European Interoperable Infrastructure Services

Study on potential reuse of system components

Final Report Study II

Version 1.1

European Commission
Direcorate-General Informatics
Directorate B - Information systems
e-Commission, Interoperability, Architecture and Methods

This report was prepared for the IDABC Programme by:

Authors' names: Koert Declercq, Anne Vincent, Thomas De Backer, Hervé Loterie,

Company's name: Deloitte

Approved by: Jean Gigot (DIGIT B2) and Serge Novaretti (IDABC)

Disclaimer

TABLE OF CONTENTS

1 Guide for the reader

This document is the final report of the European Interoperability Infrastructure Services (EIIS) Study, drafted after finalizing the Phase 3 of the EIIS study. The report should be of interest to the EIIS stakeholders, including the Information Resource Managers (IRM) of the Directorates General (DG) of the European Commission and the representatives from Member States (MS).

The executive summary provides the context and the outcomes of the study, and is thus adequate as an overview of the main elements for those readers with limited time.

A brief description of the EIIS study (definition, terminology and scope) can be found in Chapter 3.

For those interested in having more details on the study phased approach, Chapter 4 will be of interest. The reading of the annexes is also recommended. They contain a summary of the systems studied in Phase 1 (Annex 8.1), the systems selected for Phase 2 (Annex 8.2), and the detailed description of the services (Annexes 8.3 to 8.11).

Full results of Phase 3 can be found in Chapters 5 and 6. Chapter 5 presents the implementation options for each infrastructure service. Chapter 6 looks into the Service Oriented Architecture (SOA) paradigm and how the European Commission could benefit from this concept in the domain of interoperable infrastructure services.

Finally, the conclusion to the final report can be found in Chapter 7.

Further it should be noted that the intermediary deliverables of the Phase 1 and Phase 2 of the EIIS study are partially overlapping and partially complementary to this final report of the EIIS study. Due to some sensitive content of these previous deliverables these cannot be published on the IDABC website.

2 Executive summary

1. The present report is the final report of the EIIS Study, drafted after finalizing the Phase 3 of the EIIS study. This report summarizes the results of Phase 1 and Phase 2, presents the results of Phase 3 and the final conclusions of the EIIS Study II, started in 2009 by the Directorate General for Informatics at the European Commission (DIGIT) within the framework of the IDABC programme.

2. The main objective of the EIIS Study was to identify and describe common interoperability infrastructure services to support European Public Services. In that perspective, the EIIS Study has selected components in existing systems or systems in development that were best positioned to be part of the solution that could deliver these EIIS. The EIIS Study has also proposed implementation options for the EIIS.

3. The EIIS Study has defined nine European Interoperability Infrastructure Services. These EIIS are: Audit Trail and Log, Service Registry, Identity and Access Management, Document Storage, Workflow Management, Data Certification, Data Transport, Data Translation and Structured Data Storage.

4. During Phase 1, 86 systems, managed either by the European Commission or by the Member States, have been evaluated in order to discover potential system components that could provide generic interoperability infrastructure services. The evaluation of the 86 systems resulted in a list of 50 systems with potential to be reused in a European interoperability infrastructure.

5. Phase 2 provided a detailed description of each service, and for each service, a selection of system components for reuse and a description of the rationale behind this selection. In general, no more than five reusable systems have been selected for each service.

6. Based on the system selection of Phase 2, Phase 3 identified a small set of implementation options for each of the EIIS. Each scenario lists the systems selected in Phase 2 that may fit with the scenario.

7. One of the main findings of the study indicates that the implementation scenarios with federated or distributed architectural topology were preferred to the centralized ones. The scenarios proposed for Service Registry, Identity and Access Management, Data Certification, Document Storage and Structured Data Storage are examples of this preference. These scenarios align with the European principle of subsidiarity, which defines that the central authority should have a subsidiary function, performing only those tasks which cannot be performed effectively at local level.

8. From a practical perspective, it was also noticed that some of the EIIS cannot be provided as reusable services. However, the Member States and the European Commission are also interested in guidance and reusable components in that field. It is the case for Audit Trail and Log and Data translation.

9. Another important finding is that the reuse scenarios of EIIS seem complex to implement, because reuse of services itself is not easy. Reuse in information systems has always been extremely difficult, and many technologies and concepts have promised to increase the reuse: Object- Oriented, Platform independency of Java, software patterns are just some examples. All of these had advantages but faced many practical obstacles.

10. To overcome these issues, the Service Oriented Architecture paradigm (SOA) appears as the more mature approach to reuse and will probably initiate more reuse. SOA is indeed particularly suited to help government agencies deal with the obstacles to implementing the new systems that will enable them to modernize their business architecture, integrate agency service delivery, and share information across organizational boundaries.

The SOA approach is also an opportunity to have a more structured and phased approach to setup the services which support the information exchange between the Commission and the Member States. When the SOA experience and maturity increases, the type of services also changes. This principle of inherent relationship between the type of services and the maturity of a SOA architecture can be applied to the European Interoperability Infrastructure Services.

A recommendation of the Phase 3 of the EIIS study is thus to move to a more Service Oriented Architecture in a stepwise approach, that implements progressively the EIIS, according to the level of SOA maturity. Three implementation waves[1] of the services have been defined in that perspective.

11. Based on these findings and recommendations, the next step is to elaborate interoperability architecture guidelines, and agree on these guidelines. A European Reference Interoperability Architecture could be envisaged as a possible option. It would provide a framework to implement progressively the EIIS to support the information exchange between the European Commission and the Member States. This architectural work must be performed before conducting feasibility studies for the implementation scenarios. It has to be done in close cooperation with the EIIS stakeholders, in order to ensure that the future EIIS respond to their needs and effectively improve the reuse and interoperability between the Member States and the European Commission

[1] Implementation wave 1: Data Transport, Service Registry, Identity & access management, Data certification, Audit trail and log.

Implementation wave 2 : Document storage, Structured data storage, Data translation.

Implementation wave 3: Workflow management.

3 EIIS, the study

3.1 DEFINITION, OBJECTIVES

The Directorate General for Informatics at the European Commission (DIGIT) has started in 2008, within the framework of the IDABC programme, a preparatory study on the European Interoperability Infrastructure Services (EIIS).

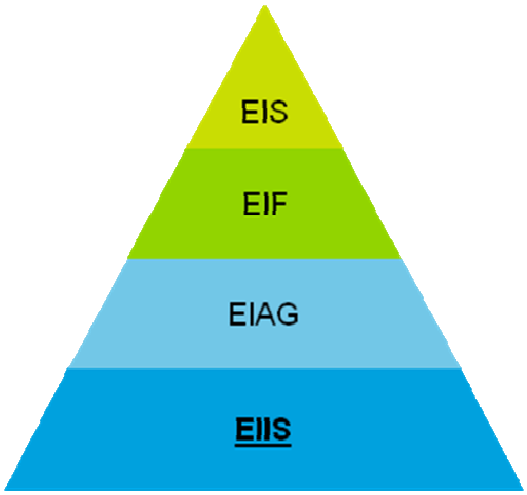
The main objective of the EIIS study was to identify and describe common interoperability infrastructure services to support European Public Services. The EIIS study has selected components in existing systems or systems in development that were best positioned to be reused as such in future systems or as a solution that delivers these (European Interoperability) Infrastructure Services.

The Information Resource Managers (IRM) of the Directorates General (DG) and the representatives from Member States (MS) formed the group of the EIIS stakeholders. The awareness about the study and its outcomes as well as the involvement of the EIIS stakeholders were important to achieve the objectives of the EIIS study. In that perspective communication activities with the EIIS stakeholders were performed all along the study.

3.2 EIIS IN THE CONTEXT OF IDABC

The European Interoperability Infrastructure Services (EIIS) must be positioned in the broader context of the main interoperability initiatives of the IDABC programme. These initiatives include the European Interoperability Strategy (EIS), the European Interoperability Framework (EIF) and the European Interoperability Architecture Guidelines (EIAG).

The EIS, the EIF, the EIAG and the EIIS form the interoperability governance pyramid in which each initiative complements the other. The best practices acquired in the lower levels are fed into the upper levels:



The EIS is at the top of the pyramid. It serves to steer the activities on cross-border interoperability by setting strategic priorities and principles.

At the second level of the governance pyramid steered by the EIS, the EIF defines the general rules and principles for governance and provides practical guidance for the implementation of European Public Services.

The EIAG constitute the third level of the interoperability governance pyramid. They are derived from the EIF and provide structured guidance for implementation. They identify specific standards and specifications for European Public Services, built on a well-defined common architecture.

The EIIS form the basis of the interoperability governance triangle; they are common or generic and reusable components of an operational interoperability infrastructure. The EIIS are linked to the different layers of the governance pyramid and reinforce the interoperability governance pyramid.

3.3 EIIS TERMINOLOGY

Interoperability is the ability of disparate and diverse organisations to interact towards mutually beneficial and agreed common goals, involving the sharing of information and knowledge between the organizations via the business processes they support, by means of the exchange of data between their respective information and communication technology (ICT) systems[2].

Zooming out, an **interoperability infrastructure** represents a set of ICT systems that support the delivery of European Public Services to administrations, citizens and businesses**.** Zooming in, an ICT system can be further broken down in different interacting **system components**, which can be seen as the parts of the system. A system component can be of technical nature (e.g. workflow engine, service register, single-sign-on module) but it can also be of functional nature (e.g. a pattern, a methodology).

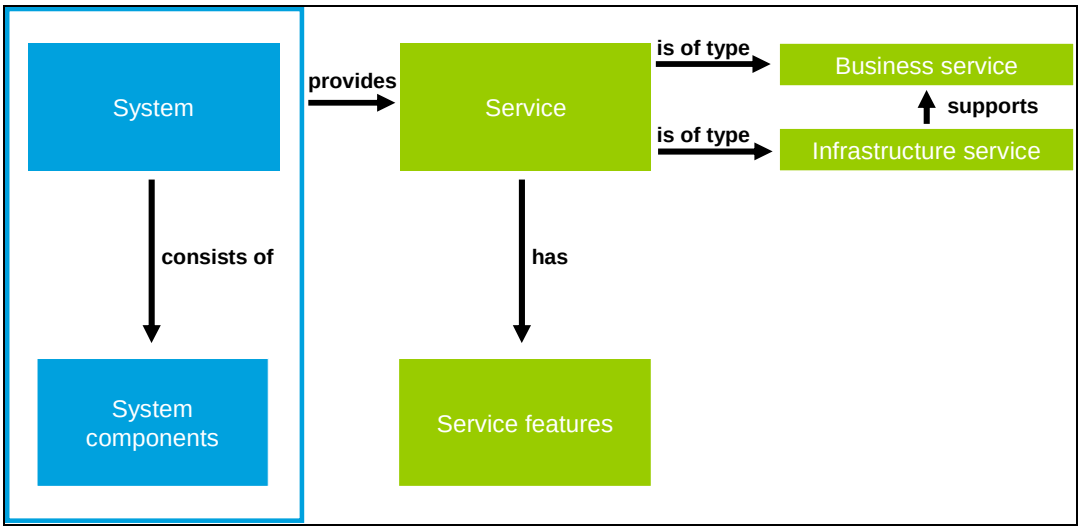
Systems and their components, as part of the interoperability infrastructure, provide **services**. The term service refers to *a discretely defined set of contiguous and autonomous business or technical functionality*, making it important to distinguish business services from infrastructure services.

A **business service** is a 'business functionality' provided by a system to support one or more business processes and which is tangible for end-users. An **infrastructure service** is a generic 'technical functionality' of a system that supports the delivery of one or multiple business services and which is not directly accessible to end-users.

A service, being of business or infrastructure nature, has one or more **service features**. A feature is a distinctive characteristic of a service.

The terms introduced above and their relationships are illustrated in the figure below:

[2] Interoperability solutions for European public administrations (ISA), OJ L 260, 3.10.2009, p. 20 *2009* <*http://eur-lex.europa.eu/JOHtml.do?uri=OJ:L:2009:260:SOM:EN:HTML*>



More details are provided in the EIIS Report Study II – Phase 1.

3.4 SCOPE OF THE EIIS STUDY

The EIIS study aims at:

- Identifying and describing common interoperability infrastructure services to support European Public Services delivery.
- Selecting system components that are best positioned to be part of the solution that delivers these (European Interoperability) Infrastructure Services.
- Proposing a minimal set of implementation scenarios for each of the EIIS.

The study will not:

- Define business services, European Public Services or e-government services.
- Study the component development feasibility which comprises: technical specification, costs & benefits, detailed milestones & schedules, functionalities description, etc.

3.4.1 INTEROPERABILITY AND REUSABILITY

To avoid misunderstandings, it is important to stress the difference between interoperability and reusability, although they are very closely related.

Interoperability is defined in 3.3 EIIS terminology. Interoperability addresses the need for:

- ***cooperation*** between public administrations aiming at the establishment of public services;
- ***exchanging information*** between public administrations to fulfil legal requirements or political commitments;
- ***sharing and reusing information*** among public administrations to increase administrative efficiency and reduce administrative burden on citizens and businesses;

leading to:

- ***improving public service delivery*** to citizens and business by facilitating the one-stop shop delivery of public services;
- ***reducing costs*** for public administrations, businesses and citizens through efficient and effective delivery of public services.

Reuse is key to the efficient development of European Public Services. Reuse means that public administrations confronted with a specific problem try to benefit from the work of others by looking at what is available, assessing its usefulness or relevancy to the problem at hand, and decide to reuse solutions that have proven their value elsewhere.

This implies that public administrations must be willing to share with others their knowledge or/and technology of their existing architecture. Reuse and sharing naturally lead to collaboration, i.e. working together towards mutually beneficial and agreed common goals.

3.4.2 REUSE BY COMMON SERVICES OR GENERIC TOOLS

Reuse can happen on a number of levels and ways. This study will focus on 2 types of reuse:

- Reuse by providing common services;
- Reuse by providing generic tools.

A **common service** has an operational aspect. This service requires a running application and an infrastructure. The infrastructure and application are maintained by a dedicated (centralized) service provider. A good example of a common service which is currently running is sTESTA, which is the European Community's own private, IP-based network. sTESTA [3]offers a telecommunications interconnection platform that responds to the growing need for secure information exchange between European public administrations.

Generic tools could be common components or similar building blocks. It is important to make the distinction with common services. A reusable tool or component is an application that requires installation on the local premises.

Examples of reusable building blocks and best practices are shared on the Open Source Observatory and Repository (OSOR)[4] and some of these components are licensed under the European Union Public Licence (EUPL)[5]. OSOR and EUPL assist, among others, public administrations to share and reuse software components or to collaborate on the development of such components.

Remark on the scope of common services:

The scope of the common interoperability services is limited to interoperability between Member States and the European Commission and between the Member States within the context of the European community policies and legislations.

European Community based interoperability means that the European Commission has received a mandate from the legislators. In the scope of the EIIS study, a mandate implies that Infrastructure Interoperability services in this application domain can be supported e.g. by funding or provisioning of these types of services.

European Community interoperability does not mean that the Commission is directly involved in all communications. For instance, the majority of the interactions within the taxation systems are MS to MS, even though the European Commission owns and funds the taxation systems.

For this scope, the EIIS study will propose scenarios to create reusable components and common services for each of the Interoperability Infrastructure Services. This will be done based on the information gathered and the analysis done during Phases 1 and 2 of this EIIS Study.

[3] http://ec.europa.eu/idabc/en/document/2097/5644

[4] http://www.osor.eu/

[5] http://ec.europa.eu/idabc/eupl

4 Phased approach

The EIIS study has been divided in two consecutive studies:

- **Study I (2008):** Study on operational interoperability infrastructure services. The purpose of Study I was to identify interoperability infrastructure services and modules of existing systems or systems in development that have a potential for providing generic service components within the context of a European Interoperability Infrastructure.
- **Study II (March 2009 – November 2009)**: Study on the potential to reuse systems and system components to provide generic infrastructure interoperability services. The Study II consists of the following phases:

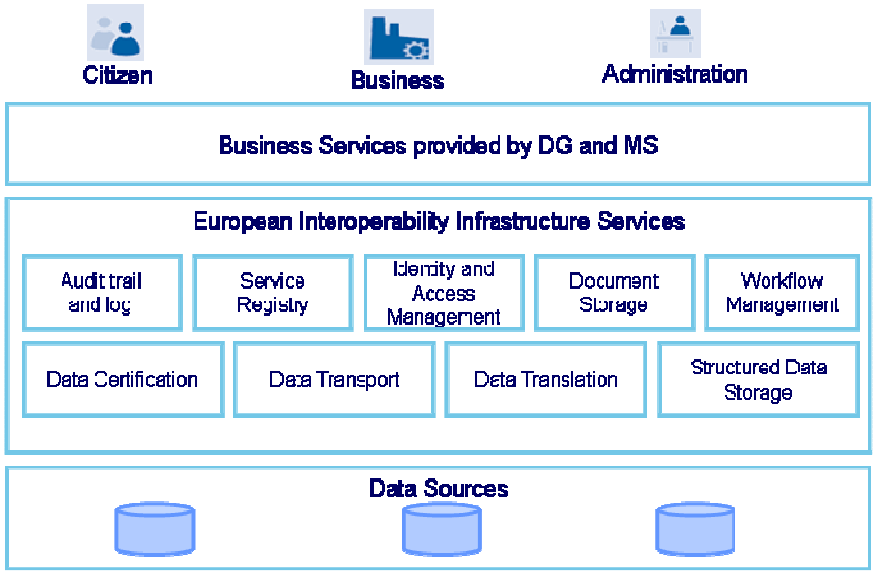
a) **Phase 1: Preparation**. The goal of this phase is to restructure the documentation available from Study I and to complete the documentation where information is missing. The results of this phase are a reference service list, and a high-level documentation of the systems.

b) **Phase 2: Description of the EIIS and selection.** The goal of this phase is to describe the key characteristics of the generic interoperability infrastructure services and to qualify and select the system components for reuse.

c) **Phase 3: Options for implementation.** To goal of this phase is to identify a minimal set of implementation scenarios for each of the EIIS.

4.1 PHASE 1 APPROACH AND RESULTS

During Phase 1 of the EIIS Study II, nine European Interoperability Infrastructure Services have been defined in a reference service list:



In this first phase, 86 systems have been evaluated in order to discover potential system components that could provide generic infrastructure interoperability services. The evaluation of the 86 systems resulted in a list of 50 systems with potential to be reused in a European interoperability infrastructure. All of the systems are managed either by the DGs of the European Commission or by the European Member States.

Among the 50 systems retained, a mapping was established between the systems and the nine interoperability infrastructure services. This evaluation document of the 50 systems and the mapping between these systems and the EIIS reference list are the basis for the selection of the

components in Phase 2. We recommend reading the **Annex 8.1** for the summary of the mapping between the systems and the service list.

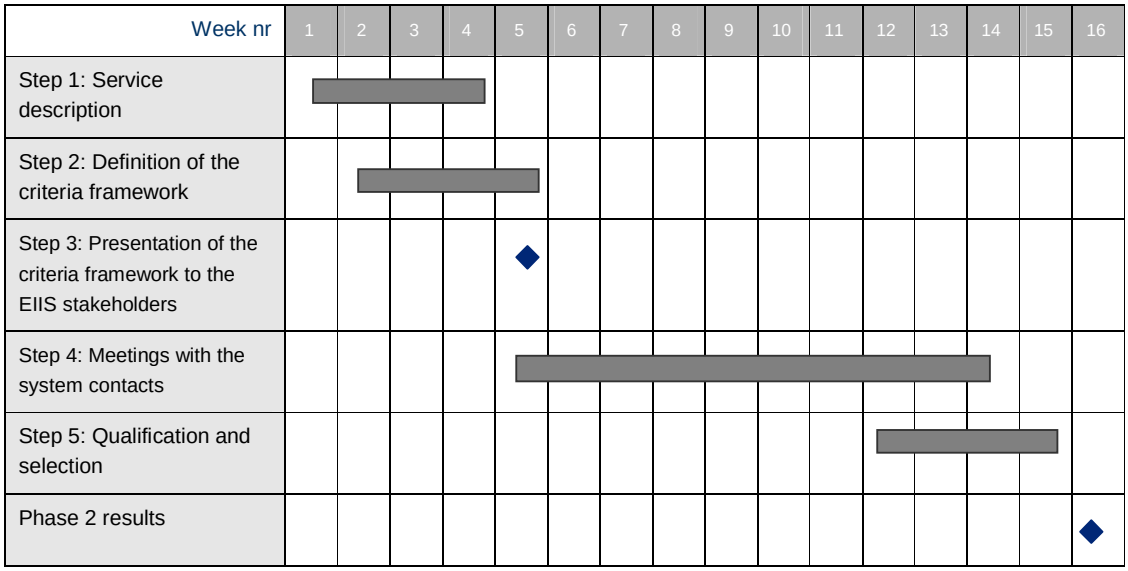
4.2 PHASE 2 APPROACH

During Phase 2, in parallel with the detailed description of the interoperability infrastructure services, it was necessary to:

- Verify whether the services identified in Phase 1 in the different systems were effectively provided by those systems;
- Assess the system components providing the different services;
- Compare the evaluated components;
- Propose a selection of components for reuse.

In order to provide a justified selection of components at the end of Phase 2, the assessment, comparison and selection of the system components must be based on a structured approach. Such approach was adopted in Phase 2 through different activities/steps.

The table below presents these activities and their high level planning in Phase 2. The activities are briefly explained after the table:



4.2.1 STEP 1: SERVICE DESCRIPTION

The goal of this step is to provide a detailed description of the interoperability infrastructure services. Phase 1 described the service requirements in general. In Phase 2, each infrastructure service has been split into multiple service features. All service features are described in detail. The description of the services (especially the service features) should be seen as the functional requirements for potential system components that will deliver this interoperability infrastructure service.

Each of the nine interoperability infrastructure services description follows the same structure:

- definition,
- service features,
- business value,

– background.

The detailed description of the services can be found in the Annexes **8.3** to **8.11**.

4.2.2 STEP 2: DEFINITION OF THE CRITERIA FRAMEWORK

The criteria framework for the system components is based on best practices on system evaluations. The criteria framework is used to evaluate the reusability of system components. It is aimed at making the system selection process transparent and objective.

The selection criteria framework consists of the following criteria:

Criterion	Criterion Descriptions
Functional fit	The functional fit criterion enables to identify gaps in functionality with the required service features.
Scalability	The scalability criterion indicates the reusability of a system component in an international context, at European level. It also qualifies the adaptability of a system for an increased need in the technical capacity (e.g. number of users, bandwidth, number of transactions,…).
Security	The security criterion assesses the balance between security versus flexibility and identifies the alignment of the system component with the international security policies set forward by the Commission and Member States (i.e. confidentiality levels, consistency, integrity of the data).
Standards	The standards criterion qualifies the alignment of the chosen standards with market best practices.
Supplier viability	The supplier viability criterion assesses the dependence of the Commission or the Member State on the supplier, which might be an open source or proprietary software supplier.
Operational maintenance	The operational maintenance criterion enables to identify the level of effort required to operate and maintain the system, e.g. number of suppliers, availability of knowledge on the market, relative cost of operations, maintenance of parameter tables,...
Training & Documentation	The ' Training and Documentation' criterion assesses the maturity of the documentation of the system and its components. Good documentation facilitates sharing of information and expertise.

In order to make the evaluation more comprehensive, the selection criteria have been grouped in three dimensions: functional fit, technical fit and operational fit. The criteria dimensions are shown in the table below:

Criterion Dimension	Criterion
Functional fit	Functional fit
Technical fit	Scalability
	Security
	Standards
Operational fit	Supplier viability

	Operational maintenance
	Training & Documentation

4.2.3 STEP 3: PRESENTATION OF THE CRITERIA FRAMEWORK TO THE EIIS STAKEHOLDERS

A Webinar was organized on 30 June to inform the EIIS stakeholders on the current status of the EIIS study, the approach for Phase 2 and the next steps of the EIIS project. More than 35 people participated in this Webinar. The participants were stakeholders from the European Commission and from the Member States.

The objective of the Webinar was to explain the criteria framework used to evaluate the systems that could be reused in a European interoperability infrastructure.

At the end of the Webinar, during the 'Questions and Answers' session, the participants had the possibility to ask more questions on the criteria framework.

4.2.4 STEP 4: MEETING WITH THE SYSTEM CONTACTS

During Phase 2, more than 30 meetings with the respective system contacts were held in order to discuss in detail the functional and technical aspects of systems and components which potentially could provide interoperability infrastructure services as identified in Phase 1.

The goal of the meeting was to verify the mapping established during Phase 1, to evaluate how suitable the system components were to deliver the interoperability infrastructure services and how suitable they were to be reused in other systems.

The meetings were structured according to the criteria framework. Each of the services identified in each system was discussed and assessed following the different criteria.

All participants in the meeting received preparation material before the meeting. Meetings minutes were written after the meeting, based on the discussion and the documentation available. These minutes were structured according to the criteria framework. They were sent to the participants for validation.

4.2.5 STEP 5: QUALIFICATION AND SELECTION OF THE SYSTEM COMPONENTS FOR REUSE

During Step 5, each system has been evaluated regarding the three evaluation criteria dimensions. This evaluation was done based on:

- The information provided by the system contacts during Phase 1;
- Meetings with the system contacts;
- The meeting minutes approved by the system contacts.

The qualification of each component enabled to compare the components and select the components with the best match with the different criteria.

The second phase started with 50 systems. The result of the meetings with the system owners and the qualification of the systems is a list of 24 systems which can be reused in a European interoperability infrastructure. These 24 systems are the basis for the definition of implementation scenarios in Phase 3 of the study.

This Step 5 is a continuation of a sifting process which started in Phase 1. The final selection should not be seen as an absolute result. The qualification is a process that keeps the systems which are relatively better than the systems which have been sifted out.

As already mentioned, the evaluation is based on the documentation available on the systems and the meetings with the system owners. While the system documentation provides ample evidence to support the reasoning why a system has a good functional fit, it is much harder to find facts on the technical or operational fit. In general, information on technical and operational fit is not sufficient to provide an inclusive and well funded rationale. For that reason, the qualification and the selection answered the following question "Is there a reason in the information available which indicates that this criterion is not satisfied?". For instance, finding evidence that a system passes the scalability criteria is difficult. However, a system 'passed' the scalability criteria, if the documentation does not provide information on scalability issues or if no proof of blocking issues could be found with respect to scalability.

It is also worth mentioning here that a qualification of components is not a quality review of components. The selection (or non- selection) of the component has to be viewed in the scope of the study. The question addressed in the study is how easy a system can be reused in a European Infrastructure Interoperability context. The study does not evaluate the quality of the individual systems.

In general, no more than five reusable systems have been selected for each service. Therefore, if a system is not selected, it does not necessarily mean that its component is not reusable in a European context, it just means that some components are better qualified to be reused.

The evaluation in the three criteria dimensions has been done within the limitation of the approach of this study. The evaluation depends heavily on the availability and cooperation of the system owners before and during the meeting, the availability and the quality of the system documentation and the timely approval of the meeting minutes. Furthermore, as in any project, the EIIS study has only finite resources. For that reason, the duration of the meetings with the system contacts was limited to one hour. For practical reasons, face to face meetings were sometimes replaced by video or conference calls. To summarize, **the selection of the reusable components has to be viewed in the perspective and the limitations of the approach adopted for the EIIS study**.

Finally, it is important to remind that the scope of the EIIS study covers only the qualification phase of the systems components, which is a first evaluation of the systems components. However, in order to assess deeper and definitely the possibility to reuse the selected components, it is absolutely necessary to do a feasibility study.

4.3 PHASE 2 RESULTS

During Phase 2, each of the 50 systems has been evaluated regarding the three evaluation criteria dimensions. The qualification of each component enabled to compare the components and select the components with the best match with the different criteria.

A first result of the Phase 2 is a **qualification and selection** of 24 systems which are possible candidates for 'reuse' in a European interoperability infrastructure, including a description of the rationale behind the selection. The overview of information systems and interoperability infrastructure services is available in **Annex 8.2.**

A second result of the Phase 2 is a more detailed description of the nine interoperability infrastructure services as compared to what has been described in Phase 1. The **service description** section contains detailed information on the services and a description of its key service features and business value. Additionally some background information is provided containing e.g. references to other studies or trends on the market. The complete description of each interoperability infrastructure service can be found in the **Annexes 8.3** to **8.11**. of this report.

4.4 PHASE 3 APPROACH

Phase 3 of the EIIS study has been executed during the period between September 2009 and November 2009. This phase aimed at:

- Identifying a small set of implementation scenarios for each of the EIIS;
- Prioritising the implementation scenarios.

An iterative approach has been adopted to develop the implementation options for the EIIS:

- Firstly, a wide range of implementation scenarios has been identified during a few rounds of internal brainstorm sessions within the EIIS team.
- Secondly, a selection of scenarios has been presented and discussed during a workshop with the EIIS stakeholders. The workshop was not a validation workshop. Its only purpose was to collect additional information on the implementation scenarios. During the workshop, the participants were divided in three breakout sessions. Each breakout session discussed and assessed three interoperability infrastructure services.
- Finally, the scenarios have been re-assessed and adapted based upon the feedback received during the workshop with the EIIS stakeholders.

The next chapter presents the results of Phase 3. For each service, the general trends and evolution of the service are presented. These trends allow to introduce the implementation scenario(s) proposed for the service, and the list of systems that could be reused to implement the scenario(s).

5 Implementation options EIIS

5.1 AUDIT TRAIL AND LOG

5.1.1 TRENDS AND EVOLUTION

For reasons of security or of regulatory compliance, public administrations are required to secure their systems to ensure the integrity, availability and confidentiality of data. Hence it is important to keep user activity logs and audit trails.

Based on the needs of security and legislative requirements, many organizations have taken initiatives to define best practices and frameworks in that domain[6]. The implementation of most of these frameworks calls for the design and implementation of effective logging and the traceability of events taking place in the information systems[7].

Today, the following trends can be observed in the field of Audit Trail and Log. Firstly, the need for traceability of data changes has increased with the increase in the volume of transactions and the opening up of organizations' boundaries. Secondly, with the increased volume of transactions, it has become a great challenge to select the information to be logged. Finally, the audit trail information has to be easily and effectively analysed by the auditors.

These trends are applicable to the European Commission and Member States systems that must deal with large volumes of generated log messages. Effectively analyzing large volumes of diverse logs can pose many challenges (huge log volumes, log format diversity, etc.). Making sense of this large volume of log data is no simple task and requires not only having the right technology, but also the people and processes in place to manage the logs.

5.1.2 IMPLEMENTATION SCENARIOS

Facing these challenges, the European Commission and the Member States need support concerning Audit Trail and Log. However, it is difficult to envisage Audit Trail and Log in an implementation scenario of a service. Audit trail and Log indeed is generally strongly related to specific systems and legislation, and a shared service does not seem realistic. For that reason, no implementation scenario for Audit Trail and Log service will be proposed. Audit Trail and Log is considered as a component linked to a specific application.

Instead, guidance and reusable components in that field are of a high interest for the European Commission and the Member States. Audit Trail and Log will thus be proposed only in terms of guidance and as a set of reusable components.

5.1.3 NEED FOR GUIDANCE FOR AUDIT TRAIL AND LOG

Referring to the previous section, the Member States and the European Commission should be provided with guidance for Audit Trail and Log. The future guidance should at least concern the following areas:

[6] For instance, the COBIT framework specifies in its control objective DS5 the use of a violation and security activity report and security surveillance as key areas to be addressed. It also emphasizes the need to review the design of audit trails while identifying the automated solutions under control objective AI1.10.
International Organization for Standardization (ISO) 17799, the international security management standard, also addresses this issue in clause 12.3, System Audit Consideration; 10.2, Security of Application Systems; and 9.7, Monitoring System Access and Use.

[7] D.K. Agarwal, CIQA, The Need for and Implementation of Audit Trails, 2006.

- Audit record creation and structure: Type of information, specific event that has to be logged, format of the logs …;
- Audit record retention: Duration of the retention period, archiving etc. Retention is often dependent on legislation. It is a challenge, as the volume of raw log data can be huge;
- Audit trail protection : The integrity, security and authenticity of the data must be ensured;
- Audit trail analysis : Frequency of analysing documentation and reporting;
- Etc.

As example of Audit Trail and Log requirements, NISPOM, the National Industrial Security Program Operating Manual, developed by the US Department of Defense, sets comprehensive standards for protecting classified data. NISPOM recommends logging and analysis of certain activities, and provides security auditing features that the systems must implement and perform (audit creation, retention, protection and analysis) [8].

There remain open questions:

- Which level of detail has to be addressed in the guidance for Audit Trail and Log?
- When can this guidance made been available?
- Etc.

5.1.4 REUSABLE SYSTEMS

The systems proposing reusable components have been studied during Phases 1 and 2 of this study. A more detailed description of the reusable components is available in the Phase 2 deliverable. The following systems have a good potential to provide reusable components offering Audit Trail and Log functionalities.

5.1.4.1 @firma (MS Spain)

@firma Audit Module provides information on the transactions that took place in the platform. The @firma Audit Module is custom-built. It is mainly relying on open source tools (Log4j) although further in-house adaptations have been made. @firma Audit Module currently uses an Oracle database, but there is no obstacle to use an open-source one. The component easily interacts with all kinds of database platforms since it uses Hibernate, an open source tool, allowing to communicate with different databases.

5.1.4.2 Data Verification System or DVS (MS Spain)

The DVS Traceability Module allows to trace the sequence of operations carried out by the system, while the Audit and Statistics Module provides usage statistics. The module is included in a Java library (.jar). It is custom-built, based on Log4j. The logs are stored in an Oracle database.

5.1.4.3 Federal Service Bus or FSB (MS Belgium)

The audit component in FSB follows the "System Agent" principle (Front-end and Back-end service agents external to the Bus). The service agents register all the incoming and outgoing information in a database. Every message is identified by three elements: ID of the FSB service consumer, ID generated by the FSB platform and submitted to the FSB service provider, ID submitted by the service provider. Those elements are logged, what enables to identify the transaction chain. To optimize the performance of logging, the “Wire & Tap” pattern has been

[8] NATIONAL INDUSTRIAL SECURITY PROGRAM OPERATING MANUAL, DoD 5220.22-M, February 28, 2006.

used. This means that when a consumer calls a service, this request (or message) is doubled and redirected to a system agent avoiding latencies for the consumer that would be caused by logging.

5.1.4.4 jCore Logging (DG AGRI)

jCore Logging was initially conceived to create logs in a standardized way for all the DG AGRI applications. The jCore Logging module distinguishes between 3 types of messages: audit for legal purpose, log and error for information purposes. It provides a service layer for creating a trail of the executed business actions for a user in an application. The service layer operates where business logic is stored in an application: Java and PL/SQL (Oracle Database). Its purpose is to log specific business actions rather than technical actions. jCore Logging automatically groups all messages generated during a same business action, so that the application owner is provided with a detailed logging overview per use case execution.

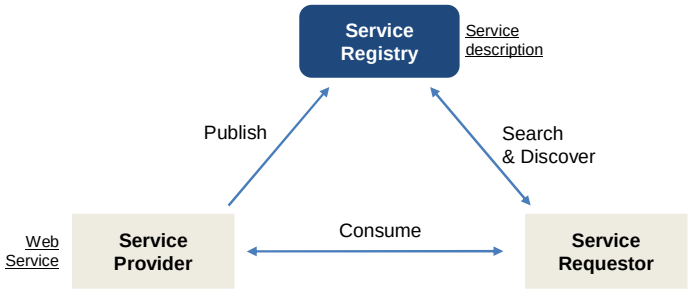
jCore is based on Java technology and open source standards. jCore Logging is based on Spring and Hibernate. It uses the standard Java framework Log4j.

5.2 SERVICE REGISTRY

5.2.1 TRENDS AND EVOLUTION

During the last decade Web services are being increasingly implemented by public administrations and private companies. Web services provide access to software systems over the Internet using standard protocols and enable e.g. public administrations to more efficiently integrate applications and improve the accessibility to business processes for citizens, businesses, partners, and internal staff.

Web services proposed by service providers are "consumed" by service requestors. To allow Web service discovery, a (public or private) Web service registry is required, as defined in the industry standard UDDI[9] for publishing and discovering services. The service provider must assure that his service is published to the registry, while the requestor or consumer wants to be able to search for available services in a flexible way.



Besides simply listing available services, a Web service registry is useful for those systems that wish to connect at runtime to service end points. Instead of hard-coding a service path in a client application, one may query the Web service registry for a particular service end point and use the registry's response to call the service.

Looking at the situation in the European public administrations, Member States and the European Commission currently lack visibility on the services offered by the different service providers across Europe. It occurs often that a public administration of a Member State and/or a Directorate General of the European Commission starts developing a certain service it wants to offer to internal staff, citizens and/or businesses, while a similar service is already offered by another service provider.

In light of this study, it is clear that a Web service registry is convenient in the context of European Interoperability Infrastructure Services. It will increase interoperability and reuse of existing services, reducing development time and improving service lifecycle management.

[9] http://uddi.xml.org/. Universal Description and Discovery Information (UDDI) is a type of registry whose primary purpose is to represent information about web services. It describes the service providers, the services that providers offer, and in some cases, the specific technical specifications for interacting with those services.

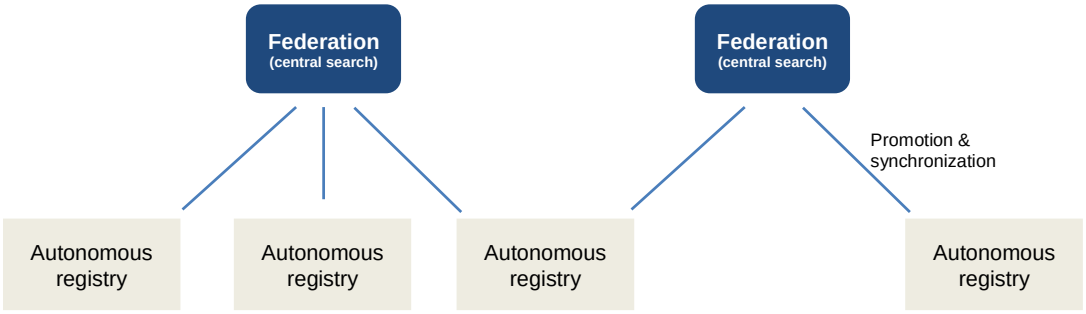
5.2.2 IMPLEMENTATION SCENARIO: FEDERATED REGISTRIES

5.2.2.1 Introduction

It is realistic to say that there will always be a number of registries (potentially of different technologies) around, perhaps even within the European Commission or within a single Member State. One central service registry at European level is hard to realize from organizational and legal perspectives. Integration between these different registries is therefore key to improve interoperability and reuse of services. Federated registries could certainly be beneficial for Europe.

Federated registries are a **collection of autonomous but cooperating (web) service registries**, which can be private or public[10]. It must be noted that one autonomous registry can participate in more than one federation. The goal of a federation is to establish a "registry community" serving either a business domain or policy area with similar and/or competing services. The combination of a registry describing drivers license services and a registry describing car registrations services is a good illustration of federated registries.

5.2.2.2 High-level model



Any autonomous registry within the federated registry architecture is able to negotiate and exchange information with other registries within the federation. Additionally, a federated registry architecture can facilitate searches across other registry indexes for relevant information regarding service endpoints for either human or system consumption. This federated model requires promotion and synchronization of information or a subset of information of the autonomous registries towards a central registry / search engine. The synchronization capabilities are crucial to satisfy the goal of supporting independent registries at local level while providing access to various registries across public administrations in Europe. The federated registry model enables the user to search for services at any single registry within the federation without manually interacting with several different registries.

A metadata retriever could automatically extract the metadata associated to the service stored. Pulling the metadata automatically in the registry has many benefits in keeping the service information adequate and up-to-date.

5.2.2.3 Benefits and challenges

Federation of registries can provide several advantages over individual registries. The main **benefits** of a federation of registries is that it enables service requestors to search multiple registries and then aggregate the returned result by using filtering and ranking techniques. Entries in one registry can be discovered by service consumers who query it via other remote registries.

[10] Discovery of Web Services in a Federated Registry Environment, Kaarthik Sivashanmugam, Kunal Verma & Amit Sheth.

Improved visibility of the services portfolio enables faster development and greater application reuse.

The **challenges** are:

- Organizing the registry in such a manner that different user communities have access, which will require solving any governance issues that emerge from it. For example maintaining the information up to date in the registry and informing the service requestors of important updates (e.g. using a notification feature).
- Setting up an authentication policy to access entries in the service registries. A common authentication policy of federated registries would bring additional value compared to autonomous registries.
- Ensuring business service visibility. If users cannot easily find these business services, the promise of "service oriented thinking" is largely lost. Services that cannot readily be found and reused essentially don't exist. For example, the current UDDI search mechanism can only focus on a single search criterion that needs to match exactly with the service name to discover the requested service. It should be possible to recognize similar names of the services and then to return service's development and runtime information to the user even when the there is not an exact match for the service names in the registry.
- Integrating the potential large scale growth of private and public registries in one infrastructure which can support discovery and publication over this large group of autonomous registries.
- Distribution of the data. Replication of entries, where all registries are equal and each registry is an exact and complete replica of each other, is not scalable. Distributing data among multiple registries based on vertical or horizontal partitions would provide increased availability and reliability[11].

5.2.3 REUSABLE COMPONENTS

The systems that have been identified in Phase 2 of the EIIS study are listed in this section.

5.2.3.1 NemHandel (MS Denmark)

NemHandel is based on UDDI. A national master UDDI holds all web service end points offered. The effective end point connection is established via replication of the master UDDI to one or more runtime UDDI instances.

5.2.3.2 DVDV (MS Germany)

DVDV received the e-Government award in 2007. DVDV is currently taken into account for the assessment phase of **SPOCS**, which is the acronym for "Simple Procedures Online for Cross-border Services". SPOCS is a pilot project launched by the European Commission which aims to remove the administrative barriers that European businesses face before offering their services abroad[12].

5.2.3.3 Federal Service Bus (MS Belgium)

Federal Service Bus has a public registry available online and an internal repository. The internal repository contains technical information to develop and maintain the services. The information in the internal repository is published automatically to the online registry.

[11] Based on research of Thaden et al., 2003, Schmidt and Parashkar, 2003.

[12] http://www.eu-spocs.eu/

5.2.3.4 X-ROAD (MS Estonia)

X-ROAD has a scalable internal service registry for machine to machine communication. An additional "registry of registries" is provided externally by the national information system RIHA. RIHA administrates and provides access to the information systems and databases of the state and local governments, and the metadata of data services offered by them. RIHA has a UDDI interface.

5.2.3.5 Other registries on the market

A long list of existing registry efforts is published on the xml.gov website[13]. Yet none of them have been federated to provide users the ability to discover relevant schemas and elements wherever they may reside on the Web.

[13] http://xml.gov/registries.asp

5.3 IDENTITY AND ACCES MANAGEMENT (IAM)

5.3.1 TRENDS AND EVOLUTION

Public authorities across Europe are increasingly offering electronic access to government services. In doing so they have been focusing mostly on national needs and means, which has led to a diverse system landscape. This situation creates interoperability barriers for electronically accessing services of public administrations across Europe. One of these barriers concerns the use of electronic identification[14]. Access to public administrations' electronic procedures often implies the need for the individuals involved to identify themselves (i.e. allowing the administration to make sure that the persons are who they claim to be by checking their personal credentials).

With regard to cross-border e-identification, a large number of initiatives have already been launched by the Commission that aim at finding solutions to interoperable e-identification at EU level, on a political, legal, organizational and technical level. What follows is a non-exhaustive list of the most important initiatives:

- Directive 1999/93/EC of the European Parliament and of the Council of 13 December 1999 on a Community framework for electronic signatures[15]
- the i2010 e-Government Action Plan[16]
- Action plan on eSignatures and eIdentification[17]
- The IDABC eID Interoperability for PEGS study[18]
- the STORK pilot project[19]
- ePractice European eID Observatory Community[20]
- OSOR.eu eID Community[21]
- …

Taking a look at some trends on the market, Gartner has mentioned at its Identity & Access Management Summit 2009 (23-24 March in London) that *there is a continuing need, in this time of economic uncertainty and budgetary constraints, for cost-effective, risk-appropriate IAM methods. This includes, amongst others, growing demand for service-based IAM offerings. Solution sets related to identity and access management are evolving from monolithic application models to composite services models. These reduce the costs of implementation and use and prepare for a more mature production-centric approach to* ***delivering IAM as a service****.*

Gartner recommends that users evaluate service-based options for extending their existing IAM solutions, rather than significantly upgrading those solutions. Those that have not deployed a mature IAM solution should consider service-based options for their future IAM solution.

---

[14] Identification is the process of using claimed or observed attributes of an entity to deduce who the entity is. The term "identification" is also referred to as entity authentication, which is defined as the corroboration of the claimed identity of an entity and a set of its observed attributes. [Common Terminological Framework for Interoperable Electronic Identity Management, November 23, 2005]

[15] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31999L0093:EN:NOT

[16] i2010 e-Government Action Plan: Accelerating e-Government in Europe for the Benefit of All [COM(2006) 173 final].

[17] http://ec.europa.eu/idabc/en/document/7791

[18] http://ec.europa.eu/idabc/en/document/6484

[19] www.eid-stork.eu

[20] http://www.epractice.eu/community/eureid

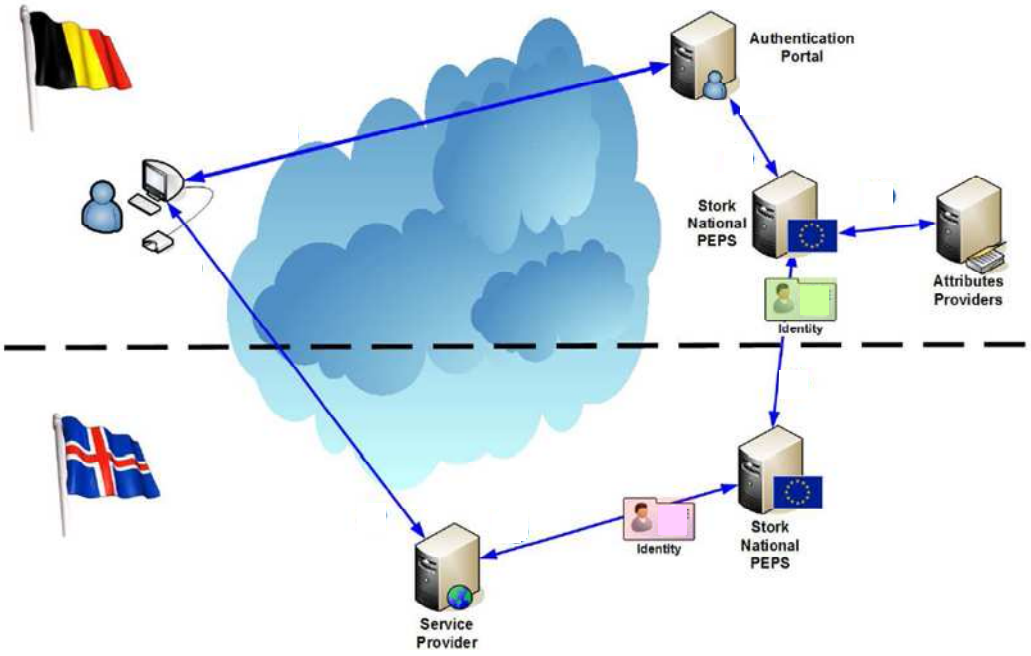
[21] http://www.osor.eu/communities/eid/eid-community

5.3.2 IMPLEMENTATION SCENARIOS

5.3.2.1 Improving Interoperability between existing IAM solutions and promoting entity authentication (identification) as a service in a federated model

As mentioned above, a number of initiatives are already ongoing in the area of identity and access management. One of these initiatives is the STORK project. **STORK** is a large scale pilot project (2007-2010) that aims at implementing an EU-wide interoperable system for recognition of eID and authentication that will enable businesses, citizens and government employees to use their national electronic identities in any Member State. By 2010, the STORK project aims to provide secure and convenient electronic systems for European citizens and businesses accessing public services in any country of the European Union.

STORK will be as technology-transparent as possible and ensure interoperability solutions can operate with existing national eID systems. STORK will rely as much as possible on open standards, such as the SAML 2.0 protocol (federated identity).



Source: eID Workshop 2009 (www.epractice.eu, 29/09/2009)

The EIIS study wants to avoid elaborating on solutions (architectures) that are already subject of another study or project. Therefore for those Member States that are looking for a common service architecture allowing citizens, businesses and/or government employees to authenticate themselves to access e-Government portals across borders, the STORK project should be the reference project to turn to. Nevertheless, this study wants to stress the following:

Only 28 out of 32 EU countries use or plan to use, an electronic ID scheme. While some countries have signed agreements on mutual recognition, eID systems differ from one Member State to another. STORK will improve interoperability between the national systems but only 13 Member States and Iceland are currently involved in the project. Therefore, it is likely that not all Member States will support eID and will be able to connect to STORK on short notice, which might be a risk for the success of STORK and its continuation after a successful pilot roll-out. Therefore the EIIS study poses the **open question** whether or not the STORK infrastructure can be extended to facilitate not only eID authentication, but also other authentication methods (as there are username/password, biometric credentials, etc.). This way the interoperable infrastructure of STORK is reused to support an EU-wide common service architecture for federating identity credentials of citizens, businesses and government employees "*collected*" by a mixture of

authentication methods. In this case, security and privacy are even bigger concerns to evaluate as they are now.

5.3.3 REUSABLE COMPONENTS

While for the service implementation options the EIIS study could not bring many new insights compared to what the STORK project is bringing, the added value of the EIIS study is there for Member States and Directorate General that do not yet have a best practice IAM solution in place and that are looking for a new solution. The following paragraphs describe possible candidates for reuse for an authentication portal, as shown on the STORK picture above.

5.3.3.1 Entity authentication (identification) taking ECAS as example

Currently ECAS is used to authenticate internal and external users to access applications within the European Commission. ECAS provides one login page, one single password, one place to collect the password, one repository to store the password, web single sign-on to access all information systems and a self-registration mechanism including CAPTCHA[22]. First time users need to self-register with a valid e-mail address before they can use ECAS.

ECAS consists of two components: a client (an API running on Weblogic, Tomcat, ColdFusion, PHP …) and a server (Weblogic). ECAS is based on open source components (CAS, Spring, Struts) and is using an open protocol. The ECAS components therefore can be possibly reused within a Member State. The code is currently only internally published. There is no compliance yet with the European Union Public Licence but the ECAS team has however the intent to have ECAS under the European Union Public License (EUPL) and to release the code on the Open Source Observatory and Repository (OSOR.eu). If it is part of EUPL, no non-disclosure agreements have to be signed.

ECAS is a candidate for reuse in the sense of:

- Using ECAS as an authentication service within the European Commission for all (new) systems;
- Promoting ECAS to Member States (e.g. under the EUPL) as a reusable open source authentication solution which the Member States can install in their premises;
- Sharing best practices.

The main barriers for reusing ECAS are related to security and support:

- ECAS relies on a username/password login as credentials. In case a username/password is stolen from a user, anyone can access the applications this user has access to. Passwords are considered one of the weakest kinds of authentication, but if the rules for the password usage and composition are good enough, it can be sufficient for most applications that deal with information that is ‘sensitive’, but not classified.
- The self-registration mechanism is currently based on e-mail validation. For persons with a Commission e-mail the risk is rather low that a person is not who he/she is claiming to be. However when persons with a generic e-mail (e.g. hotmail, gmail, yahoo, etc.) can self-register, one can absolutely not be sure of the identity the user pretends to have.
- ECAS is developed and maintained by DIGIT. A user support team is available to solve all kinds of issues. If however ECAS is promoted for reuse in the Member States, this raises questions such as:

[22] Completely Automated Public Turing test to tell Computers and Humans Apart

- Will the support remain centrally organized? What about the different time zones, languages, etc. Specifically for third line support dealing with complex issues and requiring on-site assistance, this is a problem to address.
- What about the funding?

However, the upcoming integration between ECAS and eID STORK will bring additional benefits. ECAS will provide users with the possibility to authenticate with a national eID card. This way ECAS is absolutely sure of the identity of the user during self-registration and during authentication. The process is as follows: the user will be authenticated in the STORK gateway, after which an assertion is sent from the gateway to ECAS (unidirectional). This integration between ECAS and STORK requires however changes to the ECAS clients and to the ECAS server, but after that all applications using ECAS could benefit from the eID authentication possibility. ECAS integration with STORK will be a pilot because STORK itself is a pilot project. The initial planning of Stork eID has foreseen to provide users from Austria and Belgium login possibilities with eID coupled with ECAS for the end of 2009. In 2010, users of the whole STORK network should be integrated.

However, during the analysis it has become clear that certainly not all application owned by the European Commission are based on the same authentication component/service, simply because ECAS did not exist at the moment of the development of these systems. The good news is that many systems are nowadays migrating or planning to migrate in the near future towards ECAS. This means that the added value of reusing ECAS is recognized within the Commission. Nevertheless, some systems owned by the European Commission have no intention to migrate to ECAS. It should be analyzed whether or not reusing ECAS could reduce the operational cost of these systems (maintenance, support, licenses, etc.). Certainly when these systems would opt for eID integration, migrating towards ECAS will be one of the preferred scenarios if the ECAS – STORK integration is finalized.

5.3.3.2 Entity authentication (identification) taking Modules for Online Applications (MOA) (Austria) as example

The MOA ID authenticates citizens, businesses and authorities (using the digital signature on the citizen card) and forwards the user's login information to the requesting application.

The MOA solution of Austria can be a candidate for reuse in the sense of:

- sharing best practice to realize a commonly agreed hard token standard, specifically in countries which are implementing new eID schemes;
- sharing technology experience of the eID platform, including the STORK project, with countries that still need to deploy an eID infrastructure.

The items that need to be taken into account are the following:

- The results of the eID STORK pilot project will be determining for the future strategy of eID interoperability;
- MOA is an open source project, which makes the solution independent of commercial products. However, maintenance and the development of new services in MOA are ensured by the Austrian Government. The MOA are available on an open source platform (http://www.egovlabs.gv.at). They are available for the private sector. Most of the documentation is in German;
- Austria is one of the 14 members of the STORK consortium.

There remain open questions:

- Due to the design based on web services, the MOA components should be able to support a higher load than they are currently supporting. A study would be certainly required for a very large-scale deployment.
- Is there a dependency on the Austria citizen card, or can any identity card be used?

5.3.3.3 Entity authentication (identification) taking other existing solutions as example

During Phase 2 of the EIIS study, a number of system components were selected as candidates for reuse. What follows is a list of system components that should be considered as good examples for Member States that do want to replace the existing IAM solution or require a new IAM solution. More details can be consulted in the deliverable of Phase 2:

- jCORE Security, a JAVA component providing IAM services. jCore Security is also part of the RefApp developed by DIGIT;
- X-ROAD, delivering authentication and authorization since 2001 in Estonia;
- IMI's self-registration component, which enables a person to register him- or herself via an online e-form. Additionally registered users can delegate access rights to other persons, which unburdens system administrator(s) with this task;
- The Slovenia eUprava and eVEM portal solutions support username/password or digital certificate (eID) login. eVEM is the winner of the 2009 United Nations Public Service Award for excellence in the field of public administration.

5.3.3.4 Authorization as a reusable component has not been identified

The EIIS study does not provide a scenario on the IAM service feature authorization.

At the level of the Commission, a repository for users and roles exists. This repository (CUD) supports:

- The management of information related to internal and external users;
- The assignment of access rights by mean of groups either managed manually or derived automatically based on user attributes stored in a database.

It would however not be realistic to extent this repository with users and roles of each Member State (thus realizing one giant central authorization repository or a combination of inter-related authorization repositories in a federated model). There are a number of reasons for this:

- The effort it requires to centrally manage access rights would be enormous in particular if the access rights cannot be automatically managed by deriving it from user attributes;
- It would raise questions how to define access control of applications to personal information within the central authorization repository;
- The generic defined roles are probably not detailed or flexible enough to meet the complex authorization requirements of each single application within the MS. If it is possible to define specific access rights, it would be probably difficult to reuse these in other applications.

Therefore it is recommended to keep the management of access rights a matter of the Member States or Directorate Generals. One could think of a scenario that provides a standardized (in terms of structure, objects, roles, …) authorization repository to Member States. This standardized repository can be customized and locally deployed. Additionally this repository could be equipped with a central search engine that audits to what application(s) a certain user has access to across Europe. This scenario goes however beyond the scope of this study.

5.4 DATA CERTIFICATION

5.4.1 TRENDS AND EVOLUTION

As explained earlier for Identity and Access Management, electronic communication is becoming increasingly important in many aspects of economic and public life. Public authorities across Europe have started to offer electronic access to government services. Often there is a need to provide an electronic signature allowing the administration to identify the signatory as well as to make sure that the data submitted has not been altered during transmission.

A large number of initiatives have already been launched by the Commission related to electronic signature and certification. What follows is a non-exhaustive list of the most important initiatives:

- Directive 1999/93/EC of the European Parliament and of the Council of 13 December 1999 on a Community framework for electronic signatures[23]
- the i2010 e-Government Action Plan[24]
- Action plan on eSignatures and eIdentification[25]
- ePractice European eID Observatory Community[26]
- OSOR.eu eID Community[27]
- IDABC's preliminary study on mutual recognition of eSignatures for e-government applications[28]
- IDABC's European Federated Validation Service[29] (EFVS) study

The EIIS study wants to avoid elaborating on solution (architectures) that are already subject of another study or project. In the area of data certification, the results of the EFVS study should be considered. The next paragraphs will highlight the most important conclusions of the EFVS study. For a full report, please visit the IDABC website.

5.4.2 IMPLEMENTATION SCENARIOS

5.4.2.1 European Federated Validation Service (EFVS) study

The **European Federated Validation Service (EFVS)** study was initiated by IDABC in order to assess the feasibility of specific measures to ensure the availability of a European scale federated electronic signature verification functionality.

The proposed model of Figure 2 is a federation of national validation authorities. The idea is that national registered applications can request at a national validation authority for the validation of certificates issues by the Certificate Service Providers (CSP[30]) established in their own country (step 1 and 4 in Figure 2). The national validation authority redirects the validations of "foreign"

[23] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31999L0093:EN:NOT

[24] i2010 e-Government Action Plan: Accelerating e-Government in Europe for the Benefit of All [COM(2006) 173 final].

[25] http://ec.europa.eu/idabc/en/document/7791

[26] http://www.epractice.eu/community/eureid

[27] http://www.osor.eu/communities/eid/eid-community

[28] http://ec.europa.eu/idabc/en/document/6485/5938

[29] http://ec.europa.eu/idabc/en/document/7764

[30] A CSP is also known as a Certification Authority (CA). A CSP is an organisation responsible for the issuing, renewing, suspending or revoking of digital certificates.

certificates to the validation authority of the country in which the CSP is established (step 2 and 3 in Figure 2).

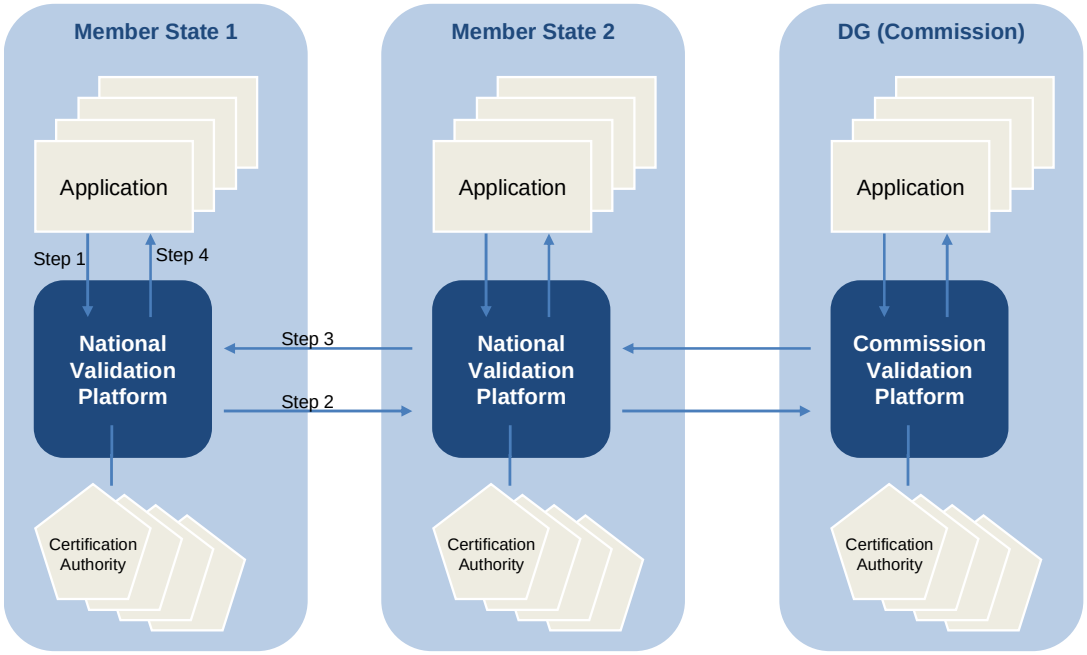


Figure 2: Model

The main **benefit** of such a model is the creation of a one-stop-shop for certificate validation avoiding the complexity for a validation component of an application to establish relationships with every CSP/CA throughout the European Union. The added value of a validation authority for cost-effective certificate validation at the national level in the public sector has already been proven by the Spanish @firma project[31].

To create a federation of validation authorities in a European context, trusted relationships between the actors of the federation need to be established. The **EFVS study**[32] analysis showed that there are currently two options for creating trust at a cross-border level, taking into account that there is a fundamental difference between signatures based on qualified certificates and those based on non-qualified certificates.

- The first option, as demonstrated by @firma, SVS and e-Notarius, is to leverage the existing trust model that has been created by the eSignatures Directive, which created the concept of the qualified certificate and made this subject to national supervision. However, neither @firma, SVS nor e-Notarius supports foreign qualified certificates due to the lack of a trustworthy source to identify CSPs issuing qualified certificates. This question is however already being addressed in the context of the **CROBIES study**.
- The second option is demonstrated by the BBS model, which consists of operating largely on the basis of a contractual framework, which does not depend on the European regulatory framework and its trust model, and can thus also be applied to non-qualified certificates. In this case, the validation authority defines its own norms and standards, which it applies to any number of chosen CSPs. This has the advantage of being applicable internationally, but it also puts much more effort and responsibility with the validation authority as a single source of trust. From an interoperability perspective, this option also creates the risk that different validation authorities apply different norms and

[31] http://www.epractice.eu/node/277227

[32] http://ec.europa.eu/idabc/en/document/7764

standards, meaning that service providers will not be able to easily compare guarantees offered by different validation authorities.

The **challenge** is that both options imply the need of multilateral agreements and/or a legal framework among the several Member States on the recognition of other countries' certificates. There must be a central body which will generate and update lists of trust of the CSPs to be internationally recognized: the Trust-service Status List (TSL).

5.4.3 REUSABLE SYSTEMS

5.4.3.1 Solution profiles of the EFVS study

The EFVS Study has collected information on twenty-two existing solutions that already provide all or some of the functionalities associated with European signature verification functionality, or that could provide valuable insights on how such an EFVS could be organised. This has been done by drafting standardised profiles of the identified solutions, focusing specifically on how each of these solutions (a) determine the validity of certificates; (b) verify electronic signatures created using these certificates; and (c) provide specific guarantees to their customers on the outcomes of these processes.

These twenty-two solution profiles were compared to each other and the solutions that most closely or completely meet the requirements were identified. On the basis of this comparison, four so-called key solutions were chosen, which offered the main functionalities expected from a validation solution. These key solutions were the @firma platform, the BBS Validation Authority, the A-SIT Signature Verification Service (SVS) and e-Notarius. @firma has also been in scope of the EIIS study.

Each of these four key solutions have been analysed in detail, including in particular the scope of the solutions, their technical approach, and the legal model behind it. While the emphasis of the analysis was on the key solutions, useful inputs from other profiles were identified and considered as well.

To view the different solution profiles, please visit http://ec.europa.eu/idabc/en/document/7764.

5.4.3.2 Additional solution profile?

The EIIS study wants to point out that the X-ROAD information system of Estonia has not been identified during the EFVS study or are at least not available as solution profile in the final deliverable of EFVS. X-ROAD is a standard data communication layer between databases and information systems that allows information systems with different underlying platforms to transfer data. X-ROAD is operational since 2001 and is widely used in Estonia. All messages exchanged with X-ROAD are signed, logged and time stamped. Signing keys are certified by the X-ROAD central agency. It is recommended to at least take away the findings and knowledge build up at X-ROAD about a data and communication layer and related certification on national level.

Further it must be stated that the ongoing IDABC/DIGIT initiative of ESSI (Electronic Signature Service Infrastructure) should also be taken into account. ESSI aims to setup a common infrastructure to be deployed at the European Commission to facilitate the introduction of electronic signatures in all kinds of exchange with internal and external partners.

5.4.3.3 Solutions delivering added value services

With added value services are meant those services that are not strictly necessary to determine whether or not a signature is valid at the present time but nevertheless deliver added value to the signature/certification validation process in a whole. These services are not necessarily delivered by the validation authority, but can be delivered by other trusted third parties or by the Certification Authorities themselves. In the EFVS study these added value services are e.g. time stamping and historical validation.

Time stamping

Time stamping is the process of securely associating a trusted time to a document in such a way that the signature (time) cannot be disputed later. The EFVS study has stated that time stamping needs to be taken into account from a legal and technical perspective in the framework of the European federated service. The lack of a clear legal framework for timestamps complicates issues surrounding the long term validity of electronic signatures and leading Member States to implement their own and diverging rules, presenting interoperability risks. Among the key solution profiles analyzed by the EFVS and the ESSI studies, **@firma** is the only non-commercial solution providing time stamping verification services.

Historical validation

Historical validation refers to the situation in which a service provider allows the validity of a signature to be checked for a significant amount of time after its creation, without necessarily resorting to time stamping. A typical example would be a validation authority that receives a signature which was created two years ago using a certificate that expired one year ago, and that would try to determine the validity of the signature at the time of creation (i.e. two years ago) based on its records of valid certificates at that time, such as e.g. archived Certificate Revocation Lists[33].

Historical validation can be optionally offered by the validation authority. A real life example of that is the **Austrian SVS/MOA solutions**, which support historical validation by default.

[33] A Certificate Revocation List (CRL) is a signed data structure that contains information about revoked certificates.

5.5 DATA TRANSPORT

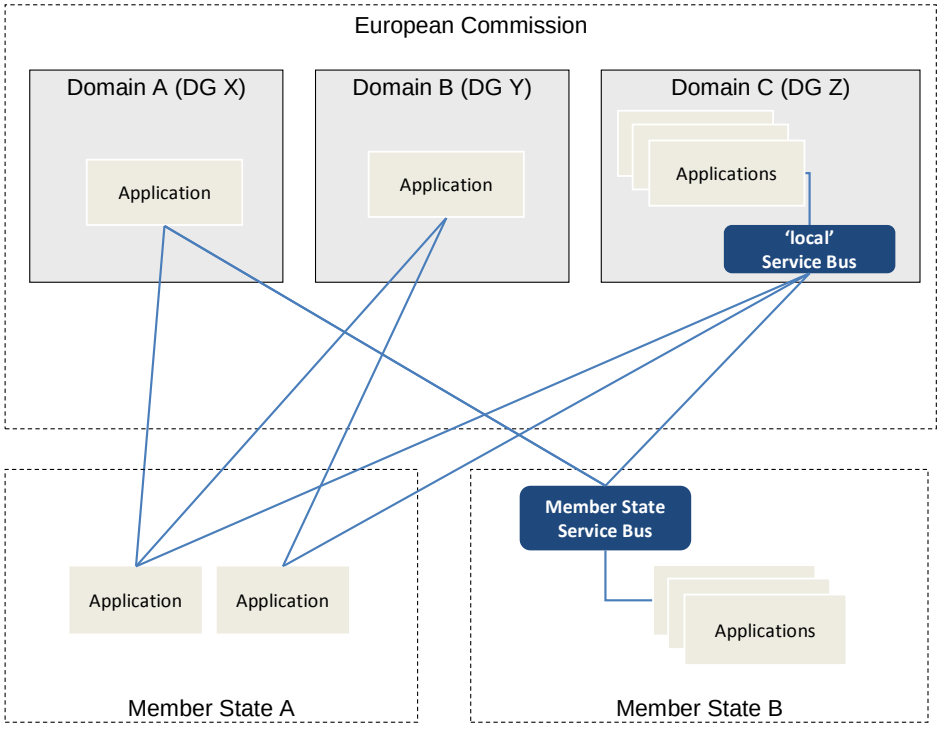
5.5.1 TRENDS AND EVOLUTION

IT culture within the public sector has long been known to be unique. The responsibilities of managing a wide range of often critical public services establish a distinct set of priorities that can't be compromised, especially when it comes to a reliance on technology. SOA adoption has grown substantially in government agencies at federal, state, and local levels. The strategic benefits of service-orientation can help overcome many of the traditional cost and efficiency-related IT problems.

Government institutions across the world at national, regional, and local levels, are significant consumers of technology. Governmental services affect us all and can be found in multiple domains. Service can be delivered in the domains of defence and national security, health, taxation, law enforcement, justice, environment, energy, social services, disaster management, and land use management.

Technology clearly needs to play an important role for any organization responsible for any one of these areas. In the public sector, automated systems perform a common, fundamental function: getting information to and from the "users." In other words, systems need to be in place to effectively share data between agencies and the public community, comprised of citizens and businesses, and between administrations.

However, as with large private corporations, information technology in government institutions has been built in silos, where the interoperability and exchangeability of information is only an afterthought, leaving service-orientation concerns by the wayside. For example, the Commission tends to structure information domains in terms of DG specific objectives. The more DG specific silos are created, the greater are the integration problems that will have to be overcome down the road, when DGs need to communicate with each other (or between user communities or Members State administrations).



This model makes the whole IT architecture and information exchange very complex. Customized applications have many point-to-point interfaces and the IT staffs in the Member States and the DGs spend a lot of time on working on interfaces.

This complexity is not specific to public administrations. It exists in many companies and limits the flexibility of the infrastructure. In many businesses, the adoption of a Service Oriented Architecture (SOA) has been considered as an enabler to create a flexible infrastructure. SOA enables organizations to build and deploy IT systems that directly serve the goals of the business faster and more easily than traditional approaches. A business services approach helps businesses and IT to establish a common language of communication, align IT with business needs, and facilitate change.

The Enterprise Service Bus (ESB) is a collection of middleware services that provides integration capabilities. It is defined by Forrester as "an intermediary that makes a set of reusable business services widely available", and is often seen as a major component typically included in a SOA. It is responsible for passing secure and reliable messages between the different components of the SOA. All of the component parts are designed to work together in a standardized and repeatable way leading to a consistent quality of service.

The ESB is central to eliminating the need for point-to-point connectors. It avoids common problems that other enterprise application integration (EAI) platforms could not. The hub and spoke EAI platform, in which all integrated applications work through a single message broker, creates a single point of failure, and constitutes a risk for a complex business system. The ESB, though, has numerous brokers, and so avoids this risk.

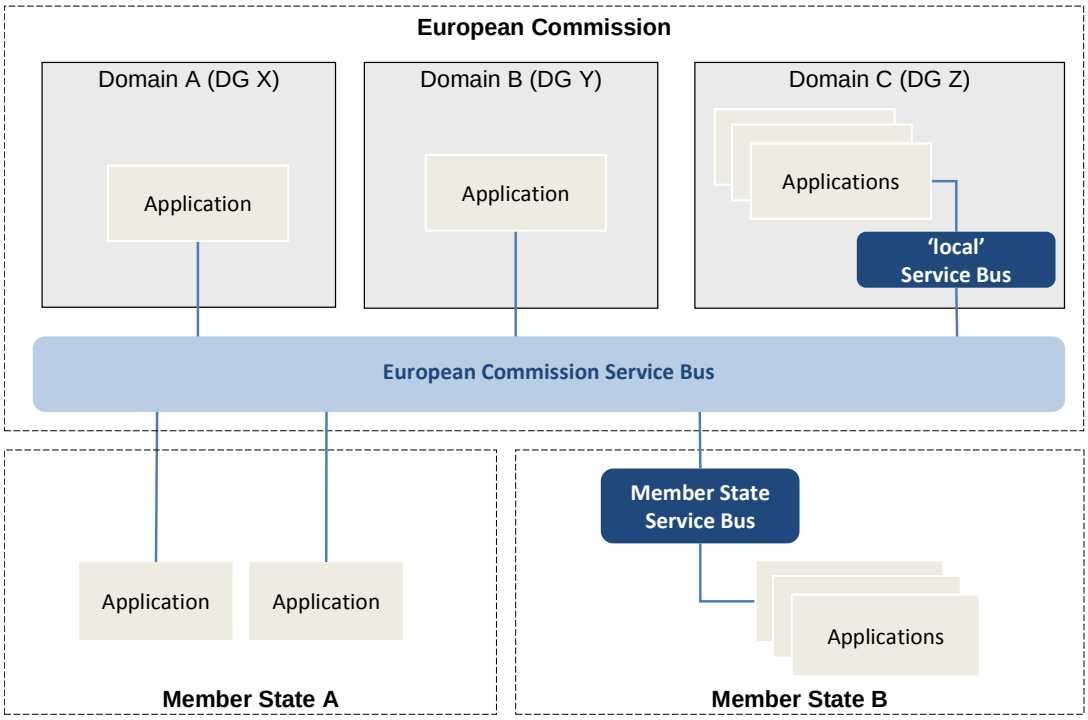
The ESB is also more suited for SOA. Unlike the hub and spoke platform, an ESB facilitates the loose-coupling of systems and the use of open-standards, two features of most successful SOA implementations.

Coming back to the specific context of the information exchange between the EC and the MS, the ideal situation would be to have a flexible and configurable intermediary at EC level, what would allow the MS to have a unique integration layer on their side. An ESB could greatly ease the burden of system integration and minimize the chore of dealing with local changes.

5.5.2 Implementation scenario: European Commission service bus for data transport

The scenario proposed for the data transport service is to have an Enterprise Service Bus for the data transport at the level of the European Commission. The Enterprise Service Bus takes care of the conversion of the data to be exchanged and allows the applications of the Member States and the European Commission to exchange without much change in the applications. It reduces the complexity of the integration of the applications. The European Service Bus allows the Member States to have a unique platform at European level, reducing the complexity of information exchange with the DGs.

In this scenario, the ESB is seen as a distributed services architecture based on Web services standards, which delivers messaging middleware, intelligent routing, and XML transformation in conjunction with a flexible security framework and a management infrastructure for configuring, deploying, and monitoring the services. The European Commission Service Bus architecture is represented in the figure below:



The ESB is a standards-based integration platform that combines messaging, Web services, data transformation, and intelligent routing to reliably connect and coordinate the interaction of significant numbers of diverse applications across an organisation.

The European Commission service bus should not be seen as a single service bus. Considering the organisational structure of the Commission this will most likely be a federation of multiple service busses. The most important step is that this is an initial step for reusing interoperability infrastructure services in a service oriented approach.

An open-source solution is a potential solution to implement this scenario. Open source has been proven particularly effective when implementing SOA. It reduces the cost of tools and often provides a range of options. Having access to the source code also eliminates concerns about vendor lock-in. Finally, open source offers a broad and helpful community and ecosystem to support the SOA initiative34.

Challenges

The main difficulty will be to find motivators for DGs to use the same common platform. DIGIT is the most logical DG to take ownership of this ESB. However the funding should come partially come from the DGs. A clear and transparent chargeback model should be designed.

5.5.3 Reusable Systems

IPCIS can be used to build a data transport service, in combination with technical services provided by e-PRIOR. CCN/CSI could also be envisaged in that way. The other systems listed can provide reusable components for data transport.

[34] Understanding the Business Benefits of an Open Source SOA Platform, A Hurwitz white Paper, 2009.

5.5.3.1 IPCIS and e-PRIOR services (DG DIGIT)

The ESB model currently implemented in IPCIS is a distributed ESB. It is already used by applications in the Member States in the context of eFP7 and e-PRIOR. To implement the European Commission service bus scenario, IPCIS should expose services to the Member States for the exchange of information between the Member States and the European Commission, or for the exchange of information between Member States, in a European context. The European Service Bus allows the Member States to have a unique platform at European level, reducing the complexity of information exchange with the DGs.

IPCIS is currently based on BEA Aqualogic Service Bus (Oracle). However, the IPCIS team is currently looking for an Open Source ESB, which could replace the existing one.

Independent of the product choice made for the EC ESB in IPCIS (open source or not), technically it is easy to expose web services to Member States. Potential difficulties lie in the domains of security, transactions and operational maintenance. To address these concerns, e-PRIOR offers functionalities that are necessary to connect systems to each other via IPCIS. e-PRIOR implements a set of technical services which support systems to communicate with each other in a secure manner. e-PRIOR services can be combined with IPCIS to implement the data transport scenario.

5.5.3.2 CCN/CSI (DG TAXUD)

CCN/CSI is the communication infrastructure of DG TAXUD. It is composed of the CCN (Common Communication Network), the trans-European network itself, and the CSI (Common System Interface), a set of protocols and application programming interfaces allowing the applications to exchange information through the CCN. The CCN is made up of a series of physical computers, the so-called "gateways", located either in the National Administrations or in the Commission premises. These gateways are interconnected in a secure way through their own communication services, and communicate with the national computers. The CCN software components (including standard products like Tuxedo, MQSeries, Apache and Sun One Directory Server) run on the gateways.

The CCN is foreseen to be more secure and to be transformed to an ESB. The plan is to deploy CCN 2.0 as a SOA platform to offer agility, scalability, availability, continuity and security to expand the interconnectivity with new Member States, other administrations, 3rd countries and possibly traders.

This SOA platform set up for CCN by TAXUD could be seen as the starting point for a European Commission ESB. The CCN should be envisaged to also function as an ESB for other applications which interact with the Member States.

Some challenges however exist when envisaging CCN 2.0 as starting point for the European ESB:

- Deploy a trans-European Systems security policy, including federated Identity & Access Management across Member States Administrations and Commission.
- Set up a central hosting environment which can meet the 24/7 high availability and continuity expectations for offering IT services to Customs and Taxation environment.
- Define a common reference infrastructure in order to foster the interoperability of national solutions across Member States.

5.5.3.3 Spring integration in Open e-PRIOR (DG DIGIT)

Spring Integration is an extension of the Spring Framework. It supports the Enterprise Integration Patterns[35] while building on the Spring Framework's existing support for enterprise integration. It enables simple messaging within Spring-based applications and integrates with external systems via simple adapters. Those adapters provide a higher-level of abstraction over Spring's support for remoting, messaging, and scheduling. Spring integration cannot play the role of an ESB. Spring Integration's primary goal is to provide a simple model for building enterprise integration solutions while maintaining the separation of concerns that is essential for producing maintainable, testable code.

In Open e-PRIOR, Spring integration has been used to build data transport services. Spring integration is used to expose web services to other applications.

5.5.3.4 Eucaris II (MS Consortium)

Eucaris II provides a direct point-to-point architectural model to exchange information on license plates and registered cars. The application is deployed in each participating Member State. All participating countries are connected to each other and are able – by means of an interface – to search in the license plate registers of the others, without influencing the national architecture set up for their registers. So there is no centralised system and no central register to be searched by the registration authorities.

Eucaris II is currently owned by the Eucaris Community (Member State consortium). The maintenance and the support are ensured by RDW (Dutch registration authority).

Eucaris II can potentially be transformed into a more generic platform that can be used for the exchange of all sorts of data and not only for information related to vehicles and driving licences. Recently the consortium of connected Member States have declared that EUCARIS should be the general exchange mechanism for other transport related data (e.g. for tachograph cards, transport undertakings etc).

The potential for reuse is in its lightweight and pragmatic architectural approach. The development cost to add additional countries is low, and the operational cost is minimal.

[35] http://www.eaipatterns.com/

5.6 DATA TRANSLATION

5.6.1 TRENDS AND EVOLUTION

Citizens and business are more free to move and operate than ever before and EU Member States and the European Commission are collaborating with each other in almost every business/policy domain, despite the different administrative, technical and linguistic backgrounds. Seamless data exchange is getting therefore ever more important in Europe. To improve data exchange, system integration and interoperability, it has been recognized that this can be achieved – amongst others – through:

- the provisioning of information from sender to receiver and vice-versa in their own language,
- the use of well defined ontologies (e.g. reliable data definitions and data relationships),
- easy-to-implement data exchange formats (e.g. XML schemas).

The above corresponds with what has been defined as the data translation service in the deliverable of Phase 2 of the EIIS study. Multilingualism, syntax translation and semantic translation will facilitate communication between public administrations, citizens and businesses. The next paragraphs will briefly touch upon some trends and evolutions of each service feature.

5.6.1.1 Multilingualism

In January 2009, the results of a "Study on Multilingualism[36]" have been published. This study presents an initial approach for dealing with multilingualism. The results clearly state that it is not sufficient to develop international solutions and merely use English as the common language. Citizens should get access to European Union legislation in their own language.

Multilingualism in the context of semantic interoperability implies that terminology and vocabulary has to be translated from source to target language. This requires that sender and receiver must have a common and ideally identical understanding of meaning of the data in all languages involved. This study on multilingualism recommends a pivot language approach as the most effective solution. More information can be found in paragraph § 5.6.2.1 Multilingualism.

5.6.1.2 Syntax

The recognized open data standard XML has become the foundation for sharing data in European networks. Applications for both information sharing and exchange are necessarily based on XML (related) standards because XML standards:

- facilitate the setup of reliable data definitions,
- enable the exchange of data in a coherent way,
- reduce the amount of effort required for syntax translations.

Nowadays there are many software on the market and/or developed within the European Commission and Member States that support syntax data transformations, referring to:

- The process of modifying the source format to the target format. Importing an incoming XML into a database, rendering an incoming XML file into a PDF document or extracting data out of a database into an XML file are a few examples.

[36] http://www.semic.eu/semic/view/snav/library/SEMIC-EU-Publications.xhtml?cid=516968

– The process of transforming the format of a data field. For example, date conversion (DDMMYY into DD/MM/YYYY), currency conversion (EUR into USD), etc. It should be noted that adaptations such as the conversion of units and dates can also be seen as a form of multilingualism.

Typically the technical barriers for data exchange in Europe are not related to syntax data transformations because technically this is not the most complex development issue of an implementation project with European scope. The technical barriers are mostly related to security, accessibility, availability, storage, etc. Additionally it is important to not confuse the technical aspect with the (complex) semantic aspect of doing data transformations, which is the subject of the next paragraph.

5.6.1.3 Semantics

The following fictive example illustrates the importance of semantic harmonization as a pre-requisite for successful data exchange between Member States and the European Commission. EUROSTAT collects statistical information from the National Institutes for Statistics (NIS). A measure collected is for example the "unemployment rate". It is realistic to say that amongst the NIS different interpretations of unemployment exist: should it include only full-time, also part-time, exclude sabbatical year?

It is important to stress out that the complexity lies in the semantic harmonization; for example aligning the different data definitions of “unemployment rate”. It is however not too complex – once there is a common data definition – to technically setup the data transformations between the “unemployment rate” data objects of the different Member States and the European Commission.

To improve semantic interoperability – which refers to common understanding of data definitions, terms and models within a certain context – the European Commission, via the IDABC programme, has launched the Semantic Interoperability Centre Europe "**SEMIC.EU**" in June 2008. SEMIC.EU supports meaningful data exchange for e-government projects. Its core feature is a repository that provides reusable interoperability assets, such as XML schema or taxonomies, for European e-government projects. SEMIC.EU promotes the reuse of syntactic (e.g. XML schemas) and semantic assets (e.g. ontologies) needed for semantic interoperability. Syntactic harmonization has proven to be an important pre-requisite for successful semantic interoperability between systems in Europe.

SEMIC.EU’s repository of reusable data definitions and data exchange formats has been growing ever since its launch and it is used as a central resource for authorities across Europe[37]. It is especially encouraging to see that national e-government bodies are consulting the Centre as a direct access path to other countries’ interoperability assets for eventual implementation of the same or similar solutions.

5.6.2 IMPLEMENTATION SCENARIOS

The following paragraphs will describe the implementation scenarios for:

– multilingualism, paragraph § 5.6.2.1 Multilingualism

– semantic translation, paragraph § 5.6.2.2 Syntax and semantic translation

– syntax translation, paragraph § 5.6.2.2 Syntax and semantic translation

[37] A workshop (21/10/2009) in Brussels has shown the potential for multiple implementations of SEMIC.EU-based repositories. During the workshop, representatives from 17 European countries met to exchange approaches and ideas for national, regional and domain-specific implementations of similar repositories. SEMIC.EU's open-source technology can be downloaded under an EUPL (European Union Public Licence) from the Open Source Observatory and Repository (http://www.osor.eu/projects/repository).

5.6.2.1 Multilingualism

Multilingualism is a special aspect of semantic interoperability and is actually out of scope of this study. It is in the scope of SEMIC.EU. Nevertheless this document will highlight some aspects of multilingualism.

Multilingualism in the context of computing indicates that an application supports dynamically two or more languages when building or running the system. The multilingualism can be applicable at three levels:

- Application level
- Metadata level
- Data value level

Application level:

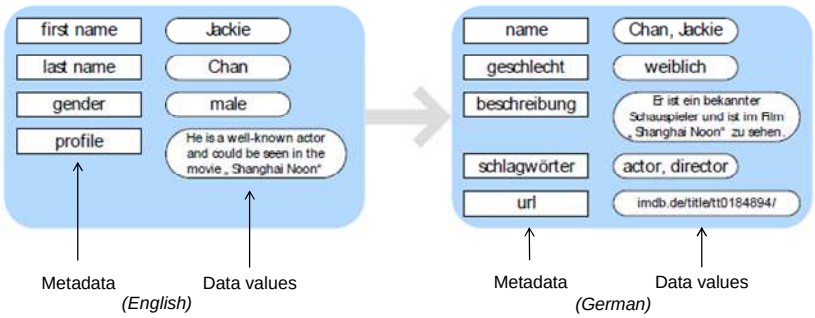
Most commercial products support multiple languages for their navigation menus. The language is initially automatically chosen based on the language settings of the PC of the end user but can be manually modified by the user. Also in case of custom development, multiple languages for the navigation menu are mostly foreseen to be developed, although the number of available languages is often reduced to the most common languages (e.g. English, French, German for the European Commission).

No reusable service scenario is put forward for this type of multilingualism, because it does not require a dynamic and recurring translation. The navigation menu is developed and translated once. Updates to the application require possible additional translations.

Metadata and data value level:

Metadata in this context can be defined as descriptive data on the raw data ("data about data"). A good example is the name and description of a data field that can be displayed in multiple languages. The data value level is referring to the raw data themselves.

An example taken from the **study of multilingualism** can illustrate the distinction:



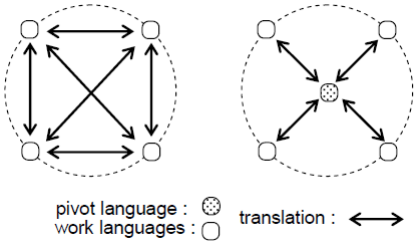
The individual's personal description contains four pairs of values. Each "value" is characterized by metadata, namely first name, last name, gender and profile (or the corresponding German descriptions). Not only these metadata require translation (gender become beschreibung) but also the data itself requires sometimes a multilingual mapping or translation:

- Two properties can be combined to form a new property. In the reverse case, a property can be split into two multiple properties;
- There are values, like the terms for the various genders, which can be part of a list of values;
- And there might be (portions of) a text that should be translated.

According to the **study on multilingualism**, the multilingual mapping should be included in a semantic interoperability asset (e.g. XML schema) to support data exchange in Europe. The

foundation for the multilingual mapping should be a **pivot mapping**[38], which implies that in an asset, a pivot schema is provided and there are multilingual mappings between the local schema of the Member States and the pivot schema.

All data exchanged as well as the defining metadata should be available in the pivot language. Usually English is used as the pivot language in the context of the European Union. The pivot mapping reduces the number of mappings. The figure below shows the principle of a pivot language. To translate from language A to language B, A is translated into the pivot language and the result is subsequently translated to language B. In general the effort is only N in the pivot case compared to N*(N-1)/2 in the non-pivot case, assuming N languages.



A quite special issue is the automatic translation of longer text parts in contrast to the mapping of simple strings or terms. These translations of longer text cannot be included in the Semantic Interoperability Assets, such as XML schema. The related techniques of machine translation are the subject of research and not a component of multilingual practice in the context of SEMIC.EU. Instead of seeking to translate texts, the study on multilingualism recommends to support the end user by using **semantic tagging**[39], which adds additional information to words, such as context information and references to other words or text resources (e.g. EUR-Lex or the DGT Multilingual Translation Memory). A word does not get its full meaning if it is pulled out of its context. Most words can not be translated without having the context surrounding it. In most tagging processes the tags that are given to words are dependent on the context. Semantic tagging is therefore an innovative approach and should be analyzed in a further study as a possible future option to support multilingualism.

The conclusion is that multilingualism is an aspect of semantic interoperability rather than technical interoperability. Therefore it is not relevant in the context of the EIIS study to define a reusable interoperability infrastructure service. The “Study on multilingualism” should be the reference for application environments that want to support multiple languages.

5.6.2.2 Syntax and semantic translation

Data within the public administrations of Member States (MSs) and the Directorates General (DGs) can be stored in various formats (e.g. flat files, XML files, spreadsheets and a variety of proprietary storage methods) and can reside on different operating systems.

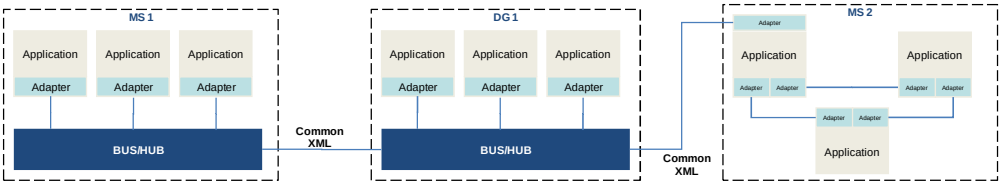
To exchange data between the public administrations of Member States and European Commission DGs, syntactic and semantic translations are required. These transformation functionalities are identified in most of the analyzed systems in scope of this study. However, it is a question whether the functions responsible for the transformations should be available as a reusable interoperability infrastructure service? This is not likely.

---

[38] The original work about semantic interoperability arises in the area of database technologies. Technologies like Extensible Stylesheet Language Transformations (XSLT), XQuery, and Regular Expressions are known in the European Commission, although these technologies use methods that are more syntactically oriented. When it comes to multilingualism, pivot mapping is recommended by the study on multilingualism.

[39] Semantic Tagging, Susanne Ekeklint, Term Paper for Natural Language Processing 1, autumn 2001, GSLT - Graduate School of Language Technology

– *Firstly* because syntactic and semantic transformations are typically executed by locally hosted middleware products, preferably on a bus or hub rather than point-to-point. The bus/hub connects to applications through adapters/connectors, which can be specific to an application or which can interact with any application through a standard communication protocol, such as SOAP. The application-specific format is converted to a common (XML) format, to avoid that every adapter has to convert data from and to every other applications' format. Additional to the syntactic transformations, semantic transformation in many cases needs to be applied as well. This requires common data definitions and mappings between the application-specific data objects with the data objects of the common XML, which is subject of semantic interoperability and thus out of scope of this study.



– *Secondly*, because syntactic and semantic transformations require integration with the application-specific environment.

To illustrate that an existing data transformation service cannot be reused as such, one might think of a solution that transforms the data received via a web portal form into an XML file. Many public administrations could (re-)use such a service. However an existing service of MS administration A cannot be reused by another administration B without adaptation. Every web form will be different in content and thus also the transformation towards the XML schema will be different. Nevertheless, the best practices how such a service has been setup can be shared by administrations A to B.

5.6.3 REUSABLE SYSTEMS

The next paragraphs need to be interpreted as a list describing best practices rather than a list of reusable services. The descriptions should provide the opportunity to the reader to get acquainted with the best practices implemented in each system. More information about the described systems can however be found in deliverable of Phase 2.

5.6.3.1 Multilingualism

IMI (DG DIGIT)

IMI provides Member States with the tools that they need to improve the implementation of Internal Market legislation. IMI has a mature system component which supports the translation of static application data and the content of free text fields. IMI provides a mechanism to export the static data which need to be translated. This greatly facilitates the interaction with the translators, which now have a structured overview of the data to be translated.

The application exports an Excel file for each of the target languages (IMI supports 23 languages). The Excel file contains a unique code for each element to be translated and the label/description of the element in a reference language (e.g. English, French or German). Once translated, the same Excel file can be used to upload the translated elements into the system. Downloading or uploading of Excel files can be done by the translators themselves. The translations need to be approved by another person before they are visible into the production system. In case there is no translation available, IMI can fall back on one of the reference languages. IMI provides web services to enable this automatic translation.

CIRCABC (DG DIGIT)

CIRCABC provides Interest Groups with a private web workspace to collaborate on common objectives and tasks, enabling the effective and secure sharing of resources and documents.

CIRCABC provides multilingualism features at the application level and at the metadata level of the stored documents:

- Application level: the user interface is available in the 23 official languages of the European Union;
- Metadata level: when a document is stored in CIRCABC, translated versions of that same document can be associated via links to the original document. This allows easy navigation from any translation to the original content and conversely (which is very useful when e.g. the original version is subject to change). Multiple versions can be displayed simultaneously. The CIRCABC's search engine retrieves information from the properties and the text body of multilingual documents.

EUCARIS (MS Consortium)

Eucaris II is a system focusing on the secure data exchange of vehicle registrations, driving licences, and the accompanying personal data. Eucaris II is based on XML and web service technology. The EUCARIS web client application supports:

- multiple languages per country/user at the level of the application and navigation menus;
- a user interface enabling an administrator to translate screen items (the data labels, metadata);
- a message interface enabling an administrator to translate coded attributes.

In addition to the Eucaris II web client multilingualism features, the core application of Eucaris II offers a web service to translate the content of the messages (based on the coded attributes) as exchanged between countries. The translated result can be used in customised clients. Therefore multilingualism is not restricted to the Eucaris II web client and the multilingual features are reusable as a service.

5.6.3.2 Syntax translation

e-PRIOR (DG DIGIT)

e-PRIOR aims at improving interoperability between the internal systems of DIGIT and the broad collection of systems used by its suppliers, in a business context of procurement, invoicing and ordering.

e-PRIOR receives documents in XML format from EC suppliers. By using XML transformers (files), the input format is transformed to a canonical format after which e-PRIOR can convert the canonical format to any output format.

The canonical format is an internal format that is based on Universal Business Language (UBL) and is used to shield the system from multiple evolutions of multiple incoming formats. This intermediary format should evolve more slowly than the input formats (Northern European subset (NES) of UBL 2.0 documents). The XML transformers are developed with XQuery or with schematron documents (XSLT), describing the syntax rules for XML documents.

TACHONET (DG TREN)

TACHOnet (stands for Digital TACHOgraph NETwork) has setup an exchange system network between the Member States to check and to guarantee the uniqueness of tachograph cards that are issued. TACHOnet reduces fraud attempts by making sure that the applicant driver does not already hold another card in another Member State.

TACHOnet delivers a central phonex and transliteration service that is used to transliterate Cyrillic, Greek or Latin characters into US/ASCII characters for countries where these ‘special or accented’ characters are not supported by their local databases. The phonex and transliteration service is a web service that has been developed in Java and Microsoft .NET; it is therefore a reusable component. A Member State can choose to implement the Phonex algorithm on its own platform

to run the service locally or can choose to 'call' the Phone Service centrally hosted at the TACHOnet data centre.

5.6.3.3 Semantic translation

X-DIS (DG ESTAT)

X-DIS (XML for data interoperability in statistics) is a Eurostat project in the framework of IDABC. It focuses on the interoperability of statistics, especially in the domains of the Principal European Economic Indicators (PEEI), and more generally business and financial statistics.

The global common approach is to use XML based standards to simplify the use of statistical data. In the context of X-DIS, Eurostat participates to the implementation and support of SDMX processes. SDMX is the ISO standard for the exchange of statistical data and metadata.

In the different work areas, Eurostat provides a set of freely available tools of which the SDMX converter is the most relevant for the data translation service. The SDMX converter allows both manual and automatic conversion between SDMX version 2.0 standard (generic, compact, utility and cross-sectional), GESMES (TS, 2.1, DSIS), CSV and other formats.

5.7 WORKFLOW MANAGEMENT

5.7.1 TRENDS AND EVOLUTION

Over the last decade there has been increasing interest in Workflow and Business Process Management (BPM) systems[40]. Many workflow management systems have been developed to support workflow automation.

As explained in an article of Margie Virdell (2003), the development of the workflow systems have come from two different originating viewpoints: *people-based business processes* (choreography) and *rules-based automation processes* (orchestration). Choreography of Web services reflects workflow's people-based roots, while orchestration of Web services reflects workflow's automation roots. Choreography, by defining behaviors for handling varied and unpredictable interactions among a set of Web services, is more complex than orchestration. Orchestration and choreography of workflows are essential parts of ongoing standards definition work at this time[41].

In regard to this, it is important to clarify the segmentation between autonomous and embedded workflow deployments. Workflows can be embedded in applications or can be deployed as independent applications interoperating with other applications. An autonomous workflow management system is a separate piece of application software that provides the workflow functionality. It defines, creates and manages the execution of workflows, without any additional application software, with the exception of database management systems and message queuing middleware. It is able to interpret the process definition, interact with workflow participants, and, where required, invoke the use of external applications[42].

Today, a large number of workflow offerings, proprietary or open source, have been developed. Workflow systems use a variety of languages based on different concepts (BPEL4WS, BPML, WSCI, etc.)[43]. Among the various solutions proposed, the differences are considerable. One reason for the lack of consensus and the complexity of what constitutes a workflow specification is the variety and the complexity of the business processes.

For that reason, most enterprises seek to limit the costs and difficulty of development of the workflow automation by adding easy-to-use tools, templates or pre-configured solutions.

These general trends and evolution must be taken into account when it is question of the eventuality of a workflow service to support the provision of European Public Services. In a general way, the Member States and the European Commission exchange information in various formats and for various reasons. It means that many processes involve different administrations and applications. It can be shown in many examples.

A first example is illustrated by the Service Directive. In that context, as of 2010, every company will have the opportunity to offer its services all over Europe to further open up the European internal market by reducing bureaucratic obstacles and cross-border constraints. It implies to handle the different steps of the entire procedure electronically.

Another example concerns the field of Customs and Taxations. Customs clearance requires daily exchanges of data between customs authorities of the Member States, between those authorities

[40] The Business Rules Management System must be distinguished from the workflow. Although a business process involves many activities and resources, a workflow is only part of a business process and involves specific automated steps. Business rules should be supported by an independent Business Process Management System. They are maintained separately by a business rule management system.

[41] Business processes and workflow in the Web services world. A workflow is only as good as the business process underneath it, Margie Virdell, e-business Architect, IBM Developer Relations, 2003.

[42] Evolution of the workflow management systems, Krasimira P. Stoilova and Todor A. Stoilov, 2006.

[43] Petia Wohed, Arthur H.M. ter Hofstede, Nick Russell, Birger Andersson1, and Wil M.P. van der Aalst, " On the Maturity of Open Source BPM Systems" , in BPTrends, June 2009

and the economic operators, and between the Commission and other administrations or official agencies involved in the international movement of goods. In that context a seamless flow of data is crucial in order to make the custom clearance efficient.

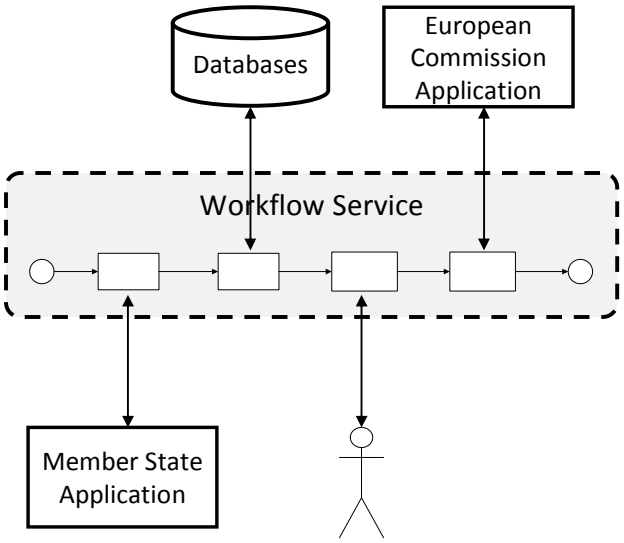
In these examples, although the domains, the business processes, the content and the actors involved in the information exchange greatly differ, the basic principle of the exchange of information stays the same. It is a workflow with the same typical steps (request for information, validation/approval, notification, list of tasks of participants etc.).

This assumption, and the general trends and evolution explained above, will be the basis of the workflow scenario presented hereafter.

5.7.2 IMPLEMENTATION SCENARIO: LIGHT CENTRAL WORKFLOW SERVICE

5.7.2.1 Introduction

The scenario for the workflow service proposes to have a central workflow system. This autonomous workflow is able to interact with participants in the Member States and the European Commission, and invoke the use of European Commission and Member States applications.



Workflow is effective to coordinate large organizational processes such as processes where European Commission and Member States administrations are both involved. With a central independent workflow service used by applications in the Member States and the European Commission, the efficiency and the transparency of the work is improved, and the cost of maintaining the applications is lower as changes to processes/rules do not require changes to Member State and European Commission applications.

When building future applications in the European Commission and the Member States, the use of this central workflow service for the processes concerned will allow to focus entirely on the process, and not on the technical specifications of the workflow. It also provides more flexibility when there are changes in the process, by avoiding having workflows fully embedded in applications.

5.7.2.2 Requisites

The workflow system must be fast, lightweight, scalable and simple to use. The workflow system has the capability to update workflows easily when processes and organizations change. It helps to standardize and monitor the business processes, and improve the collaboration between public administrations.

As already mentioned, the connectivity to other systems is critical. The workflow must be able to be integrated with other IT applications, for example external workflows in the Member States and in the European Commission but also other external components such as Business Rules Management System or notification service.

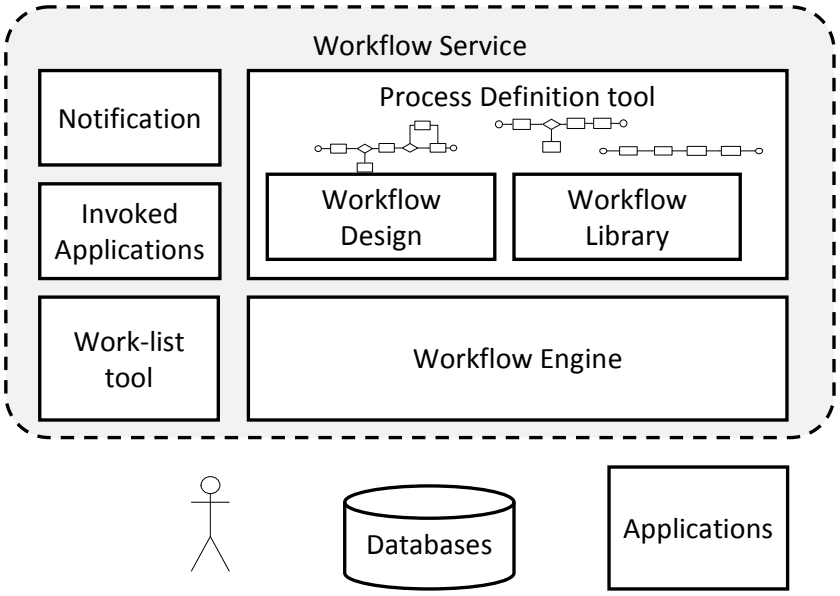
Finally, workflow systems can be made available as web services.

5.7.2.3 Potential services features

We assume that the workflow needed is a simple and light workflow, the totality of the service features presented in this section will not probably be necessary in the service. As the vision is not yet mature in the domain, it is not the right time to make this choice.

The potential workflow architecture could contain the following modules: Workflow Engine, Invoked Applications, Work-list tool, Process Definition tool (Workflow Design and Workflow Library), Notification.

The workflow service can be represented by the picture below:



The central element of the service is the **workflow engine** that runs the workflow tasks step by step. Some steps of the process require human intervention, but other functions are automated, and handled by the system. The engine is the logical center of the system, and the other modules interact with it.

Workflow is linked to applications (**invoked applications**), in addition to an effective work-list interface that allows the human operators to find out the tasks currently assigned to them (**work-list tool**).

The **Process definition tool** is used to create and change process definitions. The tool may be a component of the BPM system, a stand-alone application, or a component of a workflow management system. Process definition tools can contain the workflow design module (creation of the workflow) and the workflow library (store all the workflows created). It provides thus the ability to reuse stored workflow elements, and even entire subprocesses[44].

At a certain step of the workflow, the **notification** service could be used to notify the participant of the process.

5.7.2.4 Challenges

The workflow service presents many open points of discussion in order to define the services needed, among others:

- Who are the participants? The participants involved in the workflow can be applications, persons, Web services, and other workflows.
- What do the participants do? Some workflows are completely automatic, and some consist of manual tasks that must be performed by people. Workflows are frequently a combination of the both types.
- How to avoid the creation of a high number of workflows which are finally not used? Isn't it advisable to have a governance instance that follows up the use of the workflow in the workflow engine?
- Etc.

5.7.3 REUSABLE SYSTEMS

In most of the systems studied during Phase 1 and 2 of the EIIS, the workflows are embedded in the systems and do not exist as independent applications.

However, some systems can be proposed as reusable components to implement the workflow scenario. They have been studied during Phases 1 and 2 (for detailed information, please consult the document of Phase 2).

5.7.3.1 IPCIS (DG DIGIT)

The component selected for reuse is the future "Business Process Management" solution. It will provide business rules management, business process modelling and the business process execution, monitoring (rule engine, statistics). It will support interactions with humans (human workflow).

5.7.3.2 FIDES (DG MARE)

DG MARE is currently working on a new Open Source version of the FIDES workflow (OS – FIDES 3 workflow). The OS-FIDES 3 will have the same architecture, design and features as the existing FIDES 3 system without using any commercial products. OS-FIDES 3 workflow is Java-based and uses the Business Process Execution Language (BEPL) notation. It is conceived as light, simple, flexible and easy to use.

[44] Business processes and workflow in the Web services world. A workflow is only as good as the business process underneath it, Margie Virdell, e-business Architect, IBM Developer Relations, 2003.

5.7.3.3 NOTIS(DG DIGIT)

NOTIS can be reused for the notification feature only. NOTIS handles notifications of events which occur in Information Systems of the European Commission. The application communicates different events to the users. For instance, it informs users that a message is available or that the user has to perform a task. NOTIS consists of the following main components: NOTIS message consumer, NOTIS web application, NOTIS message counter and NOTIS Desktop Client application.

5.8 DOCUMENT STORAGE

5.8.1 TRENDS AND EVOLUTION

In most of the organizations, the management of (electronic) documents has become a high priority since many years. For that reason, electronic document and content management systems have experienced significant changes since the 1980s.

The first systems were centred on scanning and storing of paper documents as digital images. The main capabilities were capture, storage, indexing and retrieval of image files. Later, the systems enabled the organisation to manage electronic documents (created on computers, and often stored on local user file systems). Today, after roughly 20 years of improvement and innovation, electronic document management is no longer a simple document storage solution used to keep digital documents after the work has been completed. The systems have become an integrated component of the daily work[45].

Most of the today's document management systems are able to manage any type of file format and encompass namely collaboration tools, security, and auditing capabilities. However, storing and retrieving documents remains one of the major challenges in the information exchange between the European Commission and the Member States. Appropriate management and preservation of the documents pose specific problems. All of the administrations of the Member States and the European Commission are generating documents that need to be shared with other administrations, as well as being safely archived. The execution of large processes performed by public administrations and involving many employees depends highly on the treatment and transfer of documents.

On one hand, European Commission and Member States administrations have to share documents while offering services to citizens with all the legal guarantees, and thus have to follow specific rules regarding data sensitivity, transparency obligation, etc. On the other hand, they have to work with many different systems of document management existing in the different administrations.

In that context, the idea of a reusable Document Storage service to improve the information exchange between Member States and the European Commission is challenging. Such a service would allow the Member States and the European Commission to share safely their documents and have access to documents produced by other administrations. There would be a standardised access to all needed shared information within the European Union so that it would be available for use at any time within the processes.

The next sections propose 2 scenarios to implement the service.

5.8.2 IMPLEMENTATION SCENARIO 1: CENTRALIZED DOCUMENT STORAGE

Scenario 1 envisages implementing a centralized document storage service for the European Commission and the Member States. The service would allow the Member States and the European Commission to store and file the documents in a central database, and to search and retrieve the documents from this central repository.

Given the context explained in the section 'Trends and Evolution', many obstacles appear to implement this scenario, typically in the areas concerning the location, the security, the legal basis, the retrieval, the retention period, the archival storage, the distribution of the documents, etc.

Open questions

- Which kind of documents have to be stored centrally?

[45] WestBrook Technologies, SOA: The Next Technology Inflection Point, in Business Management US, November 2009.

- Is the central repository the unique source of storage for the document?
- Etc…

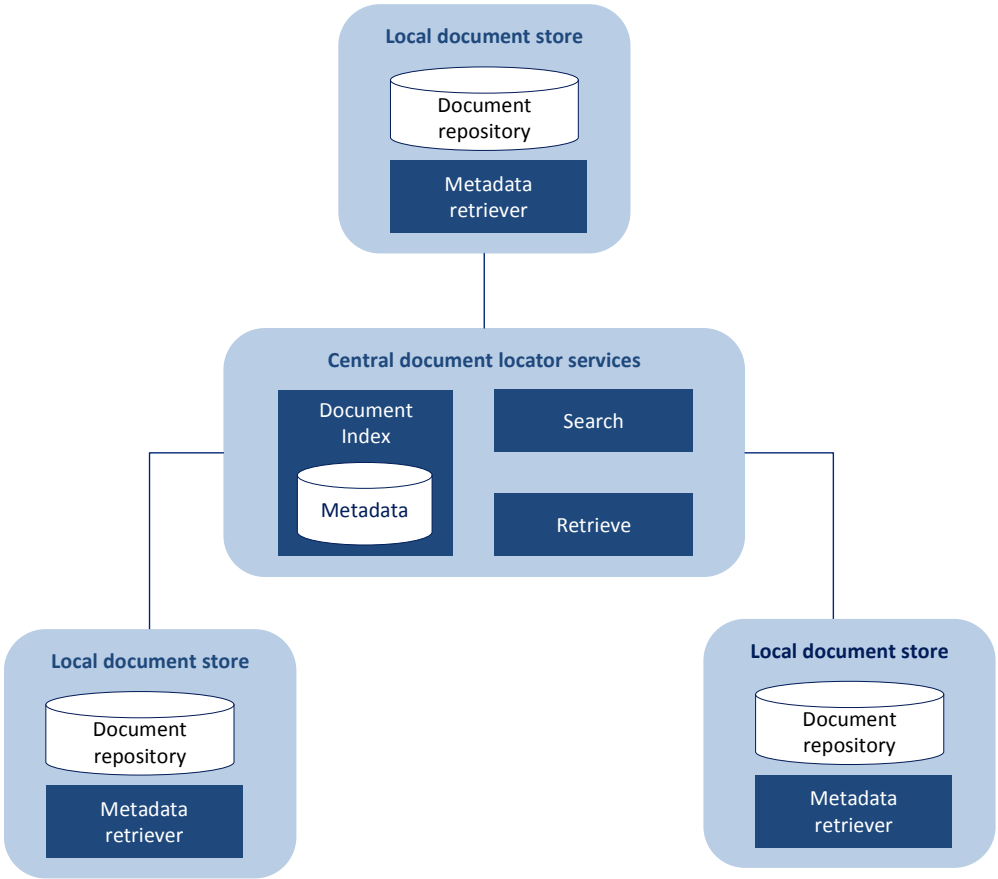
Except if the centralized document storage concerns a specific domain or sector and is related to a specific legislation, the scenario seems highly improbable and for that reason won't be further considered.

5.8.3 IMPLEMENTATION SCENARIO 2: CENTRAL DOCUMENT LOCATOR SERVICES

To overcome the issues presented in the previous scenario, Scenario 2 proposes to implement central document locator services to better leverage on the existing systems.

The scenario is based on a distributed model without a central document repository. Instead of using a centralized database to share the documents, the Member States and the European Commission keep their documents in their own repository. They manage storage, versioning and archiving following their own rules and systems.

As illustrated in the figure below, there is an interoperability application-independent layer (Central document locator services), supporting distributed document management and providing a set of (web) services. Locally, each DG and MS keeps its document storage (Local document storage). Also locally, a metadata retriever extracts the metadata associated to the documents stored. At central level, the metadata are indexed and stored so that the documents are easily retrievable. The central locator services offer the possibility to query the different local document stores whenever a document is needed. The documents need to be available on demand, and downloaded as needed. The Member States and the European Commission connect their systems so that they can retrieve and access documents stored in other locations. Other services could be added at the central document locator services if necessary.



Open questions

In that scenario, it is necessary to focus on some discussion points that have to take place before the implementation. In that scenario indeed, although the MS and the EC keep their system locally, important obstacles exist. Some of the major ones are listed here (the list is not exhaustive):

- Which set of services have to be implemented at central and local level?
- Which documents have to be exchanged between the administrations? Public Administrations produce millions of documents and it is difficult to imagine that all the documents produced have to be exchanged. In this regard, it may be interesting to start with a limited number of documents (e.g. in a specific sector).
- The concept of document itself is not always clear. What is a document in the context of sharing documents between EC and MS? Is it about paper documents, electronic documents, working documents, official documents…?. Many definition and rules exist in that area (i.e. e-Domec[46], MoReq[47]), but a decision is a preliminary step to a document storage service for the EC and the MS.
- Another obstacle is the definition of the set of metadata. Which metadata have to be associated to the documents? The metadata model of MoReq2[48] may be a helpful source of information in that domain.

5.8.4 REUSABLE COMPONENTS

The infrastructure to implement the document storages service scenario has not been found in the studied systems. However, some systems can be proposed as reusable components. They have been studied during Phases 1 and 2 (for detailed information, please consult the document of Phase 2).

5.8.4.1 CIRCABC (DG DIGIT)

CIRCABC distributes and manages electronic documents and files in any format, many languages and with version control. CIRCABC is based on Alfresco which is an Open Source Java-based Enterprise Content Management tool, with some enhancements to the original software.

CIRCABC is implemented in two different ways. On one hand, CIRCABC is designed to work on a full Open Source platform. It is offered as Open Source Software to the EU public administrations, businesses and citizens. In that sense, it could be implemented as Local Document Store (Document database) for the DG or the MS, in the context of Scenario 2.

On the other hand, CIRCABC is also deployed centrally within the DG DIGIT Data Centre. In that context, it could be used for centralized document storage at European level, for example limited to working documents, or specific sectors (Scenario 1).

---

[46] "Electronic archiving and Document Management in the European Commission". The European Commission officially started the e-Domec project, the new electronic archiving and document management policy of the Commission, in January 2002.The Secretariat-general is responsible for the normative part and is also project owner for several informatics applications of document management.

[47] Model Requirements for the Management of Electronic Records, Update and Extension, 2008, MoReq2 Specification. MoReq2 consists of a formal requirements specification for a generic electronic records management system, accompanied by testing documentation and related information. Published in 2008 by the European Commission, it is intended for use across the European Union, but can be used elsewhere. MoReq2 is generally considered a de facto standard, in Europe, but it does not have any formal status as a standard.

[48] Model Requirements for the Management of Electronic Records, Update and Extension, 2008, APPENDIX 9 TO THE MoReq2 Specification Metadata Model.

5.8.4.2 jCore Document Storage or JDS (DG AGRI)

JDS is conceived as a generic reusable component. It could also be used to implement Local Document Storage (without Metadata retriever), in Scenario 2. It is based on the Java Spring Framework and consists of 3 components: JDS-client (Java API), embedded in the client application; JDS-server which is a standalone application, responsible for serving documents directly to a browser, and JDS-database.

5.8.4.3 Hermes Repository Services (DG DIGIT)

Hermes Repository Services (HRS) have to be used for the central document storage of EC official documents. These are web services that allow local applications in DGs of the EC to connect to the Hermes common repository of documents and files. They act as an intermediate layer between the Hermes core services and client applications. The question of the reuse of Hermes within the Commission is not appropriate, because Hermes is aimed to be used in all DGs. Deployment of Hermes outside of the European Commission seems difficult because of the business rules which are specific to the European Commission.

5.8.4.4 CEH (MS Slovenia)

The CEH component is a custom-built solution for document storage, which could be implemented as Local Document Storage (without Metadata retriever), in Scenario 2. CEH is based on the Open Source Struts framework using Open Source components.

5.8.4.5 INSPIRE geo-portal (DG JRC)

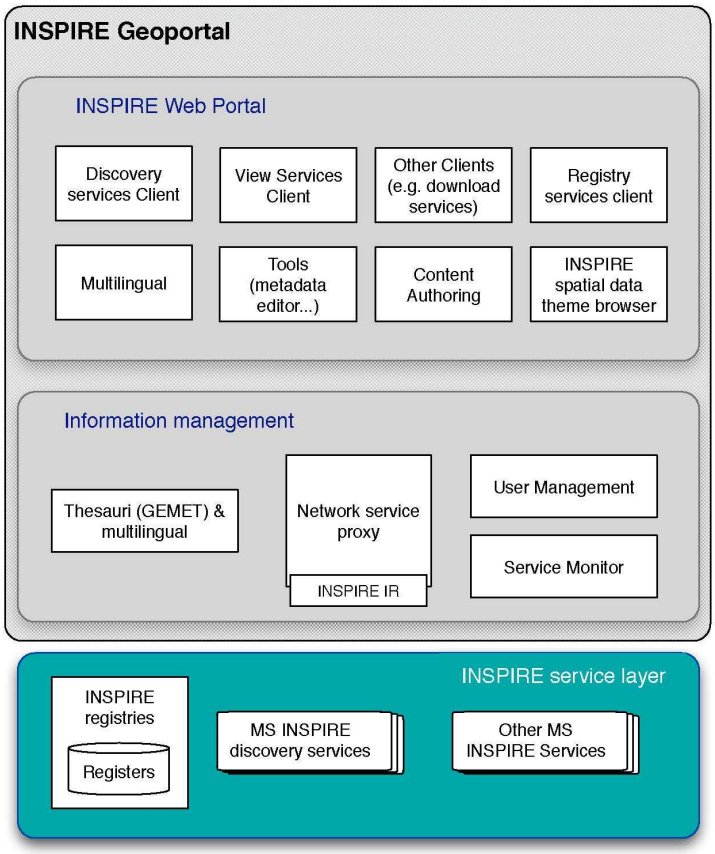
The INSPIRE geo-portal is seen as a guideline to implement the scenario of the central document locator services, because it implements distributed data management in a specific area (spatial information).

The INSPIRE Directive (2007) lays down general rules to establish an infrastructure for spatial information in Europe for the purposes of Community environmental policies and policies or activities which may have an impact on the environment. INSPIRE shall build upon infrastructures for spatial information established and operated by the Member States.

The Directive requires the Commission to establish a community geo-portal. The INSPIRE geo-portal is an Internet site providing access to the INSPIRE network services. The geo-portal does not store or maintain the data; instead the portal uses the data from the original sources in real-time. It acts as a gateway to geographic data and services, distributed around Europe, allowing users to search, view or, subject to access restrictions, download geographic data or use available services to derive information.

The INSPIRE geo-portal includes a proof of concept metadata editor compliant with the INSPIRE Regulation on Metadata. The editor implements the ISO mapping according to the INSPIRE technical guidelines. A validation procedure is also provided to allow users to validate their metadata for INSPIRE compliance.

The portal also includes registers and registry services. Certain resources need to be maintained properly and be made available online to the community for proper functioning of the infrastructure. These resources are maintained in registers, which have to have a clear and well-defined governance model. The register contents are made available in form of a registry. The registries are expected to be available through registry services.

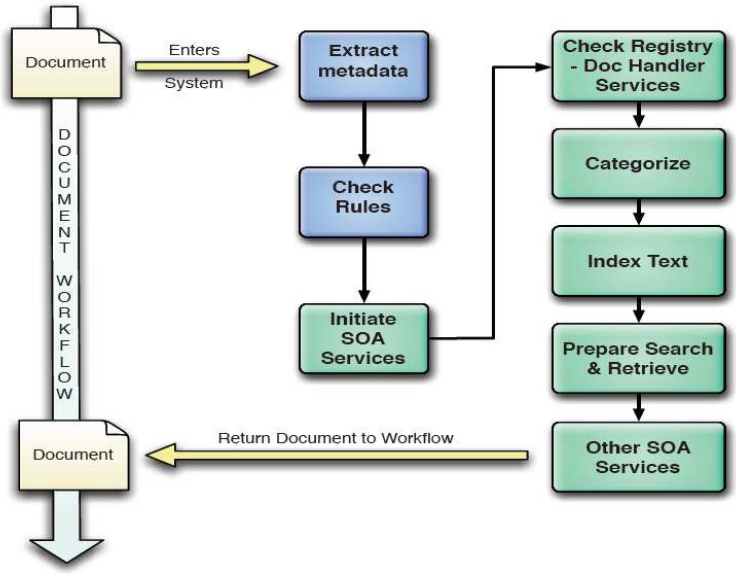


5.8.4.6 Other examples

Although not found in the systems studied, some other examples can be used as guidelines to implement the scenario of Central document locator services.

An example is given in a large Government Agency, in an article on Open Source SOA Platform[49]. The Agency turned to SOA to solve its document management issues. The Agency receives a large number of documents daily and needs to make them available to its users to perform the work of the agency. However, the Agency was under tight budget constraints. The Agency started by building J2EE services connected through the JBoss ESB. Faced with a situation in which there was no predictable document flow, the agency built a business-driven system based on services that would extract document meta data, categorize and index specific text, prepare documents for user search and retrieval, and even do translation. The figure below illustrates the approach adopted by the Agency.

[49] HURWITZ & Associates, Understanding the Business Benefits of an Open Source SOA Platform, 2009.



Other examples can be found in the Healthcare sector, or in frameworks developed for public administrations, where the document management has been a high priority since a long time[50].

[50] Examples:
Healthcare sector: TUCCI L., Record locator service a step to health information exchange, in Senior News Writer, 16 Sep 2009 (www.SearchCompliance.com).
Public administrations: PETTENATI M.C., PARLANTI D., CHINI D., PIRRI F., University of Florence, InterDataNet: Interoperability Framework to Support Collaborative Creation and Management of Official Documents in e-Government Processes, 2008.

5.9 STRUCTURED DATA STORAGE

5.9.1 TRENDS AND EVOLUTION

During the past decade, the amount of electronic data created by the organizations has grown at a staggering rate. Most organizations have responded to this trend by managing critical information in structured environments like databases and spreadsheets. While there are many differences between structured environments in terms of both complexity and functionality, all the systems share a common premise: they define schemas of rows and columns that are designed to promote efficient data storage and retrieval.

With data growing exponentially, it has become very hard to create comprehensive and accurate structured data storage. The structured data sources are often very large, containing millions of rows of data. Databases this large are expensive, and they also require greater expertise when it comes time to analyze and clean up data. Another challenge is that structured data sources frequently contain sensitive data[51].

Much of this information is stored in data warehouses, and is therefore not readily accessible using traditional search technology. For that reason, search technologies have been developed to find the information needed. For example, structured data search, also known as data discovery, is optimized for discovering data that is typically stored in databases. It provides a fast and simple self-service tool for IT to provide to business users who lack deep SQL knowledge or data modelling skills[52].

However, other common problems still exist in the structured storage, namely the lack of agreement on vocabularies and schemas. Existing information-processing methodologies require that all the communities involved in generating, processing, or consuming the same information agree on a given schema and vocabulary. Database views have been designed to alleviate this problem, yet views do not solve the schema heterogeneity problem in general[53].

In that context, an efficient exchange of structured information between public administrations remains very difficult. The objective of the "structured data storage" service is to facilitate the exchange of data by providing a simple and structured interface to access data stored in large and complex databases. This service acts as an abstraction layer between the technical data structure of a database and the functional point of view of a standard user.

5.9.2 IMPLEMENTATION SCENARIO 1: CENTRAL DATA LOCATOR SERVICES

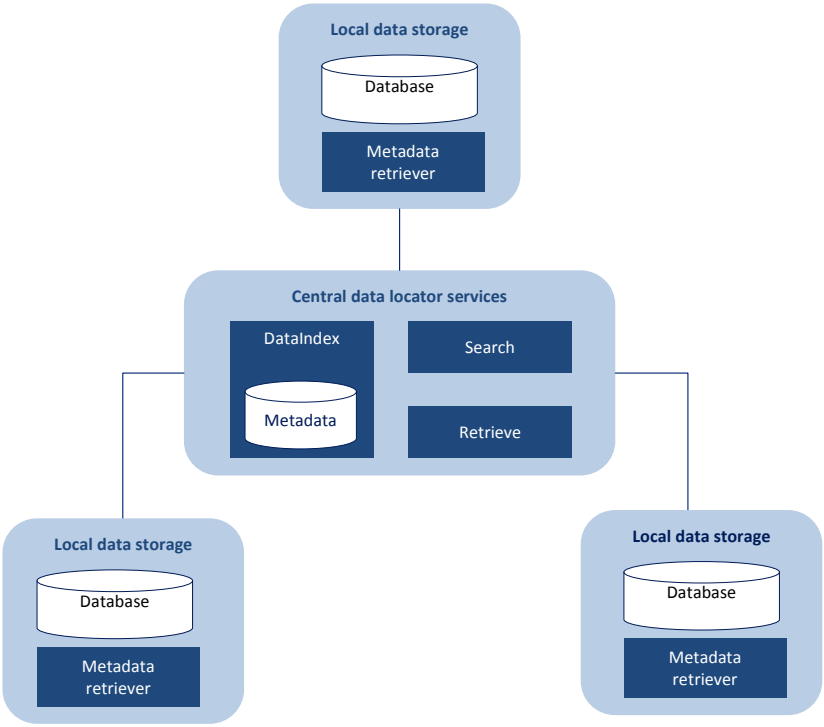
This scenario describes how central data locator services could function (following the same principle as the central document locator services).

The scenario is based on a distributed data model and is illustrated in the figure below. Locally, the Member States and the European Commission keep managing their data in their own database, following their own methods, rules and systems (**Local data storage**). Also locally, a **Metadata retriever** extracts the metadata associated to the data stored. At central level, the metadata are indexed and stored so that the data are easily retrievable. The **Central data locator services** offer the possibility to query the different local data stores whenever a data is needed.

[51] ALLEN Lauren, REEVES Angela, Discovery and Databases: Understanding the Basics of Structured Data in Litigation, in IE Discovery, 2009.

[52] BESEMER David, The Next Wave in the Search Revolution, in Information Management Magazine, December 1, 2008.

[53] FLORESCU Daniela, Managing Semi-Structured Data, December 8, 2005.



Open questions

The open questions are very similar to the open questions concerning the Central document locator services:

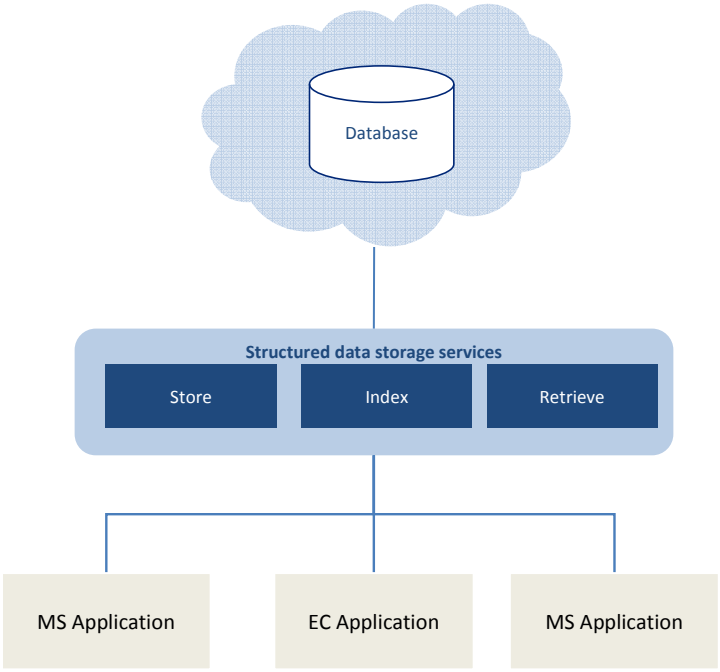
- Which set of services have to be implemented at central and local levels?
- How to define the set of metadata?
- Which kind of data have to be searched and exchanged? Does it have to concern a specific domain?
- Etc.

5.9.3 IMPLEMENTATION SCENARIO 2: MANAGING STRUCTURED DATA IN THE CLOUD

The second scenario is a cloud-based[54] scenario to manage the structured data. The emergence of cloud computing has forced companies to re-think about where their data are stored, as well as which technologies are best to support the structured data storage.

The scenario proposes, via a web service, to automatically store, index and query the data. While a traditional database is complex to design, and often requires extensive and repetitive database administration, this kind of database in the cloud requires no schema. It is cheap and quickly scalable.

[54] Cloud computing involves making computing, data storage, and software services available via the Internet, transforming the Internet into a vast computing platform.



Recently, a number of providers including Amazon and Google have introduced cloud computing infrastructure services. These services include database management services such as Amazon SimpleDB, and Google BigTable[55].

Amazon SimpleDB is a web service providing the core database functions of data indexing and querying in the cloud. By offloading the time and effort associated with building and operating a web-scale database, SimpleDB allows developers to focus on application development.

Bigtable is a distributed storage system for managing structured data that is designed to scale to a very large size: petabytes of data across thousands of commodity servers. Many projects at Google store data in Bigtable, including web indexing, Google Earth, and Google Finance. BigTable presents a simple abstraction that is useful in many circumstances and easy to scale.

Open questions

Although this scenario presents many advantages (less complex, scalable, cheap, easier retrieval of the data, and thus improvement of the information exchange), some major questions remain open, among others:

- How to guarantee that the database works properly? Database remains a core competency in most of the organizations.
- How reliable, fast, performing is it?
- How can the security and the privacy of the data be ensured?
- Etc.

5.9.4 REUSABLE SYSTEMS

The infrastructure to implement the scenarios has not been found in the systems analysed by the EIIS study. As already mentioned in Phase 2, the systems studied for structured data storage can

[55] Feuerlicht George, Govardhan Shyam, SOA: Trends and Directions, in Systems Integration, 2009.

provide patterns, methods and best practices to support the implementation of structured data storage. However, they do not provide generic reusable components for structured data storage (for detailed information, please consult the Phase 2 report).

5.9.4.1 eFP7 (DG DIGIT)

ePF7 covers the service features data model definition and historization, with the URF/PDM (Unique Registration Facility/Participant Data Management). URF is a service providing unique registration for participants in the Research Framework Programmes at proposal and contract stages. It encompasses a set of rules for data acquisition and validation. URF/PDM takes care of all the data management related to the participants to research programs. There is a unique database, and the history of previous data is recorded. If a change occurs in the data, it is logged as an additional data, so that the original data in the database is never modified. The same goes for deletion of data: the original data remains in the database although it has been deleted by the user (and appears as such for the user).

5.9.4.2 Target Data Storage (TDS) and Metadata Handler (MH) of CVD (DG ESTAT)

Target Data Storage (TDS) is a model for storage of data and metadata within Eurostat. It is the central component of the CVD architecture. TDS either stores directly the metadata in its default database system or links to specialized database systems that host the data. In that case, the TDS provides access layer components permitting the creation of requests and reception (reading and writing) of data from these specialised database systems.

Metadata Handler (MH) is a system to manage statistical metadata. Within a CVD compliant architecture, the complete set of metadata for the domain will be defined via the MH and stored in the TDS. The MH will provide a single working environment for all preparation and administration of the diverse types of metadata. The MH will contain an overall view of the existing metadata types and include a logical view of the relationships between the different types and occurrences of metadata. The degree of coupling between data and metadata is inherently tight, i.e. data and metadata interact in many aspects.

5.9.4.3 SDMX Tools (DG ESTAT)

SDMX consists of technical and statistical standards and guidelines. It also provides the tools for efficient exchange and sharing of statistical data and metadata, not only in the transmission of data to Eurostat but also in the exchange of data between CVD components.

In the context of the X-DIS project, the SDMX registry, which is part of the MH, provides support for data model definition. It can be reached via a web graphical user interface and is also available as a web service.

Eurostat provides other freely available tools that support the creation of SDMX data structure definitions and their visualisation. The Data Structure Wizard is a tool for offline development of SDMX Data Structure Definitions (DSD). It can connect to the registry and upload/download DSD to/from the central repository.

The SDMX data can be visualised with the Eurostat SDMX Visualisation Tools that provide graphical representation of the statistical data series carried by the SDMX messages.

5.9.4.4 INSPIRE geo-portal (DG JRC)

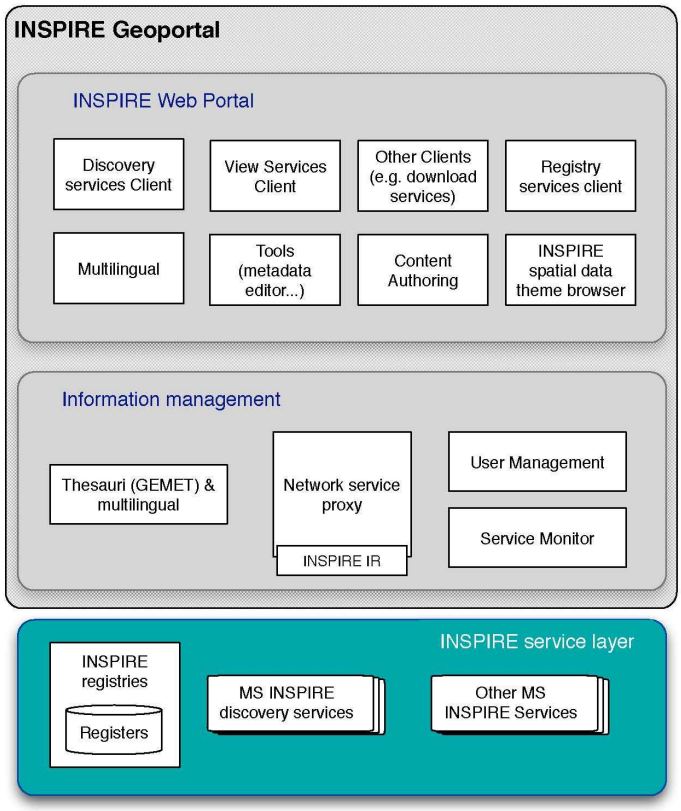
The INSPIRE geo-portal can be used as a reference architecture to implement the scenario of the central data locator services. INSPIRE implements a distributed data management system in the area of spatial information.

The INSPIRE Directive (2007) defines general rules to establish an infrastructure for spatial information in Europe for the purposes of Community environmental policies and policies or activities which may have an impact on the environment. INSPIRE shall build upon infrastructures for spatial information established and operated by the Member States. INSPIRE aims at making available relevant, harmonised and quality geographic information to support formulation,

implementation, monitoring and evaluation of policies and activities which have a direct or indirect impact on the environment.

The INSPIRE Directive requires the Commission to establish a community geo-portal. This INSPIRE geo-portal is an Internet site providing access to the INSPIRE network services. The geo-portal does not store or maintain actual data; instead the portal queries geographical data from the original sources within the Member States in real-time. It acts as a gateway to geographic data and services, distributed around Europe, allowing users to search, view or, subject to access restrictions, download geographic data or use available services to derive information.

The portal also includes registers and registry services. Certain resources need to be maintained properly and be made available online to the community for proper functioning of the infrastructure. These resources are maintained in registers, which have to have a clear and well-defined governance model. The register contents are made available in form of a registry. A registry is expected to be available through registry services.



6 Architectural perspective

Reusable components are the subject of the EIIS study. When talking about reusable components delivering interoperability infrastructure services, the Service Oriented Architecture (SOA) paradigm should not be overlooked. SOA has been widely recognised and accepted as a concept which will enable and support reuse[56].

Service-oriented architecture has equally attracted much attention in government IT circles. This is because SOA directly addresses government most pressing goals: integration of program functionality and information across organizational boundaries in a heterogeneous technology environment[57].

This chapter will look more into detail Service Oriented Architecture paradigm (SOA) and how the European Commission could benefit from this concept in the domain of interoperability infrastructure services.

6.1 SOA IN GOVERNMENT

Government institutions across the world, at national, regional, and local levels, are significant consumers of technology. Governmental services affect us all. They can range from areas of defence and national security, to health, taxation, law enforcement, justice, environment, energy, social services, disaster management, and land use management.

Technology clearly needs to play an important role for any organization responsible for any of these areas. In the public sector, automated systems perform a common, fundamental function: getting information to and from the "users." In other words, systems need to be in place to effectively share data between agencies and the public community comprised of citizens and businesses.

However, as with large private corporations, information technology in government institutions has been built in silos, where the interoperability and exchangeability of information is only an afterthought, neglecting service-orientation concerns. For example, government agencies tend to structure information silos in terms of agency-specific objectives. The more agency-specific silos are created, the greater are the integration problems that will have to be overcome down the road, when agencies need to communicate with each other (or between agency user communities).

SOA is particularly suited to help government agencies deal with the obstacles to implementing the new systems that will enable them to modernize their business architecture, integrate agency service delivery, and share information across organizational boundaries.

6.2 EXAMPLES OF SOA IN GOVERNMENT

SOA principles have been applied into practice. Below one can find two examples where SOA principles have been applied. Both examples provide a simple basic infrastructure with a basic set of services which are accessible via portal or via a service bus. Both of the examples show practical information exchange in a complex distributed public environment where there are a wide range of actors come into play.

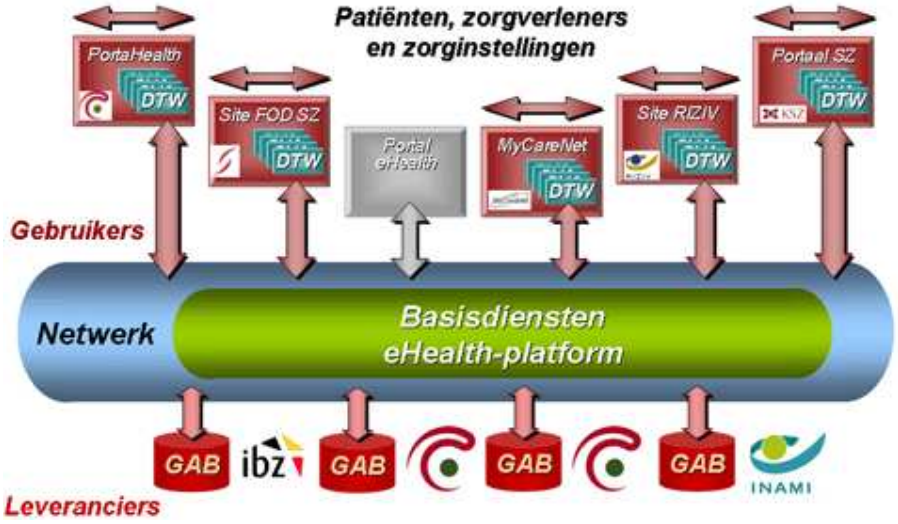
[56] The Value of SOA; Gartner Survey Update; Daniel Sholler and W Schulte, 2009.

[57] Forrester Report, Why Is SOA hot In Government?, 2006.

6.2.1 BELGIAN EHEALTH PLATFORM

The Belgian eHealth platform offers a number of basic services which can be used for free by all actors in the public health sector. These basic services can be used by the IT providers to build business services or which can be used to provide access to authentic data sources.

The basic services provided by the eHealth platform are identity & access management, audit trail & log, time stamping, electronic post office, end-to-end encoding and service orchestration.



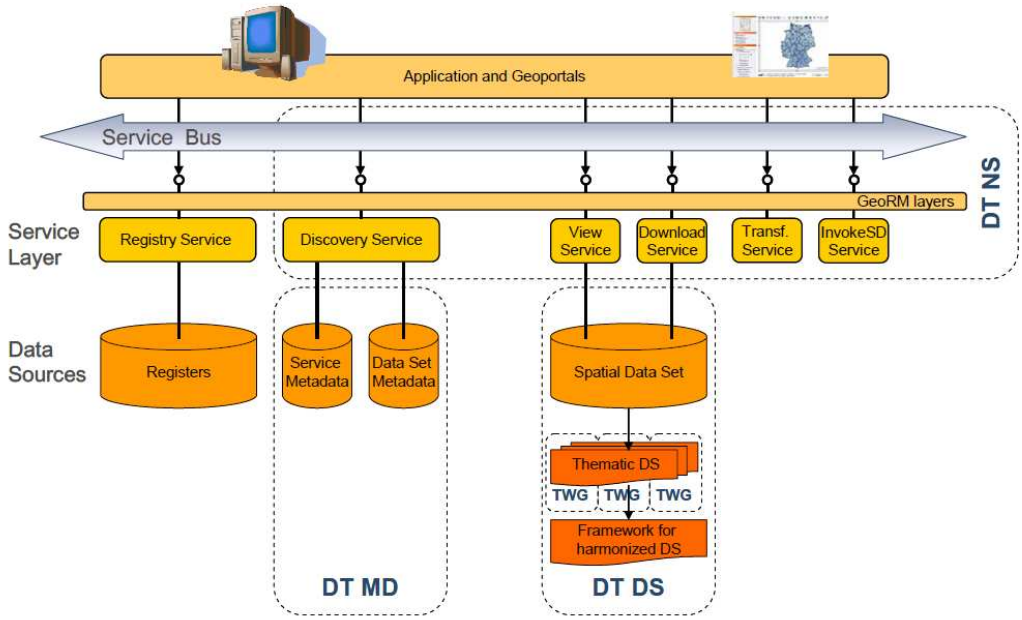
6.2.2 INSPIRE GEOPORTAL

The INSPIRE directive aims to create a European Union (EU) spatial data infrastructure. This will enable the sharing of environmental spatial information among public sector organisations and better facilitate public access to spatial information across Europe.

The INSPIRE Directive requires the Commission to establish a community geoportal and the Member States shall provide access to their infrastructures through the geoportal as well as through any access points they themselves decide to make available.

The current version of the INSPIRE geoportal is a prototype and allows for discovery and viewing of spatial data sets and services. Its aim is to identify issues related to its implementation and to access distributed INSPIRE services, to facilitate the development of the operational geoportal.

At the core of the architecture are the INSPIRE Service Types: Discovery, View, Download, Transform and Invoke. These INSPIRE Services are accessed via the Applications or Geoportals through both of them using the INSPIRE Services Bus.



6.3 SOA MATURITY MODELS

To reap the benefits of SOA (loose coupling, interoperability and reuse), a strong architectural backbone is important. But it is also necessary to assess the SOA maturity in an organization which this means: assessing the effectiveness of processes, technology choices and also the technical quality of the architecture.

There have been numerous publications about SOA maturity models[58]. The goal of this paragraph is to learn something from these maturity models on the typical activities an organisation has to perform to evolve towards a service oriented organisation.

All of the published SOA maturity models show that there is a logical evolution with common characteristics on three domains of a SOA maturity model; the SOA governance, the organisational scope of the architecture, and the type of services which are delivered through a Service Oriented Architecture.

The following paragraphs detail how the SOA maturity influences the SOA governance, the organisational scope and the type of services delivered in a SOA architecture.

6.3.1 SOA GOVERNANCE

SOA governance relates to all activities which are not immediately required to build working services, but aim to increase overall quality of the SOA and enable control in complex environments. This governance relates to people, processes and products.

[58] - Service Integration Maturity Model (SIMM), developed by IBM.
-The Open Group SOA Working maturity matrix (OSIMM)
- Microsoft Maturity Model (SOAMM)

The Deloitte SOA lifecycle governance model[59] can serve as an example of how to address the SOA governance in a lifecycle based maturity model.

The model defines six main processes which need to be executed to perform good SOA governance. The lifecycle represents one iteration, a scoped project to achieve a concrete goal in implementing a Service Oriented Architecture. SOA development takes place in many iterations of the SOA governance lifecycle. Through these iterations, the maturity level can (and should) increase.

The SOA governance processes are briefly explained below.

Create a SOA Strategy: This process specifies the organization's perspective on SOA, but also the goals set for the current iteration on the lifecycle.

Creating the SOA organization: This process defines the roles and responsibilities for SOA governance. Often this leads to the initiation of some kind of SOA governance board (or Centre of Excellence). This board should represent all the stakeholders of SOA governance.

Service portfolio management: Together with business representatives, consensus is needed about the services that will be developed. The architect should weigh IT arguments against business arguments for developing specific services. By being involved in portfolio management from the start, the architect should be able to point out services that are suitable for reuse and are therefore preferred to be developed early on. A services roadmap, listing current and planned services, is a possible product of service portfolio management.

Service lifecycle management: This process addresses the implementation, updates and retirement of enterprise services. This process should help in ensuring that all services are in scope of lifecycle management, to prevent the presence of ungoverned services.

Policy enforcement: This process concerns the design and enforcement of policies.

Service level management; SOA requires different service level management than other IT architectures, because of the finer granularity of services, compared to applications. Also, the flexible consumption of services developed by others requires good understanding of service levels by consumers.

The scope of the activities discussed in the SOA governance lifecycle above is very broad and an immediate adoption of all of them is not realistic. Imposing excessive governance procedures can degenerate in a source of unnecessary bureaucracy, which is especially dangerous for SOA-immature organizations. A well established SOA governance structure would potentially only slow down the development and reuse of infrastructure services.

The following steps however can be interesting to take and to start and initial SOA governance:

- Determine scope for proof of concept and identifying series for pilot;
- Development of technical skills and capabilities;

[59] More information on SOA governance lifecycle can be found in:
SCHEPERS T., KRATZ B. (Deloitte), SOA Governance Maturity – an Architect's View, June 2009
http://www.infoq.com/articles/soa-gov-architect-s-view.
SCHEPERS T., IACOB M., VAN ECK P. (Deloitte & University of Twente), A lifecycle approach to SOA governance,2008.

- Document best practices and maintain a list of running services.

6.3.2 SOA ORGANISATIONAL SCOPE

The SOA enablement of an organisation will not happen overnight. An organisation has to increase the scope of SOA adoption gradually, from the inside out, starting with one domain and moving on to other domains of an organisation, and finalizing to include the business partners outside of the organisation. An increased scope of adoption would therefore require an increased level of SOA governance. The organizational scope is a good indicator for the SOA maturity. Most of the SOA maturity models identify 4 levels of organisational scope: Pioneer, Process, System and Network.

- **Pioneer–** SOA consists of small-scaled projects, meant to offer flexibility to determine the best practices. These efforts are often quick-wins and incur a low organizational complexity. SOA Governance is minimal, and focuses on the scope of a project. As there is usually one owner of the SOA, control is not very complicated.
- **Process**– here an organizational unit, product or process is fully committed to SOA. Often SOA is implemented with a specific purpose, e.g. inventory management. One owner is in the lead here, and funds the majority of the SOA
- **System–** Here the SOA is controlled by multiple owners. This SOA governance needs to become centralized. SOA governance dictates standards for consequent implementation of SOA in all parts of the organization.
- **Network–** SOA is coordinated by different, actively involved, organizations. These organizations provide services for each other, which are connected to support end-to-end business processes. This means that SOA governance between these parties needs to be aligned. Without proper governance, service changes will have a major and unpredicted impact on the business operations.

The networked SOA organisation is clearly the 'to be' state for scope of the SOA organisation for interoperability infrastructure services for the European Commission. The Commission by its nature is forced to cooperate in an integrated way with national and international administrations. In order to make a SOA approach successful, it should be supported with a networked organisation governance.

6.3.3 SOA SERVICES SCOPE

When the SOA experience and maturity increases, the type of services also changes[60]. One can identify 5 types. Many take the traditional view of the SOA Maturity Model (some of which are based on the CMMI Maturity Model), which in essence classify your implementation into 5 different levels of maturity. With the growing organisational needs on services, the type of services will change**.** Five classes of services have been identified:

- SOA maturity level 1 – Initial Services
- SOA maturity level 2 – Architected Services
- SOA maturity level 3 – Business & Collaborative Services
- SOA maturity level 4 – Measured Business Services
- SOA maturity level 5 – Optimized Business Services

[60] http://www.soainstitute.org/ SOA Maturity Model: Compass on the SOA Journey; By: Theo Beack, Chief SOA Architect, SoftwareAG

1: Initial Services

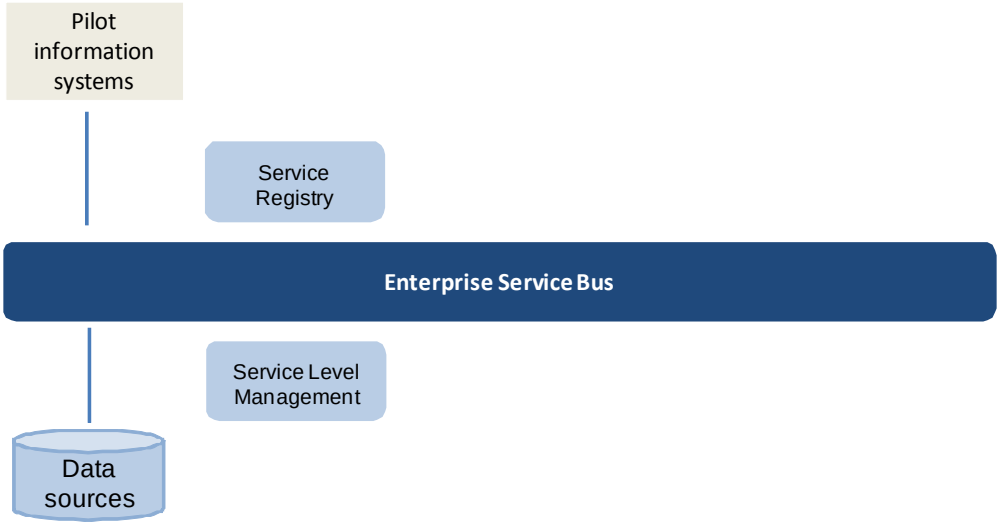
The initial services represent the initial learning and project phase of a SOA adoption. Projects here are typically done to simultaneously meet a specific need to implement functionality while trying out specific technologies and an approach to SOA. This is the level at which the most basic of SOA standards from W3C are introduced such as XML for definition of message formats, WSDL for service interface definition and SOAP for invocation.

An 'initial services' SOA architecture typically contains the following components:

An Enterprise Service Bus (ESB), which provides a standard interaction model for SOA components including Web services and relational databases as a scalable, easy-to-deploy distributed infrastructure. The ESB provides a large number of adapters to allow services implemented in disparate technologies to interchange messages allowing, for example, a .NET application to communicate with a J2EE application at a services level.

A *Service Level Management Service* which provides visibility into Web services performance and service level metrics. The Service Level Management Service can provide such monitoring for events which are directly tied to the SOA infrastructure. In other scenarios, the events generated by the Service Level Management Service could be passed on to more generic services for processing and display.

A Services Registry supporting the UDDI standard, provides a central store of service definitions across the initial projects and provides a single point of references for service developers to obtain services definitions.



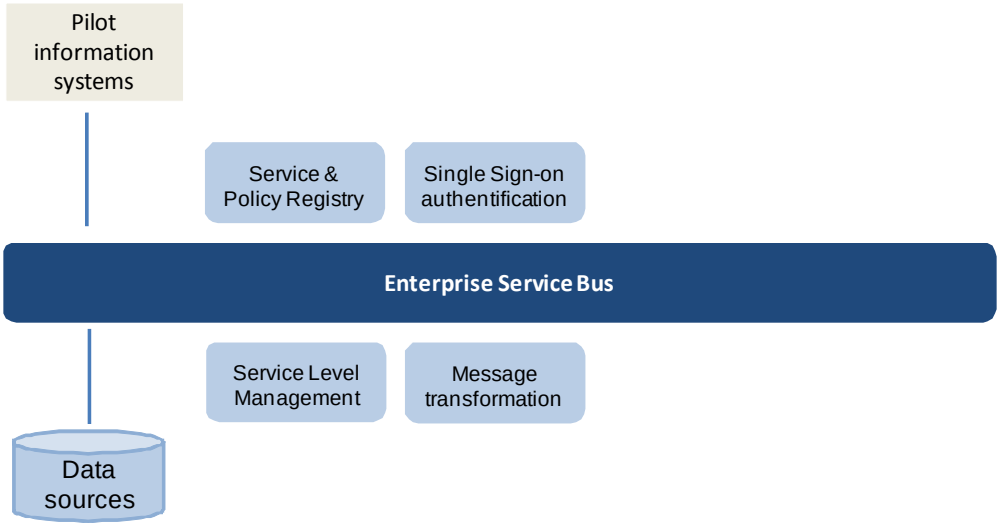
2: Architected Services

Based on the learning and feedback from the initial services scope, standard implementation technologies and platforms are defined. An 'Architected services' SOA architecture typically contains the following the following components:

A *Services and Policies Repository* extends the Services Registry to provide a repository for full storage and support of SOA governance information including policies and service definitions with lifecycle management including notifications and approvals. Usage of such a repository with both development and runtime support is key to the processes supporting Architected Services.

Message Transformation to allow the integration of services with differences in expected message contents or formats. This is often done by the invocation of XLST transforms applied to an XML message—in this example as a "mediation" function under the control of the ESB.

A *Single Sign-On Service* supporting user authentication and authorization across the organization. Such a service, could be based on the OASIS SAML standard for the exchange of authentication and authorization information.

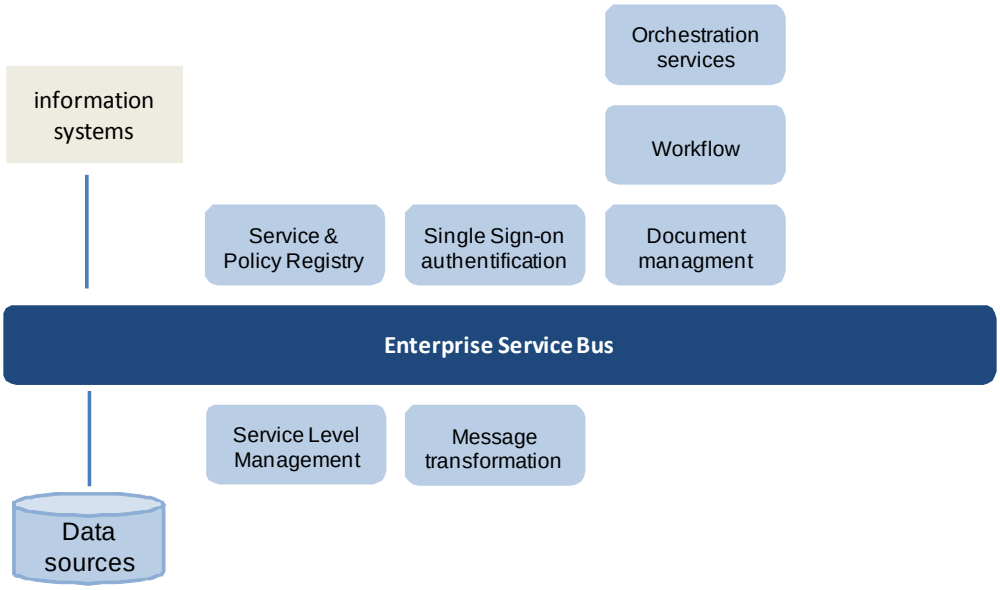


3: Business services and collaborative services

The focus of this level is on the partnership between technology and business organizations in order to assure that the use of SOA provides clear business responsiveness. Core to the value of SOA is the linkage between business process and the technical implementation of the processes.

Key to this Business Services implementation is Business Process Management (BPM) involving the management of long-running processes involving sequential messages between services. This could be done with an *Orchestration Server* which can manage the state of each process along with intermediate results.

As part of collaboration services we find workflow and document management services

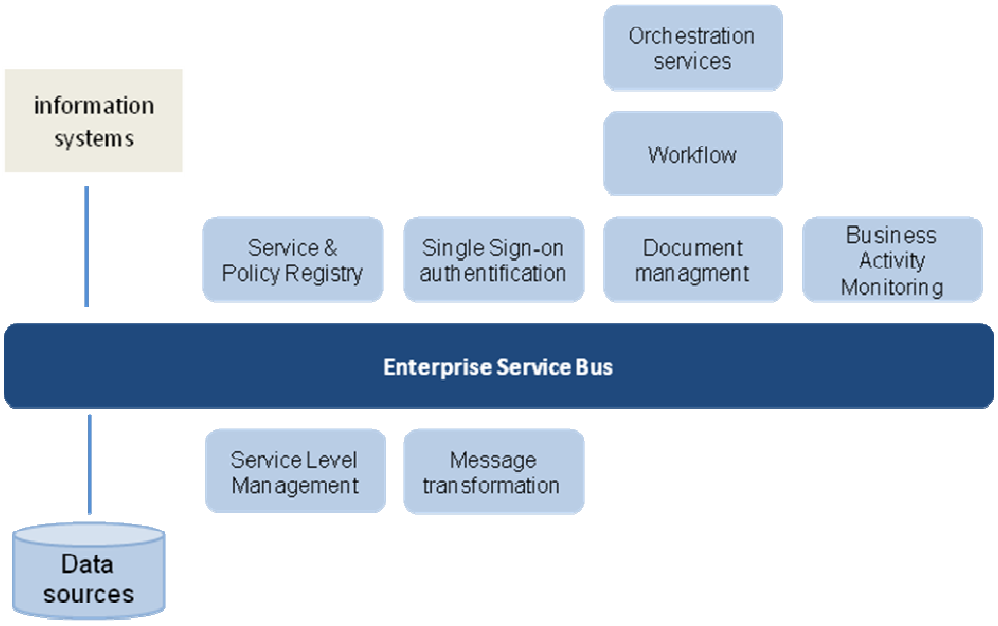


4: Measured business services

The scope of this architecture focuses on measuring and presenting the business processes at the business level so as to provide continuous feedback on the performance and business impact of the processes implemented

A 'Measured business services' SOA architecture typically contains the components:

Business Activity Monitoring component, which provides feedback to management as to real-time business performance metrics, shown here as a dashboard.



5: Optimized business Services

Optimized Business Services add automatic response to the measurements and displays of Level 4. In this way, the SOA information system becomes the "enterprise nervous system" and takes action according to events occurring at the business level according to rules *optimizing* business goals. A Business rules engine.

6.4 SOA SERVICE SCOPE APPLIED TO INTEROPERABILITY INFRASTRUCTURE SERVICES

The previous explained the relationship between the maturity of an organization and the type of service which typically can be found in a SOA architecture. This principle of inherent relationship between the type of services and the maturity of a SOA architecture can also be applied to the European Interoperability Infrastructure services. In other words we have reasons to assume that **there is a different need in the type of infrastructure services with the growth of usage of the interoperability infrastructure services**.

In order to simplify, three implementation waves of interoperability services have been defined:

Implementation wave 1:

- Data Transport services (ESB)
- Service Registry
- Identity & access management
- Data certification
- Audit trail and log (provided as a service embedded in the ESB)

Implementation wave 2

- Document Storage
- Structured data storage
- Data translation

Implementation wave 3:

- Workflow management

Member States which at this point show no interest in services of implementation wave 1 might become very interested in this after the completion of the implementation wave1. Or from the point of view of the European Commission it is not wise to invest time and money in the services of implementation wave 2 if there is not a minimum adoption of service of implementation wave 1.

7 Conclusion

Having reached the end of the third and last phase of the EIIS Study I, it is time now to draw some conclusions.

During this study, more than 80 information systems managed by the European Commission and the Member have been considered and analyzed. A number of these systems were selected as potential candidate for reuse in order to provide the nine European Interoperability Infrastructure Services (EIIS) defined by the study. Following the selection of the best candidates for reuse, one or two implementation scenario(s) have been proposed for each EIIS.

One of the main findings of the study indicates that the implementation scenarios with federated or distributed architectural topology were preferred to the centralized ones. The scenarios proposed for Service Registry, Identity and Access Management, Data Certification, Document Storage and Structured Data Storage are examples of this preference. These scenarios align with the European principle of subsidiarity, which defines that the central authority should have a subsidiary function, performing only those tasks which cannot be performed effectively at local level.

From a practical perspective, it was also noticed that some of the EIIS cannot be provided as reusable services. However, the Member States and the European Commission are also interested in guidance and reusable components in that field. It is the case for Audit Trail and Log and Data translation.

Another important finding is that the reuse scenarios of EIIS seem complex to implement, because reuse of services itself is not easy. Reuse in information systems has always been extremely difficult, and many technologies and concepts have promised to increase the reuse: Object-Oriented, Platform independency of Java, software patterns are just some examples. All of these had advantages, but faced many practical obstacles.

To overcome these issues, the EIIS Study introduced the Service Oriented Architecture (SOA) paradigm, that appears as the more mature approach to reuse and will probably initiate more reuse. The SOA approach is also an opportunity to have a more structured approach to setup the services which support the information exchange between the Commission and the Member States. When the SOA experience and maturity increases, the type of services also changes. This principle of inherent relationship between the type of services and the maturity of a SOA architecture can be applied to the European Interoperability Infrastructure services.

The SOA approach is also an opportunity to have a more structured and phased approach to setup the services which support the information exchange between the Commission and the Member States. When the SOA experience and maturity increases, the type of services also changes. This principle of inherent relationship between the type of services and the maturity of a SOA architecture can be applied to the European Interoperability Infrastructure Services.

A recommendation of the Phase 3 of the EIIS study is thus to move to a more Service Oriented Architecture in a stepwise approach, that implements progressively the EIIS, according to the level of SOA maturity. Three implementation waves of the services have been defined in that perspective.

Based on these findings and recommendations, the next step is to elaborate interoperability architecture guidelines, and agree on these guidelines. A European Reference Interoperability Architecture could be envisaged as a possible option. It would provide a framework to implement progressively the EIIS to support the information exchange between the European Commission and the Member States.

This architectural work must be performed before conducting feasibility studies for the implementation scenarios. It has to be done in close cooperation with the EIIS stakeholders, in order to ensure that the future EIIS respond to their needs and effectively improve the reuse and interoperability between the Member States and the European Commission.

8 Annexes

8.1 PHASE 1: SUMMARY OF THE MAPPING BETWEEN THE SYSTEMS AND THE SERVICE LIST

The legend below explains how the mapping should be read:

Y	YES : Means that the service has been identified in the system
E	EXTERNAL: Means that the service is identified, but is provided by a component external to the system
	[blank]: Means that the service has not been identified in the system.

System Owner	System Name	Audit trail and log	Service registry	Access control	Data certification	Data transport	Data translation	Workflow management	Document storage	Structured data storage
DG AGRI	CATS	Y		Y		Y				Y
DG AGRI	Isamm & Agrides	Y		E	Y	Y	Y	Y	Y	Y
DG COMP	ECN			E	E	Y		Y	E	Y
DG COMP	eQuestionnaire	Y		Y		Y		Y	Y	Y
DG COMP	SANI			E	Y	Y	Y	Y	E	Y
DG EMPL	EESSI (formerly PROTECTUS)	Y		E	E	Y		Y		
DG ENV	CECIS			Y		Y	Y	Y		Y
DG ENV	CITL	Y		E		Y	Y			Y
DG ENV	SEIS					Y				Y
DG ESTAT	CVD			Y		Y		Y		Y
DG ESTAT	X-DIS			Y		Y	Y			Y
DG MARE	FIDES (version 3)	Y		Y		Y	Y	Y		Y
DG MARKT	IMI	Y		Y		Y	Y	Y		Y
DG REGIO	SFC2007			Y	Y	Y		E	E	Y
DG TAXUD	CCN / CSI	Y	Y	Y	Y	Y			Y	
DG TREN	TACHOnet	Y		Y	E	Y	Y	Y		Y
DIGIT	CIRCABC			Y	Y	Y	Y	Y	Y	
DIGIT	ECAS	Y		Y						
DIGIT	eFP7	Y		Y,E	Y			Y	Y	Y
DIGIT	eID (Stork)			Y	Y					

System Owner	System Name	Audit trail and log	Service registry	Access control	Data certification	Data transport	Data translation	Workflow management	Document storage	Structured data storage
DIGIT	e-PRIOR	Y		Y,E	Y	Y	Y	Y	Y	
DIGIT	ESSI			E	Y					
DIGIT	Hermes	Y		E		Y		Y	Y	
DIGIT	IDABC Certification Services (PKI, Time-stamping, …)			Y	Y					
DIGIT	IPCIS		Y				Y	Y		
DIGIT	Notis			E				Y		
Secretariat-General	Argus			Y				Y		
Secretariat-General	eGreffe				Y	Y				
MS Austria	Directory Service		Y							
MS Austria	ELAK							Y	Y	
MS Austria	Electronic Delivery Service				Y			Y		
MS Austria	Modules for Online Applications (MOA)			Y	Y	Y				
MS Austria	Portal Group			Y		Y				
MS Belgium	Federal Service Bus	Y	Y			Y	Y			Y
MS Consortium	EULIS	Y		Y		Y	Y			Y
MS Consortium	PEPPOL		Y	Y	Y				Y	
MS Denmark	eINVOICING		Y		Y	Y	Y		Y	
MS Denmark	NemHandel		Y		Y				Y	
MS Estonia	X-Road	Y	Y	Y	Y	Y	Y			
MS Germany	De-Mail			Y	Y				Y	
MS Germany	DVDV		Y		E	Y				
MS Germany	RISER		Y			Y	Y			
MS Germany	Virtual Post Office			Y	Y	Y				
MS Germany	Xgenerator						Y			
MS Netherlands	EUCARIS II	Y	Y	E	Y	Y	Y			

System Owner	System Name	Audit trail and log	Service registry	Access control	Data certification	Data transport	Data translation	Workflow management	Document storage	Structured data storage
MS Slovenia	eVEM (One Stop Shop)		Y	E		Y		Y	Y	
MS Slovenia	State Portal of the Republic of Slovenia (e-Uprava, e-SJU)			Y		Y				
MS Slovenia	The Slovenian workflow engine, CTDS, ESB and register of services		Y			Y	Y	Y		Y
MS Spain	DVS	Y		E	E	Y				
MS Spain	Validation Authority Service @Firma	Y			Y	Y				

8.2 PHASE 2: OVERVIEW SELECTED INFORMATION SYSTEMS

The table below gives an overview of the information systems and the mapping with the interoperability infrastructure services they deliver.

systems	Audit trail and log	Service Registry	Identity and Access Mgt	Data Certification	Data Transport	Data Translation	Workflow	Document Storage	Structured Data Storage
@Firma VA	Candidate			Candidate					
Argus							Other		
Austria MOA			Candidate	Candidate	Other				
CCN / CSI					Candidate				
CIRCABC						Other		Candidate	
CVD									Candidate
De-Mail				Other				Other	
DVDV		Candidate							
DVS	Candidate								
ECAS			Candidate	Other					
EDMA								Other	
EESSI					Candidate				
eFP7			Other	Other			Other	Other	Candidate
eID (Stork)			Other						
ePRIOR	Other					Candidate	Other	Other	
ESSI				Candidate					
EUCARIS II	Other			Other	Candidate	Other			
eVEM (One Stop Shop)							Other	Candidate	
Federal Service Bus	Candidate	Other			Candidate				
FIDES (version 3)							Candidate		
FUSEFRAME	Other					Other			
Hermes	Other							Candidate	
IDABC PKI				Other					
IMI			Other			Candidate			
IPCIS					Candidate		Candidate		
ISAMM					Other	Other			
jCore	Candidate		Candidate		Other			Candidate	
NemHandel		Candidate		Other	Other				
Notis							Candidate		
SFC2007			Other	Other			Other		
Slovenia eVEM/eUprava			Other						
Statel (Edamis)					Other				
TACHOnet	Other		Other		Other	Other	Other		
Virtual Post Office	Other		Other	Candidate	Other				
X-DIS					Other	Other			Candidate
X-Road	Other	Other	Other	Other	Candidate				

8.3 SERVICE DESCRIPTION: AUDIT TRAIL AND LOG

8.3.1 DEFINITION

The audit trail and log service chronologically records information about the usage of European Public Services. It collects data to examine how and when events occurred, who accessed a system and what actions he or she performed during a given period of time.

The logged information can be the exchanged information between the system and the users of the system (incoming and outgoing messages), the log-on data, the transaction content and properties-time, checks and other actions performed by the users as well as actions performed by system administrators, or automated actions initiated by the system.

Audit trail and log records data generated by system processes and which do not correspond to specific user actions, and actions taken by identifiable and authenticated users.

8.3.2 SERVICE FEATURES

8.3.2.1 Security Checks

'Security checks' are part of the regular examination of the actions performed by the users, in order to assess the effectiveness of security mechanisms. Review of audit data allows to detect system misuses (security violation), anomalies (actions not expected) or intrusions. Security checks can highlight:

- Repeated attempts by users of the system to bypass security mechanisms
- Users assuming privileges greater than their normal ones
- Installation of unauthorized, potentially damaging software
- Modification or deletion of sensitive information
- Etc.

Security checks also verify the compliance with the organization's security policies.

Security checks can be used reactively to assess the impact of a security violation, or to proactively discover potential security holes.

8.3.2.2 Usage statistics

The audit trail data provides the necessary information to develop usage statistics. Usage statistics track the events, analyse and report information. Usage statistics can be used for various purposes, such as performance analysis of the system and usage pattern monitoring. Queries on usage statistics can provide information on e.g.:

- Volume of information exchanged
- Number of people accessing the system
- Types of actions performed by the users
- Types of users having accessed the system
- Etc.

8.3.2.3 Error logging

'Error logging' records errors, warnings and other significant events. Errors can be filtered and categorized in order to indicate their significance, and to decide on appropriate actions and facilitate the resolution of issues.

8.3.2.4 Data lineage

The lineage of the data is very important for the European administrations that deal with sensitive data. They must be able to determine where the data come from, where they flow, and how they are transformed.

The data lineage is the ability to trace origin and ownership of the data. It provides information on the source and destination of data, as well as the transformations applied to the data. It allows the reconstruction of a previous activity. It can also serve to recover lost transactions.

8.3.3 BUSINESS VALUE

Audit trail and log primarily serves security purposes. The objective is to monitor the actions done by the users. The audit trail shows whether business rules are being followed and ensures that unauthorised activity can be identified and traced. It holds users of the system accountable for their actions.

The analysis of the audit data can help to provide high quality, accurate, and complete data that can be used by business users, for reporting purpose, to e.g. change or improve an existing process. Audit trails need to be available and convertible to a human readable form.

The electronic document management is a concrete example that shows the use of the audit trail and log service. It allows to track all the major stages of the document (e.g. identification, receipt, creation, modification, addition, digitisation, filing, destruction, signing, transmission, transfer and archiving)[61].

As a general rule, it is important to select the audit information that is necessary to avoid that the volume of audit trail information becomes too large and causes performance issues.

8.3.4 BACKGROUND

The European Commission and the Member States are dealing with extremely sensitive data. In the context of interoperability, an Administration must be able to exchange information with other Administrations (whether a Member State or the Commission), as well as Businesses and Citizens of the European Union, in a secure and controlled manner.

The large amount of data, systems, users (administrations, businesses and citizens) makes the security of the data a critical point. The exchange of messages, email and documents between the systems, which in some cases actually results in direct access to authentic data sources, is intended to meet specific security requirements. As described in the draft EIF v2.0 the audit trail and log service is part of these security requirements. It allows to maintain a legal audit trail of the exchanged data for evidentiary purposes.

[61] HAN Vision Document – Adonis D(2006)2276

8.4 SERVICE DESCRIPTION: SERVICE REGISTRY

8.4.1 DEFINITION

A service registry is a central registry which provides a description of available services. The registry presents, for each service:

- how to use them
- their current status
- their physical locations

A service registry maintains the catalogue of available services in a service oriented context. Service producers publish services and register them into the registry thus consumers are able to find them.

An enterprise may have one or more service registries which can be merged to one enterprise service registry. This is called a federated service registry.

8.4.2 SERVICE FEATURES

8.4.2.1 Service Catalogue

The service catalogue provides a list of the available services and should provide the following capabilities:

- Replication
- UDDI support
- Business services dashboard for web-based configuration
- Service browsing to maximize reuse and identify service dependencies (e.g. specific parameters of the service that need a prior call to another service)
- QoS management dashboard (using service contract information)

The service catalogue content is composed of multiple entries corresponding to each available service. Each entry is at least described by:

- A unique service name
- A service description
- Service attributes such as inputs, outputs, usage guidelines…
- A service location(s) that provide(s) the technical address(es) of centralized and decentralized services.
- Service contract information
- Service security (security level of the service)

8.4.2.2 Service versioning

Service versioning allows operating different versions of a service. It avoids updates of system configurations of service consumers each time a service changes. Versioning services include change and approval management, change notification, service versioning maintenance.

8.4.3 BUSINESS VALUE

The direct business value of a service registry is difficult to demonstrate. A service registry should be seen as a tool encouraging the reuse of existing services. This implicitly creates business value by:

– enhancing the development speed,

– facilitating interoperability

– and decreasing the maintenance cost.

The development speed is linked to the reuse of already existing services. The information stored in a service registry should be structured to enable the discovery of reusable services for technical and for business-oriented users. For business-oriented users, it should contain information such as service business logic and usage guidelines.... For technical-oriented users, the registry should contain information on service' dependencies and input and output parameters.

A service registry offers service browsing that permits the identification of service dependencies. A meta-data description will greatly improve the interoperability. Clear documentation of existing services, its input and output parameters, the standards used will improve the possibilities of exchanging information between systems.

A service registry provides enough information on reuse potential and can also offer service contract information which can include usage cost of that service. The design of new applications can be based on objective cost information available in the service registry. This can help to make a cost effective new application because the business can compare the cost of creating a new service (and running it) to the cost of integrating an existing service (and running it).

8.4.4 BACKGROUND

Member States and European Commission systems – especially when they are put in production or if they are still in development - are struggling with common issues when they build large scale systems.

Member States are currently not aware of services and components available outside. They suffer from this lack of visibility on available components. Some Information Systems have solutions that other system builders are looking for. Although it is also part of this study to discover such services, on the long run, up-to-date list of such services should be provided by a service registry.

A common topic is that more and more systems have to be interoperable. It is not a wish coming from nowhere. Quite often, it is a direct need coming from the regulation at the European Commission level. Interoperability is often a painful process if systems are already in place. Having a common registry where Information System specifications can be seen by everyone is a key feature of a service registry.

Finally, smaller Member States tends to look for reusable solutions as much as possible. The main reason is the size of their development team which is not big enough to design, build and maintain a solution from scratch. Instead, they only have integration capabilities. A service registry for various business and infrastructure services would be a key partner for them.

For those reasons, a service registry is convenient in the context of European Interoperability Infrastructure Services.

8.5 SERVICE DESCRIPTION: IDENTITY AND ACCESS MANAGEMENT

8.5.1 DEFINITION

A definition set forward by the Identity & Access Management Working Group[62] in 2005 is as follows:

> *Identity and Access Management (I&AM) is the set of business processes and supporting infrastructure which verify that users are who they claim to be (authentication) and provide them with access to the right resources (authorisation). In order to authenticate users, there is a need to create, maintain and use digital identities. Likewise, to authorise users, there is a need to create, maintain and enforce permissions for accessing resources.*

In other words, Identity and access management (IAM) encapsulates all the processes, policies, and technology solutions that manage digital identities and specifies how digital identities are used to access resources[63].

This infrastructure service includes

- Entity authentication: the mechanism needed to manage controlled access of entities (a user and, by extension, a legal entity, an ICT service or an other) to applications, including single sign-on capabilities enabling the user to avoid further authentication and simplify user navigation.
- Authorization: the mechanism to define what access privileges an entity has within the application by defining roles and groups.

Data authentication, which verifies origin and integrity of data, is not part of this "identity and access management" infrastructure service. Data authentication will be treated in the "data certification" infrastructure service.

8.5.2 SERVICE FEATURES

8.5.2.1 Entity authentication

As defined in the Common Terminological Framework for Interoperable Electronic Identity Management, entity authentication is the corroboration of the claimed identity of an entity and a set of its observed attributes. In other words, authentication checks the identity of the entity by verifying whether or not the credentials are valid.

Entities can be identified based on:

- knowledge, what you know (e.g. username password),
- possession, what you have (e.g. a smart card, a token, an identity card, etc.),
- a personal characteristic, what you are (biometrics, e.g. fingerprints),
- location, where you are (e.g. network address or phone number),
- or by a combination of these. A typical example of a two-factor authentication mechanism consists of the combination of password and fingerprint authentication.

Entity authentication can be unilateral or mutual.

- Unilateral authentication provides assurance of the identity of only one entity.

[62] Identity and Access Management - Scope and Requirements, Version 1.0, DIGIT, April 2005.

[63] Forrester TechRadar™: Identity And Access Management, Q2 2008, June 18, 2008.

- Mutual authentication provides assurance of the identities of both entities.
- Entity authentication can be by individual or by group. Group authentication allows establishing an understood level of confidence that an individual possesses a specific attribute that does not provide ties to the user's identity. The IDA Authentication Policy posed that when sectored networks are implemented, the information is generally exchanged between "functional areas" and not by named individuals.

A choice has to be made between the different authentication mechanisms, depending on the desired authentication level. The stronger the authentication is set up, the higher the confidence that an entity is who/what it claims to be.

The authentication of an entity can be broken down in 2 main phases:

- the registration phase, which is the process of

1. applying for credentials via a Registration Authority and Credentials Service Provider;
2. verifying the claimant, the entity that applied for credentials;
3. registering and delivering the credentials.

- The authentication phase itself, which is the process of proofing possession of the correct identity credentials.

The registration phase is out of scope of the EIIS study. The authentication phase is the key feature of the identity and access management service.

8.5.2.2 Authorization

Authentication simply establishes identity. Authorization is on the other hand about what this identified entity is allowed to do or what access privileges it has. As defined in the Common Terminological Framework for Interoperable Electronic Identity Management, authorization refers to

- the permission of an authenticated entity to perform a defined action or to use a defined service/resource;
- the process of determining, by evaluation of applicable permissions, whether an authenticated entity is allowed to have access to a particular resource.

Authorization cannot be seen as independent from authentication. Permission to perform activities is granted or denied based on the result of authentication. Once an entity is authenticated, it may be authorized to perform different activities, each of which is referred to as a role.

With role based access management, permissions are not assigned directly to an entity. Permissions are acquired through their role(s), making the access right maintenance (i.e. adding, changing, deleting roles and assigning individual entities to roles) less complex. Role based permissions are often variations or extensions of three basic types of access: read, write and execute.

In the best case, authorization should be auditable. For instance, in case of an identity audit or security breach, it should be possible to map out who has access to what applications.

8.5.2.3 Related concepts

This paragraph will touch upon a number of concepts related to authentication and authorization: single sign-on, identity federation and electronic credentials.

Single-sign on

According to Forrester, Web Single Sign-On (W-SSO) is the most highly adopted and implemented technology amongst business users. W-SSO allows a user to log in to a Web

application and then move to another application without being prompted again for authentication. ECAS is an example solution that provides W-SSO within the European Commission.

Identity Federation

Identity Federation is a dominant movement in identity management today. Identity Federation provides a means for business partner to agree on and establish a common, shared name identifier to refer to an entity in order to share information about the entity across organizational boundaries. This type of sharing can help reduce identity management costs as multiple services do not need to independently collect and maintain identity-related data (e.g. passwords, identity attributes)[64].

Electronic credentials

Electronic credentials within the European Commission are often based on X.509. X.509 is a standard for a public key infrastructure (PKI) that specifies, amongst other things, standard formats for:

- digital certificates that are e.g. used to electronically sign documents, binding together a public key with an identity. The certificate is issued by a certificate authority (CA);
- certificate revocation list, which is a list of certificates that have been revoked or are no longer valid.

Digital signatures can be used to:

- authenticate the source of messages,
- ensure the integrity of the message, to be sure that it has not been altered during transmission. The latter is a subject of the "data certification" infrastructure service.

8.5.3 BUSINESS VALUE

With electronic exchanges of information taking place across Europe every day, security is becoming an increasingly important issue. Security is tackled by ensuring that exchange takes place over a secure channel, and also by guaranteeing that the identities of the persons exchanging information are known and authenticated.

The benefits of a common framework and solution for identity and access management would be:

- less redundant efforts during implementation and operation,
- more confidence between parties that exchange information,
- more efficient and aligned procedures for maintaining access rights,
- etc.

8.5.4 BACKGROUND

Notwithstanding the fact that there is at the moment of writing no directive in place for eIdentification, the European Commission actively supports both politically and financially activities that aim at finding solutions for interoperable eIdentification at EU level.

A large number of initiatives have therefore been initiated in the context of e-Identification (Identity and Access Management):

- IDA(BC) has developed an authentication policy in 2004. The authentication policy is a series of recommendations and guiding principles for the establishment of appropriate

[64] Source: eID Interoperability for PEGS.

authentication mechanisms for the Member State administrations and EU institutions in IDA(BC) sectoral networks. This policy has been a reference document for further studies.

- The project "eID Interoperability for PEGS", delivered in 2009, aimed at proposing a solution to the legal, technical and organisational issues related to the creation of an interoperable pan-European identity management infrastructure. The main challenge for the eID Interoperability for PEGS project was to propose a general interoperability architecture for authentication solutions used or planned in e-government applications without affecting Member States' own existing infrastructures.
- STORK (www.eid-stork.eu) is a large scale pilot in the ICT-PSP (ICT Policy Support Programme), under the CIP (Competitiveness and Innovation Programme), and co-funded by the European Union. It aims at implementing an EU wide interoperable system for recognition of eID and authentication that will enable businesses, citizens and government employees to use their national electronic identity in any Member State.
- On 28 November 2008 the European Commission adopted the "Action plan on eSignatures and eIdentification. The action plan aims at assisting Member States in implementing mutually recognised and interoperable electronic signatures and eIdentification solutions, in order to facilitate the provision of cross-border public services in an electronic environment.
- The IDABC European Federated Validation Service project aims at constituting a platform that would provide trust in cross-border transactions involving the usage of eSignatures. Cross-border recognition of nationally issued digital signatures for security of data exchange requires interoperability at legal, operational and technical levels. The project will provide the necessary tools for the establishment of trust between different issuers of certificates and for the technical validation of eSignatures.

8.6 SERVICE DESCRIPTION: DATA CERTIFICATION

8.6.1 DEFINITION

Data certification can be defined as:

- the process of signing an electronic information (which could also be an e-mail, a file or a data source);
- the process of verifying if the origin and integrity of information are what they are expected to be (data authentication);

based on certificates issued by different Certification Authorities (CA).

This infrastructure service includes the creation, validation, and extension of advanced electronic signatures as *front-end services* in conformity with the requirements of the EC Directive. Validation of certificates and time stamping are *back-end* services to provide these front-end services, and may optionally offer also a direct client interface.

This infrastructure service does **not** include

- a certificate management system and certification authority in charge of issuing and managing the signature certificates and provisioning the signature creation devices (e.g. smartcards);
- the process of verifying the 'correctness' of the data. It is possible that the origin and integrity of the data are approved, but that the data themselves are factually wrong.

8.6.2 SERVICE FEATURES

8.6.2.1 Creation of electronic signatures

A system with an electronic signature creation service properly implemented is capable of signing electronic information (which could be of any format: e-mail, document, flat file, XML file, etc.) digitally. An electronic signature can have the same legal value as a handwritten signature. Multiple signatures can be added to a message.

The "European Union Directive 1999/93/EC addresses three forms of electronic signatures:

- "Simple electronic signature" (SES). This one has a wide meaning. It serves to identify the signing person and to authenticate data. It can be as simple as signing an e-mail message with a person's name or using a PIN-code.
- "Advanced electronic signature" (AdES). This form of signature needs to be uniquely linked to the signatory and it needs to be linked to the data in such a way that any subsequent change in the data can be detected.
- "Qualified electronic signature" (QES) consists of an advanced electronic signature based on a qualified certificate (QC) and created by a secure signature-creation device. It offers the highest level of security ensuring that the data come from their purported sender and that the transmitted data have not been altered.

It depends on the context and the required level of trustworthiness of the data what type of signature is best suited. For example, documents meant for internal use within the Commission will not always require AdES or QES.

Electronic signatures come in different formats, of which the most common are:

- XML Signature or XML-DSig, which is a W3C recommendation that defines an XML syntax for digital signatures. It is geared towards signing XML documents. It is used by various Web technologies such as SOAP, SAML, and others.
- Cryptographic Message Syntax (CMS) which is used to sign and/or encrypt messages under a PKI. It is based on the syntax of PKCS#7.
- CAdES (CMS Advanced Electronic Signatures) is a set of extensions to CMS signed data making them suitable for advanced electronic signature.
- XAdES (XML Advanced Electronic Signature), which specifies precise profiles of XML-DSig for use with qualified electronic signature in the meaning of European Union Directive 1999/93/EC.

8.6.2.2 Validation of electronic signatures

A system with an electronic signature validation service can provide this service to:

- external parties of the European Commission for any electronic signature on electronic documents generated by the European Commission, e.g. Member States, suppliers;
- internal staff of the Commission for any electronic signature either received by the Commission or generated within the Commission.

The validation service for signatures will return the status of all signatures on the document, and validity of the certificates used. An overall assertion covering all signatures is returned as well as individual assertions for each signature and certificate (e.g. valid, revoked, unknown). A signature is verified by the signer's certificate. Signature and certificate policies will determine most aspects of a signature quality, including use of smart card or other hardware token and legal aspects.

8.6.2.3 Signature extension

The state of an advanced electronic signature can evolve in time. One should make sure that electronically signed documents can remain valid for long periods, even if underlying cryptographic algorithms are broken. Technically speaking, this will happen through an **extension mechanism** that will incorporate additional information as part of the signature's unsigned properties.

Multiple extended AdES signature formats are defined for use with qualified electronic signature in the meaning of European Union Directive 1999/93/EC:

- AdES-T, adding a time-stamp on the basic form generated by the signer in order to provide evidence on the existence of the signature before a trusted time;
- AdES-C, adding complete validation reference information to the AdES-T form;
- AdES-X, builds up from AdES-C by adding time-stamp on the certificates and validation data references;
- AdES-X-L, builds up from AdES-X and incorporates the complete certificate values and validation (revocation) data values; and ultimately
- AdES-A is the most extended form: built up from AdES-X-L adding an archive time-stamp, it offers long term validation of the signature and therefore long term archiving capability.

The extension can be applied on any AdES signature form by both the signer and/or the verifier relying on the central service provided by the signing platform that will take care of all the complexity involved in such process.

8.6.2.4 Validation of certificates

A system with a certificate validation service is capable of handling the complete validation of certificates on behalf of a client application.

When the signature platform receives a digital certificate (e.g. a certificate related to a signature) it has to decide whether or not this certificate can be trusted and decides whether or not it can be accepted in the current application context. For this, the certificate has to be validated against a set of validation rules, specific to that context, and called a "Validation Policy".

The validation of a certificate against a validation policy implies several steps. The first one, also called "path discovery" or "path building", consists in building a complete certificate chain, starting from the validated certificate and up to a trusted CA certificate. The list of trusted CA is a configurable parameter part of the validation policy and serves as input to the path building process. The resulting constructed certificate chain is called a "certification path". This path building process can also accept additional configuration parameters further refining the scope of accepted certificates. The flexibility of such configuration is up to the service implementation.

Once a valid certification path has been discovered, the second step of the validation process can be performed. This step consists in a full revocation status check of the complete certification path. This check is done against the various Certification Service Providers (CSP) revocation services available like the CRLs/Delta CRLs or OCSP servers.

8.6.2.5 Trusted time stamping

A trusted time stamping service provides trusted time stamps for keeping securely track of creation and modification time of documents. Not even the creator and/or owner of a document are capable of changing unnoticeably the document once it has been time stamped as it would break the (digital signature of the) time stamp.. A trusted time stamp is a time stamp that is issued by a trusted time stamping authority (TSA).

A time stamp is created by sending a hash of a given set of data to the TSA. The TSA adds a time mark to the hash and digitally signs the combination of hash and time mark. The latter is sent back to the requester together with the signature of the TSA. Any party that trusts the TSA signature

can then verify by replaying the time stamp signature that the document was not altered after the date indicated by the time stamp.

8.6.3 BUSINESS VALUE

The business value of exchanging electronically signed documents is amongst other:

- A decrease in the exchange of paper based documents;
- The facilitation of electronic document storage. The documents can be time stamped and securely signed.
- Electronic signatures provide non-repudiation, meaning that a person who signed cannot say he/she did not sign a message (with exception of theft of the private key);
- Electronic signatures (especially the AdES and QES) are more difficult to forge than the handwritten signatures and can have the same legal value as handwritten ones.

8.6.4 BACKGROUND

Electronic communication is becoming increasingly important in many aspects of economic and public life. Public authorities across Europe have started to offer electronic access to government services. Often there is a need to provide an electronic signature allowing the administration to identify the signatory as well as to make sure that the data submitted has not been altered during transmission.

In 1999, the e-Signatures Directive (EU Directive 1999/93/EC[65]) was adopted to promote the legal recognition of electronic signatures and to ensure the free circulation within the single market of e-signature products, equipment and services. However there are still interoperability problems that currently limit the cross-border use of e-signatures, because Public Authorities have been focusing mostly on national needs and means, which has led to a complex system with different solutions .

Therefore, the Commission proposed in its Communication “A single market for the 21st century Europe” of 20 November 2007 to adopt an “Action Plan on e-signatures and e-authentication[66]”, which was published in November 2008. This Action Plan aims to assist Member States in implementing mutually recognised and interoperable electronic signatures and e-identification solutions, in order to facilitate the provision of cross-border public services in an electronic environment.

8.7 SERVICE DESCRIPTION: DATA TRANSPORT

8.7.1 DEFINITION

The goal of this service is to exchange data in a reliable way by providing standardized transport capabilities.

This service facilitates seamless communication between systems for collecting and delivering data. It does not store the data centrally. Each system independently handles its own data and, when required, draws data from the database and sends it to another system.

[65] http://europa.eu/eur-lex/pri/en/oj/dat/2000/l_013/l_01320000119en00120020.pdf

[66] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2008:0798:FIN:EN:PDF

8.7.2 SERVICE FEATURES

8.7.2.1 Set of standardized protocols and formats

This feature defines a set of protocols and formats adapted to business and performance requirements. It enables communications between systems which are compliant with the standardized set of protocols and formats defined within the Data Transport service.

8.7.2.2 Decentralized data exchange

It provides capabilities supporting a point-to-point architectural model suitable for interoperability. These capabilities include information on the detailed information model thus enabling the other parties to understand the format the exchanged data whether this is static (agreed upfront between parties) or most frequently dynamic (provided by metadata). Although there is no central data storage for enabling the data exchanges, multiple architectural pattern of this service can exist:

- Point to point: it makes the endpoint smart enough to avoid centralized management of data exchange. It needs a referential interface or framework to be implemented at each endpoint. Each endpoint must know the address of the others. A change in the data exchange protocol often needs to upgrade all the endpoints at the same time.
- Hub and spoke: it manages data flow centrally to make the exchanges possible although it does not store data but caches them. It needs a referential interface or framework to be implemented at each endpoint. A central hub offers global data exchange logging capabilities and the ability to work easily with various endpoint referential interface versions.
- Enterprise service bus (ESB): it avoids the need of a referential interface at the endpoint. Therefore, the endpoint can keep a specific interface to communicate with the others – the ESB takes care of the protocol conversion. Nevertheless, the endpoints and the ESB have to be configured to speak together and this can be sometimes be cumbersome.

8.7.3 BUSINESS VALUE

A data transport service is a service which reduces the design and the maintenance of the exchange of data. This creates business value by:

- reducing implementation time of new business service;
- increasing the reliability of data exchange;
- and being more flexible on data format.

Setting up data exchange architecture between various administrations whether they are Member States or European Institutions can raise a lot of issues. Amongst those issues are managing different transport protocols, handling different formats or ensuring the architecture do not make centralization mandatory for information exchange by requiring a central indexing server.... The data transport service avoids creating a too complex architecture model which could end up in an unmanageable situation when new business service should be provisioned.

Using a common data transport service reduces interoperability issues such as data quality losses or data corruption. Interoperability issues are often caused by a lack of a reliable protocol that provides assurance about the delivery or the completeness of the exchanged data or by an inefficient synchronous protocol. The data transport service provides a trustable and efficient way of exchanging information over a network so that the business can rely on the quality of exchanged data.

Developing and maintaining customized data transport protocols consumes precious time of developer resources. Moreover, the data transport and content are often tightly coupled. Developers cannot focus on the structure and waste precious time on making the exchange

possible. By delegating this task to the standardized data transport service, they don't have to take care of all the technical issues about the transport the data or adapting the transport when they change the data structure. Therefore they can focus on the structure and the business can be more flexible on data structure changes (which is often seen by the business as the data content).

8.7.4 BACKGROUND

MoReq / MiReG defines the data transport service as a data exchange facilitator leaving aside the data content interoperability.

Back-offices from government administrations and European Institutions process complex local information. Those systems have often been designed to be tightly coupled internally, providing for very little or no interaction with external IT systems. Opening the data content of these systems is already a challenge in itself.

The need to share information with others entities (already open or not) is rising everyday because of European regulations or because a pragmatic cross-border collaboration is needed (resulting in bilateral agreements).

Developing and implementing a complete data exchange always requires a high effort of all parties involved. A data transport service will decrease the development time by providing a ready-to-use architecture for exchanging data.

8.8 SERVICE DESCRIPTION: DATA TRANSLATION

8.8.1 DEFINITION

To facilitate data transfers between systems (using their own data format, data model and data encoding), the data translation service includes semantic translation, syntax translation and multilingualism capabilities.

8.8.2 SERVICE FEATURES

8.8.2.1 Syntax translation

Syntax translation can be interpreted in two ways:

- The process of modifying the format of a source file into the format of a target file. Importing an incoming flat file (e.g. csv format) into a database or rendering an incoming XML file into a PDF document are two examples.
- The process of transforming the format of a data field. For example, date conversion (DDMMYY into DD/MM/YYYY), currency conversion (EUR into USD), etc.

With syntax translation, no semantic alteration of the source data occurs; only the 'format' changes.

8.8.2.2 Multilingualism

Multilingualism in the context of computing indicates that an application supports dynamically two or more languages when building or running the system. The multilingualism can be applicable at three levels:

- **Application level**: most commercial products support multiple languages for their navigation menus. The language is initially automatically chosen based on the language settings of the client PC but can be manually modified by the user.

- **Metadata level**: metadata in this context can be defined as descriptive data on the raw data ("data about data"). A good example is the name and description of a data field that can be displayed in multiple languages.
- **Data level**: this level is referring to the raw data themselves that are stored in database tables. If the data content is supported in multiple languages, this means that the translation of the content is physically stored. During dissemination, the data are presented in the language of the user making use of lookup tables.

The practical aspects of multilingual mapping of data include:

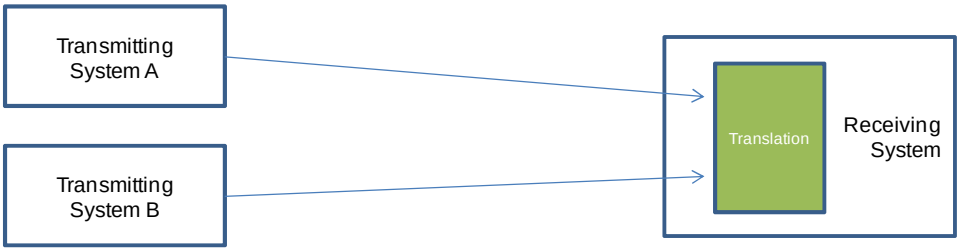
- structural changes implemented by techniques such as schema mappings;
- the translation of controlled vocabularies implemented by techniques such as code lists, multilingual thesauri, taxonomies, and/or ontologies .

Similar to syntax translation, multilingualism does not alter the source data.

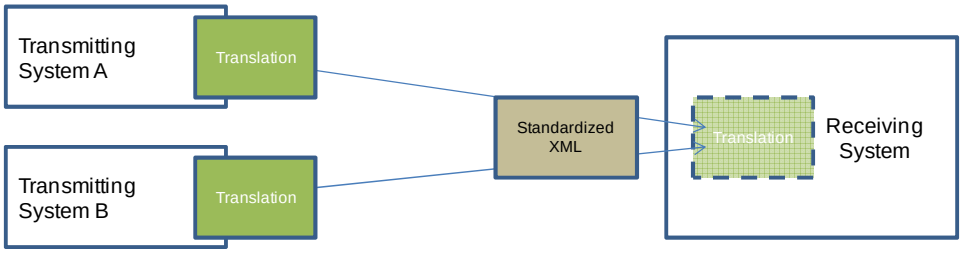
8.8.2.3 Semantic translation

Semantic translation is a technical service "translating" information in such a way that the receiving system interprets this information in the same sense as intended by the transmitting system. Two main scenarios are possible:

- Non-standardized data sets are delivered to the "receiving" system where they are altered to make them coherent for end-users.



- This scenario is typically encountered in a data dissemination environment such as, for instance, a data warehouse / business intelligence environment. It is however not the preferred scenario when e.g. 33 national countries need to deliver non-standardized data to a Directorate general.
- The data in the "transmitting" system(s) are translated to a predefined standard data set which has commonly been setup for the transmitting and receiving systems. This scenario is more applicable in a context of interoperability.



- This requires unanimous agreement of all partners involved in data definitions and data presentation. The European Commission and Member States are therefore striving for semantic interoperability, which refers to common understanding of data definitions, terms and models within a certain context. Syntactic harmonization (e.g. using open data standards such as XML, XBRL, SDMX) has proven to be an important pre-requisite for successful semantic interoperability between systems in European infrastructure.

Semantic translation, in contrast with syntax translation and multilingualism, does change the data content compared to the source data .Therefore it is recommended to keep an audit trail of the changes applied.

8.8.3 BUSINESS VALUE

A fictive example can illustrate the business value of having semantic translation capabilities: EUROSTAT collects statistical information from the National Institutes for Statistics (NIS). A measure collected is for example the "unemployment rate". It is realistic to say that amongst the NIS different interpretations of unemployment exist: should it include only full-time, also part-time, exclude sabbatical year?

Considering the examples of the previous paragraph, scenario 1 would mean that EUROSTAT should convert the different definitions that exist locally, so that the data are consistent with the definitions handled by EUROSTAT.

Scenario 2 would mean that each NIS should be aware of the common definitions handled by EUROSTAT and that they deliver their data according to these definitions.

8.8.4 BACKGROUND

Syntax

As explained in the "European Interoperability Architecture Guidelines" the recognized open data standard XML has become the foundation for sharing data in European networks. Applications for both information sharing and exchange are necessarily based on XML (related) standards. XML standards

- facilitate the setup of reliable data definitions,
- enable the exchange of data in a coherent way,
- reduce the amount of effort required for syntax translations.

Thus XML facilitates the communication between public administrations, citizens and businesses.

Multilingualism

In January 2009, the results of a "Study on Multilingualism[67]" have been published, which clearly state that it is not sufficient to develop international solutions and merely using English as the common language. Citizens should get access to European Union legislation in their own language. This detailed study on multilingual issues in data exchange recommends a pivot language approach as the most effective solution. The recommendation of this study should be taken into account when developing a reusable multilingualism service. For more information, please consult the study online.

Semantics

In June 2008, the European Commission, via the IDABC programme, launched the Semantic Interoperability Centre Europe "SEMIC.EU". SEMIC.EU supports meaningful data exchange for e-government projects. Its core feature is a repository that provides reusable interoperability assets, such as XML schema or taxonomies, for European e-government projects. SEMIC.EU promotes the reuse of syntactic (e.g. XML schemas) and semantic assets (e.g. ontologies) needed for semantic interoperability.

[67] http://www.semic.eu/semic/view/snav/library/SEMIC-EU-Publications.xhtml?cid=516968

8.9 SERVICE DESCRIPTION: WORKFLOW MANAGEMENT

8.9.1 DEFINITION

According to the Workflow Management Coalition[68], a workflow can be defined as the "automation of a business process – in the whole or partially – during which documents, information or tasks move from one participant to another according to a set of predefined rules". The participant can be a user, a work group, or a system.

Workflow management spans a number of different concepts: workflow definition, workflow implementation… Workflow management orchestrates interactions between workflow participants (human and systems) and provides each participant with the information that is necessary to complete his or her task.

8.9.2 SERVICE FEATURES

8.9.2.1 Core workflow engine

The core workflow engine runs workflow tasks step by step.

Once the workflow and the related steps are defined, the workflow can be instantiated the required number of times.

Generally, 2 types of workflow can be distinguished:

- Procedural workflow (also called production workflow or managing workflow), which corresponds to known business processes and uses predefined sequences to guide and control processes. The procedural workflow is more or less fixed.
- Ad hoc workflow where the participants are involved in the decision and determine the process sequence on the fly. The ad hoc workflow is dynamic.

The workflow engine can be managed with an administrative console by the workflow responsible to initiate the workflow or to verify the status of an instance in a workflow.

A workflow engine is able to initiate automatically other services (business or infrastructure) that are defined as steps. It can also manage temporary data collected during the workflow (user input or results of an external service call).

A typical workflow could include:

- Validation of data by a human (human action),
- Tasks to be accomplished by the participants of a process,
- Automated sending (service call) of the data to a set of e-mail addresses (built dynamically),
- Reminders, deadlines, delegation and other administration functionalities,
- Monitoring and documentation of process status,
- Incorporation of data processing tools (such as specific applications) and documents (such as office products).

[68] Workflow Management Coalition (WfMC), founded in 1993, is the acknowledged professional association, formed to define standards for the interoperability of workflow management systems.

8.9.2.2 Alerts and notifications

Alerts and notifications are used to notifiy the participant of the process, at a certain step of the workflow. It can be a message that a certain process step has been reached or that a certain error has occurred. The status can also be monitored proactively by the workflow owner. The system can eventually send reminders if no action has taken place during a predefined period.

8.9.2.3 Graphical representation

Graphical representation provides tools for designing and displaying process. It allows to analyze the workflow and evaluate its performance:

- The workflow flowchart enables to analyze the status of a running instance.
- The performance report allows to evaluate the average performance of all instances of a certain workflow and look for improvement of the process execution (KPI).

8.9.2.4 Business rules management

Business rules management aims at creating and maintaining the business rules of the workflow engine.

Each workflow has a business owner responsible for creating, documenting and maintaining the workflow (workflow owner). This owner defines and creates the steps and selects a responsible for each of them (step owner).

8.9.3 BUSINESS VALUE

Workflow management helps to improve efficiency and quality of work of the administrations. By automating the processes and establishing a procedure that is consistently followed, unnecessary steps are eliminated, and every participant is fully accountable in the process. Workflow management also makes easier to improve process performance.

Workflow management can support the improvement of the service by involving the administrations, businesses or citizen in the entire process. It increases the transparency to the European citizen and enables the traceability of the procedure, allowing to know how tasks are structured, what is the state of a file, who is in charge of the different tasks, etc. This is particularly helpful when a process includes steps in different administrations (multiple DGs, Member States).

8.9.4 BACKGROUND

In the context of European Public Services and interoperability, the improvement of the processes, the traceability of the procedures and the transparency to the citizens are important preoccupations at European and Member State levels. There is a need to share, reuse and exchange data within and between Member States administrations

In that perspective, the workflow is part of the General Public Services Conceptual Model (GPSCM) defined by the EIF v 2.0 (draft). It orchestrates the basic European Public Services, enabling a collaborative approach of the administrations.

8.10 SERVICE DESCRIPTION: DOCUMENT STORAGE

8.10.1 DEFINITION

Document storage is used to store and manage documents, providing features at each stage of the document life cycle: creation, retrieving, reviewing, versioning, distribution, publishing, archiving and eventual destruction.

This service facilitates collaboration between different contributors to the document life cycle.

8.10.2 SERVICE FEATURES

8.10.2.1 Archiving

Archiving ensures that the electronic documents and files are preserved in a reliable format for the entire time span for which the information is required for (short-, medium- and long-term). Long-term can be defined as "long enough to be concerned with the impacts of changing technologies, including support for new media and data formats, or with a changing user community. Long-term may extend indefinitely"[69].

Archiving ensures that the documents can be retrieved, accessed and read again in the future if necessary.

The way information is stored is important for ensuring its longevity**.** Filing, consisting of attaching a document or group of documents to a specified activity, is often used, following a filing plan. The filing plan organizes the files of the institution, according to a logical and hierarchical structure. The files can be easily retrieved and accessed and the traceability of the documents is improved. For the different file types, retention rules can be defined for duration, location and format conditions.

8.10.2.2 Versioning

Versioning allows tracking of changes between different versions of documents. The documents are checked in or out of the system, allowing users to:

- retrieve previous versions;
- review older versions of content;
- compare different versions of documents;
- remove selected older versions.

8.10.2.3 Metadata

Metadata provides data descriptions of the properties of a document (e.g. in the form of an XML metadata file). They give additional information on the content of a document. They are typically stored for each document and may, for example, include the date the document was stored, the identity of the user storing it, the document identifier and the format of the document.

Metadata can be added by the user. They may also be extracted from the document automatically.

Metadata can be used by a search engine to help users in locating documents by identifying probable keywords or providing full text search capability. Indexing metadata can be critical for rapid retrieval of the documents.

8.10.2.4 Distribution

Distribution ensures that the document is delivered to the right people and that the document is available as long as required by the regulations in place. Distribution of document is associated with the authorization to access the document: it must ensure that only authorized users can read the documents.

[69] Consultative Committee for Space Data Systems. (2002). Reference Model for an Open Archival Information System (OAIS). Washington, DC: CCSDS Secretariat, p. 1-1

Distribution is often linked to workflow. For the effective distribution of documents among the interested parties, the document storage may support collaborative workflows. It can be a manual workflow, which requires a user to view the document and decide who to send it to, or a rule-based workflow, which allows an administrator to create a rule that dictates the flow of the document through an organization.

Distribution can also concern publishing a document. In most cases, a workflow process is associated with the publication of a document. Typical steps of a publishing workflow are: proofreading, peer or public reviewing, authorizing, approving etc.

8.10.3 BUSINESS VALUE

The Member States and the European Commission handle – i.e. produce, receive and hold - millions of documents per year. All the activities and decisions in the political, legislative, technical, financial and administrative fields ultimately lead to the production of documents (official and non-official). In this context, document management can be quite complex and time-consuming. Document storage helps to organize the documents in a logical manner. It preserves the institution's memory, facilitates the exchange of information, provides proof of operations carried out and meets the department's legal obligations. A well-designed document storage service promotes finding and sharing information easily.

Document storage allows to improve the quality of work of the Commission departments and the Member States. It also increases transparency to the European citizen, enabling the European Commission or the Member States to provide, at any time, information on the matters for which they are accountable.

8.10.4 BACKGROUND

The long-term preservation of documents has been a constant preoccupation within the European Commission and the Member States.

In the context of interoperability, the long-term preservation of information is essential. The 2005 Manchester Ministerial Declaration explicitly recognised eID and eDoc (cross-border recognition and long-term archiving of electronic Documents) as "key-enablers" for the implementation and provision of European Public Services.

The draft EIF 2.0 reminds that "the long-term preservation of information held by administrations in electronic form is a horizontal concern which stretches beyond any particular set of applications or any particular sector, some aspects of which are regulated by EU law" [70].

In that perspective, the EIF recommends the use of standards for metadata, document formats, storage media and maintenance procedures that are adequate for long-term preservation purposes.

At European level, it is worth mentioning the MoReq initiative (Model Requirements for the electronic records management). MoReq was developed in the framework of the IDA programme. The MoReq specification, which first became available electronically in 2001, was published by the European Commission in early 2002. Since 2008 MoReq2 is available. MoReq2 proposes some recommendations for document management, when Document Management facilities are part of the Electronic Record Management System.

Within the European Commission, the e-Domec project aims at introducing an electronic archiving and document management system based on common rules and procedures for all Commission

[70] Directive 2006/24/Ec of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC.

departments. The e-Domec project officially started in January 2002. e-Domec is now part of the Procedures of the Commission by Commission Decision 2002/47 on document management. It allows any document connected with the Commission's official functions to be managed, stored and retrieved securely at all times.

8.11 SERVICE DESCRIPTION: STRUCTURED DATA STORAGE

8.11.1 DEFINITION

Today, many developers correlate the word “database” with Relational Database Management Systems (RDBMS). While RDBMS offerings provide deep functionality, for many use cases, they introduce more complexity (and more cost) than is necessary. For many interoperability projects and applications simply want to store, process, and query data, without many requirements about managing schemas, maintaining indexes, tuning performance or scaling access to their data.

The "structured data storage" service facilitates the exchange of data by providing a simple and structured interface to access data stored in large and complex databases. This service acts as an abstraction layer between the technical data structure of a database and the functional point of view of a standard user

The structured data service removes the need to maintain a schema, while your attributes are automatically indexed to provide fast real-time lookup and querying capabilities. This flexibility minimizes the performance tuning required as the demands for your data increase.

This approach reduces the specific technical knowledge to use the data base. Eg. It eliminates the administrative burden of data modelling, index maintenance, and performance tuning, or automatically indexing your data.

Authentication mechanisms are provided to ensure that data is kept secure from unauthorized access.

8.11.2 SERVICE FEATURES

8.11.2.1 Data model Definition

The data model describes the characteristics of the data in the data storage. Examples of these characteristics are business descriptions, technical formats of data objects, relationships between different data objects (data model), validation rules …

8.11.2.2 Store and Retrieve Data

A simple set of functions should be available to sore and query data. Also a batch upload functionality should be available. E.g a functionality which allows to upload data by XML files.

8.11.2.3 Historization

Historization means ensuring that all relevant changes to stored data are tracked and recorded and that proper time stamps are assigned to these changes. This feature ensures that the content and structure of the data at any given point in time can be restored.

8.11.3 BUSINESS VALUE

The value of the 'Structured Data Storage' service lies within the provisioning of access to complex technical database and making it understandable for users.

The metadata describes the functional meaning of stored data. The metadata are necessary to be able to exchange data between different users or organisations.

Privacy or security regulations require full traceability of access and changes to the structured data. A system which keeps track of access and changes is therefore better positioned to be part of a public interoperability infrastructure.

The Structured Data Storage service facilitates authentic sources provisioning.

8.11.4 BACKGROUND

Interoperability within the European Commission is closely linked to the availability of authentic data sources (E.g.: VAT, information about citizens...).

Metadata and data storage management have been the core of two documents published by IDABC:

- MoReq (Model Requirements for the Management of Electronic Records): these are the functional requirements for electronic records management;
- MIReG (Managing Information Resources for e-Government): it extends the Dublin Core metadata set to meet the additional and specific requirements of public administrations. MIReG is intended to define both the proposed metadata set and a framework for its effective use.

The Structured Data Storage service also intends to be suitable for the EIF Principle "Preserve Information over time". For some European Public Services for which the data sources are not owned by purely national authorities, such as sectoral services, or services offered by non-governmental agencies or the private sector, the question of preservation of the related data sources becomes a European question. The preservation includes the use of standards for metadata, document formats, storage media and maintenance procedures that are adequate for long-term preservation purposes. This also encompasses the transfer of digital records between (to and from) operational sources and archives.

Just as in a physical archive service, a user expects that any item that is checked-in is "well taken care of", so that upon consultation, the same item in the same condition is returned. The analogous function for electronic based preservation services could be termed *maintenance of the archived information*."