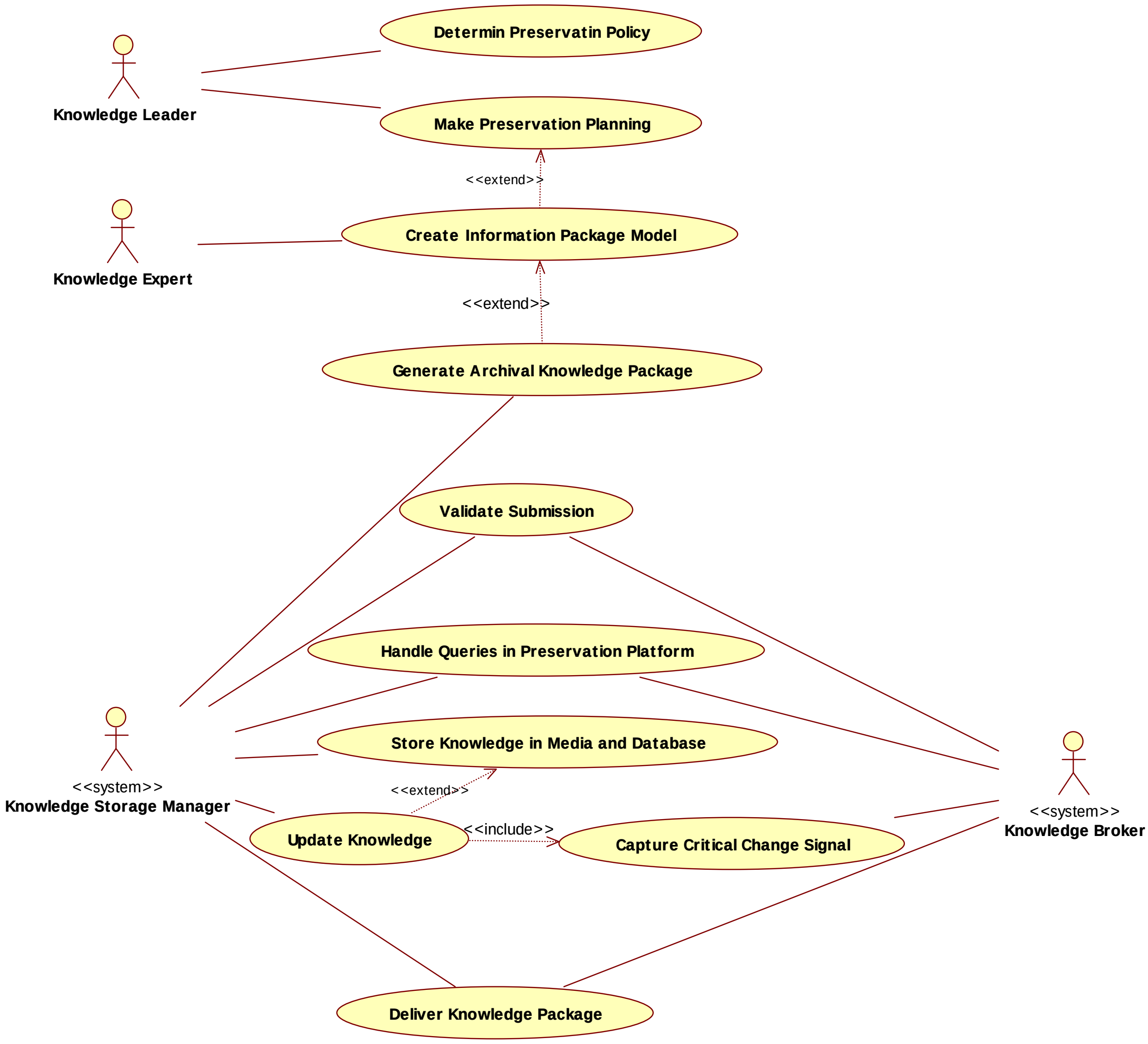

Figure 4.3: Use Case diagram of Digital Preservation Layer

4.3.2.3. Requirements of Mediation Layer

Mediation Layer is the integration of methods and tools for dynamic communications. This layer connects the previous two layers together. For the purpose of dynamic perspective, we acquire Service-Oriented Architecture (SOA) theory and aspect in our proposal. Thus, the digital preservation platform connects to the knowledge source through enterprise knowledge bus (EKB). The use cases of this layer are shown in Figure 4.4.

The Knowledge Leader also makes EKB plans in this layer. The planning includes the model transfer rules and the configuration of the dynamic connections. As there are different forms of knowledge between the other two layers, the primary job of this layer is the creation and modification of model transfer rules, which allow the knowledge of different form to be regenerated and recognized by either layer. At the same time, the configuration of the connection between layers has to be done by negotiating with the other two layers. Once the connection and the model transfer rules are determined, the Knowledge Broker manages the communications between the other two layers, and delivers knowledge, queries and the critical change signals. Sometimes, a critical change signal will also appear as a knowledge package or as a query message when being submitted.

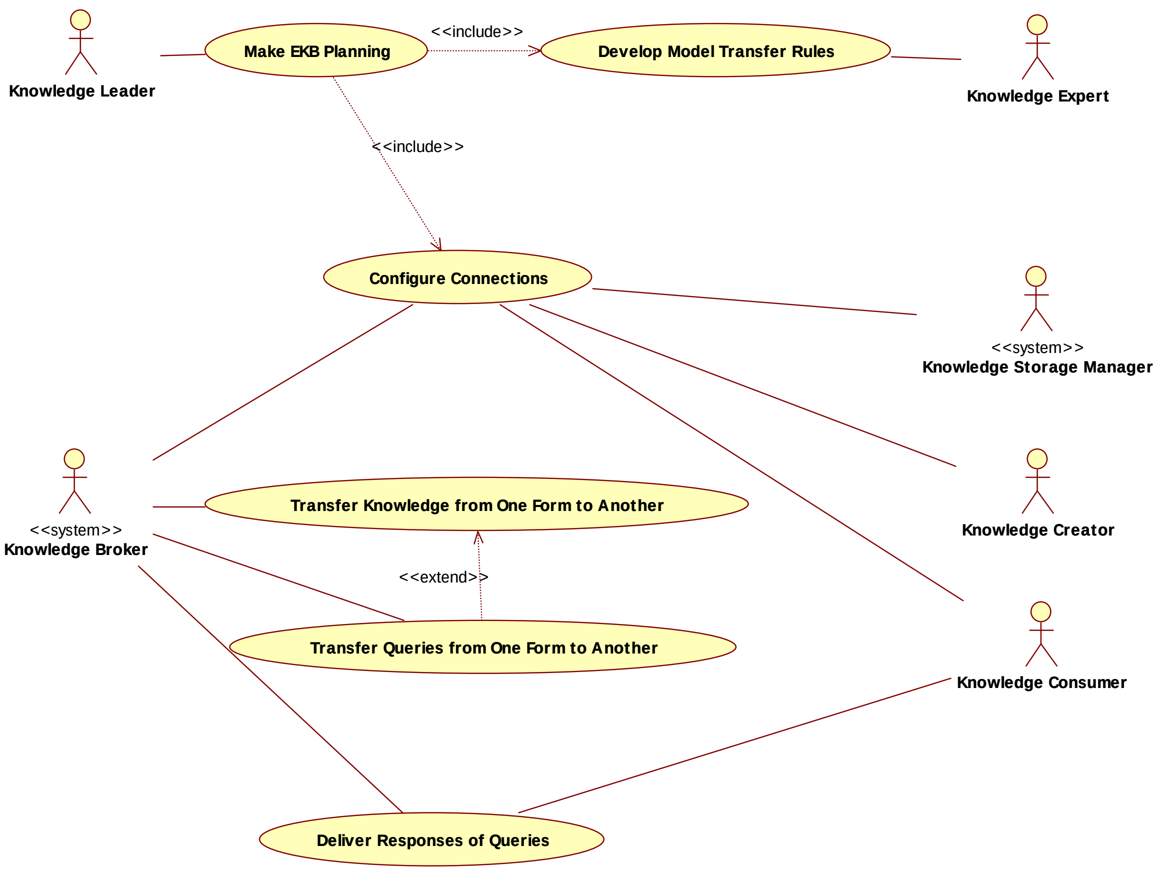

Figure 4.4: Use Case diagram of Mediation Layer

4.3.2.4. Communications between Layers

Through the synthesis of the previous use cases, we have noticed that there are some communications between different layers. In order to make the use case diagram readable, we will not put all the roles and use cases together to show the interactions and communications between layers. Instead, we extract the activities of interactions between layers. Figure 4.5 and Figure 4.6 show the communication between Mediation Layer and the other two layers. The interactions mainly are exchange of information. The dynamic feature is that all cross-layer information and the knowledge are transferred through the Knowledge Broker (and through the EKB of Mediation Layer), without worrying about the formats and data structure. In other words, both KM and digital preservation approaches will pay more attentions on their businesses and benefit from the existing methods, technologies and tools. This kind of architecture looses coupling of different information systems and makes the digital preservation platform and preserved knowledge be stable in certain extent.

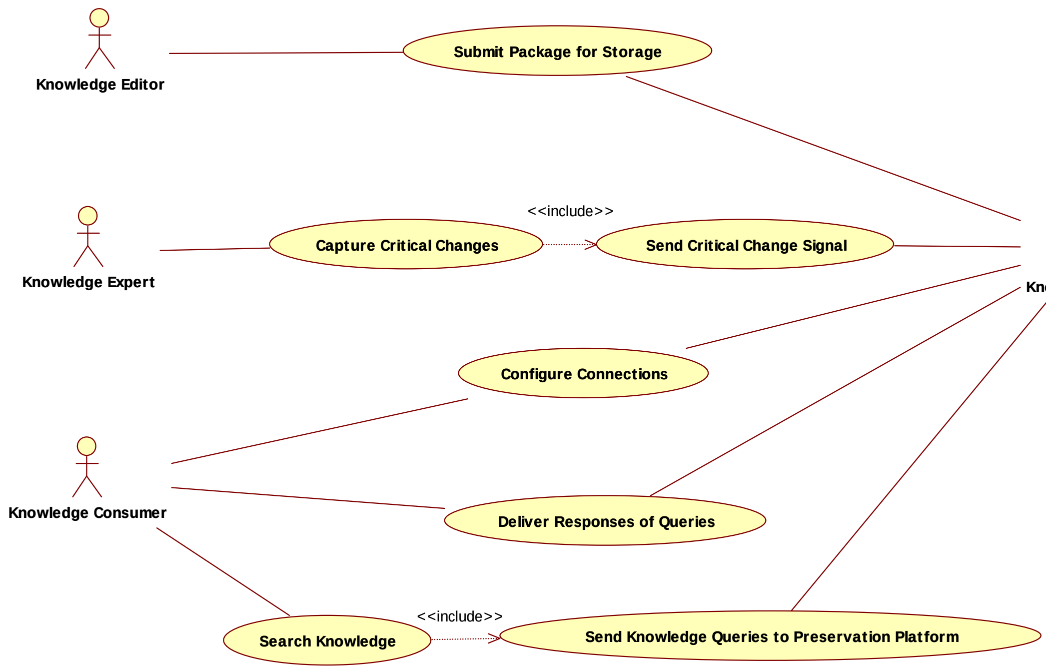

Figure 4.5: Use Case diagram of communication between Enterprise Layer and Mediation Layer

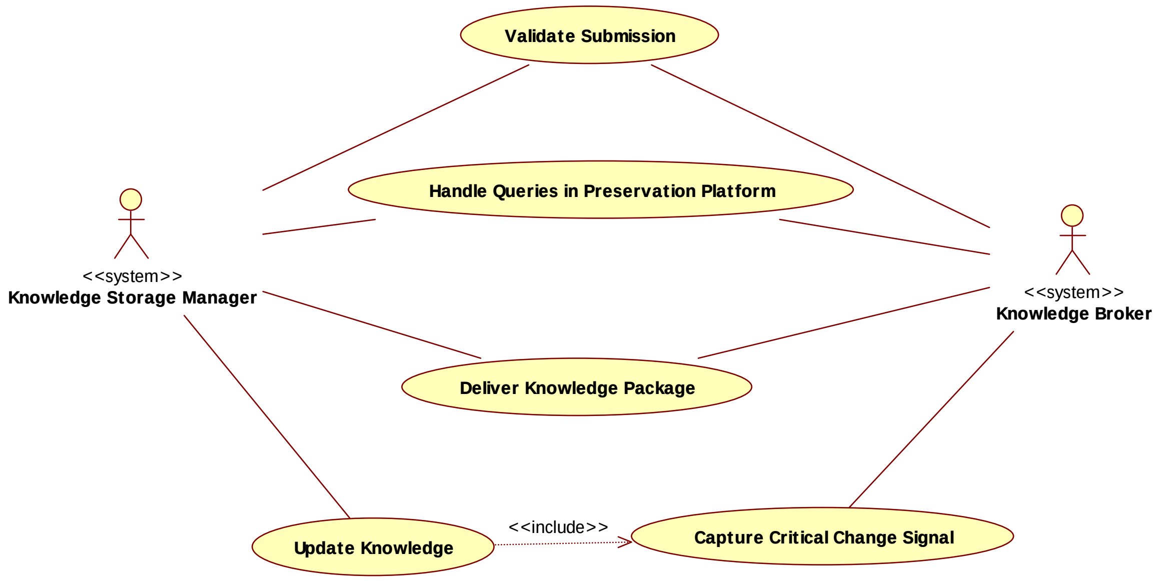

Figure 4.6: Use Case diagram of communication between Digital Preservation Layer and Mediation Layer

However, in some circumstances, the Knowledge Consumer communicates with the digital preservation platform directly. In our previous research on digital preservation platforms, there is a browsing feature in most of the digital preservation platforms. Thus, the Knowledge Consumer may connect to digital preservation platform for browsing of preserved knowledge (Figure 4.7). Nevertheless, when the Knowledge Consumer finds appropriate knowledge in digital preservation platform, the knowledge objects may also need to go through the knowledge transfer process of Mediation Layer in order to be used correctly in Enterprise Layer (e.g. need to be generated into certain form).

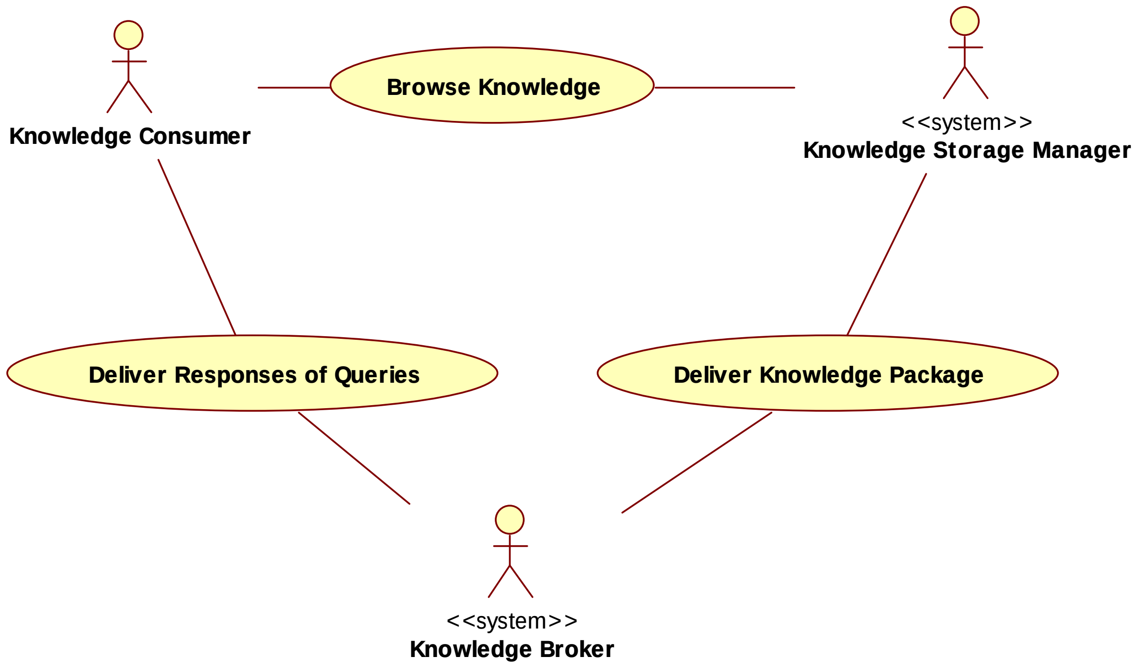

Figure 4.7: Use Case diagram of direct communication between Enterprise Layer and Digital Preservation Layer

These usage requirements that we have introduced in this section lead to the functional design for the MadPK architecture. We will introduce the functional design in the following sections.

4.4. Functional Design of MadPK

The structural design of MadPK is represented by Architecture Model, while the functional design is realized by Function Models. As we have stated in **Chapter 2**, the functional design appears as the Function Model of our extended CommonKADS methodology. In order to identify the Function Model, we have to identify the basic functionalities and the sub-functionalities that construct the MadPK architecture. We have defined the relations between the Function Model and the Architecture Model as:

- The functions are basic elements within the architecture, and can be seen as the decomposition of the architecture;
- The functions are organized and determined by the architecture;
- The functions are determined by the duty of the architecture, as well as the requirements of long term knowledge preservation.

4.4.1. Enterprise Layer

Figure 4.8 shows the Function Model of Enterprise Layer, while we consider the other two layers as black boxes, which provides sufficient capacities for appropriate inputs and outputs. Generally, the Knowledge Acquisition and Knowledge Reuse are both directed by the Knowledge Model that defined in KM Planning. According to the KM strategy developed by KM Planning, KM Acquisition captures knowledge from information systems or domain experts, and then submits it into Mediation Layer. The Knowledge Reuse function is triggered when an end user attempts to use preserved knowledge or when the critical changes are detected in KM Planning. After Knowledge Reuse is triggered, it sends queries to Mediation Layer and waits for the responses. When the responses (i.e. query results, error messages, etc.) are sent back from Mediation Layer, Knowledge Reuse handles the response objects, get use of the query results, sends more queries or terminates the knowledge reuse process.

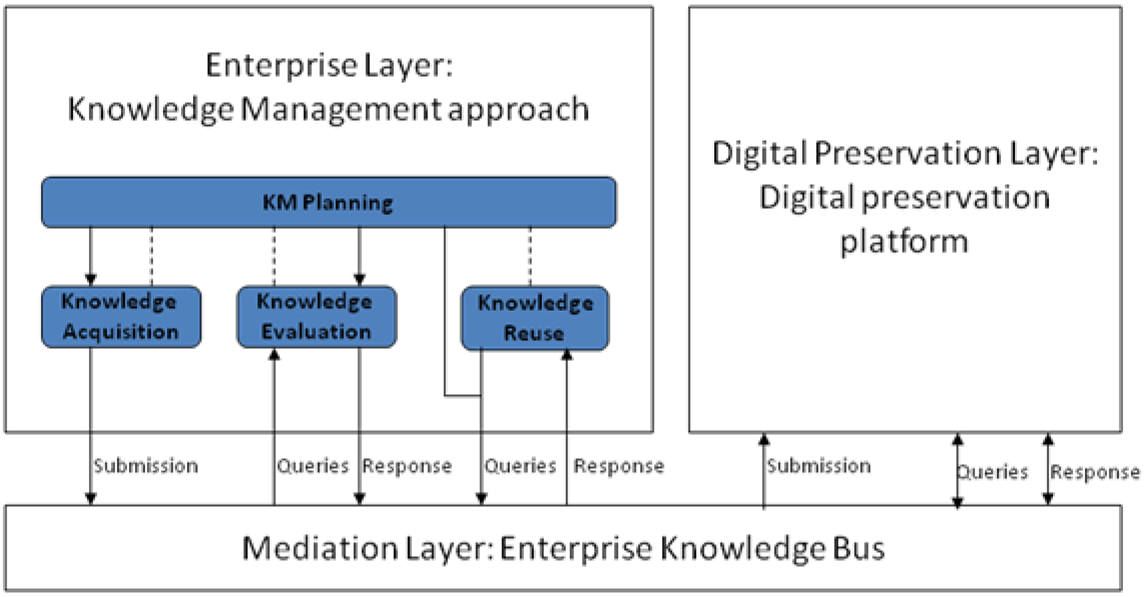

Figure 4.8: Function Model in Enterprise Layer

4.4.1.1. Knowledge creation

We have introduced the KM approach in **Chapter 2**, and we have stated that the functional aspect "strategy alignment" is the first and primary task in KM approach, for it will lead to the following each step or function in a KM project.

Business process management (BPM) and KM is two major parts in company. As the trend of enterprise has changed from industrial economy to knowledge economy, KM is required to be able to interoperate with BPM. Automating business process is able to preserve business rules. The repository of business rules makes companies easy to change business logic and to manage process knowledge. In other words, one requirement of coordinating business logic and knowledge is automating process culture (process automation culture).

If considering a knowledge related project, different business processes have different knowledge intensities. Some processes may mostly benefit from the knowledge sharing culture, while others mainly benefit from the automating process. No matter what sort of business process is concerned in KM, the knowledge workflow must have correlations with the business processes. Both the knowledge and business process are supervised by the corporation strategy and rules. Of course, in corporation strategy, BPM and KM have different objectives. Therefore, in the knowledge creation process, we have adapted the PPO model concept, and we acquire not only the data object (product), but also capture the organizational and business process information, which is associated with the data object (Figure 4.9).

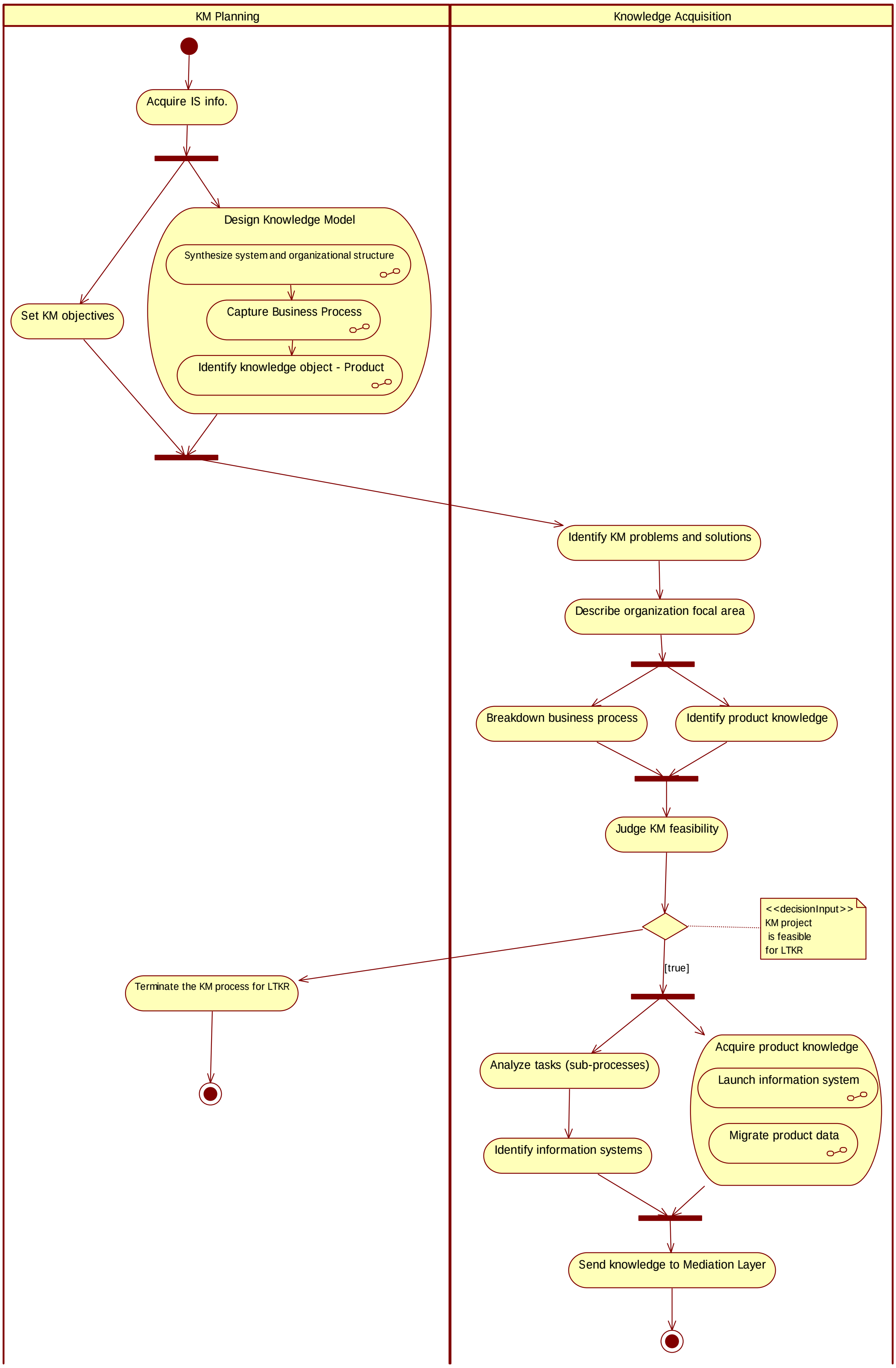

Figure 4.9: Activity diagram of knowledge creation

This process produces several documents. According to the Context Level model set, which we have introduced in **Chapter 2**, the sub-function of "design knowledge model" in KM Planning determines the detailed structure of the Organization, Task and Product Models. Based on this structure, the function of Knowledge Acquisition performs interview to the Knowledge Creator and extract knowledge from the knowledge source. After the extraction of organizational and environmental information, Knowledge Acquisition function will judge the feasibility of the KM process (performed by the role of the Knowledge Leader). If the knowledge objects are not fit for long term preservation, the KM process will be terminated. For example, a KM objective of a long term preservation project is preserving knowledge for reengineering and redesign. By analyzing the information systems for workshop planning and scheduling, the Knowledge Leader considers the scheduling information is not significant for reengineering and redesign in long term. Thus the KM process in this information system may be terminated.

The acquired knowledge (several documents) will be packaged and be submitted to Mediation Layer. In Mediation Layer, the Knowledge Broker will transfer the knowledge into Information Package form and submit the knowledge to digital preservation platform. The Knowledge Broker will send notifications to Enterprise Layer to inform the Knowledge Editor, in order to confirm the submission or request for resubmission.

4.4.1.2. Knowledge query

We place "knowledge query" as the title of this section, rather than "knowledge reuse". The reason why we use knowledge query is that the requests from Enterprise Layer are not always from the Knowledge Consumer. In some circumstances, when the critical changes are identified in Enterprise Layer, the Knowledge Expert will send query requests (critical change signals) in order to check and update the knowledge in digital preservation platform. In either case, Enterprise Layer throws queries into Mediation Layer and waits for responses from digital preservation platform. After the Knowledge Expert sends the critical change signals, a backward query would be received from digital preservation platform, in order to get enough information for the knowledge update. In other words, the sending query processes by the Knowledge Consumer and by the Knowledge Expert are identical (Figure 4.10). In any case, Enterprise Layer sends "something" to Mediation Layer.

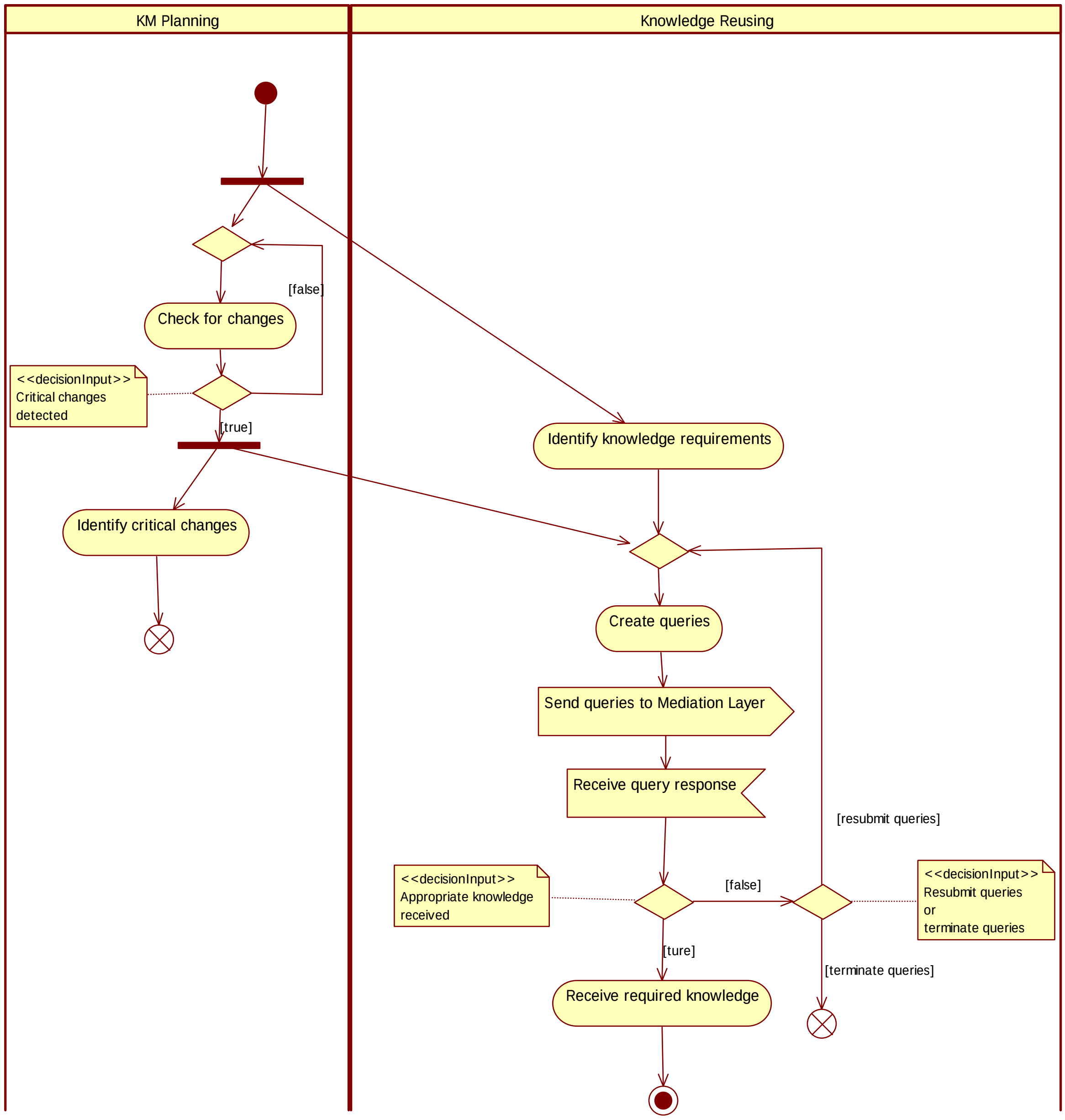

Figure 4.10: Activity diagram of knowledge query

In the function of Knowledge Reuse, after sending queries, responses will be sent back from Mediation Layer. At times, the required knowledge may not be found in digital preservation platform, and the responses message will suggest changing some information on queries. This happens when the Knowledge Consumer want to reuse the preserved knowledge. In this circumstance, the Knowledge Consumer will choose re-send modified queries or stop the query process.

4.4.1.3. Knowledge evaluation

The Knowledge Evaluation function plays in Enterprise Layer. However, its purpose is to evaluate whether to terminate the knowledge, which is preserved in digital preservation platform. This is a dynamic feature for MadPK in Enterprise Layer. The Data Management function of Digital Preservation Layer checks the preserved Information Package periodically, and filters the Information Packages, which are not been used in a designated long term (e.g. one year, etc.). From the Digital Preservation Layer side, the Knowledge Storage Manager synthesizes the summaries of these filtered Information Packages and throws the Descriptive Information of these Information Packages as queries to the Knowledge Broker. The Knowledge Broker transfers the queries into appropriate forms and sent them to Enterprise Layer. This is how the Knowledge Evaluation function is triggered.

The Knowledge Evaluation function handles the queries from digital preservation platform and tries to retrieve knowledge by the Knowledge Creator from current knowledge source (i.e. information systems or domain experts).

- If the Knowledge Creator provides certain knowledge corresponding to the queries, the Knowledge Expert will decide whether the preserved knowledge in digital preservation platform should be updated or not. If yes, new submissions will be made and be thrown to the Mediation Layer; if not, a simple notification will be sent to indicate that the preserved knowledge is up-to-date and still be valuable in Enterprise Layer. In other words, the knowledge should be kept.

In some case, there is not corresponding knowledge retrieved according to the queries, because the checking period (designated long term in Digital Preservation Layer) may longer than the product life cycle of a certain product in this organization. Of course this is the main reason why we need to perform long term preservation. When there is no query results from the Knowledge Creator, the Knowledge Expert must go back to check the original documents of KM planning. Here is a manual function or human task, which should be performed by the Knowledge Expert and the Knowledge Leader. They will decide whether the preserved knowledge, which is not been used in a designated long time, should be preserved for longer time or should reach its obsolescence.

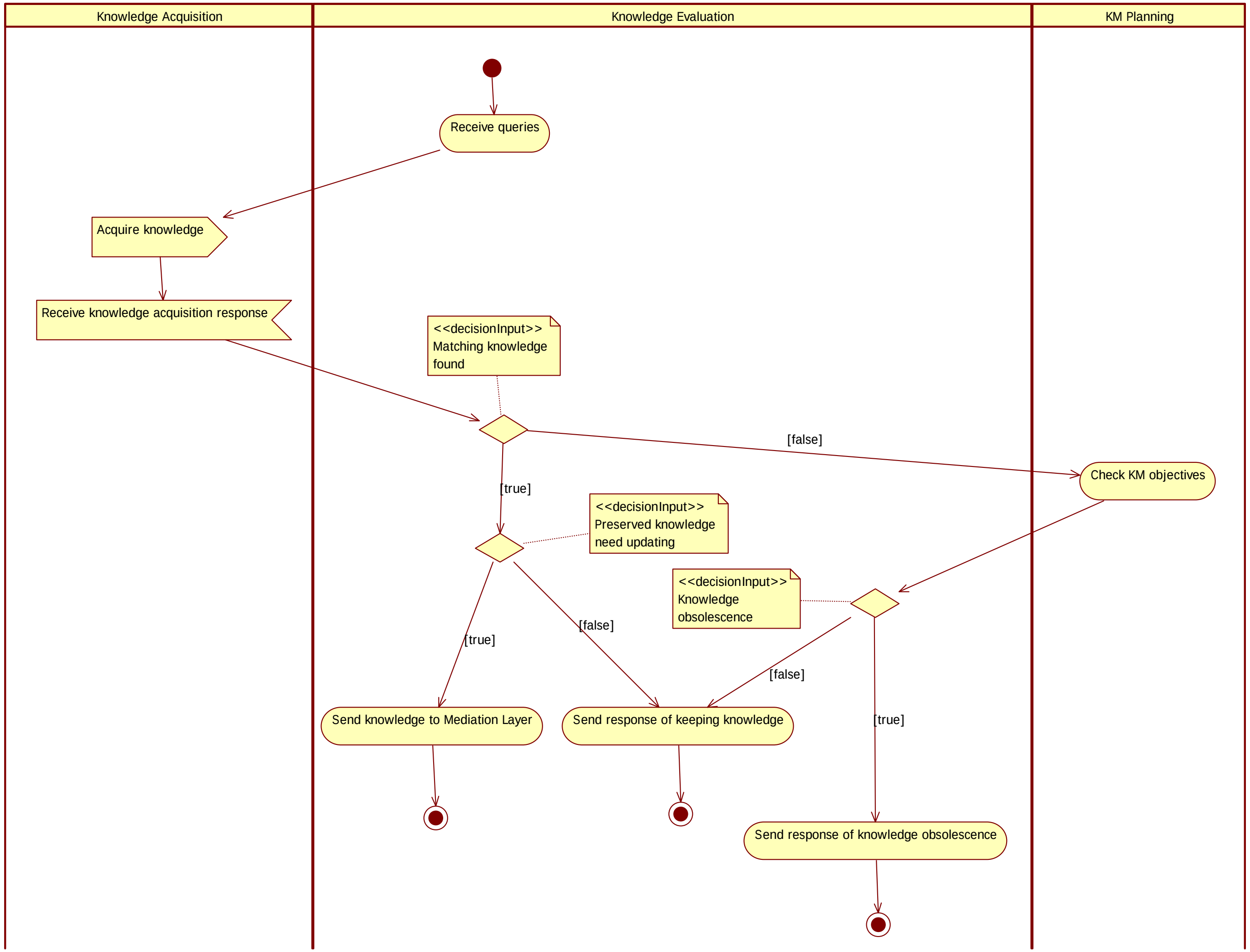

Figure 4.11: Activity diagram of knowledge evaluation

The submission for long term preservation comes from the KM processes in Enterprise Layer, with pre-performed analysis and KM objectives. However, the Digital Preservation Layer could not see the KM processes and may argue the efficiency of utilization of the resource for maintaining Information Package, which is not been touched for a long time. Sometimes, due to the organizational changes or technical issues, it is quite possible that certain Information Packages, which should have been updated or terminated, are not handled correctly in time. Therefore, the Knowledge Evaluation function is necessary and is one significant feature in MadPK.

4.4.2. Digital Preservation Layer

The Function Model of the Digital Preservation Layer has already presented in **Chapter 3** (Table 3.3). We notice that the main knowledge objects exchanged in digital preservation platform are submission information package (SIP), archival information package (AIP), and dissemination information package (DIP). Therefore, we distinguish the whole process of OAIS data workflow into four phases: "Receiving submission and validating SIP", "Generating and storing AIP (with Descriptive Information)", "Generating DIP and sending results", and "Updating AIP". Thus, we describe the digital preservation platform not only from a functional viewpoint, but also from a business process viewpoint. This will help us in the following development by adapting BPM (Business process management) and SOA (Service-Oriented Architecture) technologies and applications.

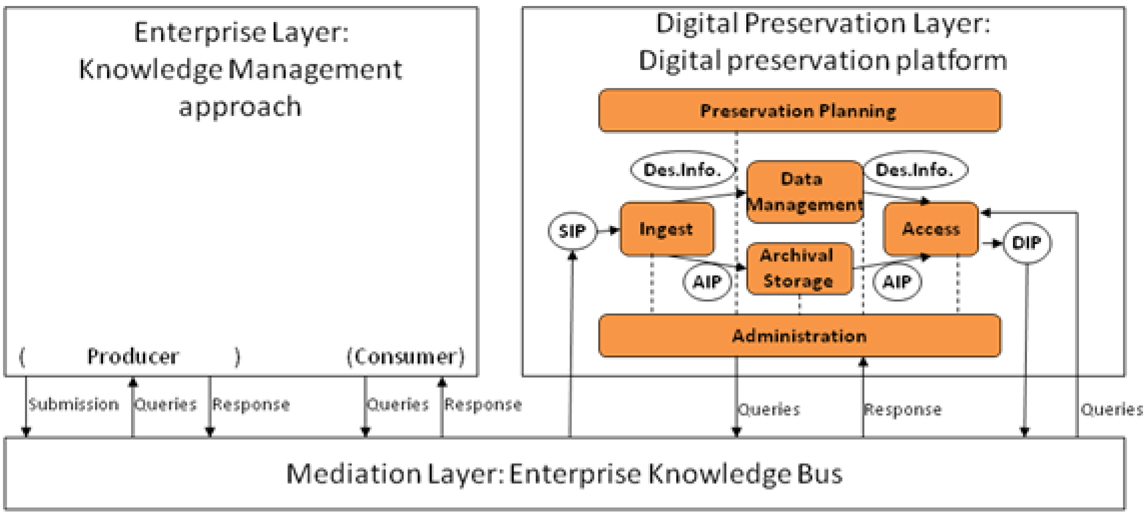

Figure 4.12: Function Model in Digital Preservation Layer

4.4.2.1. Receiving submission and validating SIP

As we have introduced previously, we identify the "Producers" as the production information systems or domain experts in the KM approach, but they submit knowledge or data in their own knowledge model, which is not an exact Information Package. The Ingest function of the digital preservation platform gets knowledge in Information Package form. (Figure 4.13)

- When acquiring SIPs, there is a sub-function to validate and audit the submission (sub-function: Validate Submission). Although SIPs are actually produced by going through the KM approach, within an OAIS, these SIPs should be proved and validated. This sub-function of "validate submission" will check the submission to make sure that there is no error (the definition of error depends on the policies made by the Knowledge Expert, e.g., missing description information, wrong format of package title, etc.) in the submission. However, whether errors are detected or not, the digital preservation platform will send a Confirmation of Receipt to the role who controls the KM approach, in order to confirm that the SIPs have been received. Then in case of errors resulting from the submission, a "re-submit" request will be included in the Confirmation of Receipt.
- In the OAIS recommendation, despite the fact that it has developed multiple functions to deal with multiple situations, we still need to provide more extra feature in our real implementation. In this receiving submission case, Knowledge Expert, the Knowledge Editor and the Knowledge Broker will have to negotiate and make an agreement on a definite lead time of sending Confirmation of Receipt. If the definite lead time has expired, the Knowledge Editor will once more send the submission and inform the OAIS of the resubmission in the Descriptive Information of the SIPs.

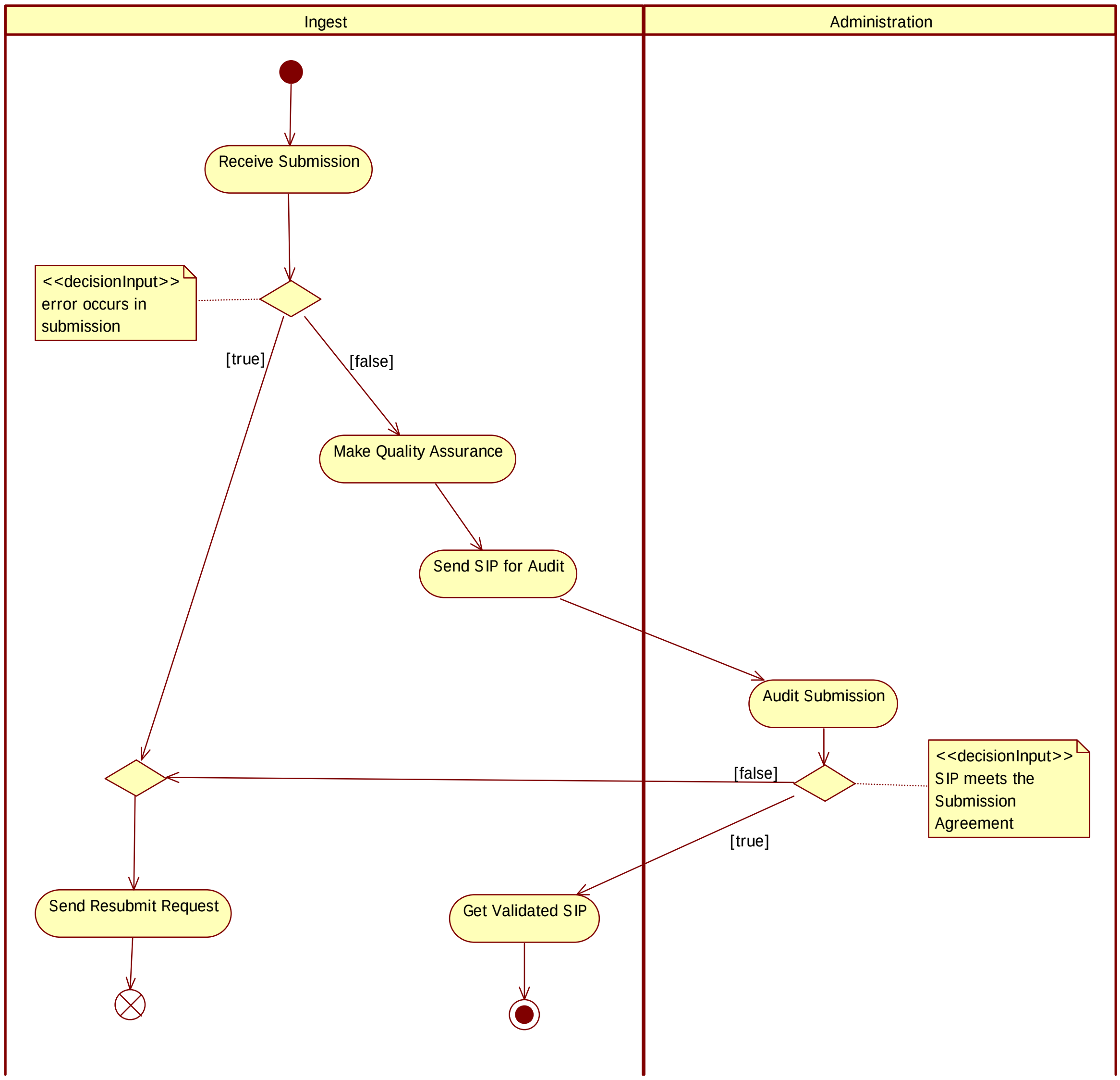

Figure 4.13: Activity diagram of validating SIP

- Once the SIPs (without errors) are obtained, all these SIPs will be sent to go through the Quality Assurance process. For digital submissions, these mechanisms (of Quality Assurance) might include Cyclic Redundancy Check (CRC) or checksums. Checksum is one simple way ensuring the security of the transferred digital objects. Checksums are associated with the data files or written in the system logs [CCSDS 650.0-B-1; ISO 14721:2003].
- Before they are utilized to generate AIPs, the SIPs have to be verified by the sub-function of "audit" in the entity Administration. The "audit" sub-function makes sure that the submissions meet the specifications of the Submission Agreement, which is negotiated by both the Knowledge Editor and Knowledge Expert [CCSDS 650.0-B-1;

ISO 14721:2003]. After the audit process, an Audit Report will be sent to the entity Ingest to generate AIP, and a Final Ingest Report will be generated, too.

4.4.2.2. Generating and storing AIP (with Descriptive Information)

The validated SIPs are transferred into Archival Information Packages (AIPs) for storage (sub-function: Generate AIP). Besides, descriptive metadata (Descriptive Information) should be extracted from the AIPs (sub-function: Generate Descriptive Information), simultaneously AIPs will be retrieved by this Descriptive Information. (Figure 4.14)

- Although proper SIPs are obtained and validated by the previous process, more information on existing information packages in the OAIS is still needed for generating new AIP. Thus to generate AIPs, validated SIPs as well as reports from Data Management are required. There is a Generate Report sub-function in Data Management, which will deal with all the report requests. Once the Generate AIP function has sent report request to the entity Data Management, the Generate Report sub-function will provide report, which may include summaries of archive holdings by categories, or usage statistics for access to archive holdings. With both the SIPs and the reports from the entity Data Management, the Generate AIP sub-function transforms SIPs into the AIPs.
- Generating AIPs may involve the conversions of the file formats, the data representation or just a simple re-organization of the content in the SIPs. The content of the report from Data Management is used to generate Descriptive Information that complete the AIPs. The mapping between SIPs and AIPs is not one-to-one, and it depends on the Data Formatting and Documentation Standards as well as the type of SIPs. According to the OAIS recommendation, this mapping between SIPs and AIPs can be one-to-one, many-to-one, one-two-many, many-to-many, one-to-zero (**Annex 4**).

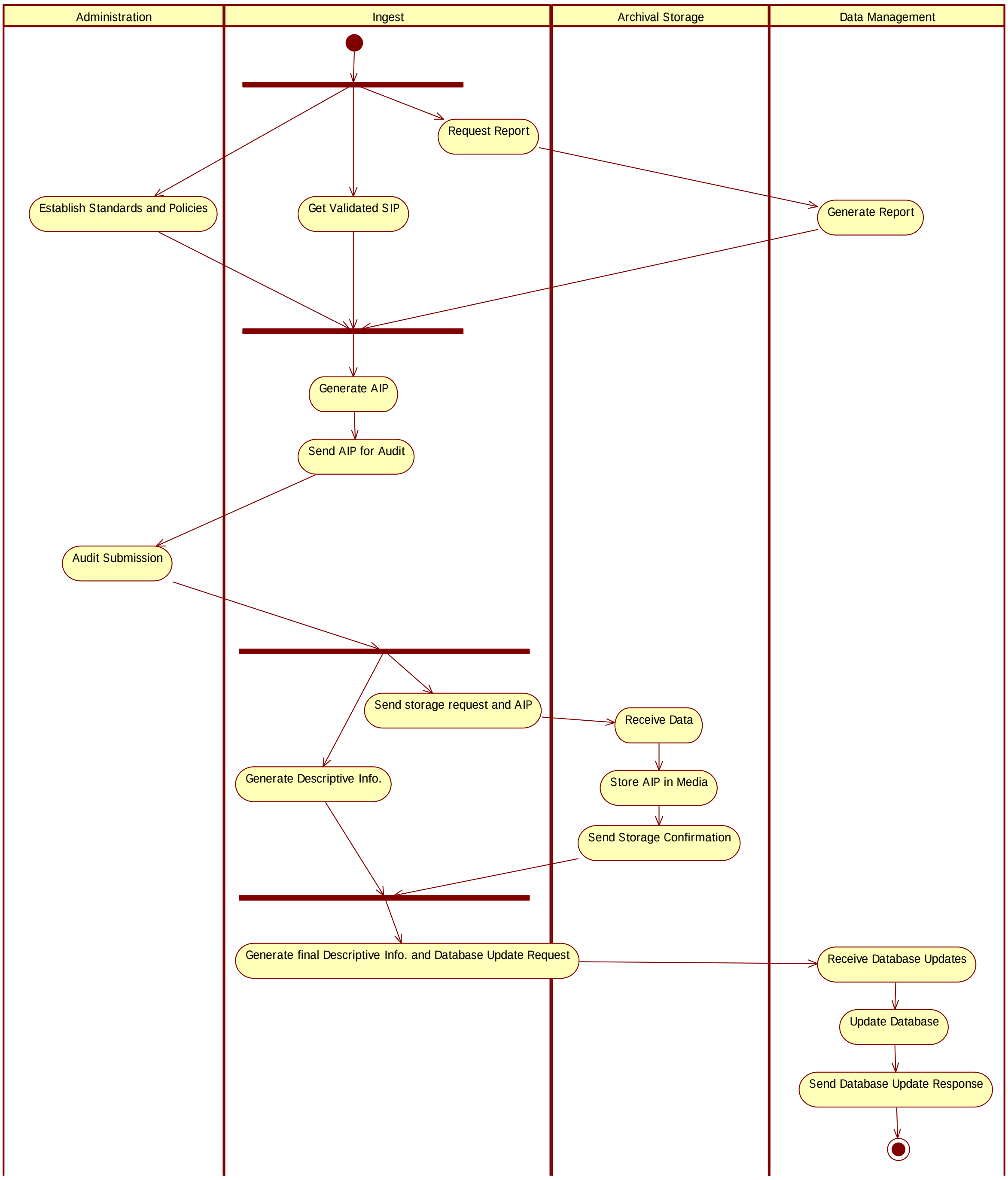

Figure 4.14: Activity diagram of generating and storing AIP

- Similar like the SIPs, the AIPs also need to be verified by the sub-function of "audit" in Administration. The AIPs will go through the same process of auditing to make sure that the AIPs meet the specifications of the submission agreement. After the audit process, an Audit Report will be sent to the entity Ingest to confirm that AIPs are approved and ready to be stored.

- After the AIPs are generated and verified, the Ingest process will go through two parallel paths. One is contacting the Archival Storage entity and store the AIPs in certain sort of media; the other is getting Descriptive Information and transferring it to the entity Data Management to update the database.
- When transferring the AIPs to Archival Storage, a Storage Request will be attached to the AIPs. The Storage Request may represent an electronic, physical, or virtual transfer. Moreover, in the Storage Request, the frequency of utilization of the data objects comprising the AIP may be indicated. Thus, after the AIPs and the Storage Request have been sent to the Receive Data function in Archival Storage, appropriate storage devices or media will be selected for storing the AIPs according to the requirements. This function performs the real physical storage activities and stores the AIPs. After the storing of AIPs has been accomplished, the Receive Data function will send a Storage Confirmation message to Ingest, including the storage identifications of the AIPs.
- The result of the Generate Descriptive Information function proposed in the OAIS recommendation has more content than the extracted Descriptive Information from AIPs. Still, it is not the final Descriptive Information that will be stored in the OAIS database. The Generate Descriptive Information function extracts Descriptive Information from AIPs, collects Descriptive Information from other sources (information for searching and retrieving) and mixes them together. After the Storage Confirmation message has been sent by Data Management, the storage identification information also needs to be added into Descriptive Information. In other words, the generating of final Descriptive Information will go through the three following phases: extracting it from AIPs, collecting it from other sources, and obtaining it from Storage Confirmation.
- As soon as the final Descriptive Information is generated, it will be sent into the sub-function of "archive data' in Data Management, along with a Database Update Request. Although the Descriptive Information requires the Storage Confirmation from Archival Storage, Data Management updates may take place without a corresponding Archival Storage transfer when the SIP contains Descriptive Information for an AIP already in Archival Storage. In this case, since the AIP is already stored in Archival Storage, a Storage Confirmation with storage identification is still available. The Receive Database Updates is partial of the sub-function

“maintain storage”, which adds, modifies or deletes information in the Data Management persistent storage. In any case (updates succeed or fail), the “maintain storage” sends a Database Update Response back to Ingest, indicating the status of the update.

4.4.2.3. Generating DIP and sending results

This process corresponds to the Access entity in the OAIS reference model. In the Knowledge Reuse process, end users send queries to the digital preservation platform, and search for appropriate information. According to the knowledge model we have defined in our KE approach, the descriptive information of one information package represents all the critical information corresponding to the data content. Therefore, when searching, only the predefined descriptive information is searched, while other information related to the data content is packaged in the information package. (Figure 4.15)

- After searching of information, the preservation platform should convert information package again (i.e. convert AIP to DIP) and deliver the DIP to end user. This is because inside of the digital preservation platform, the information package form is fit for knowledge and storage management. As the supported information may not be sufficient for end users, the conversion process is necessary. The other reason for the conversion of information package is that we try to develop business process models which are dynamic and reusable. Besides, the conversion process could also be use in the knowledge changing/updating process.
- Three categories of Knowledge Consumer requests are distinguished: Query Requests, Report Requests and Orders. Query requests are performed in the Data Management entity and the responses of query requests are query results of the stored information (AIPs). Report requests could be considered a combination of a series of queries, and the query results need to be structured as a “formal” report, whose format is defined by end users. Orders represent the stored information held by the Archival Storage entity, and the AIPs will be generated as “formal” DIPs according to the requests. These three categories are transferred by KM approach After receiving the requests, estimates should be done in order to determine whether the existing resources are available for performing the requests and assuring the users are authorized to access and receive the requested items, etc. At last the Knowledge Consumer will be informed whether the requests are accepted or not. The response for the Knowledge

Consumer can also be a report, illustrating the information about the estimates or the reasons for the rejection of requests.

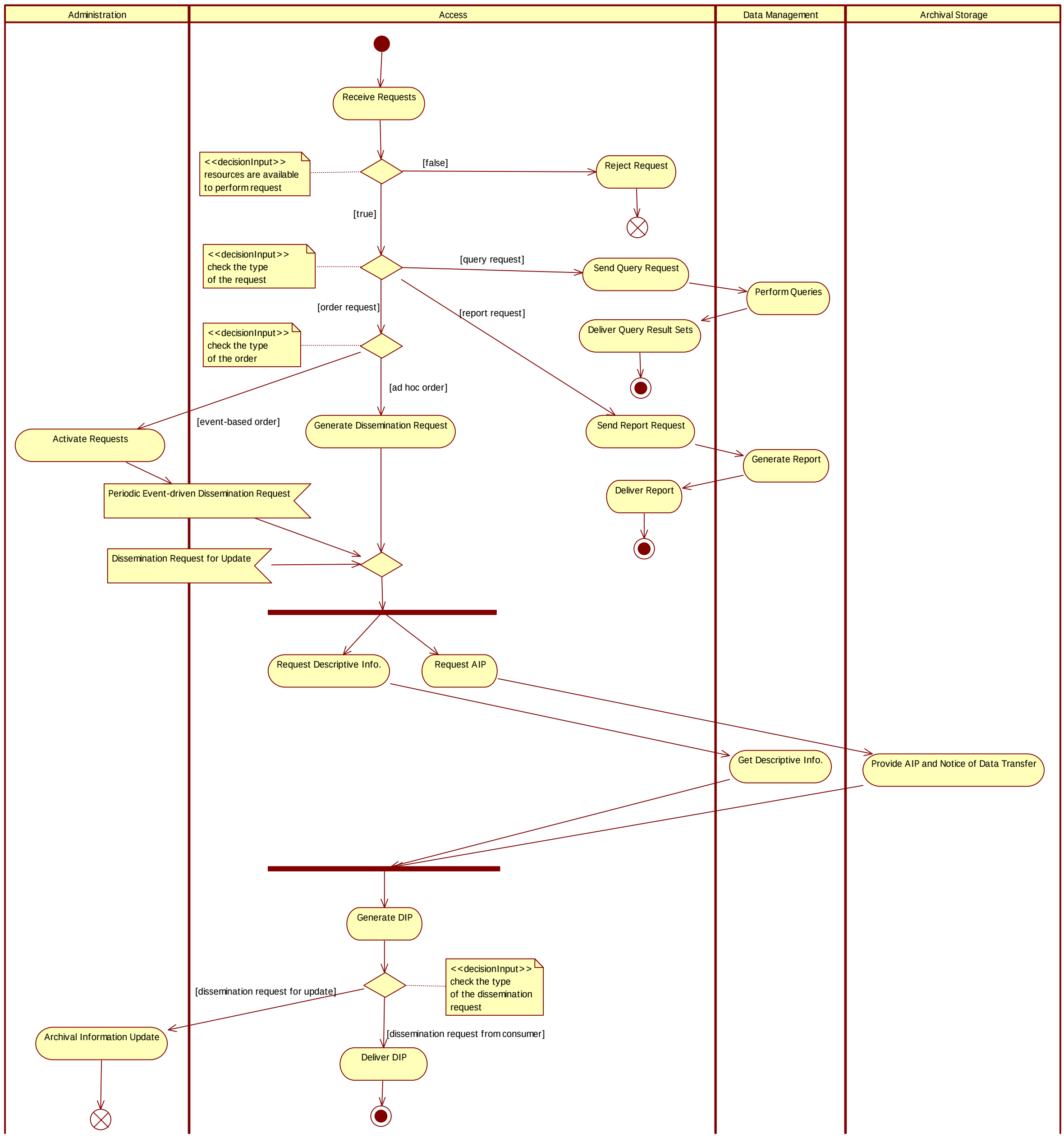

Figure 4.15: Activity diagram of generating DIP

- The type of requests must be checked. Both a Query Request and a Report Request are simpler than an Order. When an accepted request is a query request, the digital preservation platform just sends the query/report request to the entity Data Management, and within Data Management there are sub-functions performing queries among the stored Descriptive Information and/or generating specific formatted reports using the query results. These queries are performed quite fast and efficiently. And the results sets or reports will be sent back to the Knowledge Broker, and consequently to the Knowledge Consumer.

- The Dissemination Requests, which are used to request the DIPs, may not come from the Knowledge Consumer but from internal activity of the digital preservation platform. Generally, an Order from the Knowledge Broker can be an Ad hoc Order, which is executed only once, or an Event-based Order that will be maintained by the entity Administration. When Access receives and accepts an Ad hoc Order, it will generate a Dissemination Request. And when Access receives and accepts an Event-based (Event-driven) Order, it will send the order to the entity Administration. The entity Administration holds the event-driven requests and checks the request and the storage periodically in order to make sure the event-driven orders are still fulfillable. And if the designated "event" is reached, the entity Administration generates the Dissemination Requests and sends them to the Access entity.
- The Generate DIP sub-function generates DIPs from the AIPs, which come from the Archival Storage entity, according to the Dissemination Requests. The reason why we need both the AIP and Descriptive Information to generate DIP is that the AIP contains less Descriptive Information than the final Descriptive Information stored in Data Management. To acquire AIP, the Generate DIP sub-function calls the Provide Data sub-function in Archival Storage with an AIP Request. And Provide Data sub-function will provide the requested AIPs and transfer them to the entity Access or to a check area. Meanwhile, a Notice of Data Transfer will be sent so as to Generate DIP and make sure the quest is accomplished. As we have discussed before, if the receiver could get the notice in a designated time period, the request would be sent once more. To acquire Descriptive Information, the Generate DIP sub-function sends a report request to Generate Report sub-function in Data Management. The latter will send back a report with all the required Descriptive Information. The Generate DIP sub-function will generate DIP with all the required data, and the mapping between AIP and DIP is just similar as that between SIP and AIP. In other words, the detailed generating DIP process depends on the policies of the digital preservation platform and the Dissemination Request.

- As the Dissemination Request come from the outside of the preservation platform, The Deliver Response sub-function handles both on-line and off-line deliveries of responses (DIPs, result sets, reports and assistance) to the KM approach. As we model the information packages workflow, the delivery can be considered as the end of the workflow. The delivered response will be handled in the Mediation Layer.

4.4.2.4. Updating AIPs

Updating AIPs in the OAIS is a self-submission cycle process. When the updating process is triggered, the entity Administration sends a Dissemination Request to Access and gets the resulting DIPs. The contents of the DIPs will be updated and submitted to the digital preservation platform itself. When Ingest catches the "DIPs", it recognizes them as SIPs, which are submitted to the digital preservation platform.

- Although we can consider the archival information updating process as the learning evolution within an OAIS (based on the Preservation Planning), the process also goes through the normal submission process (only the SIPs are from the submitted DIPs of the OAIS). In the OAIS recommendation, there is no Confirmation of Receipt requested in this kind of SIP submission.
- Although there may not need a Resubmit Request when SIPs inside the OAIS, we may need to confirm the receipt of the packages. Thus, in implementation, sending a Confirmation of Receipt in both situations when receiving SIPs should be sufficient. Nevertheless, when within the OAIS, we have a huge amount archival information to be updated; the sending confirmation information will be an insufficient function wasting also a huge amount of resources.
- The Figure 4.16 shows the sketch of the archival information updating process. This is not a formal activity diagram, but it can show much clearer the self-submission cycle of the digital preservation platform.

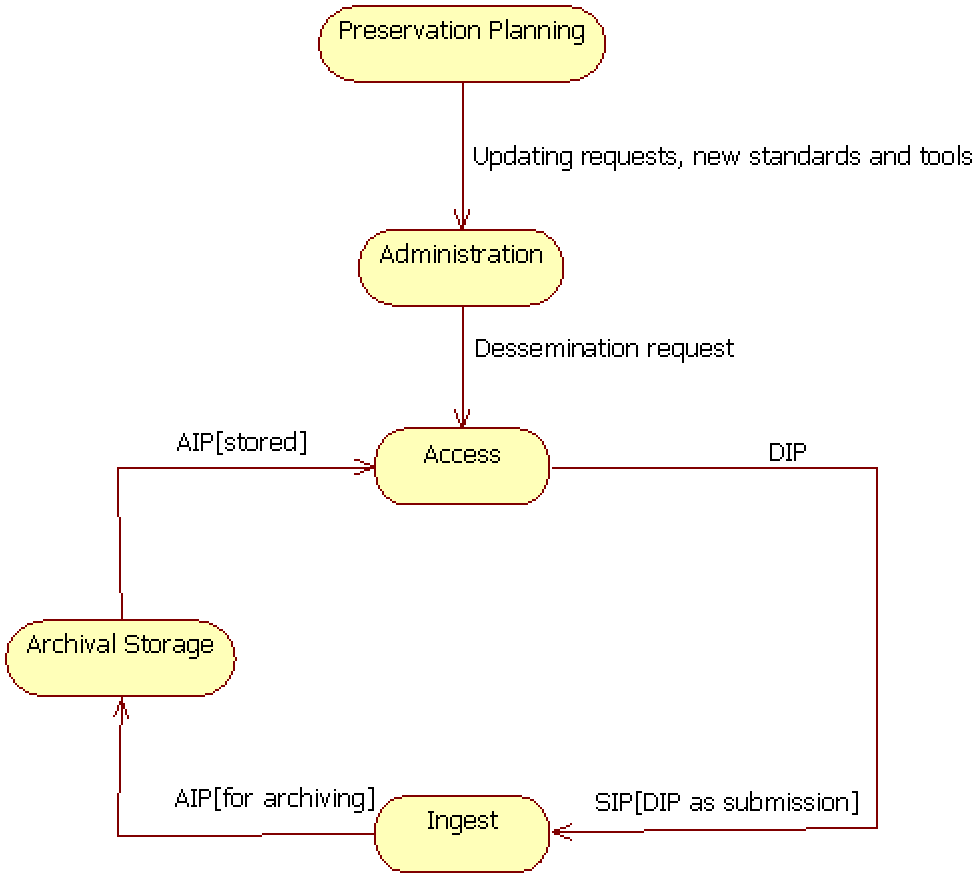

Figure 4.16: Self-submission archival information updating cycle

4.4.3. Mediation Layer

This layer should be developed based on the development of the previous layers. The purpose of Mediation Layer is to enable and enhance the interoperability between Enterprise Layer and Digital Preservation Layer. Interoperability is "the ability of two or more systems or components to exchange information and to use the information that has been exchanged" (IEEE Standard Computer Dictionary). As the digital preservation process includes multiple systems and digital models, we propose to adapt Service-Oriented Architecture (SOA) to support our multi-layer structure of MadPK. SOA configures entities (services, registries, contracts, and proxies) to maximize loose coupling and reuse [McGoven, 03]. SOA stresses interoperability, the ability of systems using different platforms and languages to communicate with each other. Each service provides an interface that can be invoked through a connector. With SOA, we will define the interfaces in terms of protocols and functionalities. This enhances not only the interoperability between digital preservation platform and information systems in enterprises, but also the interoperability between the digital preservation platform and web technical services. The former is the critical point when implementing a preservation platform into an enterprise, and the latter is a solution to overcome the issue of rapid development of technologies, which is against the capacity of knowledge retention.

The Mediation Layer connects the other two layers together, in a dynamic way. In our proposal, we build an Enterprise Knowledge Bus (EKB), which handles the knowledge objects and handles the communications between Enterprise Layer and Digital Preservation Layer. As the services provided by Mediation Layer are all dedicated for knowledge objects, we propose the EKB instead of the Enterprise Service Bus (ESB). The EKB also aims to adapt service concept to reduce the coupling of the platform and consequently to enhance the interoperabilities.

The Mediation Layer may develop multi-level ontology framework to support the structure and the evolution of existing models as well as to ensure the alignment between knowledge models in the other two layers. Ontology is one way to make the digital preservation to be dynamic. However, in our design, we do not specify the methodology for establishing model transfer rules in Mediation Layer, because the model transfer rules should be differ and specified according to the knowledge source types and company's KM strategy. Therefore, we just propose the functional design of Mediation Layer (Table 4.3).

Architecture Info.	**Function**	**Sub-function**	**Associated Model**
Mediation Layer	EKB Planning	Develop Model Tranfer Rules	Organization Model
			Task Model
			Product Model
			Knowledge Model
			Transformation Model
		Configure Connection	Transformation Model
	Knowledge Integration	-	Organization Model
			Task Model
			Product Model
			Knowledge Model
			Transformation Model
	Knowledge Distribution	-	Organization Model
			Task Model
			Product Model
			Knowledge Model
			Transformation Model
	Knowledge Retrieval	Transfer Query	-
		Capture Response	-

Table 4.3: Function Model of Mediation Layer

The functions are published as services according to the connection between the Enterprise Layer and Digital Preservation Layer. The main functions of the Mediation Layer are:

- EKB Planning: EKB Planning should configure the connections between the other two layers. Although the information systems in enterprises and digital preservation platform share and connect to an EKB, the connection configuration should be performed when there are organizational or technical changes of these information systems. In fact the basics of the connection configuration is the model transfer rules, which identify the methods of transferring enterprise knowledge into Information Package form or conversely.
- Knowledge Integration: Knowledge Integration transfers the knowledge from the form in Enterprise Layer to the form in Digital Preservation Layer (Information Package). We call this function "integration", because the formal form of knowledge object in digital preservation platform is Information Package, and the knowledge from Enterprise Layer can be various and in different forms.
- Knowledge Distribution: the Knowledge Distribution function seems like the converse of the Knowledge Integration, but it is not. It provides knowledge for the Knowledge Consumer by distributing and reediting the Information Package. Most of the time, there is no need to transfer the Information Package into certain for knowledge reuse, because normally preserved knowledge in Information Package form is readable and reusable by end users. Therefore the Knowledge Distribution function sometimes is just a connector to transfer the query results.
- Knowledge Retrieval: this function handles queries between the other two layers and sometimes invokes Knowledge Integration or Knowledge Distribution for transferring knowledge.

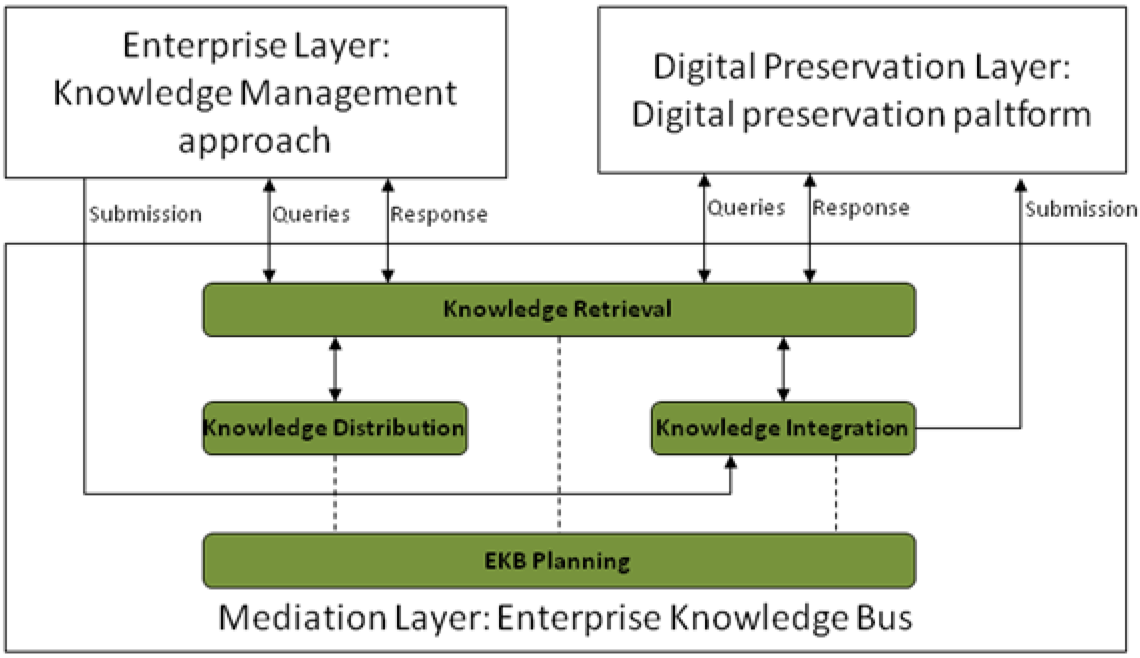

Figure 4.17: Function Model of Mediation Layer

Figure 4.17 shows the Function Model of Mediation Layer. The EKB may represent a piece of software that lives between the business applications (information systems) and enables communication between them. We use EKB to replace the direct contacts with the applications on the bus, so that all the communications take place via EKB. This needs the Knowledge Broker to handle the knowledge model definition and model transfer rules in EKB. We have defined Knowledge Integration and Knowledge Distribution models in EKB, and they recognize the input knowledge and encapsulate or reorganize it into an acceptable knowledge object for either layer. The logic of invoking functions or services in Mediation Layer is shown in Figure 4.18.

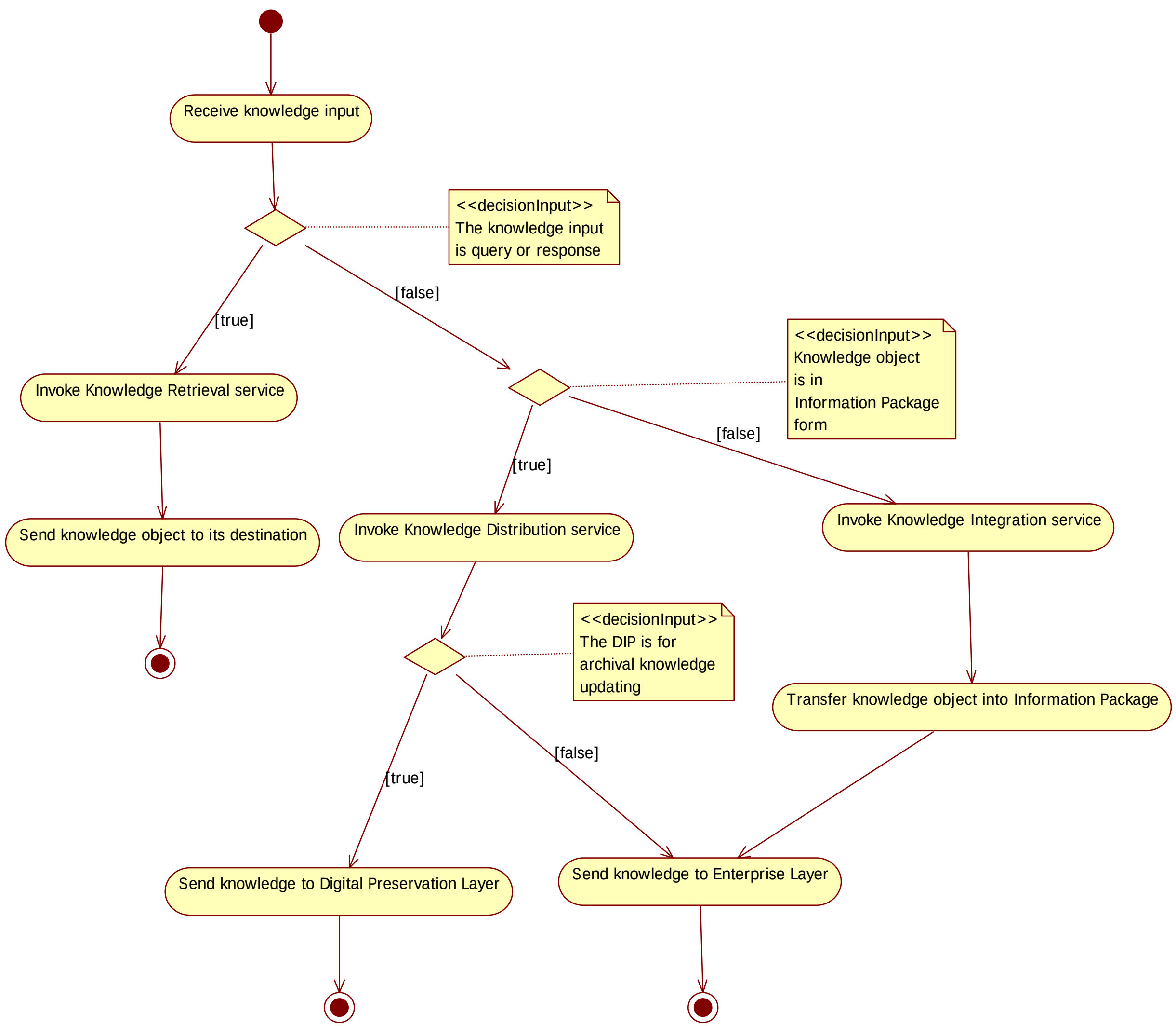

Figure 4.18: Activity diagram of Mediation Layer

One knowledge input of the EKB is encapsulated as a knowledge object, or a simple message. The message that EKB receives can be the queries or the responses to the queries. When receiving these messages, EKB just clarifies the submitter and the destination, and sends the message directly. Nevertheless, when EKB receive encapsulated knowledge object package, it should estimate the form of the package, in order to use model transfer rules to transfer it.

- If a knowledge object package is in an Information Package form, which is the formal structure in digital preservation platform, EKB invokes the Knowledge Distribution function or service, in order to distribute the Information Package to a correct destination (this destination is not always the Enterprise Layer, and sometimes it is the Digital Preservation Platform itself, e.g. in "updating AIPs" process).
- If a knowledge object package is not in Information Package form, EKB must invoke the Knowledge Integration function or service, in order to construct the knowledge object in an Information Package way, based on the model transfer rules.

The Function Model of Enterprise Layer and Digital Preservation Layer have already been shown in **Chapter 2** and **Chapter 3**. Table 4.4 shows the general view of the Function Model of the three layers of MadPK, including Mediation Layer. The functions, sub-functions and associated models in Table 4.4 are identical regarding Table 2.4, Table 3.3 and Table 4.3. Table 4.4 illustrates all the proposed functions together, and sorted by layers.

Architecture Info.	**Function**	**Sub-function**	**Associated Model**
Enterprise Layer	KM Planning	Develop KM Strategy	-
		Capture Critical Change	-
		Design Knowledge Model	Organization Model
			Task Model
			Product Model
	Knowledge Acquisition	Capture Organizational Info.	Organization Model
		Identify Business Process	Organization Model
			Task Model
		Idnetify Product Info.	Organization Model
			Product Model
	Knowledge Evaluation	Handle Query	Organization Model
			Task Model
			Product Model
		Determine Knowledge Obsolescence	-
	Knowledge Reuse	Create Query	Organization Model
			Task Model
			Product Model
		Distribute Knowledge	-

Architecture Info.	Function	Sub-function	Associated Model
Digital Preservation Layer	Ingest	Validate Submission	Knowledge Model
		Generate AIP	Knowledge Model
		Generate Descriptive Information	Knowledge Model
	Archival Storage	Archive Data	-
		Maintain Storage	-
		Provide Data	-
		Terminate Data	-
	Data Management	Archive Data	-
		Generate Report	-
		Perform Query	-
		Filter Obsolescent Data	-
	Administration	Audit	Knowledge Model
		Configure Platform	-
		Configure Policy	-
		Authorize Data Evaluation	-
	Preservation Planning	Develop Preservation Strategy	-
		Capture Critical Change Signal	-
		Design Information Package Model	Knowledge Model
	Access	Handle Query	-
		Generate DIP	Knowledge Model
		Deliver Response	-
Mediation Layer	EKB Planning	Develop Model Tranfer Rules	Organization Model
			Task Model
			Product Model
			Knowledge Model
			Transformation Model
		Configure Connection	Transformation Model
	Knowledge Integration	-	Organization Model
			Task Model
			Product Model
			Knowledge Model
			Transformation Model
	Knowledge Distribution	-	Organization Model
			Task Model
			Product Model
			Knowledge Model
			Transformation Model
	Knowledge Retrieval	Transfer Query	-
		Capture Response	-

Table 4.4: Function Model of MadPK

4.5. Conclusion

We have introduced the detailed design of Architecture Model and Function Model for MadPK (Multi-layer Architecture for Dynamic Preservation of Knowledge) in this chapter. According to the long term knowledge preservation requirements, we analyze the usage of the MadPK. We have identified the roles play in MadPK, according to the actors in KM approach.

Then, interactions between the MadPK roles are presented, and the interactions of roles consequently lead to the interactions and data communications between the layers of MadPK.

According to the use cases of each layer, we propose Function Model within MadPK. In Enterprise Layer, KM approach is performed. The functional design achieves the knowledge creation, knowledge query, knowledge reuse, and knowledge evaluation processes. In Function Model of Enterprise Layer, the Context Level model set (Organization Model, Task Model and Product Model) is adapted as formal documentation structure of extracted knowledge. In Digital Preservation Layer, long term digital preservation approach is performed. The functional design in this layer achieves the submission, archival storage, retrieval of archived knowledge, and knowledge update processes. In Function Model of Digital Preservation Layer, the Concept Level model (Knowledge Model) is adapted as the formal form in repositories. At last in the Mediation Layer, the connections and communications of the other two layers are achieved. The functional design in this layer enables the interoperability of the other two layers and thus achieves the dynamic preservation goal.

Chapter 5. Development and Case Studies

5.1. Introduction

In this chapter, we develop the designs that we have proposed in the previous chapters. During the discussion of the development, we will use case studies to prove the availability of our design on the KM approach, the digital preservation platform and the MadPK architecture. In this way of presentation, we will prove that the MadPK architecture is able to connect the two approaches together in a dynamic way, and thus achieve a dynamic preservation process for the long term knowledge preservation.

5.2. Environment of Development

5.2.1. Business Process Management

Business processes are at the heart of what makes or breaks a business, and what differentiates an enterprise from the competition. Business processes that deliver operational efficiency, business visibility and agility give an enterprise the edge by enabling it to conduct business in a low cost, dynamic way, and to consider the changes as the opportunities for business. A business process is a group of internal activities in enterprises. The purpose of a business process is normally to produce the designated products or predefined services. A business process can include sub-processes or activities. The reason why we introduce Business Process Management (BPM) is that we try to acquire the benefits from the BPM applications in order to achieve the dynamic goals in our research work on long term knowledge preservation.

Business processes are triggered by certain business events or by other business processes. Informally, we can say that a business process is a set of activities, which are executed sequentially. Here, we must introduce the definition of "business function" here. A business function is one specific area that an enterprise or an organization focuses, according to its goals. A business function is considered as a group of activities. In other words, every activity in an enterprise or an organization is one part of a business function. Thus, since a business process is a series of activities, it is also one part of a business function.

In the previous chapters, we have already developed the models, not only in the viewpoint of functionality, but also in the viewpoint of business process. BPM design will cover the implementation of all the proposals and designs. In our research work, the proposed business functions are based on the long term preservation requirements that we have discussed in the previous chapters. In the MadPK architecture, each layer consists of certain specific business processes.

5.2.2. BPM / SOA Environment

The reasons why we have used the SOA concept to design the MadPK architecture have already been stated in **Chapter 4**. However, during the development of the Implementation Level, we have to indicate SOA again, because we have utilized the Oracle BPM/SOA suite [Buelow et al. 10] for the development of our research work. The reason why we choose this to help us in our research is that the Oracle BPM/SOA provides a full set of tools in BPM modeling, SOA design and application generating.

The technology of SOA is based on services. A service has the feature of loose coupling, which provide relative independence to other software. The flexibility offered by loose coupling protects the company or organization, who implements the technology of SOA, from excessive costs when business or technical requirements change. In our research work, the significant challenge of long term preservation is the long term changes of technologies, tools, business and operators of the information which is produced by enterprises. In our proposal of dynamic preservation, we aim to develop an architecture, which is to minimize and deal with the affects of the long term changes.

We utilize business process modeling for the purpose of validating the predesigned models (that is, the structural, functional and data models) and implementing the KM and digital preservation approaches. In the business-process models, the functions that include dynamic features are defined and deployed, thanks to the SOA principle and components, which collaborate with BPM approach (Figure 5.1).

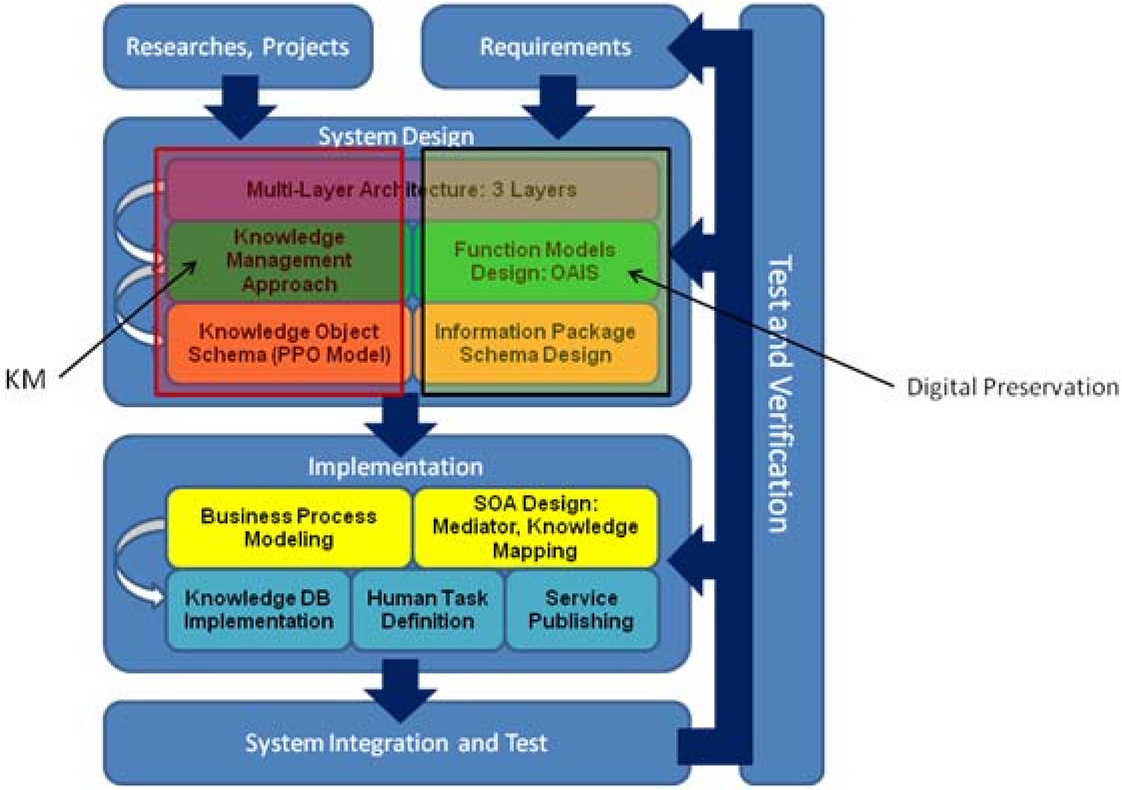

Figure 5.1: Model-based design process and designated models for MadPK [Teng et al. 11]

From our model-based design process, we will implement the KM approach and the digital preservation platform in parallel. We will use similar technologies to implement these two layers. Figure 5.2 shows the framework of our case studies. The reason why we introduce the case studies is that the KM approach should include different characteristics according to the business domain and information systems, which are the knowledge sources.

- Enterprise Layer and Digital Preservation Layer connects by the EKB (Mediation Layer), thus, in the case study, one major mission is to develop the knowledge objects according to different knowledge models in both Enterprise and Digital Preservation Layers, and identify the mapping between them in the EKB.
- After the development of knowledge models and knowledge model transfer mapping, the knowledge objects in the digital preservation platform will be static as Information Package form.
- Enterprise Layer communicates with an ERP system, which in our case is an open source ERP system called FrontAcc. The KM approach will be performed based on the functionalities of the ERP system.

- KM interacts with the domain experts. In our case, the domain experts are in manufacturing department.

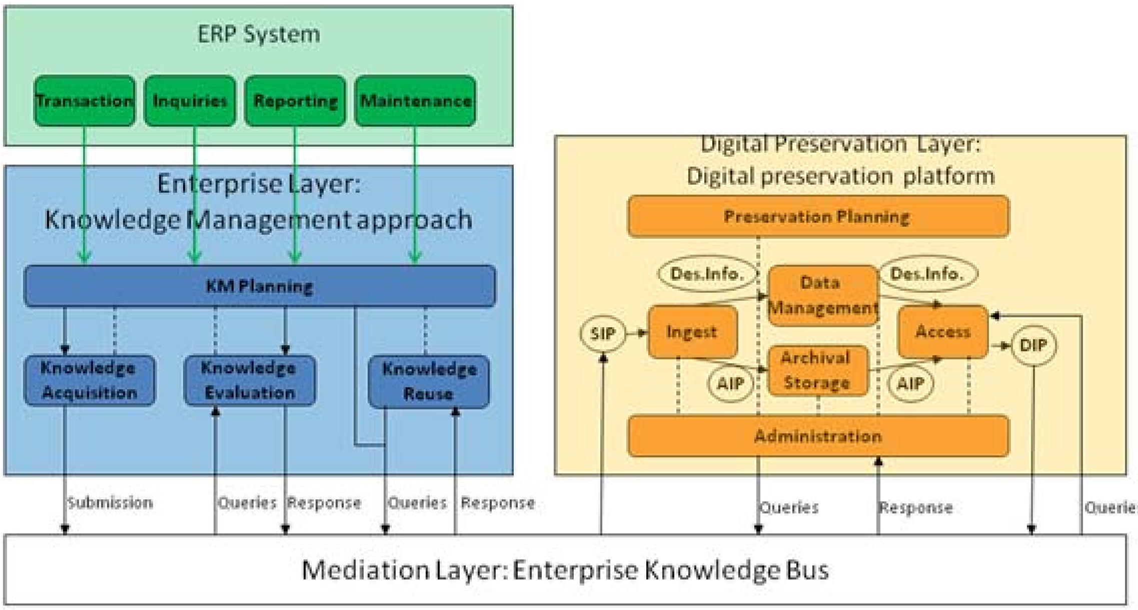

Figure 5.2: Case study framework: integration of an ERP system in MadPK

5.3. Development of MadPK

5.3.1. Knowledge Identification and Conceptualization

When performing the KM approach in Enterprise Layer, the Knowledge Expert does not need to consider the knowledge retention issues. The predefined KM process will lead the acquired knowledge into a designated knowledge model form. Thus, in this layer, we have to:

- Define formal KM processes for knowledge identification and conceptualization, based on KM methodology model sets (Figure 5.3). The KM process identification is in the business viewpoint.
- Define knowledge object structure according to KM Context Level model sets, as well as the ERP system requirements. The knowledge object structure determination is in the data viewpoint.

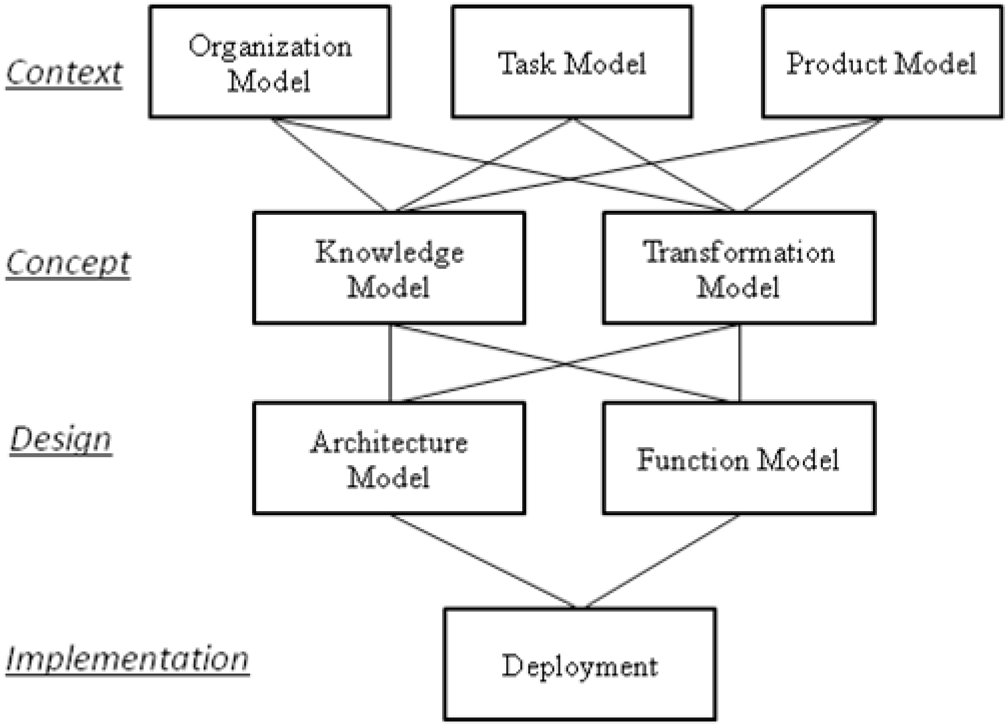

Figure 5.3: KM model sets

Although the identification of KM processes is in the business viewpoint, our KM approach is developed based on the model sets. Thus, the KM processes should adapt the Context Level models. Figure 5.4 shows that the Context Level models (Organization Model, Task Model and Product Model) are composed through the KM processes. At the end of this process, comprehensive knowledge has been collected from the ERP systems and domain experts. Figure 5.4 represents the implementation of the design of "knowledge creation" process in **Chapter 4**. We notice from Figure 5.4, the Organization, Task and Product Models are fulfilled and the organization, task and product information is captured through this knowledge creation process.

We notice that in Figure 5.4, the interactions with database are all performed by services. These services are called by human tasks and receive inputs. Then the inputs go through a mediator and communicate with the database (Figure 5.5). In Figure 5.5, the external references are the database adapters, which operate the databases.

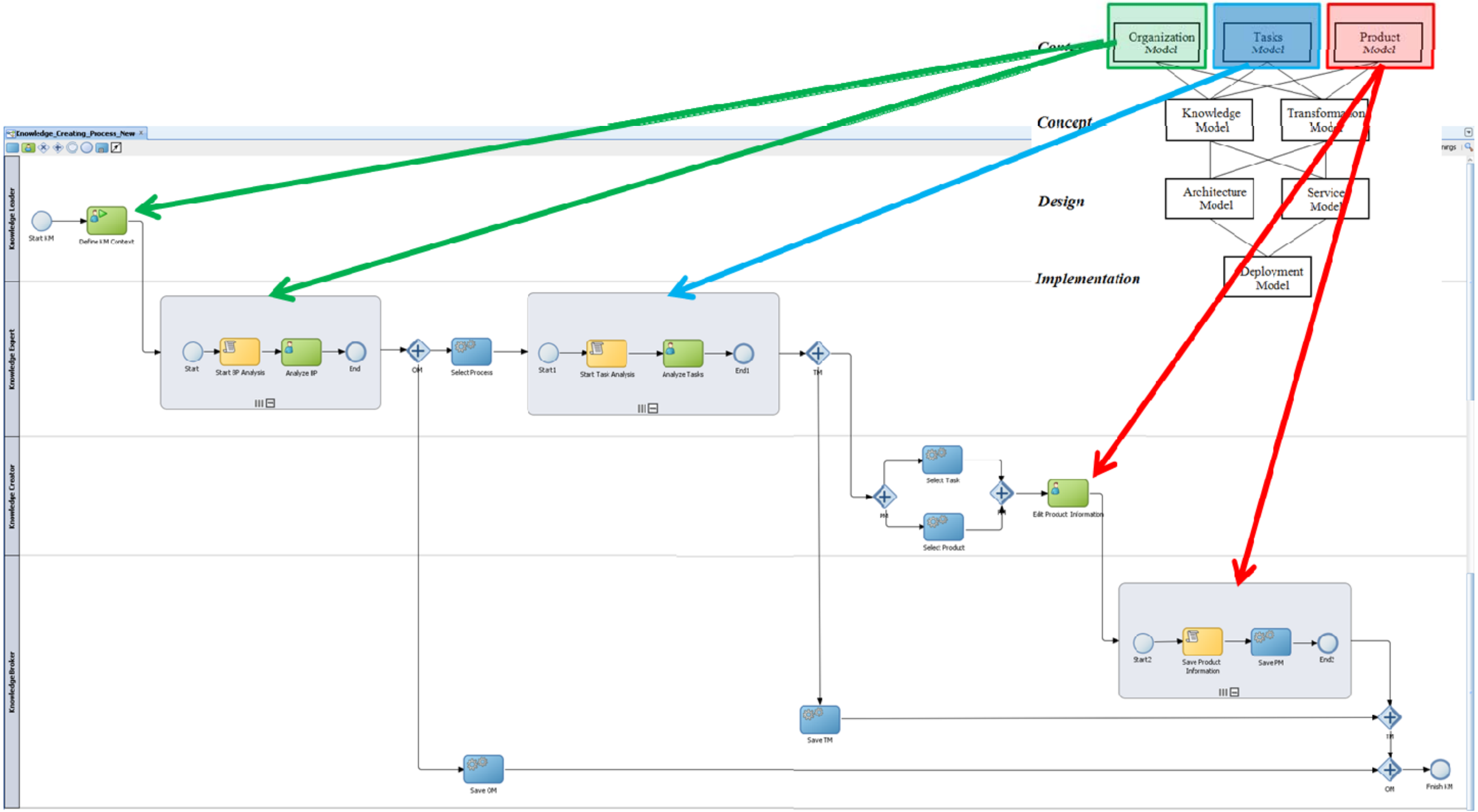

Figure 5.4: Knowledge creation process

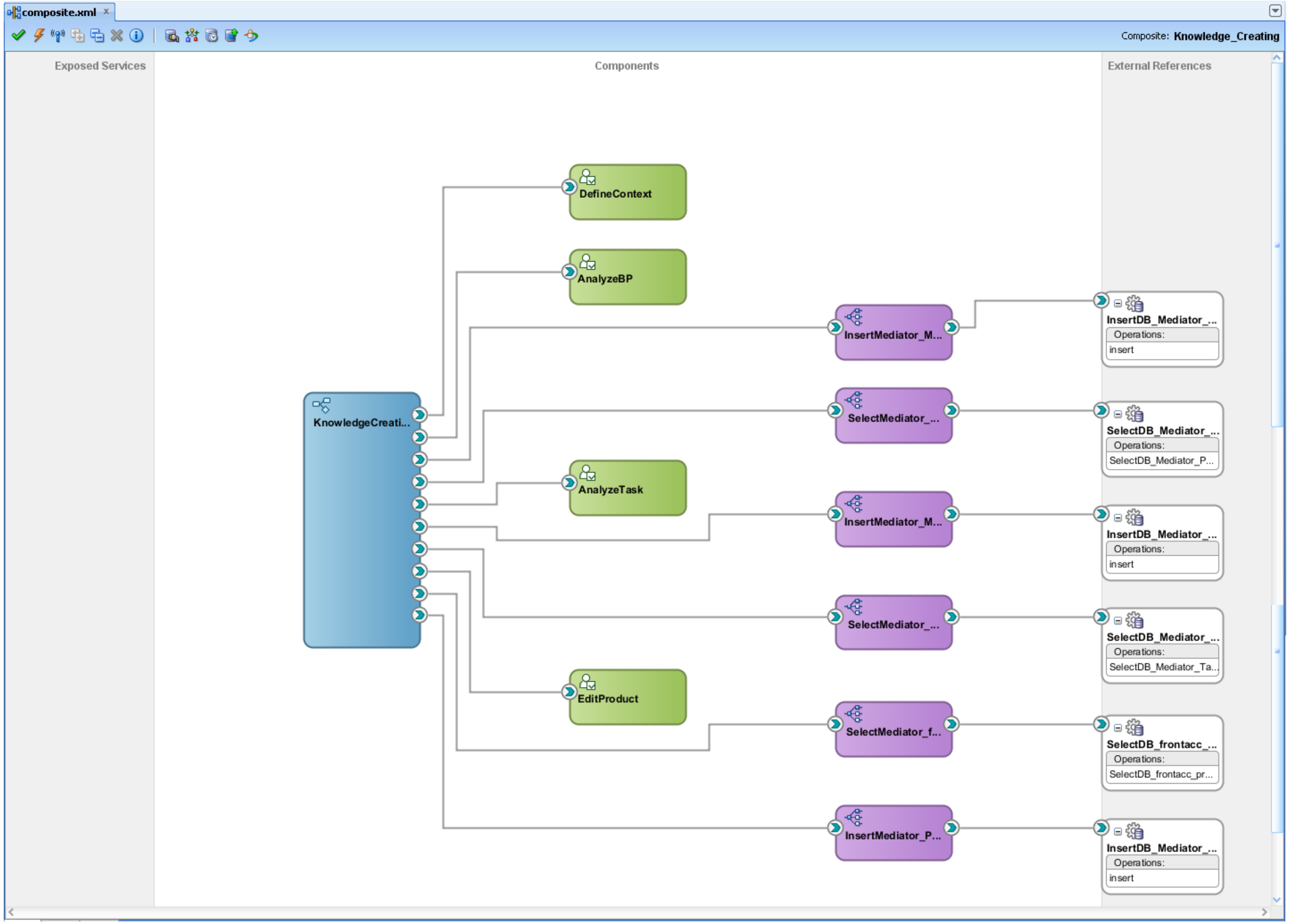

Figure 5.5: SOA composite of knowledge creating process

A simple knowledge object model structure, which is used in our case study, is shown in Figure 5.6. The knowledge model is simple, but fits for all requirements of the extended CommonKADS model sets.

- KM fundamental information: states the objectives of this KM approach;
- Organization Model: represents the organizational structure of the designated organization by decomposing the organization into several departments.
- Task Model: describes the detailed information of the tasks. According to the business function of each department, the major processes are introduced. The major processes are decomposed into tasks. Therefore, the Task Model consists of process information and task information.
- Product Model: represents the product information. In the ERP system, the product information is mainly the Bill of Material (BOM). Thus the Product Model consists of basic product information, the BOM and the component information.
- Resource information: represents the resources in this organization. There are four sorts of resources in this organization:
  - Human resource: the information of staff or labor in this organization;
  - Technical resource: the information systems and technical tools in this organization;
  - Location resource: the information of workshops and other locations in this organization;
  - Service resource: the physical services and the web services, which supports the operations in this organization.

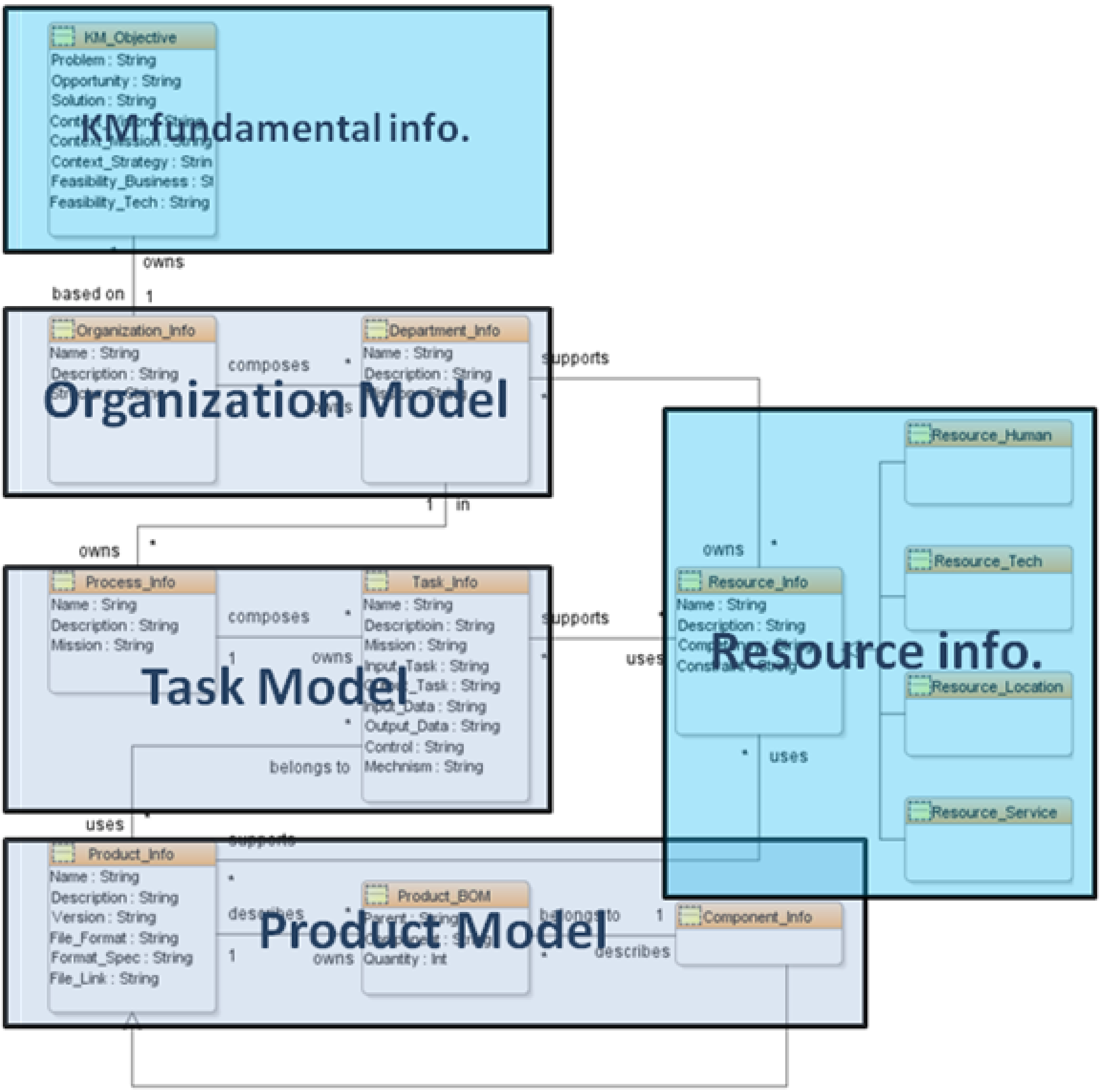

Figure 5.6: Context Level knowledge model

5.3.2. Knowledge Transfer

The output of the KM approach is mass of information in the form of Context Level model set (of the extended CommonKADS methodology). In order to archive the knowledge in the digital preservation platform (i.e. an instance of OAIS), the knowledge should be reorganized and packaged into the Information Package form (i.e. Knowledge Model of the extended CommonKADS methodology). Both the Knowledge Model and Transformation Model of the extended CommonKADS methodology would be used in Mediation Layer (according to the MadPK architecture). Here we have a fuzzy boundary: the Mediation Layer is one layer between the Enterprise Layer and Digital Preservation Layer, but in implementation it can be covered by the KM methodology. Thus, it can be developed together with the KM approach (i.e. Enterprise Layer). The reason why we always consider the Mediation Layer is one separated layer is that it provides mature and static services for knowledge model transfer, while other parts of the KM approach could always change over time according to the organizational and technical status in enterprise.

We call the Mediator Layer the Enterprise Knowledge Bus. In fact it provides services that perform knowledge model transfer, and the knowledge is both the input and output of this "bus". For instance, Figure 5.7 shows that one service transfers knowledge from one form to another. The service has one interface, which is a service object on the "Exposed Service" lane. In the "External References" lane, the emitter object is based on the form of knowledge object, which we have introduced as Context Level model set in the previous section. The receiver object is based on the Knowledge Model, in other words, the Information Package form. However, although we have known both the structures of knowledge objects, we have not directly set the mapping (i.e. Transformation Model of the extended CommonKADS methodology). In order to develop a correct mapping service, or in the end user point of view to choose a correct service from the variety of services in the EKB, we have to determine the long term knowledge preservation objectives. The long term knowledge preservation objectives are determined during the KM planning phase and settled by the domain experts. In the digital preservation platform side, the knowledge is encapsulated according to a formal knowledge model (e.g. Information Package of OAIS reference model).

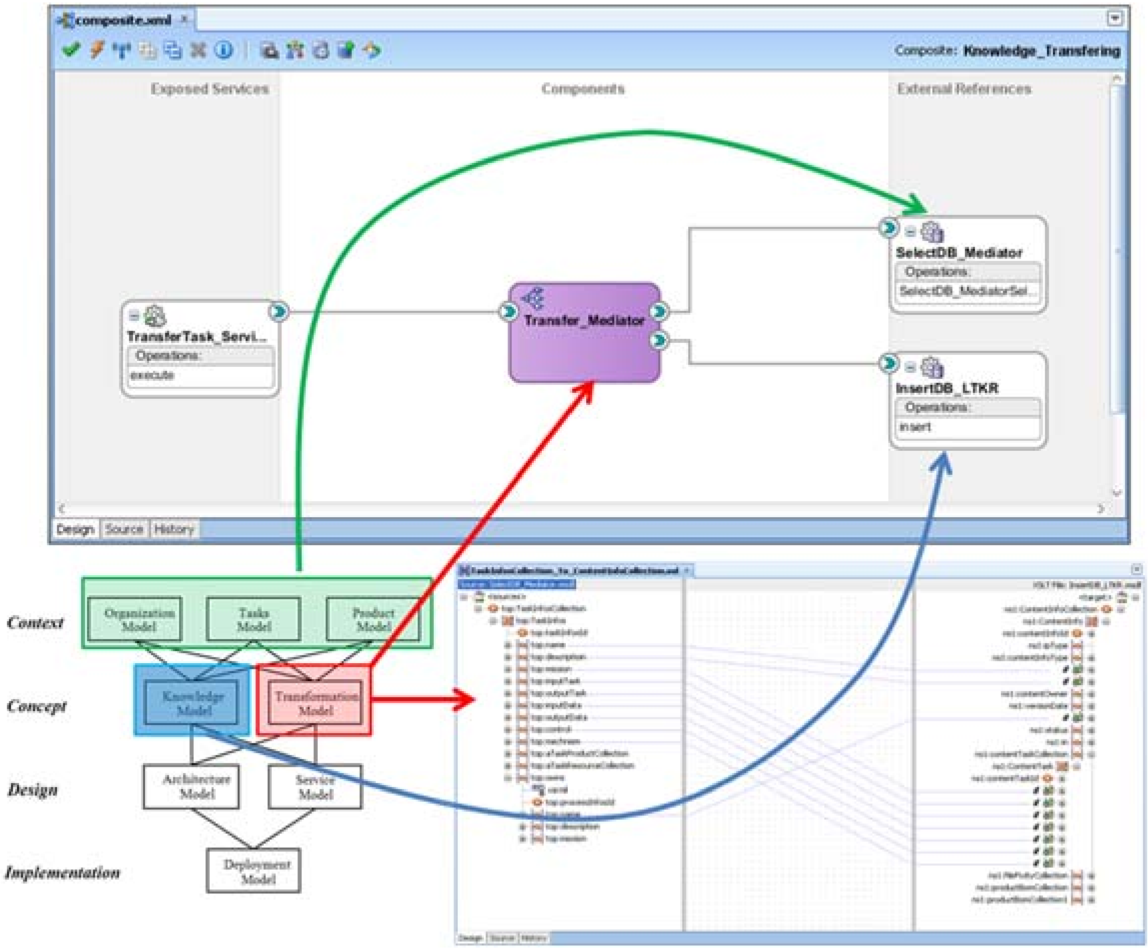

Figure 5.7: Service: knowledge transfer mapping

Generally, in our design, we propose three kinds of long term knowledge models. The first one is product-oriented knowledge model, the second one is task-oriented knowledge model and the third one is separated knowledge model.

- Product-oriented knowledge model: this is easy to understand, and the product (object, item in business processes) is the central knowledge object in the knowledge model. The other information (organization and process information) associates, explains or shows correlations of the product object. For instance, organization objects, which have correlations with one product, will be the metadata of this product, and process object, too.
- Task-oriented knowledge model: one small task or process is considered as the central knowledge object in the knowledge model. The other information (organization and product information) associates, explains or shows correlations of the task or process object.

- Distributed knowledge model: this kind of knowledge model archives product-central package, as well as task-central package. In the metadata of the Information Package, the correlations of the tasks and products are drawn. This kind of knowledge model seems more sufficient than the previous two, for it can archive comprehensive product and task information. However, according to the KM objectives, choosing either of the previous two kinds of knowledge model will simplify the knowledge acquisition process. The KM objectives depend on the information system, and the long term business perspective of the enterprise.

According to the knowledge model type, we define the Information Package model as Figure 5.8. Either task or product information can be archived as central content information.

- Content Information: represents the product information or task information. If either information is considered as the central information, the other one is archived as associated metadata.
- Preservation Description Information (PDI): is the comprehensive metadata to describe the content in this Information Package.
- Descriptive Information: is the identifier of this Information Package. The Descriptive Information could be short summary of the content with some key words. The type of the Information Package (i.e. SIP, AIP or DIP) is also stated in the Descriptive Information.
- Packaging Information: is the links between these data objects. The links connect the data objects together and form a whole package.

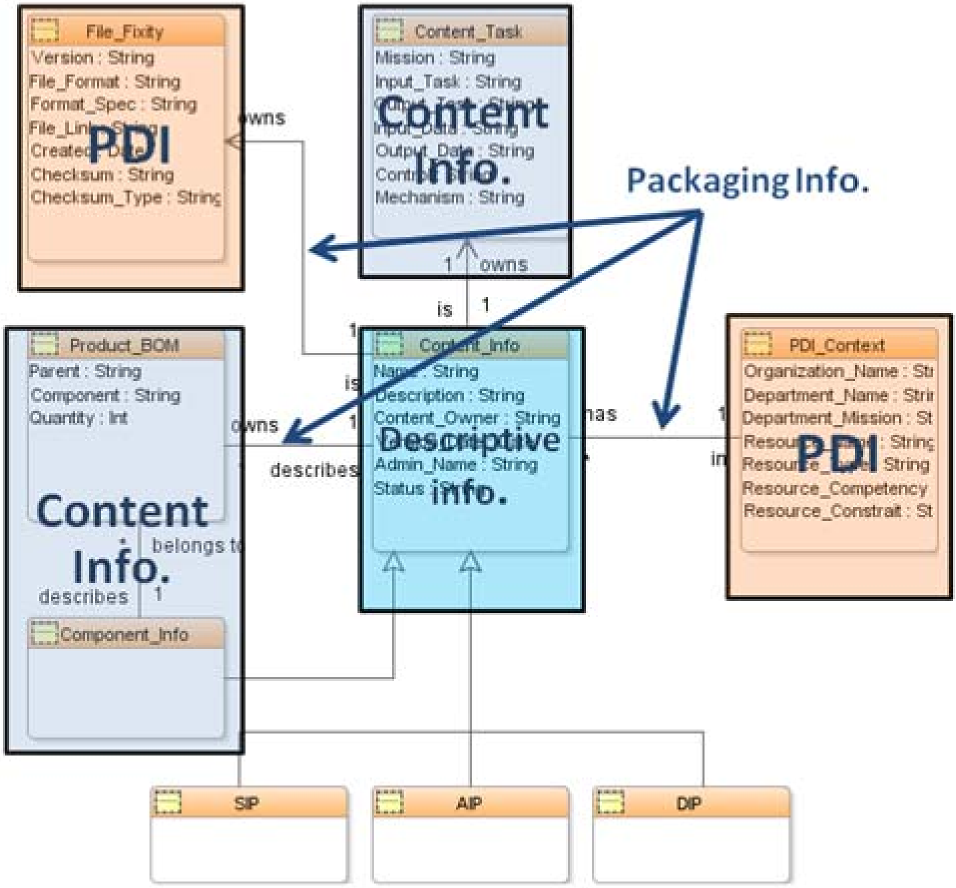

Figure 5.8: Information Package model

5.3.3. Knowledge Ingest and Preservation

This process represents the OAIS functional entity "Ingest" (Figure 5.9). The knowledge transfer service in Mediation Layer stores the Information Packages into the database of the digital preservation platform, with a label of "SIP". The duty of the "Ingest" process is to capture all the available SIPs, complete them and submit them. That is why in the definition of MadPK roles, we need a Knowledge Editor role. Knowledge Editor edits and completes the SIPs. Normally the mission information is versioning and systematic data.

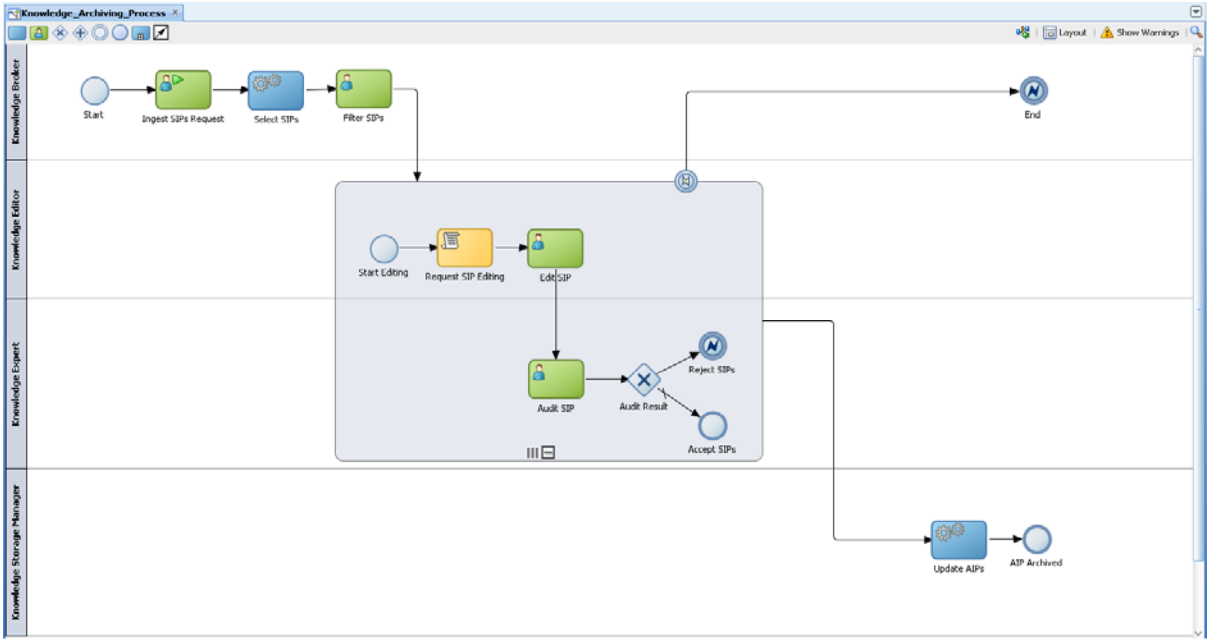

Figure 5.9: Knowledge archive process

5.3.4. Knowledge Update and Obsolescence

The maintenance of Information Package requires the associations from the Enterprise Layer. We have argued that the Enterprise Layer will go through the knowledge evaluation process when getting triggered from the digital preservation platform. In fact the trigger is the knowledge update and obsolescence process of the Digital Preservation Layer. Although the Enterprise Layer makes decisions on all these actions (knowledge update and knowledge obsolescence), the majority of the processes are performed in the Digital Preservation Layer.

The reasons why we put the knowledge update and obsolescence in the same process are that they are all periodical processes and they perform similar actions. The triggers of both the processes are timers: the Information Packages that have not been updated during a designated time t_u has the risk of losing chain of evidence; the Information Packages that have not been required during a designated time t_o may not be needed anymore, in other words, the life cycle of the content information of the AIPs has the risk of expiring. In our proposal, t_u is estimated by the Knowledge Expert according to the business changing frequency of the organization. While t_o is estimated by Knowledge Expert on the lifecycle of the content product or other content information.

The knowledge obsolescence is not mentioned in the original OAIS reference model. However, it is quite possible that some AIPs are no longer needed by its creator or any other end users.

When t_u or t_o is reached, the Knowledge Storage Manager sends the AIPs' information to the Knowledge Expert in Enterprise Layer. The returned message suggests Digital Preservation Layer, either keeping the AIPs, or performing actions on the AIPs. In our implementation, we use a simple way of reflecting the decision related to the knowledge update or obsolescence: if the updating/obsolescing request is rejected, we keep the AIPs; if the updating/obsolescing request is approved, we delete the AIPs. Because in Enterprise Layer, the decision of approving update means that there is enough current information to form a new version of the AIPs. The approving updating decision also triggers another process of knowledge identification and conceptualization, too.

Figure 5.10 shows the knowledge maintenance process, which includes knowledge update and knowledge obsolescence. From the business processes in Figure 5.10, we notice that the maintenance process depend on the logic of decision-makings. We have developed a simple rule for maintenance in our case studies, and in real enterprises, a more complex rule base can be implemented according to the complexity of the long term preservation strategy of the enterprise.

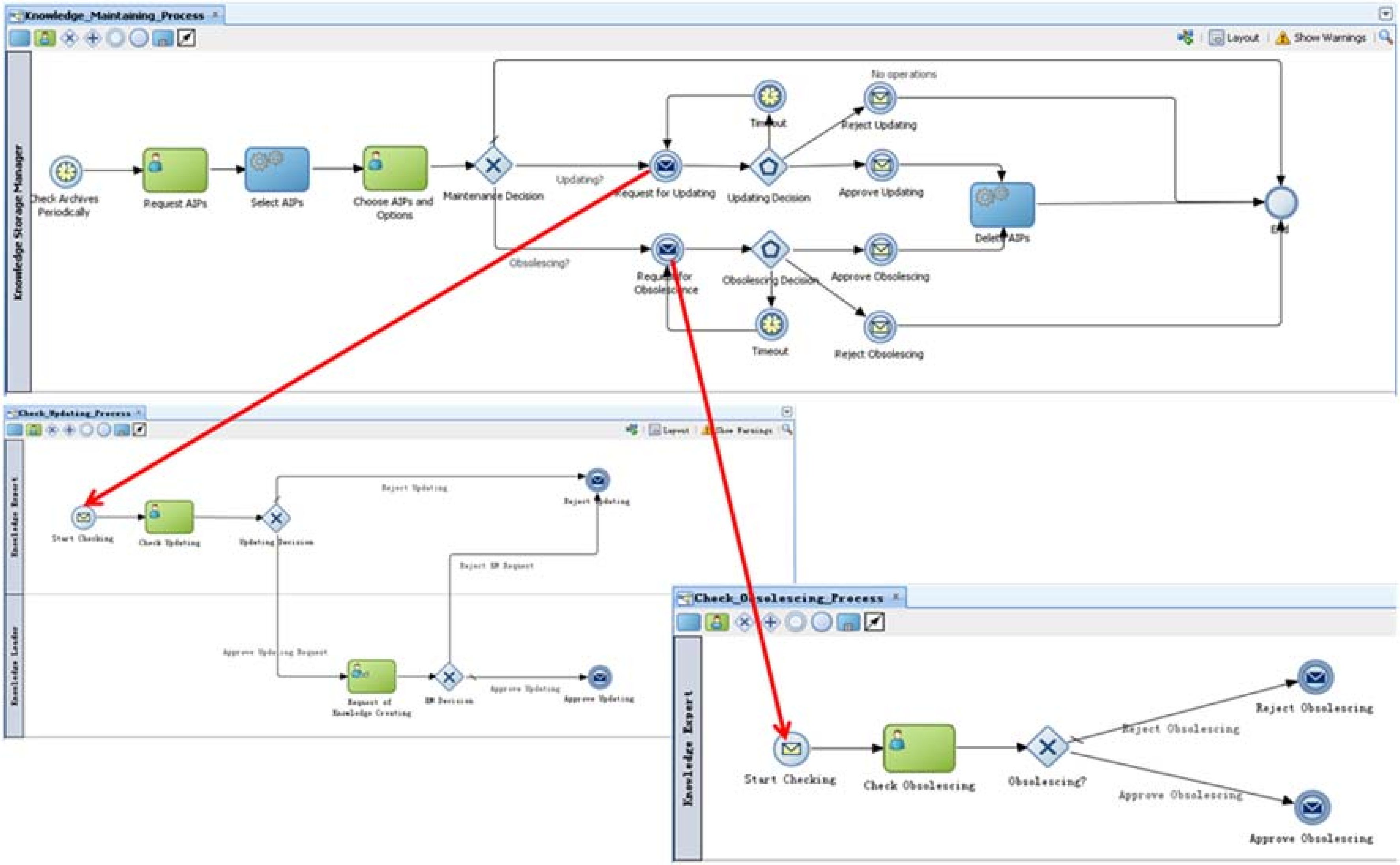

Figure 5.10: Knowledge maintenance process

5.4. Case Study on MadPK Environment

5.4.1. Scenario: knowledge identification and conception on manufacturing department

The manufacturing department needs to perform long term preservation, because the organization uses an open source ERP system, and the system changes frequently. From the enterprise side, first of all, the identification of organization, process and product knowledge should be done.

- Organization: the most common model of description of organization is a tree model. The objects in organization include “department”, “workshop”, “warehouse” and “individual”, which include “worker”, “customer”, “supplier”, etc. Normally the tree model is described in a specific database schema of knowledge source. We just get the data by database querying. In the same time, we keep the organizational tree model itself as an object, and as a part of the knowledge for organization.
- Process: the business processes in enterprise will merely described if we just look into the database schema. Thus after the identification of business processes, we should use modeling method and language (e.g. IDEF0) to define the processes in enterprise. Processes will be modeled in different levels: process, sub-process, task. Activities of processes will finally be decomposed into small tasks. In this way, we have created database schema for models and store them. We will store the models themselves at the same time. For instance, if we use IDEF0 to describe the processes in enterprise or in specific information system, each activity will be modeled like Figure 5.11. The information of activities can be stored in database.

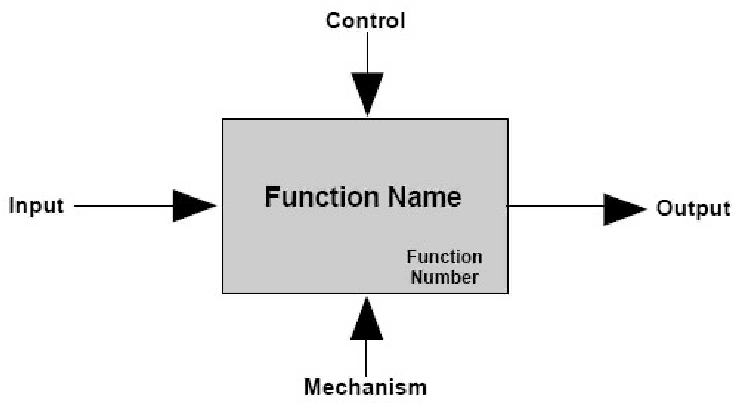

Figure 5.11: IDEF0 box format

- Product: as in the information system of enterprise, there is already definition of product, bills of materials, definition of component, definition of materials, etc. The information will be extracted from knowledge source directly and stored directly.

The worksheet of KM problem and objective is shown in Table 5.1.

Organization Model: OM-1 &OM-5	**KM Problem and Solutions**
Problems and Opportunities	**Problems:** 1. Changing of ERP system frequently, because of using open source platform 2. Data should be preserved even between different system structure 3. Manufacturing work order format should be preserved, because the workshop process will not change frequently
	Opportunities: 1. Separated digital preservation platform, thus data could be safe when updating ERP system 2. Known database schema because of open source platform
Organizational Context	**Vision:** 1. MadPK project, dealing with changes (long term or even short term)
	Mission: 1. Develop knowledge model in Enterprise side according to current ERP system database schema 2. Determine knowledge model transfer mapping 3. Establish knowledge retention network
	Strategy: 1. Knowledge structure in digital preservation platform should be keep static 2. For specific department (e.g. Manufacturing), knowledge model in KM approach should be keep static, so that the mapping service could be re-used 3. Domain expert interacts the KM approach
Solutions	**Solutions:** 1. Establish MadPK project 2. Determine knowledge models and knowledge transfer mapping rules

Organization Model: OM-1 &OM-5	KM Problem and Solutions
	3. Perform KM approach whenever changing ERP system
Feasibilities	**Business Feasibilities:** 1. Knowledge Expert interacts, so feasible on KM approach
	Technical Feasibilities: 1. Multi-layer architecture design complete

Table 5.1: KM problem and solutions – Organization Model: OM-1 & OM-5

This worksheet contains information not only in OM-1, but also in OM-5, this is due to the definition of the Human Task "Define KM Context" in Figure 5.4. Similarly, the OM-2 and OM-3 worksheets are filled in as shown in Table 5.2 and Table 5.3.

Organization Model: OM-2	Organizational Information
Organization	**Organization Brief:** 1. Name: Lyon2; Description: The Lyon2 manufacturing
Structure	**Department:** 1. Name: Manufacturing; Description: The manufacturing department; Mission: manufacturing
Process	**Process:** 1. Name: Transactions; Description: managing work orders; Mission: managing work orders 2. Name: Inquiries; Description: inquiries; Mission: check and modify work orders 3. Name: Reporting; Description: Reporting; Mission: Create reports 4. Name: Maintenance; Description: Maintenance; Mission: Manage BOM and work centers
Resource	**Resource:** 1. Name: FrontAcc ERP system; Description: Open source ERP; Competency: ERP; Constraint: Only accounting

Table 5.2: Organizational Information – Organization Model: OM-2

Organization Model: OM-3	Process Breakdown
1. Transactions	T1-1. Create work order
	T1-2. Edit work order
2. Inquiries	T2-1. Search costed Bill of Material
	T2-2. Search inventory item
	T2-3. Search work orders
3. Reporting	T3-1. Print BOM
	T3-2. Print work order
4. Maintenance	T4-1. Edit BOM
	T4-2. Edit work centers

Table 5.3: Process Breakdown – Organization Model: OM-3

The full worksheets of case study are shown here, in order to validate the design in **Chapter 3**. However, the worksheets will differ according to different KM objectives of the organization. Our Knowledge Model is task-oriented; Table 5.4 shows the detailed analysis of one task in our case study.

Task Model: TM-T1-1	**Task Analysis**
Name	Create work order
Description	Create work order
Mission	Input work order information
Input Task	none
Output Task	none
Input Data	Work order information
Output Data	Work order
Control	Manufacturing manager, RH
Mechanism	FrontAcc ERP system

Table 5.4: Task Analysis – Task Model: TM-T1-1

The special part of the KM worksheet is the Product Model. In this case study, the KM objectives include to maintain the work order structure in order to install new ERP system. Therefore one of the Product Model is the structure of work order (Table 5.5). This work order object is also the input/output data of certain tasks (Table 5.4).

Product Model: PM-1-1	**Knowledge Asset Analysis**
P1. Work order	Name: work order
	Description: 1. Reference 2. Type 3. Item 4. Destination Location 5. Quantity 6. Date 7. Labour Cost 8. Credit Labour Account 9. Overhead Cost 10. Credit Overhead Account 11. Memo
	Version: Date
	File Format: Database
	Format Specification: 1. MySQL Database 2. Known schema
	File Link: none

Table 5.5: Knowledge Asset Analysis – Product Model: PM-1-1

The KM approach is performed by domain expert, but it has not to be performed by hand and sheets. We have constructed user interface according to the structure of the worksheets for the Human Tasks. Figure 5.12 shows a screen shot of one task (i.e. collecting information of OM-1 & OM-5). The Human Task is followed by a service that archives the collected information into database. In fact, even if we did not define a comprehensive user interface, we may always input the information through the interface of the service itself (Figure 5.13). This way is not quite complex and can always be used to test the services we have designed. However, in real operation of the KM approach, we should always use the user interface associated with human tasks, so that the KM approach will go through the business process we have defined and be controlled by the designated business rules.

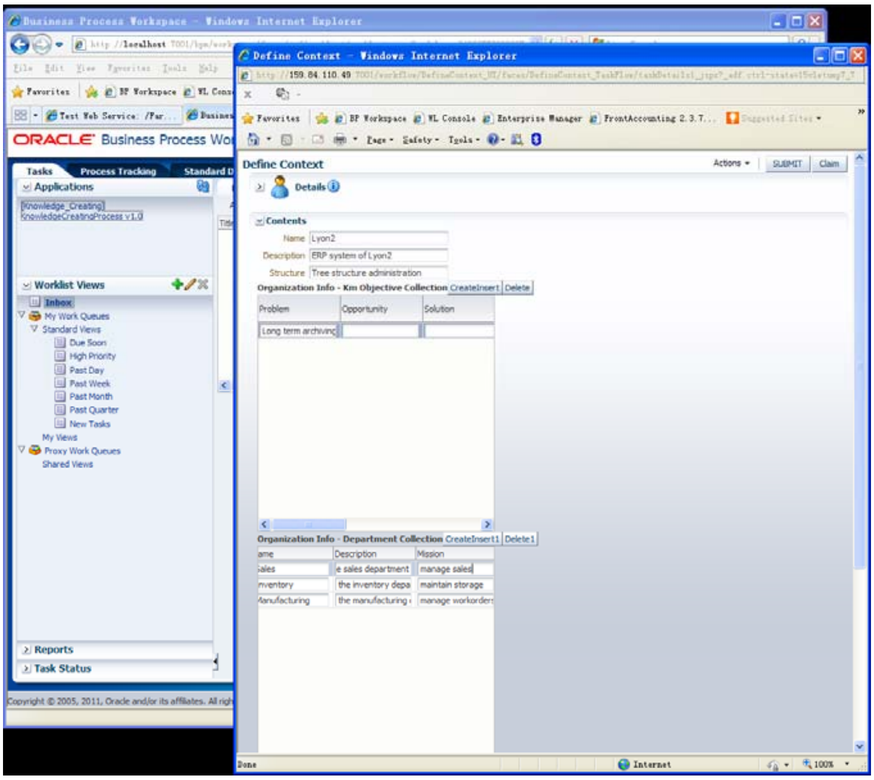

Figure 5.12: Screenshot of Human Task: KM – collecting organizational information

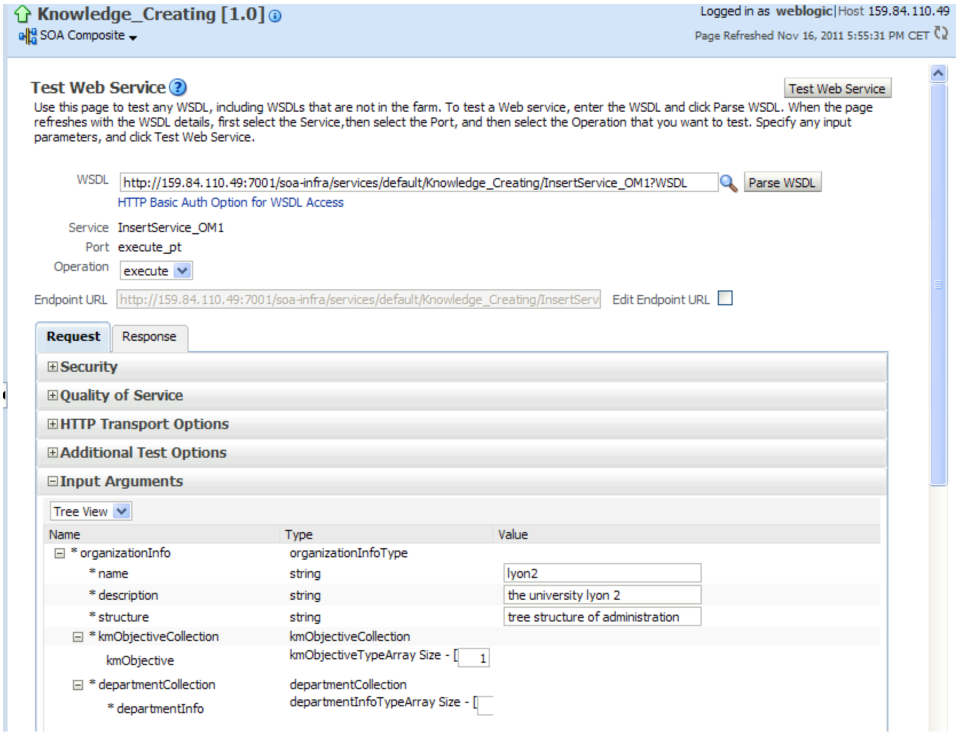

Figure 5.13: Screenshot of Service: KM – collecting organizational information

5.4.2. Scenario: knowledge transfer of task-oriented knowledge on manufacturing department

The task information is transferred into the content table of the database of the digital preservation platform. The organizational context information is stored in Preservation Description Information (PDI) tables. A more visual presentation of the task-oriented knowledge transfer, which represents as the mapping, is shown in Figure 5.14. The mapping shows that:

- KM fundamental information is not transferred, because the information contains decisions regarding the KM approaches. Normally we do not need it for knowledge retention.
- Product information is not transferred in this Information Package, because this is a task-oriented knowledge transfer. The product is just mentioned in the reference information of the task. However, the product information will be transfer into another Information Package as a product-oriented knowledge transfer.

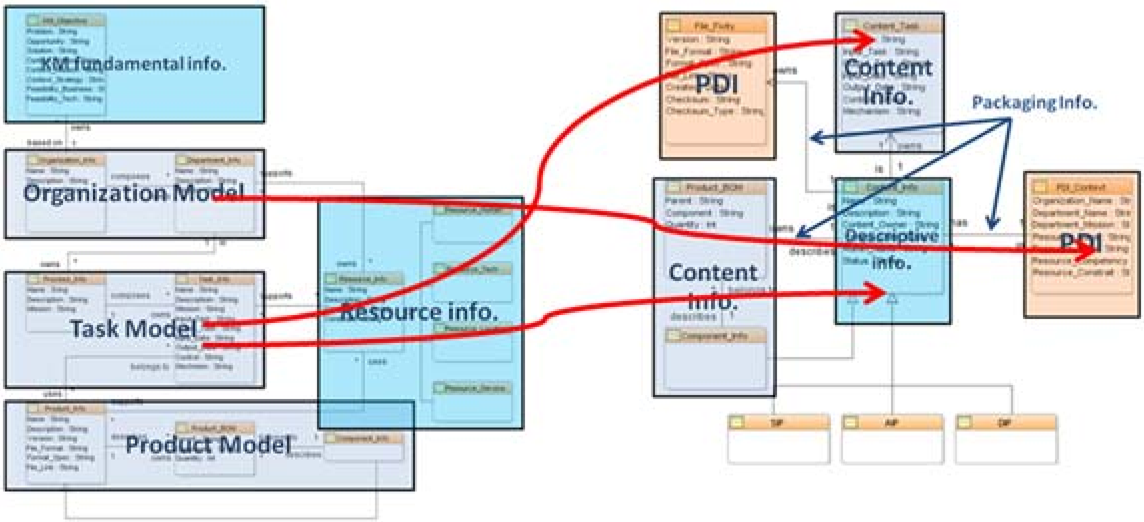

Figure 5.14. Task-oriented knowledge transfer: from KM to digital preservation

5.5. Conclusion

In this chapter, we discuss the development and implementation of our design on the MadPK project: multi-layer architecture, knowledge retention processes. Through the case study, we validate the previous designs and prove that the KM approaches and digital preservation approach can be connected dynamically by services, especially knowledge model transfer services. Of course, the aim of our research is not on detailed and comprehensive knowledge models, but on the MadPK architecture and the construction of reusable functional models and process models. The MadPK architecture is proposed based on the existing long term preservation requirements and existing KM and digital preservation technologies. Ideally, our proposed MadPK architecture can be seen as a generic reference for implementing a long term preservation project in any field.

Conclusion

This thesis presents the research work on proposals of architecture for long term digital preservation of designated knowledge. The objective of this architecture aims improve the scalability, interoperability and sustainability of digital preservation approach/platform in the context of long term knowledge preservation.

The architecture consists of knowledge management (KM) approach and digital preservation approach/platform in the perspective of long term preservation. Regarding long term knowledge preservation, the correlations between KM and digital preservation are tight. Thus we have developed a Multi-layer – Architecture for Dynamic Preservation of Knowledge (MadPK), which is the dynamical combination and coordination of KM and digital preservation. MadPK answers the requirements, which is introduced in previous long term preservation related researches and proposed by our analysis on existing methodologies and technologies.

Contributions

In this thesis, we have achieved the following research results as contributions on the field of methodologies and architectures for long term knowledge preservation:

- The study on KM methodological approach and functional features of tools. The functional analysis of KM methodologies and tools would help to understand the KM approach actual goals and benefit on modeling of additional KM requirements.
- The study on technologies and competencies of existing digital preservation platforms. The actual preservation platforms have already provided functionalities that support long term knowledge preservation. The existing preservation approach also leads to additional requirements of long term preservation.
- The identification of requirements for long term knowledge preservation projects. Through synthesis of the research on KM, digital preservation and long term preservation projects, we identify the gaps between existing approach and long term preservation goals, and the gaps are the key points for developing a more sufficient architecture.

- A proposal of dynamic preservation, which overcoming the requirements of long term preservation we have discovered. Our way of dealing with long term changes and threats in a long term preservation project is to make the preservation "dynamic". "Dynamic" is an ideal perspective, but there exist really some methodologies or technologies (e.g. SOA, etc.) that will support dynamic preservation
- A methodology on KM for long term preservation, which concerns the data object definition of digital preservation. KM methodologies are various, but in order to perform our model-based design approach, we have extended a model-sets based methodology (i.e. CommonKADS) in our research work. The extension of KM methodology includes the software engineering concepts, so this extension aims to transfer the KM approach into applications.
- A construction of a digital preservation approach and platform. The platform is based on the Open Archival Information System (OAIS) reference model, thanks to the precious works of research on digital preservation. OAIS is one sufficient reference in the long term preservation field, which is already proved by many projects' results. Moreover, we add some features (e.g. knowledge obsolescence, etc.) on it, in order to answer the long term preservation requirements that we have identified.
- A Multi-layer Architecture for Dynamic Preservation of Knowledge (MadPK). MadPK is a global logical framework of our research work. It consists of all the proposals we have argued above.

Limitations

As our research work is done through a model-based design approach, our focuses on the architectural construction, business process and functional design. In other words, our research project focuses more on system level and process level than data level. The limitations of our research are:

- No implementation of knowledge transferring methodology or technology (e.g. ontology, etc.). The knowledge objects in our case study are not too complicated. However, if meeting complex business process in enterprise, the results of KM approach would be complicated, and the knowledge model transfer from KM approach to preservation approach ought to adapt methods from, for instance, ontology.

- No adaptation of specific data preservation standards (e.g. STEP for CAD and 3D data, etc.) on knowledge transferring approach. We have proposed a “generic” approach on long term knowledge preservation. Nevertheless, in data level, it is better to adapt standards (e.g. LOTAR project), in order to ease the implementation of long term knowledge preservation and enlarge the scalability of the digital preservation systems/platforms.

Perspectives

We have several perspectives of this research work. These perspectives are based on the objectives and results of the research work, and due to the temporal limitation, technological limitations and organizational limitations we have discussed. The temporal limitation prevents our research from expanding to wider scope in digital preservation in production environment. The technical limitation prevents our research work from integrate more other technologies. The organizational limitations determine that our research work is done mostly on academic environment. We hope that some of the limitations will be broken out and the research work will be in the next level in research, as well as in industrial domains.

- From our preliminary research on digital preservation related projects, we noticed that there are still many ongoing projects, which try to solve the long term preservation problem from different point of view. Some of the researches concern the limitations we mentioned above, and we intend to introduce these methodologies and technologies (e.g. OntoSTEP, etc.) into MadPK, in order to improve the implementation of our proposal in data level.
- Our proposal focuses on the long term preservation of production related knowledge and data. Thus we need to implement this architecture with each type of information systems (e.g. PDM system, MES, etc.), which are used in a product’s life cycle. The PPO design concept has already been studied and developed since the recent years, this is one reason we have integrated the PPO design model into our KM methodology. However, we need to test this KM approach in various information systems, in order to improve the interoperability of the functional designs.
- One other perspective is that we need to test our proposed architecture in complex industrial environment. In complicated situation, data and information has intricate and mixed correlations. These correlations require more supporting methodologies and technologies (e.g. ontologies for acquiring knowledge and transferring knowledge,

etc.). Also, in real industrial environment, the original data structure and data correlations may not easy to access regarding the policy on the control of authorization and business secrets. The architecture, which can be proved in technical dimension, needs even more supports in organizational and legal dimensions.

- In our implementation of the MadPK architecture, the long term knowledge preservation process is well defined, so that the services that we have defined are not too agile, also due to the limitation of our experiences on service design. We need to improve the service design regarding our proposed functional design and architecture design.

References

- A-

Ahmadian, M. Nazari, Z., Nakhaee, N., Kostic, Z. 2005. "Model based design and SDR." The 2nd IEEE/EURASIP Conference on DSP Enabled Radio, Southampton, UK, 19-20 September

Alavi, M., Leidner, D., 1999. "Knowledge management systems: issues, challenges, and benefits", Communications of the Association for Information Systems 1 (1999) 7

- B -

Balafas, P., Jackson, T.W., Dawson, R.J. and Wilson, D., "Accelerated Knowledge Management (AKM): Developing Rapid KM Methodology for an Adaptive Service-Oriented Business Model", Business Innovation in the Knowledge Economy - Abstracts from the IBM & Stratford-Upon-Avon Conference, ISBN: 1857213563. June 2003, pp. 10-12

Banerjee, R., 2004. "A Fool with a Tool Is Still a Fool…", Knowledge Management Tools and Techniques: Practitioners and Experts Evaluate KM Solutions. United States of America. ISBN: 0-7506-7818-6, 2004. pp. 283-292

Barthelme, F., Ermine, J-L., Rosenthal-Sabroux, C., 1998. "An architecture for knowledge evolution in organizations", European Journal of Operational Research 109 (1998), pp. 414-427

Becerra-Fernandez, I., 1999. "This is rocket science: The Knowledge Repository of Kennedy Space Center" Presentation at the Delphi's International Knowledge Management Summit (IKMS'99), San Diego, CA, March, 1999

Benmahamed, D., Ermine, J-L. and Tchounikine, P., 2005. "From MASK Knowledge Management methodology to learning activities described with IMS – LD", Professional Knowledge Management, Springer (Ed.), 2005, pp. 165-175

Blažič, A.J., Klobučar, T. and Jerman, B.D., 2007. "Long-term trusted preservation service using service interaction protocol and evidence records". Computer Standards & Interfaces, Volume 29, Issue 3, March 2007, pp. 398-412

Borghoff, U.M., Rödig, P., Scheffczyk, J. and Schmitz, L., 2006. *Long-Term Preservation of Digital Documents – Principles and Practices*. Springer, 2006. ISBN-10 3-540-33639-7, ISBN-13 978-3-540-33639-6

Buelow, H., Das, M., Deb, M., Palvankar, P. and Srinivasan, M., 2010. *Getting Started with Oracle BPM Suite 11gR1*. Packt Publishing. September 13, 2010. ISBN-13: 978-1-84968-168-1.

Bures, V., 2005. "Methodology of knowledge management implementation", Proceedings of m-ICTE 2005 - Third International Conference on Multimedia and Information & Communication Technologies in Education. Cáceres, Spain

- C -

CASPAR project: http://www.casparpreserves.eu/

Castro, J., Kolp, M. and Mylopoulos, J., 2002. "Towards requirements-driven information systems engineering: the Tropos project", Information Systems, Volume 27, Issue 6, September 2002, pp. 365-389

CCSDS (Consultative Committee for Space Data Systems) Secretariat. 2002. *Reference Model for an Open Archival Information System (OAIS).* CCSDS 650.0-B-1. Blue Book(i.e., "CCSDS recommended standard"). Issue 1. Washington , DC, USA

Chen, D. and Daclin, N., 2006. "Framework for enterprise interoperability", IFAC 2nd International Workshop on Enterprise Integration, Interoperability and Networking (EI2N'2006), Bordeaux, France, March 20-21, 2006

Chen, D. and Vernandat, F., 2002. "Enterprise Interoperability: A standardisation View", Enterprise Inter-and-Intra Organisational Integration, Eds. K. Kosanke et al., Kluwer Academic Publishers, ISBN 1-4020-7277-5. October 2002, pp. 273-282

Chen, D. and Vernandat, F., 2004. "Standards on enterprise integration and engineering – A state of the art", International Journal of Computer Integrated Manufacturing (IJCIM), Volume 17, n°3, April-May 2004, pp. 235-253

Chen, H., 2001. *Knowledge Management Systems: A Text Mining Perspective*, Tucson, AZ: The University of Arizona.

Chen, H. and Chau, M., 2004. "Web Mining: Machine Learning for Web Applications," Annual Review of Information Science and Technology, 38, pp. 289-329

Chen, H. and Fuller, S.S., 2005. Friedman, C., Hersh, W., *Medical Informatics: Knowledge Management and Data Mining in Biomedicine (Operations Research/Computer Science Interfaces).* Springer-Verlag New York, Inc. Secaucus, NJ, USA ©2005, ISBN:038724381X

Comparison Repository Software for Personal Digital Archives: http://www.paradigm.ac.uk/workbook/repository/comparing-findings.html

Cuel, R., 2003. "A New Methodology for Distributed Knowledge Management Analysis", Proceedings of I-KNOW'03, pp. 531-537

- D -

Dave, P., 2007. "The second generation of KM", Knowledge Management, November, pp. 86-88

Deming, W.E., 1994. *The New Economics for Industry, Education, Government*, 2nd edition. MIT Press, Cambridge, MA

DigitalPrservationEurope (DPE) project, 2006-2010: http://www.digitalpreservationeurope.eu/

Dunham, M.H., 2002. *Data Mining: Introductory and Advanced Topics*, New Jersey, USA, Prentice Hall.

- E -

Europea Commission: CORDIS/FP7: TeLearn – DigiCult projects: http://cordis.europa.eu/fp7/ict/telearn-digicult/digicult-preservation_en.html

Ermine, J-L., Chaillot, M., Bigeon, P., Charreton, B. and Malavieille, D., 1996 "MKSM méthode pour la gestion des connaissances", Ingénierie des systèmes d'information, AFCET-Hermès, 1996, Vol. 4, n° 4, pp. 541-575

- F -

Fayyad, U. M., Piatetsky-Shapiro, G., and Smyth, P., 1996. "From Data Mining to Knowledge Discovery in Databases," AI Magazine, 17(3), pp. 37-54

- G -

Geryville, M. H., 2008. *Une architecture d'échange et de partage d'information produit : transformation et adaptation des connaissances dans un contexte d'entreprise étendue*. Doctoral thesis. University Lumière Lyon 2, September 25, 2008: http://theses.univ-lyon2.fr/documents/lyon2/2008/geryville_mh

Giaretta D., 2009. "Digital Preservation Challenges, Infra-structures and Evaluations", Proceedings: Indo-US workshop on International Trends in Digital Preservation, March 24-25, 2009, pp. 1-9

Glandney, H.M., 2007. *Preserving digital Information*, Springer-Verlag Berlin Heidelberg.

Gladney H.M. and Lorie R.A., 2005. "Trust worthy 100-year digital objects: durable encoding for when it's too late to ask", ACM Transactions on Information Systems 2005; 23(3), pp. 299–324.

- J -

John Stark Associates (JSA), 2010, 2PLM, Volume 12 Number 24, 1 March 2010 : http://www.johnstark.com/2plm320.html

Jennex, M.E., 2008. *Current Issues in Knowledge Management*. IGI Global

- K -

KEEP project: Keeping emulation environments portable, 2009: http://www.keep-project.eu/

Kesavan, M., 2009. "Data Repository and Preservation Storage Architecture Using Open Solutions", Proceedings: Indo-US workshop on International Trends in Digital Preservation. March 24-25, 2009, pp.70-76

Kohlbacher, M., 2009. "The Effects of Process Orientation on Customer Satisfaction", Product Quality and Time-Based Performance. Paper presented at the 29th International Conference of the Strategic Management Society, Washington DC, October 11–14, 2009

Krima, S., Barbau, R., Fiorentini, X., Sudarsan, R. and Sriram, R.D., 2009. *OntoSTEP:OWL-DL Ontology for STEP*. National Institute of Standards and Technology, the USA, May 2009

- L -

LiWA project: Living web archives, 2008: http://www.liwa-project.eu/

LOTAR project, 2003. *White Paper for Long Term Archiving and Retrieval of Product Data within the Aerospace Industry (LOTAR).* Technical Aspects of an approach for application

Lubell, J., Kassel, B. and Rachuri, S., 2009. "Descriptive Metadata Requirements for Long-term Archival of Digital Product Models", Proceedings: Indo-US workshop on International Trends in Digital Preservation, Pune, India, March 24--25, 2009, pp. 39--45

- M -

Margolis, B. and Sharpe, J., 2007. *SOA for the Business Developer: Concepts, BPEL, and SCA*, First Edition, MC Press 2007 (328 pages), ISBN: 978-1-58347-065-7

- N -

Noël F. and Roucoules L., 2008. "The PPO design model with respect to digital enterprise technologies among product life cycle", International Journal of Computer Integrated Manufacturing, vol. 21, Issue 2 (Mar. 2008), pp. 139-145

Ntioudis, S., Apostolou, D. and Mentzas, G., 2007. "SAKE: A Methodology for semantic-enabled Agile Knowledge-based E-Government", Proceedings of EURO XXII Conference on Operational Research, Prague, 8-11 June 2007

Ntioudis, S., Apostolou, D. and Mentzas, G., 2007. "A Collaborative Method for Agile Knowledge Management in Public Administrations", Proceedings of TED 07 Conference on e-Government, Enhancing Public Administration Back Offices, Poznan, 27 April 2007

- O -

Open Access Repositories in New Zealand, https://eduforge.org/projects/oarinz/

Open Planets Foundation (OPF): http://www.openplanetsfoundation.org/

Orsvarn K., Waern A., Gala S., Olsson O., Hanssan H.A., Hook K., Gustavsson R., Holm P., Van de Velde W., Breuker J., and Duursma C., 1995. *Specifications of*

CommonKADS – knowledge management and engineering methodology. Swedish Institute of Computer Science, (1993--1995)

- P -

PLANETS project: Preservation and Long-term Access through Networked Services, 2006-2010: http://www.planets-project.eu/

PrestoPRIME project, 2009-2012: http://www.prestoprime.org/

PROTAGE project: Preservation organizations using tools in agent environments, 2007: http://www.protage.eu/

- R -

Rao, M., 2005. *Overview: The social life of KM tools. Knowledge Management Tools and Technologies: Practitioners and Experts Evaluated KM Solutions*. Elserier Butterworth-Heinemann, 2005. ISBN: 0-7506-7818-6

Repository Comparison, 2006: http://outgoing.typepad.com/outgoing/2006/03/repository_comp.html

Rubenstein-Montano, B., Liebowitz, J., Buchwalter, J., McCaw, D., Newman, B. and Rebeck, K., 2001. "SMARTVision: a knowledge-management methodology", Journal of Knowledge Management. Volume 5 . Number 4 . 2001, pp. 300-310

- S -

Saaksvuori, A. and Immonen, A., 2008. *Product Lifecycle Management*. Third Edition. Springer-Verlag Berlin Heidelberg

Schreiber, A.T., Akkermans, H., Anjewerden, A. De Hoog, R., Shadbolt, N., Van de Velde W. and Wielinga, B., 1999, "Knowledge engineering and management: The CommonKADS methodology, The MIT Press.

Schwotzer, T., 2004. "Modelling Distributed Knowledge Management Systems with Topic Maps", Proceedings of the 4th International Conference on Knowledge Management (I-KNOW '04), Graz / Austria June 30 - July 2, 2004

SHAMAN project: Sustaining heritage access through multivalent archiving, 2007: http://shaman-ip.eu/shaman

Sharp, P., Eardley, A. and Shah. H., 2003. “Visual Tools within MaKE - A Knowledge Management Method”, Electronic Journal on Knowledge Management, Volume 1 Issue 2 (2003), pp. 177-186

Strodl, S., Becker, C. and Neumayer, R. and Rauber, A., 2007. “How to choose a digital preservation strategy: Evaluating a preservation planning procedure”, Proceedings of the 7th ACM IEEE Joint Conference on Digital Libraries (JCDL 2007) June 18-23, 2007, Vancouver, British Columbia, Canada. Copyright 2007 ACM 978-1-59593-644-8/07/0006, pp. 29-38

Sureephong, P., 2009. *Characterization of the information and knowledge exchange within the enterprise networks: a Knowledge Management System for SMEs Cluster for a handicraft cluster case in Thailand*. Doctoral thesis, University Lumière Lyon 2 & Chiang Mai University.

- T -

Teng, F., Moalla, N. and Bouras, A., 2010. “A Knowledge Engineering Methodology for Long Term Knowledge Retention (LTKR) in Product Lifecycle Management (PLM) Scope”, International Conference: Competitive and Sustainable Manufacturing, Products and Services (APMS 2010), Cernobbio, Como, Italy, 11-13 October 2010

Teng, F., Moalla, N. and Bouras, A., 2011. “A Model-Based Platform for Product Knowledge Preservation”. INCOSE, INSIGHT: “AFIS Doctoral Symposium: Systems Engineering research challenges in French universities”, Volume 14 Issue 4, December 2011, pp. 28-29

Teng, F., Moalla, N. and Bouras, A., 2011. “A PPO Model-based Knowledge Management Approach for PLM Knowledge Acquisition and Integration”, PLM 11: The IFIP WG5.1 8th International Conference on Product Lifecycle Management, Technische Universiteit Eindhoven, The Netherlands, 11-13 July 2011

Teng, F., Moalla, N., and Bouras, A., 2012. “Dynamic Preservation Approach Design in Digital Preservation Platform”, Ingénierie des Systèmes d’Information, ISSN : 1633-1311, Revue des sciences et technologies de l'information, Hermès Sciences Pub.

Timsina, P., F. Teng, N. Moalla, and A. Bouras. 2010. “Classification Framework for Digital Preservation Platforms: An Evaluation Approach for Digital Preservation Platforms in

Long Term Preservation Scope." Paper presented at the International Conference on Software, Knowledge, Information Management and Applications, Paro, Bhutan, 25–27 August 2010.

- Z -

Zhang, Q., Lim, J. and Cao, M., 2004. "Learning and knowledge creation in product development: a LISREL analysis", International Journal of Product Development, Vol. 1, No. 1, 2004. pp. 107--129.

Annex 1. Analysis of KM Methodologies

A1.1. KM methodology – AKM

The research of AKM methodology [Balafas et al. 03] is part of the preliminary work on KM methodologies. The analysis result of AKM methodology is shown in Table A1.1.

The reason why we have chosen AKM is that it aims to develop KM methodology for long term perspectives. This methodology has been carried out at the Danwood Group in Lincoln and in collaboration with the Department of Computer Science, Loughborough University.

Number	Processes/Activities	How to/Phases	Outputs	Why/Objectives	Tool(s)	Deployment
Step 1	**Identify the Restrictive Business Processes (RBPs)**	Analyse the business process matrix. Identify for each RBP the stage in which the restriction occurs.	RBPs	To identify the business processes that form a restriction to higher performance and achievement of organisational objectives	Analysis tools, business process matrix	
Step 2	**Refine RBPs by Distributing Applied-Knowledge**	Acquire and apply knowledge that has proven to be useful in other areas	Refined RBPs	To attempt to improve the RBPs	Analysis tools, knowledge based system (e.g. EULE)	
Step 3	**Reorganise Non-RBPs for Adaptation to Refined RBPs**	Reorganise non-RBPs according to refined RBPs	Reorganised non-RBPs	To only undertake optimisation when necessary, in order to improving the bottom-line	Modelling tools, analysis tools	
Step 4	**Reengineer RBPs through New Knowledge Development with Accelerated Knowledge Development Cycle (AKDC)**	Knowledge Acquisition (AKDC stage)	Knowledge from a variety of sources (internal and external)	To acquire knowledge	Internal search engies and content-specific alert management tools, internal expert-finding tools, external knowledge acquiring methods	Deploy the internal tools with organization and (if needed) acquiring external knowledge by conducting an external survey, acquiring a knowledge-rich company, subjecting employees to external training, hiring an employee, purchase data sets, monitoring the technological advances, purchasing a patented process, gathering knowledge via competitive intelligence, etc.
		Knowledge Filtering (AKDC stage)	Filtered knowledge	The knowledge from multiple internal and external sources could occur overload without filtering	Knowledge making methods, ontology	By using knowledge making methods and/or ontology to make meta-knowledge for filtering knowledge
		Knowledge Adaptation (AKDC stage)	Adapted and modified knowedge and KM activities in the organization	AKDC knowledge that has been acquired and filtered has to be adapted to the organisational environment and modified inorder to potentially create new knowledge that will address RBPs	Modelling tools (e.g. system thinking tools)	By tying in with business process modelling, knowledge modelling should be done as a support tool used for knowledge adaptation in the AKDC
		Knowledge Distribution (AKDC stage)	Distributed knowledge	Knowledge needs to be distributed and shared throughout the organisation, before it can be exploited at the organisatinal level.	Classic knowledge delivering methods, expert-finding tools	Organising frequent re-training, keeping procedural documents up-to-date, using expert-finding tools to achieve tacit-to-explicit conversion
		Knowledge Embedding (AKDC stage)	Embedded and reviewed knowledge	This stage may result in reducing the restrictions on the performance of the business processes in question and more importantly generate new observations that may be applied elsewhere and eventually embedded in other RBPs.	Techniques for testing new knowledge (e.g. after-action-review technique)	This stage could be identified as the action part of the after-action-review technique and the knowledge review stage corresponds to the review part of the same technique
		Knowledge Review (AKDC stage)	Evaluation of embedded knowledge and back flow to knowledge acquisition, essential information for justification of investment	The evaluation process helps towards deciding whether the new practises that have been developed have found a better way to deal with the restrictions on the business processes involved.	Review methods and/or tools	Knowledge flows back to the knowledge acquisition stage, thus completing the cycle.

Table A1.1: Analysis result of AKM methodology

A1.2. KM methodology – DKM

The analysis result of DKM methodology [Cuel 03; Schwotzer et al. 04] is shown in Table A1.2.

The reason why we choose DKM is that it tries to manage knowledge in a autonomous way and introduces an concept of Knowledge Node (KN), which could be potential knowledge model for long term knowledge preservation.

Number	Processes/Activities	How to/Phases	Why/Objectives	Tool(s)
Step 1	**Understand the main picture fo the firm**		To understand what happens within the firm, which kinds of procedures are developed, the nature of coexistent relations and communications among organizatinal units, etc.	A series of questionnaires and a sort of ethnographic interviews
Step 2	**Unveil Knowledge Nodes (KNs) and theire relations**	Determine the knowledge owner	To reify formal or informal organizational units which exhibit some degree of	
		Determine the system of artifacts		
		Determine a shared conceptual schema		
		Determine processes of meaning negotiation/coordination		
Step 3	**Validation**		To validate the first results through focus groups, or meetings with workers involved in the organization activity	

Table A1.2: Analysis result of DKM methodology

A1.3. KM methodology – MASK

The analysis result of MASK methodology [Ermine et al. 96; Barthelme et al. 98; Benmahamed et al. 05] is shown in Table A1.3.

The reason why we choose MASK is that it provides comprehensive approaches of knowledge capitalization: knowledge analysis and modeling. And the efficiency of the MASK methodology is proved by many KM projects.

Number	Processes/Activities	How to/Phases	Why/Objectives	Tool(s)	Deployment
Step 1	**The systemic analysis**	Produce a definition and description model of the different processes.	To link the domain knowledge wth the operational situation of the system	The systemic analysis	The analysis is a top-down functional analysis where each activity can be hierarchically decomposed into some lower levels sub-activities. The language is close to the widely spread SADT.
Step 2	**The ergo-cognitive analysis**	Produce control flow of the system	It is a representation of the strategy used to solve the problems due to the knowledge system	The ergo-cognitive analysis	The MASK language used to build the task model gives the task scheduling description using a hierarchical recursive decomposition of a high level task into lower levels sub-tasks -- control flow. The control flow is characteristic of the problem solving strategy and can be graphically presented
Step 3	**The psycho-cognitive analysis**	Make a conceptual model that gives the "static" aspect of the knowledge.	To enhance the properties of the concepts and enhance the relationships between the concepts	The psycho-cognitive analysis	Using semantic network and object-oreinted diagram to produce the conceptual model.
Step 4	**The historical and evaluation analysis**	Describe the evolution of objects and techniques among years and experiences.	To reprensent the evolution of a generation: mutation, alteration, bifurcation, interruption, etc.	The historical and evaluation analysis	Techniques and objects will be classified in order to represent their evolution among milestones.

Table A1.3: Analysis result of MASK methodology

A1.4. KM methodology – MaKE

The analysis result of MaKE methodology [Sharp et al. 03] is shown in Table A1.4.

The reason why we choose MaKE methodology is that it concerns the Information System (IS) development with KM approaches.

Number	Processes/Activities	How to/Phases	Outputs	Why/Objectives	Tool(s)
Step 1	MaKE First Step	Process to facilitate creation of knowledge definition for context to which MaKE is applied	Knowledge Definition Template	To define the knowledge used in the organization	
Step 2	MaKE Direct and MaKE Executive	Application of SolSkeme (Sharp 2002b) in an updated form. This involves prioritising Knowledge Targets and articulating means of achieving them	Knowledge Tree, Knowledge Targets Pyramid, Knowledge Blocks	To illustrate and analyze the knowledge and activities in the organization	Knowledge Tree, Knowledge Targets Pyramid, Knowledge Blocks
Step 3	MaKE Measures	Presentation of results and applying measures	Linking Overview	To provide a summary of Knowledge Targets and an indication of the types of cations identified to improve KM	Linking Overview

Table A1.4: Analysis result of MaKE methodology

A1.5. KM methodology – SAKE

The analysis result of SAKE methodology [Ntioudis et al. 07-4; Ntioudis et al. 07-6] is shown in Table A1.5.

The reason why we choose SAKE methodology is that it describes the whole KM/KE approaches from analysis to deployment in the public administration environment. The approaches it describes and the connections of each step in KM/KE is helpful in our research work, which aims to develop an architecture and methodology for long term knowledge preservation.

Number	Processes/Activities	How to/Phases	Why/Objectives	Tool(s)	Deployment
Step 1	**Knowledge as-is analysis**	knowledge audit Selection of most promising focus area(s) and target solution: processes; people; knowledge KM business case	To identify the current state of knowledge infrastructure	CommonKADS method	
Step 2	**Description of selected focus area**	Task analysis (process beakdown) Community & people analysis Knowledge assets analysis	To select focus areas or business process identified in Step 1		
Step 3	**Detailed Knowledge sources analysis**	Detailed specification of identified knowledge resources Pilot-specific extensions of the SAKE ontologies Business process semantic analysis and modelling	The outcomes of this step provide essential information that will assist the agile response of the public administration to changes with the use of the SAKE system	Semantic analysis methods and/or tools, ontology editor, OntoGov Service Modeller tool	
Step 4	**Deployment of the SAKE solution**	Enhancement of bp models CoP building Content annotation	To gather the necessary feedback, as far as functionality issues are concerned	Analysis and modelling methods and/or tools	Semantic anlysis, modelling of business processes, knowledge networks
Step 5	**Delelopment of a trial plan**	Detailed time and resource plan Assessment criteria Identify roles and actions for each stakeholder	To make concise plan for the trial of the SAKE solution	Decision making methods	The plan should include a detailed time and resource plan, with identified roles and actions for each invoved stakeholder
Step 6	**Evaluation of the SAKE solution**	Qualitative and quantitative Evaluation framework for decision making quality assessment in public administrations Decision making qualit ontology instantiation	To drive a qualitative and quantitative evaluation of the performance of the implemented SAKE System on trail.	Evaluation methods, tools	

Table A1.5: Analysis result of SAKE methodology

A1.6. KM methodology – SMARTVision

The analysis result of SMARTVision methodology [Bubenstein-Montano et al. 01] is shown in Table A1.6.

SMARTVision presents a more micro-view of a specific KM methodology, based on the existing KM methodologies. Different from some existing KM methodologies, which is more strategic, SMARTVision has detailed description of each step of implementation of KM in designated organizations. The research on SMARTVision is helpful for implementing KM methodology in our research work, and it discovers more threats and challenges in deploying a KM methodology into applications.

Number	Process/Activies	How to/Phases	Outputs	Why/Objectives	Tool(s)
Step 1	Straegize	Perform strategic planning	Business needs analysis document	To review the current IT infrastructure and document the metrics to be used for measuring success of the knowledge-management procedure	Analysis tools, meetings of decision makers and technical experts
		Perform business needs analysis	Cultural assessment and incentives document	To review the current culture of the organization and outline approaches for encouraging knowledge sharing within the organzation	
		Conduct cultural assessment and establish a motivation and reward structure			
Step 2	Model	Performance conceptual modelling: Conduct a knowledge audit	Knowledge audit document	To survey the status of knowledge in the organization. Emphasis is on identifying core competencies and weaknesses	Knowledge map, modelling tools, decision making tools
			Visual prototype	Knowledge map showing taxonomy and flow of knowledge	
		Performance conceptual modelling: Do knowledge planning	Knowledge-management problem plan	To specify the initiatives and programs that will be used to meet knowledge-management goals	
			Requirements specifications document	To identify the technological requirements for the knowledge-management system (i.e. hardware and software)	
		Perform physical modelling			
Step 3	Act	Capture and secure knowledge	Knowledge-acquisition document	To contain the methods and presumptins used in the process of acquring knowledge for the knowledge-management system based on the findings in the knowledge audit and the knowledge-management program plan	Information system, database, archiving tools and methods, modelling tools
		Represent knowledge	Design document		
		Organize and store knowledge in the knowledge-management system		To contain the knowledge calssification and encoding system as well as high-level knowledge mapping into a computer system (i.e. file structures)	
		Combine knowledge			
		Create knowledge	Visual and technical knowledge-management system prototypes	To present screen-mockups and technical design of the knowledge-management system	
		Share knowledge			
		Learn knowledge and loop back to "Capture and secure knowledge"			
Step 4	Revise	Pilot operational use of the knowledge-management system	Evaluation methodology and results document	To evaluate and review the KM system. This document will evaluated the fitness of the developed KM system for implementation in the transfer phase. Critical analysis of the completed KM system, which includes the determination of whether the program is ready for transfer and will be completed, and recommendations to cotinue development will be evaluated. The documentation of the evaluation methodologies used for the review and the documented results of the review are required.	Evaluation methodologies, modelling tools
		Conduct knowledge review			
		Perform knoledge-management system review	Knowledge-management system prototype II	To construct a pre-procuction, fully functional release of the KM system	
			User's guide for knolwdge-management system	The methods and procedures developed for the KM system are compiled into a guide for use as a training cocument and the coordination of stadard practices. The guide should describe both internal system processes and how the system interacts with the environment	
Step 5	Transfer	Publish knowledge	Maintenance document for KM system	To follow the completion of the final version of the KM system. To describe the general maintenance and create change process for the system	Information system, knowledge sharing tools, post-audit methods and tools
		Coordinate knowledge-management	Fully functinal KM system	To deliver and install the final KM system	
		Use knowledge to create value for the enterprise	Post-audit document	To follow the completed transfer of the KM system and complete a follow-up audit of the entire process. To include all lessons learned, user experiences, best/worst practices and proposed changes to the methodology and /or KM system. To include proposals for new initiatives and enhancements for the system	
		Monitor knowledge-management activities via metrics	Lessons learned document	Lessons learned and other appropriate learning functions will be formatted and loaded into the appropriate corporate memory location for dissemination throughout the organization	
		Conduct post-audit			
		Expand knowledge-management initiatives			
		Continue to learn and loop back through this Step 5			

Table A1.6: Analysis result of SMARTVision methodology

A1.7. KM methodology – KM-Beat-It

The analysis result of KM-Beat-It methodology [Bures 05] is shown in Table A1.7.

KM-Beat-It is quite similar in the KM approaches as the common KM methodologies. However, in the development of KM-Beat-It, some strengths and weaknesses of KM implementation are stated as basis of this methodology. Thus KM-Beat-It really considers the issues of KM implementation, which are also our concerns in long term knowledge preservation project.

Number	Processes/Activities	How to/Phases	Why/Objectives	Tool(s)
Step 1	**Assembly of a realization team**	1.creation of an interest about KM by top management and/or owners of the organization, 2. weighing up of real the possibilities and capabilities to start up the process of KM implementation, 3. decision about implemntation of KM, 4. nomination of team members from the top managemetn, emloyees and external experts, 5. explanation of the presence of sigle team members and definition of their team role, 6. definition of time of employment for each member	To acquire the supoort of the top management and/or owners of the organization and assembly of a realization team that will deal with and will be responsible for the whole process of KM implementation.	Meeting of top management and experts
Step 2	**Analysis of initial state**	1. creation of a survey of knowledge resources, 2. description of knowledge comprised in identified knowledge resources, 3. definition of knowledge processes, 4. analysis of current state of organizational culture, 5. description of organizational processes, 6. finding out the current state of organizational culture, 7. linkage of acquired results, 8. analysis of stengths and weaknesses of the current state in organization	To create an integrated view on the current state in the organization from KM perspective and specification of its strengths and weakness	Analysis tools, meetings of realization team and members from each division of the organization
Step 3	**Creation of a knowledge strategy**	1. definition of a required state, 2. comparison of the current and required state and identification of main gaps, 3. creation ot the list of KM activities, 4. selection of activities, 5. elaboration of plans and projects, 6. creation of knowledge strategy, 7. identification of KM metrics and their relations to the system of organization's metrics	To create a knowledge strategy that will support business strategy and identify particular knowledge activities, which will support the achievement of business and KM goals	Decision making tools or methods
Step 4	**Realization of KM activities**	realize the KM activities according to the outputs of the previous steps	To conduct different actitities, projects or plans leading to KM	Tools or activities to realize KM

Table A1.7: Analysis result of KM-Beat-It methodology

A1.8. KM methodology – CommonKADS

The analysis result of CommonKADS methodology[Orsvarn et al. 95] is shown in Table A1.8.

CommonKADS offers a structured approach to break down and structure knowledge engineering process. CommonKADS provide model-sets for creating requirements specifications for knowledge system.

Number	Processes/Activities	How to/Phases	Outputs	Why/Objectives	Tool(s)
Step 1	Context level	Coping and feasibility study: analysis	OM: 5 worksheets	Identify problem/opportunity areas and potential solutions; Put them into a wider organizational perspective.	Organization Model
		Coping and feasibility study: synthesis	OM: 5 worksheets	Decide about economic, technical and project feasibility; Select the most promising focus area and target solution.	Organization Model
		Impact and improvement study: analysis	TM: 2 worksheets; AM: 1 worksheet	study interrelationships between the task, agents involved, and use of knowledge for successful performance; what improvements may be achieved here	Task Model, Agent Model
		Impact and improvement study: synthesis	TM: 2 worksheets; AM: 1 worksheet	Decide about organizational measures and task changes; Ensure organizational acceptance and integration of a knowledge system solution	Task Model, Agent Model
			Summary: 1 worksheet		
Step 2	Concenp level	Knowledge identification	Knowledge Model	survey the knowledge items; prepare them for specification	
		Knowledge specification		complete specification of knowledge except for contents of domain models	
		Knowledge refinement		Validate knowledge model; Fill contents of knowledge bases	structured walk-troughs; software tools for checking the syntax and find missing parts; paper-based simulation; prototype system
			CM: 2 worksheets	specifies knowledge/information transfer procedures; top-level control over task execution; additional communication tasks	
Step 3	Artefact level		specification of a software architecture; design of the application within this architecture	Specify the architecture of implementation of the KM project	

Table A1.8: Analysis result of CommonKADS methodology

Annex 2. Analysis of KM Tools

A statistics has been done on the commercial KM tools. The work is simple, but messy, because we try to get as many samples for our research on the functionalities of KM tools.

We have done functional analysis of 78 KM software tools, which are developed by 33 different software companies. The reference taxonomy of functional features is provided by Banerjee (Table A2.1) [Banerjee 04].

<table>
<tr><th>Technology</th><th colspan="2">Tool Type / Functional Feature</th></tr>
<tr><td rowspan="6">Knowledge Ceneration</td><td rowspan="3">Meeting Support Tools</td><td>Structuring</td></tr>
<tr><td>Visualizing</td></tr>
<tr><td>Polling</td></tr>
<tr><td colspan="2">Group Decision Support Software</td></tr>
<tr><td colspan="2">Data-mining</td></tr>
<tr><td colspan="2">Online Analytical Processing (OLAP)</td></tr>
<tr><td rowspan="4">Knowledge Codification</td><td colspan="2">Knowledge Repositories</td></tr>
<tr><td colspan="2">Document Management Systems</td></tr>
<tr><td colspan="2">Text-mining</td></tr>
<tr><td colspan="2">Taxonomy Generators</td></tr>
<tr><td rowspan="3">Knowledge Retrieval</td><td colspan="2">Retrieval Systems</td></tr>
<tr><td colspan="2">Search Machines</td></tr>
<tr><td colspan="2">Navigators</td></tr>
<tr><td rowspan="3">Knowledge Transfer</td><td colspan="2">Online Collaboration</td></tr>
<tr><td colspan="2">Online Coordination</td></tr>
<tr><td colspan="2">Training Tools</td></tr>
</table>

Table A2.1: Classification of KM tools [Banerjee 04]

The functional analysis is shown in Figure A2.1.

Company/Author	Name of tool		Structuring	Visualizing	Polling	Assessment of decision making	Datamining	OLAP	Knowledge repositories	Document management systems	Text mining	Taxonomy generators	Retrieval systems	Search machines	Navigators	Online collaboration	Work coordination	TOTAL
IBM	Think Tools	Knowledge generation	1	1	1	1	1	1										6
	DB2 OLAP Server	Data analysis	1	1	1	1	1											5
	Lotus Notes	Client-server, Email management							1	1						1	1	4
MIV	GroupVision	Knowledge generation	1	1	1	1	1											5
Ventana	GroupSystem	Knowledge generation	1	1	1	1	1											5
Brio	Brio Enterprise Server	OLAP	1	1	1	1	1											5
Seagate	Holos	OLAP	1	1	1	1	1											5
Oracle	Express Development Tools	OLAP	1	1	1	1	1											5
Documentum	Enterprise Document Management System	Knowledge codification							1	1								2
PC DOCS	DOCS Fusion	Knowledge codification							1	1								2
	DOCS Fulcrum	Knowledge codification							1	1								2
Open Text	Live Link	Knowledge codification									1	1						2
Geofrey Bock	SemioMap	Knowledge codification									1	1						2
Compassware	Infomagnet	Retrival system											1	1	1			3
Excalibur	Retrieval	Retrival system											1	1	1			3
Verity	Verity Information Server	Retrival system											1	1	1			3
Microsoft	Microsoft Net Meeting	Net Meeting														1	1	2
Sony	Trinicom 500	Net Meeting														1	1	2
Intel	Team Station	Net Meeting														1	1	2
	Proshare	Net Meeting														1	1	2
Rosetta Technologies	PreVIEW	File management														1	1	2
Filenet(IBM)	Visual Workflow	information management														1	1	2
5280 Solutions LLC	Dynamic Filer™	Document capture											1	1				2
	Uconnect® Rapid Application Integration	System integration								1		1				1	1	4
	Dynamic Payables™	Financial							1	1	1	1					1	5
	SharePoint Consulting: Corporate Communications, Workflow and Quick-start	Consulting	1	1	1	1	1	1								1	1	8
A2iA Corp.	A2iA CheckReader™	Document capture	1	1		1												3
	A2iA DocumentReader™	Document capture	1	1		1												3
	A2iA FieldReader™	Document capture	1	1		1												3
	A2iA AddressReader™	Document capture	1	1		1												3
adenin TECHNOLOGIES	IntelliEnterprise	Intranet within enterprise	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	15
	IntelliFolks	Intranet within enterprise			1	1	1	1								1	1	6
Adlib Software	ExpressConversion Server	Document conversion	1	1												1	1	4
	ExpressRecognition Server	Document recognition	1	1												1	1	4
	ExpressPublishing Server	Document publishing	1	1														2
	ExpressEnterprise Server	Document sever in enterprise	1	1					1	1	1	1	1	1	1	1	1	11
Autonomy	IDOL	Intelligent Data Operating	1				1	1	1	1	1	1	1	1	1	1	1	12
	Etalk	Customer interaction			1			1								1	1	4
	Cardiff	Intelligent documentation	1	1		1							1	1	1			6
CA	CA Records Manager	Records management							1	1	1	1						4
	CA Message Manager	Mailbox management							1	1	1	1						4
Captaris	RightFax	Electronic document delivery							1	1	1	1	1	1	1			7
	Alchemy	Document management							1	1	1	1	1	1	1			7
	Workflow	Business process management	1	1	1	1	1	1								1	1	8
Chiliad, Inc.	Discovery/Alert Enterprise Platform	Knowledge sharing											1	1	1	1	1	5
	Discovery/Alert Real-Time Filtering and Alerts					1	1	1										3
	Discovery/Alert Concept Recognizer										1		1	1	1			4
	Discovery/Alert Knowledge Discovery		1	1			1	1										4
	Discovery/Alert Geospatial Service															1	1	2
Collexis Holdings, Inc	BioMedExperts	Expert finding											1	1	1	1	1	5
	Collexis Search	Searching		1						1			1	1	1			5
	Experts Profiling	Expert profiling										1	1	1	1			4
	Knowledge Dashboards	Data analysis	1	1	1		1	1										5
	Mediator	Text-mining		1							1		1	1	1			5
Comintell	Knowledge XChanger	Knowledge exchanging							1	1	1	1	1	1	1	1	1	9
	Matrix Analyzer	Information analysis		1										1				2
Concept Searching, Inc	conceptClassifier for SharePoint	Document Classification										1			1			2
	conceptClassifier	Document Classification										1						1
	conceptSearch	Text-mining											1	1	1			3
	conceptSQL	Data capturing							1	1								2
	conceptTaxonomyManager	Taxonomy management							1	1	1	1						4
Connotate Technologies, Inc	Agent Community GEN2	Communication	1	1	1	1	1	1					1	1	1	1	1	11
	Agent Studio	Knowledge sharing	1	1												1	1	4
	Agent Library	Knowledge sharing											1	1	1	1	1	5
	Agent Portal	Knowledge sharing	1	1	1	1	1	1										6
Coveo Solutions Inc.	Coveo G2B™ Information Access Suite	Management and Searching											1	1	1			3
	G2B™ for Email	Mailbox management											1	1	1			3
	G2B™ for Intranets	Intranet within enterprise											1	1	1			3
	G2B™ for CRM	Knowledge sharing											1	1	1			3
	G2B™ for Multimedia												1	1	1			3
	G2B™ for Custom Applications												1	1	1			3
Ektron, Inc.	Ektron CMS400.NET	Web content management	1	1	1	1	1	1		1	1	1				1	1	11
EMC Corporation	Archiving Solutions	Archiving							1	1	1	1						4
	Business Continuity and Availability Solutions								1	1	1	1						4
	Collaboration Solutions					1	1	1								1	1	5
	Compliance Solutions		1	1	1	1										1	1	6
	Content Management Solutions															1	1	2
	Recovery Management Solutions								1	1		1			1			4
TOTAL			27	29	17	22	19	14	18	21	17	20	26	27	27	28	29	
																	Standard Deviation	2.61419191
																	Max	15
																	Min	1
																	Average	4.37179487

Figure A2.1: Functional analysis of KM tools

Annex 3. Evaluation of Digital Preservation Platforms

Here illustrates the evaluations of digital preservation platforms (i.e. DSpace, Fedora Repository, EPrints) by criteria:

	Platforms		
	DSpace	***Fedora***	***EPrints***
Notice of similarity of submission	0	0	1
Keep Versions	0	1	1
Acquiring data and metadata for new version from old version	0	1	1
Log of versioning	0.5	1	1
As a total	0.125	0.75	1

Table A3.1: Comparison on Versioning

	Platforms		
	DSpace	***Fedora***	***EPrints***
Software requirement	0.75	0.5	0.75
Copy	0.75	0.75	0.75
Repository installation steps	0.5	0.25	0.5
As a total	0.67	0.5	0.62

Table A3.2: Comparison on System Development

	Platforms		
	DSpace	***Fedora***	***EPrints***
Scale up	0.5	0.75	0.5
Scale out	0.75	0.75	0.75
Architecture	0.5	1	0.75
As a total	0.58	0.84	0.67

Table A3.3: Comparison on Scalability

	Platforms		
	DSpace	***Fedora***	***EPrints***
OAI-PMH	1	1	1
SRW/SRU	1	0	0
SOAP	1	1	0
Bulk import and export	0.75	1	0.5
Integration with the other web pages	0.25	1	0.25

	Platforms		
	DSpace	***Fedora***	***EPrints***
As a total	0.8	0.8	0.35

Table A3.4: Comparison on Interoperability

	Platforms		
	DSpace	***Fedora***	***EPrints***
Data transmission	1	1	1
Server security	0.75	1	0.5
Roles and Authentication	1	0.5	0.5
As a total	0.92	0.83	0.75

Table A3.5: Comparison on Security

	Platforms		
	DSpace	***Fedora***	***EPrints***
Archival media and database	0.5	1	0.25
Storage hierarchy	0.75	0.5	0
Backup and disaster recovery of archived content	0.5	1	0.25
As a total	0.59	0.84	0.17

Table A3.6: Comparison on Archiving and Database Management

	Platforms		
	DSpace	***Fedora***	***EPrints***
user interface	1	0.25	0.75
Authorization	1	0.25	0.5
Individuation of user interface	0.25	0	0
submission report	0.5	0.75	0.75
workflow	0.75	0.75	0.5
As a total	0.7	0.5	0.5

Table A3.7 : Comparison on Submission

	Platforms		
	DSpace	***Fedora***	***EPrints***
Configuration of UI	1	0.25	0.75
Configuring system policies	0.5	0	0.5
Configuring module of information package	0.5	1	0.5
Configuring archival strategy	0.25	1	0.25
As a total	0.56	0.56	0.5

Table A3.8 : Comparison on System Configuration

	Platforms		
	DSpace	*Fedora*	*EPrints*
Writing Plugins or other packages	0.5	0.5	0.5
Alter the digital object type including metadata	0.75	1	0.5
Documentation and understanding of code	0.5	0.75	0.5
As a total	0.58	0.75	0.5

Table A3.9: Comparison on Working with Code

	Platforms		
	DSpace	*Fedora*	*EPrints*
Complex inter object relationship	0.25	0.75	0
Referenced metadata	0	1	0
Support content model	0.5	0.75	0.25
Realistic learning curve of system	0 .75	0.5	0.75
Stability monitoring of data and metadata	0.75	0.75	0
As a total	0.45	0.75	0.2

Table A3.10: Comparison on Archival and Administrative Concerns

	Platforms		
	DSpace	*Fedora*	*EPrints*
Multi-language	1	0.75	0.75
UNICODE	1	1	1
As a total	1	0.87	0.87

Table A3.11: Comparison on Globalization

	Platforms		
	DSpace	*Fedora*	*EPrints*
Search engine	0.75	1	0.75
Browser	0.75	0.75	0.75
As a total	0.75	0.87	0.75

Table A3.12: Comparison on Searching adn Browsing

	Platforms		
	DSpace	*Fedora*	*EPrints*
Development community	1	1	0.75
User community	1	0.75	0.5
Supports for the user	1	0.75	0.5
As a total	0.88	0.81	0.56

Table A3.13: Comparison on Community and Support

Annex 4. Conditions of generation of information packages

The generating AIP may involve file format conversions, data representation conversions or reorganization of the content information in the SIPs. The content of the report from Data Management is used to generate Descriptive Information that complete the AIPs. The mapping between SIPs and AIPs is not one-to-one, and it depends on the Data Formatting and Documentation Standards and the SIPs' type. According to the OAIS recommendation, this mapping between SIPs and AIPs can be one-to-one, many-to-one, one-two-many, many-to-many, one-to-zero. And the mapping between AIP and DIP is just similar as that between SIP and AIP.

One SIP – One AIP

A financial department collects all of the year's electronic fiscal records into one CD and submits the CD and one SIP. The archive stores the CD as one AIP. In production process, a design of one part is sent as one SIP and will be stored as one AIP.

In production process, a design of one part is sent as one SIP and will be stored as one AIP.

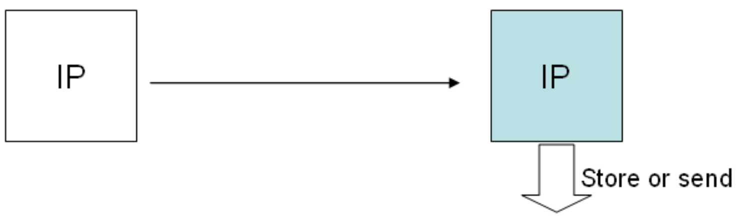

Figure A4.1: One SIP – One AIP

Many SIPs – One AIP

In production process, financial report may be sent as one SIP every week, and the company concerns only the statistics report of finance of the year, so after the whole year's financial reports as well as the annual financial statistics report are arrived, they are stored as one AIP.

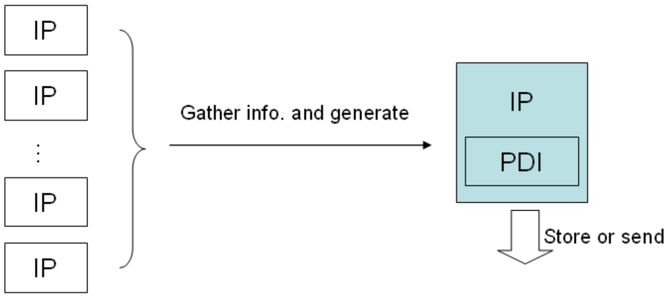

Figure A4.2: Many SIPs – One AIP

One SIP – Many AIPs

In company, all the financial information of a year is submitted as one SIP. The archive has to separate some sensitive or confidential information, and stores the information as two AIPs: confidential and public. This way is easy to control the access to the stored information.

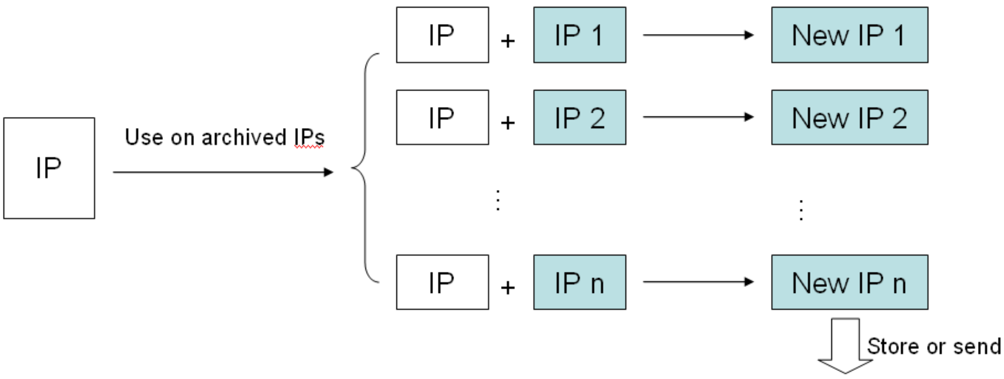

Figure A4.3: One SIP – Many AIPs

One SIP – No AIPs

One company produces a new kind of port standard, and some of the previous parts may support this port. So the information of this information port is sent as one SIP, and the archival system add the proper information to existing archival content, PDI and Descriptive Information. At last the archival packages are updated.

Many SIPs – Many AIPs

This is more useful when generating DIPs. When one consumer wants the technical information of two kinds of cars as well as the comparison of the two kinds, the system will gather the information from multiple AIPs concerning these two cars and gather them to generate the car information reports including all the information of the parts of the cars, and to generate the comparison report for consumer.

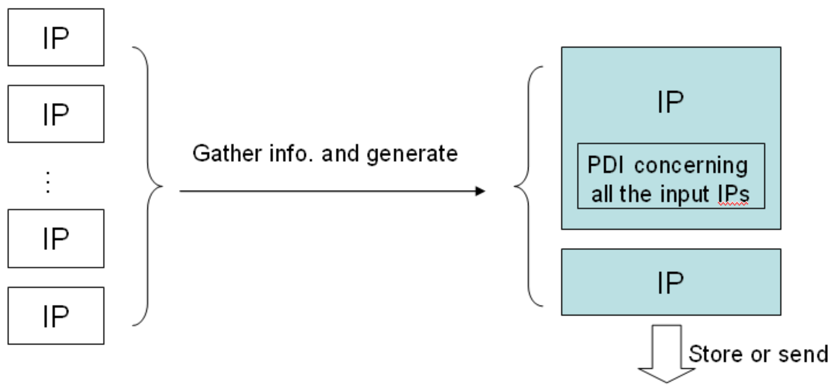

Figure A4.4: Many SIPs – Many AIPs: Gathering and Generating

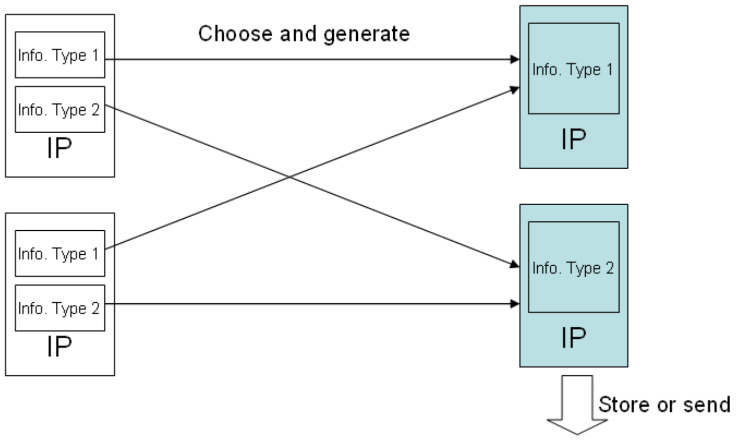

Figure A4.5: Many SIPs – Many AIPs: Choosing and Generating