

# Interoperability in Cooperative Environments for Public Administration: Metadata and Intelligent Document Issues and Applications

Alberto Polzonetti, Francesco De Angelis, and Barbara Re

**Abstract**—Today, e-Government is an emergent multidisciplinary research field that has the aim to support the delivery of electronic information and services to citizens, businesses, and other stakeholders. This vision should be based on an effective cooperation between Public Administrations (PAs) that need to be more and more organized to deliver value-added e-Government services. Such cooperative environments should interconnect several PAs using interoperability architectures exploiting the service-oriented paradigm. In this paper we would highlight the role that metadata plays in applicative cooperation presenting some guidelines and an architecture able to assure interoperability between informative systems to favor the modernization of Public Administrations reducing time and costs for its functioning.

**Index Terms**—Ontology, Metadata, Model driven, Informatio Sharing, Interoperability, e-Government

## I. INTRODUCTION

ICT in the government sector plays a basic role for a better delivery of government services to citizens, business and organizations, and for a more efficient management of the governance. For these reason we addressed interoperability and cooperation among them as fundamental aspects to promote the governance, the service delivery, and to support decision-making inside the public sector.

Public Administrations promote innovation to automate their activities, to improve their cooperation and to provide faster and more efficient access to the services they provide [10].

Heterogeneous PAs and heterogeneous users foster the need to cross-organizational collaboration whose achievement is possible through interoperability at software level and a proper management of knowledge. In this paper, the importance of standards for information representation in the Public Administration setting were clarified. Metadata standardization is the best manner to define semantics in cooperative decentralized systems in Public Administrations. They allows a better accuracy in the management of semantic aspects of cooperation, reduction of interdependencies, semantic transparency, and a fair distribution of applicative logic allowing integration of large scale applications and Public Administration information systems.

We propose an ontology-based methodology and

infrastructure for the delivery of intelligent documents within Public Administrations. PAs can access a common repository to find ontologies describing a document and set up their information systems to manage the entire life cycle of the document: user interface creation, data storage and management of the iter of compilation. To do this, we deal with the concept of Intelligent Document, a document that is able to describe its structure and to specify the procedural iter for its filling. This can be seen as a result of a cooperation between PAs involved in the delivery of services to the citizens.

We presented also a modular system for e-Government services distribution that can be easily updated, extended and integrated with the existing information system of Public Administrations.

## II. STATE OF THE ART ON INTEROPERABILITY

Cooperation in Public Administration is a complex topic characterized by a lot of different characteristics. In this section we give an overview on interoperability and related studies at different levels.

### A. Government and non-governmental Initiatives

According to interoperability challenge for the Public Administration we are aware of different government initiatives have been already implemented. The most important framework considers as the de-facto standard in the interoperability area is the European Interoperability Framework [17]. It defines a set of recommendations and guidelines for e-government services so that Public Administrations, enterprises and citizens can interact across borders, in a pan-European context.

At the same time the vast majority of countries have been developed their own frames to host e-government solutions as following.

- SAGA (Standards und Architekturen in eGovernment Anwendungen) in Germany [14]. It describes standards, procedures and methods for the use of IT in PAs and makes recommendations offered by Public Administrations particularly on the design of e-government services.
- e-GIF (e-Government Interoperability Framework) in the United Kingdom [15]. It defines the minimum set of technical policies and specifications governing information flows across government. These cover interoperability, data integration, content management, metadata and digital services access.
- CCI (Le Cadre Commun d'Intéropérabilité) in France [16]. It introduces the rules for computer

interoperability between government agencies of government in France under the General Repository for Interoperability Program ADELE.

- FEAF (Federal Enterprise Architecture Framework) in USA [18]. It provides a common methodology for information technology acquisition, use, and disposal in the Federal government.

In Italy, as well as in the other countries, an interoperability framework has been adopted. We will provide some details in the following subsection.

Several international organisms involved in standardization issues have also devoted efforts to the domain of e-government and interoperability. We mention the e-government Interest Group from the W3C, the European Committee for Standardization, a committee of OASIS (Organisation for the Advancement for Structured Information Standards) devoted to the study of the applicability of their own technologies to e-government, the Object Management Group, the Digital Government Research Center founded in 1999 by the National Science Foundation of the USA and the NESSI i-Gov working group working since February 2008 just to cite a few.

#### B. Italian Scenario: SPCoop and ICAR project

For what concern the Italian scenario starting from 2003 the CNIPA (National Center for IT in Public Administration) - now named DigitPA - started the coordination of a nationwide bottom-up consensus operation, from basic telecommunication services to advanced applications cooperation [13]. Such initiative involved 300 representative of central and local Public Administrations, universities and research centres, and Italian companies. The outcome has been a set of technical and organizational recommendations on the SPC<sup>1</sup> (Public Internet-working System). It included SPCoop for the application cooperation among PAs.

Few years later the Italian Government issued a Law Decree, namely the Digital Administration Code (Law Decree n. 82/2005) with the following aims: (i) improve the rights of citizens and enterprisers on Public Administration; (ii) put citizens and enterprisers at the center of PAs services; (iii) promote digital signatures and legal validity; (iii) enable digital contracts, payments and accounting; and (iv) development, acquisition and reuse of software in PAs. CAD defined also the role of SPCoop and SPC. Under the SPCoop umbrella some regional projects on e-government have been lunched. All the administrations have to review their way to work in line with organizational and management aspects evinced by SPCoop and SPC.

The biggest project was Interoperability and Applicative Cooperation among Regions (ICAR). It started in June 2006 with 17 partners including 16 of 19 Italian Regions. The ICAR project is co-funded in the ongoing second phase of the Italian e-government plan for Regional and Local Authorities which addresses the establishment of the so-called SPC. The project has the following objectives:

- to establish the secure interconnection of regional Public Administration networks following the rules of SPC;

- to guarantee data exchanges and application cooperation across all PA in different regions;
- implement and test standard protocols and formats for data exchange in a number of critical application/business domains for the delivery of services to end users by Regional authorities.

#### C. EU-funded projects

There have been a number of EU-funded projects in this area with promising outcomes. Following some of them.

- 1) eGOV (IST 28471). One-stop government, GovML language.
- 2) ATHENA (IST 507849). Global IOP framework for enterprise networks.
- 3) EU-PUBLI.COM (IST 35217). Definition of a 'Unitary European Network Architecture' to achieve PA interoperability.
- 4) NeOn (IST 027595). Semantic interoperability, service-oriented open infrastructure and methodology.
- 5) ONTOGOV (IST 507237). Semantic platform for composition, reconfiguration and evolution of e-government services.
- 6) SATINE (IST 002104) Semantic-based Interoperability Infrastructure for Integrating Web Service Platforms to Peer-to-Peer Networks.
- 7) SmartGov (IST 35399, smartgov.e-gov.gr) A knowledge-based platform that assists PA employees to generate online transaction services.
- 8) SemanticGov project and Access-eGov project.
- 9) TERREGOV (IST 507749) Interoperability for local and regional governments.
- 10) USE-ME.GOV (IST 2294) Platform for mobile government services.
- 11) SemanticGov (IST 027517, www.semantic-gov.org) Infrastructure for enabling the semantic web services in PA.
- 12) OPUCE (IST 034101, www.opuce.tid.es) Infrastructure for collaborative and dynamic loosely coupled services.
- 13) SUPER (IST 026850, www.ip-super.org) Modular architecture for semantic BPM.
- 14) STASIS (IST 034980, www.stasis-project.net) eEconomy services, semantic interoperability.

According to the provided description the scenario is mature enough to support interoperability and applicative cooperation. We believe this is the time to combine and integrate their results and enhance them towards a more uniform European-wide interoperability panel involving local administrations. The Marche Region experience shows the ability of regional administrations to aggregate local administration in term of technology and governance.

### III. DOMAIN KNOWLEDGE AND E-GOVERNMENT REQUIREMENTS

Domain requirements intend to highlight aspects of a particular application domain that on one side are typically not known to BP and technical experts, and that on the other side are quite obvious to domain experts. Missing to report domain requirements generally results in a project failure or in low quality systems. The discovery of domain requirements is particularly critical in the e-government

1 [http://www.cnipa.gov.it/site/it-IT/Attività/Sistema\\_Pubblico\\_di\\_Connettività\\_\(SPC\)/](http://www.cnipa.gov.it/site/it-IT/Attività/Sistema_Pubblico_di_Connettività_(SPC)/)

domain where the software to implement is directed to citizens with ample differences in the capability of operating with ICT and in the distrust they have with respect to such technologies. In developing services for the PA becomes then mandatory to consider and codify requirements aiming at removing such possible hurdles in service acceptance. Nevertheless such knowledge is not in the mind of BP or technical experts and instead it can be provided by people with expertise in social and anthropological fields. Such experts have to be involved in the requirements discovery phase, and helped in expressing the requirements they think the software system should satisfy.

In our work we considered such an aspect and, in collaborations with experts in the domain, which revised our work, we tried to identify specific quality requirements for GDS delivery processes. In particular we tried to highlight those aspects that can be considered independent from a specific service under development. The result has been the definition of a quality framework for the delivery process of GDS that has five different dimensions, where each dimension foresees different quality levels. In particular, the framework consider coordination, control, sharing, transparency and inclusion dimensions with the meaning reported in the following.

#### A. Coordination

With the term *coordination* we mean the capability of a Public Administration, involved in the execution of a business process, of using different information and communication technologies in order to interact with other PA to provide a service to the citizens.

*Lack of Coordination* is the lowest possible level of coordination that is observable within a PA according to two different possible situations (with reference to the delivery of a specific service). In the first case direct interactions between administrations, which are participating to the delivery of a service, are not precisely established. Indeed in such a case the formal definition of a business process can not be observed and the approach can not be applied. It is in general the citizen that, knowing the PAs involved in the provisioning of the service, drives the process, physically moving back and forth from one administration to the other. Lack of coordination can also emerge when, given wrong business process specifications, interactions activities could results in blocking conditions.

*Communication* is implemented by the PA without a fully integrated electronic systems. In this case the civil servants involved in the business process have different degree of freedom in the execution of tasks. So, for instance, the request coming from another PA via e-mailing system are processed by a civil servant according to his/her knowledge and expertise. The reception of a message do not automatically activate a task within the business process neither it is tracked by the system.

*Collaboration* enables PA to take part to the GDS delivery with a fully automated Business Process. Request from other participating PAs enter the organization using ICT technologies and are handled in an automated manner without requiring the intervention of a civil servant. It is worth noting that in some cases regulation and laws could impose human intervention such as for instance when document signature is requested. Even if such activities

enables a degree of freedom to the civil servants, that contribute with their knowledge, they are fully supported by the system.

*Semantic Integration* implements the highest level of coordination implementing collaboration mechanisms enriched with semantic support. Explicit formal specification of the reality related to the delivery process is shared between the participants to guarantee the understandability of the communications. In this case the system supports and enables also the civil servants during the activities making explicit the knowledge requested to fulfil the task.

#### B. Sharing

With the term *sharing* we identify the way in which the PA handles and shares citizen data with other administrations in order to participate in the delivery of a specific service according to its scope of responsibility and avoiding data redundancy and misalignment among PAs data storages.

*No Sharing* is the lower level of the GDS sharing and is observable when the administration keeps track of all citizen data and does not try to retrieve it from the right sources. Among the various issues that this way of organizing a BP brings we should certainly mention data redundancy and possible misalignment among PA data storages.

*Data Sharing* this level the Business Process implement an automatic way of retrieving citizen related data interacting with the specific PA that is in charge of maintaining the needed information.

#### C. Control

With the term *control* we refer to activation policies applied to drive one step after another the business process from its start to its final fulfilment. There is the case in which a GDS announces itself to interested citizens. Differently from an approach in which the service delivery only waits for a request from an interested citizen. The two different paradigms have profound impact on usage. Too often GDS are not used since they are not known by citizens. According to this classification we distinguish three different levels.

*Reactive control* is implemented by Public Administrations that wait for citizens requests before they start the service delivery. Citizens awareness on the “to do list” is the only driver for the service activation and delivery.

*Proactive control* enables the GDS to announce its availability through direct communications to interested citizens also providing precise references to the access point of the service itself. This is for instance the case in which a tax payment service sends an e-mail to the citizen before the deadline, specifying also a specific link in which the user will find all the necessary information to proceed with the payment. Certainly proactive control does not make sense for all different kinds of services. There are services which are inherently reactive. Nevertheless if possible proactive control can greatly foster service usage.

*Creative* refers to the presence of activities related to the promotion of related, and maybe relevant, services. In such case the PA implements services that inform the citizens of all the other services in which they may be interested (or that the citizen has to activate). For instance a citizen that

starts the procedure for getting married may be interested in services related to the provisioning of low rate mortgages, sustained by the municipality. Clearly in case related services have been implemented following the proactive paradigm the citizen will also receive information on the corresponding access point.

#### D. Transparency

With the term *transparency* we mean the ability of the administration to make citizens aware of the delivery process execution in term of activities and people in charge to govern it, improving citizens perceived trust and inclusion.

*No Transparency* is the lowest level of GDS transparency and is observable when the activities of a given service are not visible outside of the administration. Therefore a PA implementing this level of transparency makes the citizens completely unaware of the process execution. The citizen can just activate the service and wait for its end. Clearly the citizen, in particular in case of a long lasting BP, could feel frustrated given the lack of feedbacks on the execution.

*Activity Aware* is observed when the administration implements BPs tracking mechanisms. In general it is highly desirable to make the citizen aware of the activities that have been already carried on and of the activities that need to be completed. So in general the activities composing the BP and their organization are made visible to the citizens. Obviously the granularity of visible activities can be variable and a right balance should be found in order to not overwhelm the citizen with not so relevant information. The citizen will certainly feel much more comfortable with such a kind of GDS delivery and will not feel so much affected by long lasting BP if it periodically receives still alive communications.

*Role Aware* at this level activity aware transparency is implemented and enriched by the explicit identification of a civil servant responsible for the activity related to the process. This is the civil servant that is in charge of monitoring and controlling the valid execution of the service (and of executing it in case the process is not fully automated). In this way if the citizen feels that something is not going as he/she expects he can directly interact with a civil servant.

#### E. Inclusion

With the term *inclusion* we mean the ability of the administration to provide service to the citizens considering their different abilities. Economic, social, geographical condition as well as physical inability has to be considered by BP developers. In our investigation we underline three major source of diversity that mainly impact on the delivery process.

*Channel Inclusion* refers to different ways that can be implemented to access the service. Users may interact with service via many and heterogeneous devices, such as PC, wap phone, PDA, ... (see [19], [20] for a review of e-government access devices).

*Profile Inclusion* refers to the service capabilities to support citizens physical diversities. It is very relevant for disadvantaged people that are often excluded and marginalized by the introduction of ICT which do not consider inability issues. Similar effects can be observed if we refer to people with poor ICT skills and with low

education level.

*Language Inclusion* refers to the ability of the service to be used by people with different nationalities and switch among languages during the service delivery. This is particularly interesting toward the transnational service delivery enabling a fruitful provision on the European e-government pan.

#### F. Business Process Definition in the e-Government Domain

The genesis of a business process to be applied within the PA sector is rather complex and its definition typically requires the involvement of many different stakeholders. Particularly interesting is the definition of those BPs aiming at describing the cooperation of different offices belonging to different PAs. So for instance the BP permitting to move the citizen main residence requires that at least the two municipalities share a common understanding of the actions that it is necessary to put in place in order to reach the final goal (i.e. the moving of a citizen to a new municipality).

In general, limiting ourselves to the case of GDSs that require the involvement of different offices within the same country, a BP is defined and realized according to a multi-stage process where at least the following roles/steps can be identified:

- the central administration or even the law establishes the effects that a service should produce
- a governing board, composed of domain experts, is appointed to define a high level process that permits to the involved PAs to provide the service producing the effects established by law
- a PA involved in service provisioning implements the process possibly employing ICT

The approach we propose can be fruitfully employed by governing board to identify possible flaws in the BP definition which will result in limited usage by citizens. At the same time PA involved in the provisioning of a service can use our approach to judge the quality of the provided service. In particular a single PA can substitute the specification of the applied internal process within the global definition of the BP so to check if a good quality level can be reached.

#### IV. METADATA IN COOPERATIVE ENVIRONMENTS

The term "metadata" represents a set of information about data. This informal definition refer to the Greek term "meta" (who means pertaining) and from Latin "data", plural of "datum" (that means information). In literature there are several different definitions of metadata.[6,8,12] We refer here to the most common and relevant:

"Machine-understandable information about Web resources or other things" (Tim Berners-Lee, 1997) [1]

"Data associated with objects which relieves their potential users of having to have full advance knowledge of their existence or characteristics. A user might be a program or a person" [4].

"Structured data about resources that can be used to help support a wide range of operations" [3]

It is clear that the distinction between data and metadata is not intrinsic but depend on the context in which we utilize metadata. The metadata them-self are data and therefore

they can be stored as usual data (in the resource they represent, or in another), and can be described by other metadata, and so on. The number of metalevel to specify depends from the features of applications and from different specific needs. Typically, when we argue about metadata we make a distinction on the basis of a classification on five levels (Figure 1).

**Layer 1**, Instance Metadata comprises the raw data as values, and does not refer to anything other. They are represented by data instances that represents the effective value-oriented elements that are not of interest if considered in isolation.

**Layer 2**, Syntactic Metadata are the first explicit metadata that are needed by a machine to “know about” the data. These metadata are inherent to the types of digital information, data types, language formats, messages or documents length, source, bit rate, encryption, and so on.

**Layer 3**, Structural Metadata gives form and structure to “units” of data. This is simply a way to understand the structure of data and the three main way to represent them using on hierarchical, relational or object-oriented paradigm.

**Layer 4**, Referent Metadata provides linkage between different data models. They describe how data should be converted from a schema to another in the same language, or one schema to another in different languages, or again, one schema to another at very different abstraction levels.

**Layer 5**, Domain Metadata provides contextual data to other data and information. This capability is a crucial requirement for information sharing architectures to convert

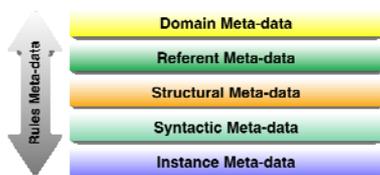


Figure 1. Metadata Stack

data meaning across multiple contexts. Typically the use of ontologies [5] as contextual metadata gives application a reference point to perform transformation and interpretation of data from different and seemingly incompatible heterogeneous contexts.

**Layer 6**, Rules is a transverse layer that cuts across all the discussed layers. Rules are used to constrain the semantics of metadata specifications and models at any abstraction level. Like structural and syntactic metadata they can be embedded in application code also if recent efforts try to alleviate this problem with specific XML-based languages. Business rule represent definitions of business terms, data integrity constraints, mathematical derivations, logical inferences, etc.

In this contest, the capability of metadata to be understood both from humans and from machine result the right choice for the reduction of time and costs related to information research and comprehension. Today, in the Public Administration setting there is a process of administrative decentralization of competencies and we observe a transformation of administrative procedures from uni-agency to multi-agency. The first represents independent activities able to perform specific tasks that involve only a single administration, while the other represents the result of

a cooperation among various administrations that cooperate for reach a common task (typically the delivery of a service required by a final user). The introduction of recognized standards appear in this contest a way to realize a full interoperability.

## V. INTEROPERABILITY FRAMEWORK AND ARCHITECTURE

### A. Semantic interoperability

Metadata plays a crucial role in semantic interoperability that is defined as:

*“a dynamic enterprise capability derived from the applications of special software technologies (such as reasoner, inference engines, ontologies, and models) that infer, relate, interpret, and classify the implicit meanings of digital content without human involvement – which in turn drive adaptive business processes, enterprise knowledge business, rules, and software application interoperability” [9].*

Clearly, this capability rely on data that as the foundation for any information sharing program.

This enhanced notion of data include semantic and context then we turn data into information using metadata. Several capabilities can emerge using metadata:

*Data Interoperability* enables data to maintain its original meaning in multiple contexts using data meanings as the basis for transformations.

*Process Interoperability* enables business processes to be expressed in terms of another inferring meaning from contextual data and applying it in a different process/context.

*Services/Interface Interoperability* enables a service to discover, bind and use (communicate) with a new service without predefined custom glue code.

*Application Interoperability* enables the interactions of methods and transactions between heterogeneous software applications to achieve platform independence. Taxonomy Interoperability enables any category in the taxonomy to be expressed in terms of other categories exploiting meaning in category definition.

*Policy Interoperability* enables businesses to protect valuable resources using security mechanisms and rights management.

*Social Network Interoperability* enables people in different communities to network, infer and discover connections through previously unknown contacts.

Interoperation of these components lead to a high level of dynamism with potentially the feature of an autonomic environments like: (i) self-configuring (of interfaces service descriptions, etc), (ii) self-optimization (of transactions, routing, queries, etc.) (iii) self-healing (recovery of error flows, etc.) and (iv) self-protection of data.

### B. Characteristics of a framework

A framework for interoperability that can constitute the infrastructure for an e-Government information system should be defined - according to Pollock and Hodgson – as “a highly dynamic, adaptable, loosely coupled, flexible, real-time, secure, and open infrastructure service to facilitate a more automated information-sharing framework among diverse organizational environments”. We mention in the definition several characteristics:

*Dynamic* relies on the infrastructure ability to support the

configuration of information assets in previously unforeseen ways without human intervention.

*Real time* relies in the capability of building new modes of communications among services and systems on demand, without change the infrastructure.

*Loosely Coupled* relies to the principle of indirection used in systems design and architecture.

*Highly Flexible* relies on the capability of reconfiguration of components.

*Security* needs to exist at every layer of a complete technology architecture. Network security, application security, data security must be considered and properly enforced inside the architecture.

*Open* refers to the feature concerning the duration in the time of the architecture.

*Service Oriented* relies on the architectural style used that is supposed to be service-based.

*Information-Centric* relies on the application of knowledge management principles that should be applied to a computing environment to formalize the description and management of data.

*Autonomic* refers to the capability of the framework to evolve on its own, typically performing configurations and maintenance in automation without human intervention.

Clearly, interoperability frameworks that fulfill all the presented characteristics may be the panacea in e-Government information systems and in general in middleware technologies. Our proposal have an high level of abstraction but should encompass as much as possible these feature to bring near the optimal solution.

### C. Architecture

Our proposal works in a distributed environment in which available information systems must be integrated to easily enable information sharing over the Internet.

In Figure 2 we propose a general system architecture. Legacy application systems will be wrapped into subsystems that enable interoperability with other subsystems in a predefined scenario. A subsystem role (and its ontological description) will guarantee the delivery of appropriate system functionality between processing modules - using web service technologies. Subsystems use a document repository to store multi-format documents generated by existing solutions. These documents form the input for a ontology engineering framework.

The Interaction module, acting as a local mediator, guarantees the interactions between roles associated with system functionalities. This means that incompatibilities at message level are handled by the interaction module that tries to mediate interaction between roles. On the other hand, it represents a part of a global mediator between legacy systems and other subsystems involved into the distributed processing. In this setting, interaction module uses a shared metadata repository and exploits the potential of a shared dictionary respect to a given local conceptualization.

Communications will use the Communication channel for service message exchange. It represents the unique medium for interaction among subsystems roles and is represented by the underlying network infrastructure. The communication channel will also support the distribution of knowledge around the service ecosystem. The global coordination activity is performed by a coordination engine

and will be able to manage global process specifications. In this context, a specification document according to a conceptualization specifies process behavior and manages document flow among subsystems.

The basic functionalities of the engine are represented by two modules: the process interaction module and the message interaction module. The first manages process specifications and is able to manage data and documents to satisfy a specified goal. This is the core of the coordination engine that uses process specification and manages the interaction between subsystems. At the same time it manages the roles involved in each subsystem and their relationships. A common schema message allows the interaction module to work effectively. Schemas and ontology are stored in a repository (ontology library) that contains all the relevant information to identify messages that are used by the interaction module to solve inconsistency problems. Moreover, it supports process interaction module managing the schema of messages exchanged among subsystems.

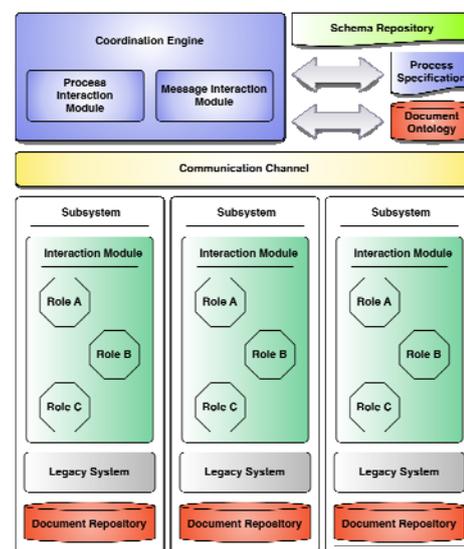


Figure 2. Interoperability Architecture

It is worth noting two forms of use (see Figure 3): (1) when a definition requires one or more process elements, these will be found by the element discovery engine (and include decomposition) using the document ontology and (2) when new knowledge is extracted from a legacy component (or other multi-format document), it will be distributed across the network by a propagation engine using the communication channel and stored in local document ontology libraries.

Finally, interoperability will be achieved because process definitions specifies business interaction among roles involved in the whole system. The specification by a business user, coupled with semantic decomposition into process element, provides a novel business integration infrastructure.

With e-Government diffusion Public Administrations will face with a new dimensions concerning technology. We propose a complete integration characterized by open and dynamic distributed systems based from strong heterogeneous resources and legacy information systems. It is clear that e-Government domain must resolve problems

that arise in all those contests. To do this a constructive comparison between technical and managerial capabilities of different administrations should take place.

This is not a trivial task because require an high cooperation level for the definition of intra-administration common strategies. Moreover, cooperation become a complex challenge when information exchange is vehiculated using services but new opportunities for interoperability arises from the use of service-oriented architecture. These should be relevant for e-Government because integration of processes and information will drive integration of large scale applications and Public Administration information systems.

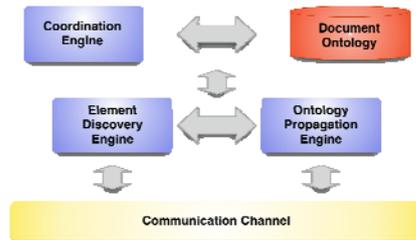


Figure 3. Ontology propagation and element discovery

## VI. ONTOLOGY-DRIVEN DEVELOPMENT OF INTELLIGENT DOCUMENTS

In e-Government, interoperability and applicative cooperation compete for the development: (1) back-offices process integration for service provisioning, and (2) distribution of services with suitable quality attributes. Public Administrations increase their shared knowledge (Pollock and Hodgson, 2004) stimulating the realization of infrastructures that allow applications to suitably cooperate. Such cooperation would certainly benefit by a formal representation of the reality of interest.

In this section we propose a model driven document management for Public Administrations that is based on ontologies [5] to describe and analyze the relations between concepts [11].

The importance of ontologies to describe the structure of a document can be found in [7]. An ontological description of electronic documents has been explored in [2] for the so-called "Intelligent Document" that describes both the document structure and the procedural iter for its filling. Within an intelligent document we can identify different parts - different zones - the compilation of which is responsibility of different PAs by means of suitable workflows. We use ontologies also to describe the type of data and information within the different zones. This provides a strong support for filling documents but also for their analysis and generation of suitable interfaces or documents form. We also discuss how our ontology-based documents formalization supports the entire life cycle of the document: from the user interface creation to the management of the document compilation and its storage. From the ontology, indeed, we can retrieve a form to represent the user interface of the document. We can also manage the iter of compilation using the information contained in the process model of the document. Moreover, the document persistence is also managed by means of an ontology management system that maps the document

ontology to a relational database.

Besides providing a semantic management of electronic documents, the general ontology-based framework we discuss in the domain of Public Administrations, provides a solid support to automatic administrative procedures and allows information retrieval and knowledge management according to the demands of a single Public Administration. PAs can take advantage of the use of electronic documents because of a (i) consistent decreasing of the costs of press, elaboration, distribution, delivery, etc., (ii) consistent decreasing of compilation errors, (iii) possibility to share and/or to access the shared information.

Moreover, from the other side, citizens and companies, as end-users, can take advantage of the use of electronic documents because of a (i) consistent decreasing of the time needed for manual compilation, (ii) electronic dispatch of the documents, (iii) decrease of the delivery costs (postal, fax, etc). This short and certainly incomplete analysis shows the necessity to create a process of workflow management that contemplates the fruition of the electronic document inside the application domain. The approach that has been used till now is not enough to answer our questions, we don't have to think to an Intelligent Document as a simple form but as the result of an orchestration of services offered by the community of PA's that participates in its filling.

### A. Document Specify Ontology

We describe a simple ontology that represents the concept of Intelligent Document.(fig. 4)

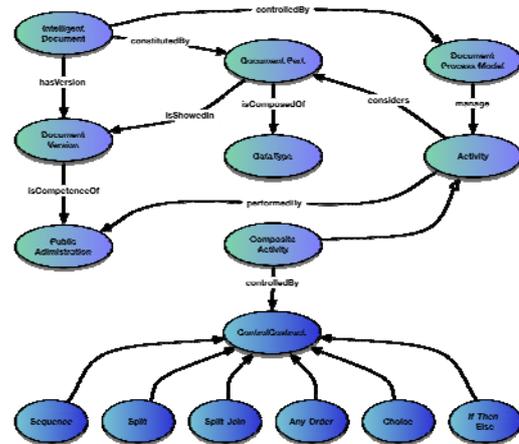


Figure 4. Intelligent Document Ontology

This ontology can be merged with the ontology describing the structure of a document - a document-specific ontology - in order to detect the forming parts of the document itself. In fact, an Intelligent Document can be divided in several parts shown in different versions/ views. Each view has a version identifier and each PA involved in the document filling process can modify only the specific part that is of its own competence.

The filling process is itself part of the intelligent document definition. It performs activities (represented by the Activity concept) that are a couples composed by a DocumentPart and a PublicAdministration.

### B. Ontology-driven Intelligent Document Compilation

All forms of engineering rely on models to better understand complex, real-world systems. Ontology-based models provide abstractions from physical world allowing

developers to reason about systems at different abstraction levels. We now discuss how ontologies can be used to guide the process of building on-line intelligent documents and to drive the storage of relevant information inherent in the document itself. Our approach is based on the architecture shown in Figure 5. We model the document using the intelligent document ontology and the document-specific ontology (describing document related concepts). This model is used in the Document Persistence and in Document Engine modules of our architecture. Document persistence module is constituted by an ontology management system that is formed by an import module, a backend storage and a query engine. The import module is responsible of the mapping from ontology to relational database. The main advantage of this approach instead of classic database is the use in conjunction with a query engine. Querying ontologies can benefit of inference mechanisms and it is possible for the system to answer complex queries using deduction.

Moreover, document engine has two tasks performed by Forms Generator and Compilation Manager.

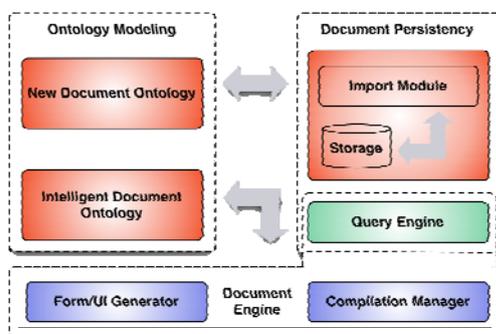


Figure 5. Architecture of the ontology-based information system for intelligent document

The first generates forms - user interface of on-line intelligent document - based on ontologies. These forms can be viewed and utilized with a browser. The Compilation Manager module is responsible for the control of the compilation process by taking into account the workflow that has to be performed to fill the document. The workflow is, of course, document-dependent and contains all the information about the procedural iter of the document filling.

Summing up, in order to deliver a new kind of document we must specify the ontology that describes the document itself together with its iter of compilation. This ontology is used by the information systems of the involved PAs (those that are involved in the delivering of the document). Each PA generates the forms needed for the document filling and is able to detect the tasks to be performed on the basis of the process model describing the activities. The compilation process is managed by the "Document Engine" of the PA that starts the iter of compilation as orchestration of services of the involved PAs, querying both the local knowledge base and external ones and interacting with users by means of the generated forms.

## VII. CONCLUSION

In this paper we have shown an ontology-based methodology and infrastructure for the delivery of intelligent documents within Public Administrations. PAs can access a common repository to find ontologies

describing a document and set up their information systems to manage the entire life cycle of the document: user interface creation, data storage and management of the iter of compilation.

The models and the conceptual architecture that we have discussed derive from a careful analysis of the e-Government application domain. We deal with the concept of Intelligent Document, a document that is able to describe its structure and to specify the procedural iter for its filling. This can be seen as a result of a cooperation between PAs involved in the delivery. Document based cooperation has the potential to improve efficiency among PAs but also satisfaction for final users as citizens and firms

## REFERENCES

- [1] Berners-Lee T. Design issues for the world wide web, <http://www.w3.org/DesignIssues/>, 1997
- [2] Corradini F., Polzonetti A., and Pruno R. E-Government administrative and semantic cooperation: the role of "Intelligent Documents". In *Electronic Government: 4th International Conference*, pages 150–157, Copenhagen, Denmark, August 2005. EGOV 2005
- [3] Day M. Metadata for digital preservation: A review of recent developments. *Research and Advanced Technology for Digital Libraries*, pages 161–172, 2001.
- [4] Dempsey L. and Heery R. Metadata: a current view of practice and issues. *Journal of Documentation*, pages 145–172, 1998.
- [5] Gruber T. A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2):199–220, 1993
- [6] Hovy, E., Philpot, A., Klavans, J., Germann, U., Davis, P., and Popper S. Extending Metadata Definitions by Automatically Extracting and Organizing Glossary Definitions, in *Proceedings of the NSF's dg.o 2003* (Boston, MA, May 2003).
- [7] Klischewski R. Towards an ontology for e-document management in public administration, the case of schleswig-holstein. In *36th Hawaii Int. Conf. on System Sciences*, Hawaii, 2003
- [8] Maedche, A., Zacharias, V.: Clustering ontology-based metadata in the semantic web. In: *Proceedings of PKDD 2002*. (2002) 348–360
- [9] Pollock J.T. and Hodgson R. *Adaptive Information: Improving Business Through Semantic Interoperability, Grid Computing, and Enterprise Integration* Wiley-Interscience, 2004.
- [10] H. J. Scholl, "Interoperability in e-Government More than just smart middleware.," presented at 38th Hawaiian International Conference on System Sciences, Waikoloa, Big Island, HI, 2005.
- [11] Sowa, J.F. "Knowledge Representation: Logical, Philosophical, and Computational Foundations. Brooks Cole Publishing Co., Pacific Grove, CA, August 1999
- [12] Widener P., Eisenhauer G., and Schwan K.. Open metadata formats: Efficient XML-based communication for high performance computing. 10th IEEE International Symposium on High Performance Distributed Computing, pages 371–380, 2001
- [13] R. Baldoni, S. Fuligni, M. Mecella, and F. Tortorelli. The Italian e-Government Service Oriented Architecture: Strategic Vision and Technical Solutions. In *Proc. of 6th International EGOV Conference*, pages 79–88, 5, 2007.
- [14] KBSt. Saga. Web available, 2008. <http://www.kbst.bund.de/182/SAGA.htm>.
- [15] UK GovTalk. e-GIF. Web available, 2008. <http://www.govtalk.gov.uk/>.
- [16] French Government. Adea. Web available, 2008. <http://www.adae.gouv.fr/adele/>.
- [17] D. Enterprise and I. I. Unit. European interoperability framework for pan-european e-government services. Web available, 2008. <http://europa.eu.int/idabc/en/document/2319/5644>.
- [18] POPKIN Software. FEAF. Web available, 2007. <http://government.popkin.com/frameworks/feaf.htm>.
- [19] W. Ebbers and W. Pieterse. The multi-channel citizen: a study of the use of service channels by citizens in the Netherlands. In *Electronic Government, 6th International Conference, EGOV Proceedings*, 2007.
- [20] Willem Pieterse and Jan van Dijk. Channel choice determinants; an exploration of the factors that determine the choice of a service channel in citizen initiated contacts. In *dg.o '07: Proceedings of the 8th annual international conference on Digital government research*, pages 173–182. Digital Government Research Center, 2007.