Proc. Int. Conf. on Dublin Core and Metadata for e-Communities 2002: 7-17
© Firenze University Press

District Architecture for Networked Editions: Technical Model and Metadata

Antonella Farsetti
Firenze University Press (FUP) - Università di Firenze
e-mail: antonella.farsetti@unifi.it

Valdo Pasqui
Centro Servizi Informatici dell'Ateneo Fiorentino (CSIAF) - Università di Firenze
e-mail: valdo.pasqui@unifi.it

Abstract

District Architecture for Networked Editions (DAFNE) is a research project funded by the Italian Ministry of Education, University and Research aiming to develop a prototype of the national infrastructure for electronic publishing in Italy. The project's initial target concerns the scientific and scholarly production in the human and social sciences. The organizational, legal, technical and business aspects of the entire digital publishing pipeline have been analysed. DAFNE system will support the request-offer chain by promoting the integration between the digital library and the electronic publishing districts. In this paper we present the main results of the project's first year of activity. First a quick outlook about the actors, objects and services is presented. Then the functional model is examined bringing out the distinction between information content and digital objects. Afterwards the technical model is described. The system has a distributed architecture, which includes three categories of subsystems: Data Providers (i.e. the publishers), Service Providers and External Services. Data and Service Providers interact according to the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Finally DAFNE metadata is discussed. Metadata permeates the whole publishing chain and DAFNE metadata set is based on already defined domain-specific metadata vocabularies. Dublin Core Metadata Initiative (DCMI) and Publishing Requirements for Industry Standard Metadata (PRISM) are the main reference standards. Open Digital Rights Language (ODRL) and Open Archival Information System (OAIS) are the two other relevant models which complete DAFNE metadata specification.

Keywords: *electronic publishing, data and service provider, descriptive metadata, rights management, digital objects, OAI-PMH.*

1. Introduction

During the last few years the publishing industry has been greatly affected by Internet and its new technologies. All the main publishers propose Web portals to organize their catalogues and services, including the possibility to read on-line or to download publications in electronic format. Many journals have a digital on-line version or they are published only in electronic format (i.e. born digital). The standardization of metadata schemas, in particular thanks to Dublin Core Metadata Initiative (DCMI), and the wide acceptance of the Protocol for Metadata Harvesting (PMH) promoted by the Open Archives Initiative (OAI), are pushing the creation of Internet-based services in the publishing market. The consolidation of XML standards and tools is leading to a more structured organization of the contents, independent of proprietary formats such as Adobe PDF and MS Word. Furthermore, the quick diffusion of e-commerce platforms is promoting more flexible trading models in the publishing market. Finally, in the academic and research communities relevant initiatives aim to promote the "free" diffusion of the scholarly knowledge by the creation of e-prints services, mainly based on institutional and thematic archives.

In this framework, the Italian publishing state-of'-the-art is still in its infancy. Moreover, some sectors such as the economy and the social sciences are self-consuming knowledge-generation contexts. This is a niche market whose authors and consumers are the same actors. In particular the distribution of scientific journals has its main focus on subscribers, special clients and specialized bookshops.

District Architecture for Networked Edition (DAFNE) is a research project funded by the Italian Ministry of Education, University and Research (MIUR) which aims to develop a prototype of the national infrastructure for electronic publishing in

Italy. The project takes into account the organizational, legal, business and technical aspects in order to define a set of services and tools for supporting the subjects involved in the digital publishing pipeline. The main goals of DAFNE are: i) to improve the efficiency and efficacy of the request-offer chain; and ii) to promote the integration between the digital library and the electronic publishing districts.

Firstly the project aims to analyse the scholarly and scientific production in the human and social sciences.

The project's participants are five relevant Italian companies: Ariadne (ICT), Bassilichi (e-commerce), Editrice Bibliografica (publisher), Casalini (bookseller) and the consortium "Parco Scientifico e Tecnologico Galileo" (ICT). The scientific committee includes three main Italian universities (Bologna, Firenze and Padova), the "Istituto di Teoria e Tecnica dell'Informazione Giuridica" of Consiglio Nazionale delle Ricerche (CNR), the Biblioteca del Mulino and the Biblioteca Nazionale Centrale di Firenze (BNCF). The University of Florence is participating in the project with the Firenze University Press responsible for metadata specification and the CSIAF responsible for the technical design.

DAFNE started in October 2001 and will last for three years. Three reference models have been defined in the first year: legal, organizational and technical. Relevant metadata standards have been analysed and a draft proposal for DAFNE reference metadata has been elaborated. Both the metadata set and the business model will be finalized by the end of the current year.

2. Actors, objects and services

2.1 Actors

In the electronic publishing pipeline one end includes a set of processes such as content acquisition and generation, digitisation and layout editing. At the other end of the pipeline there are the business transactions with the consumers (i.e. end-users). The DAFNE organization model has stressed the need for an Internet-based system to support the consultation (metadata and contents) and the on-line sale of electronic publications. DAFNE aims to provide tools and services to create a bridge between the publishers and the end-users in order to improve their relations: consumers can evaluate the contents of works more in depth and publishers can take into account the preferences of their clients. In this context the main actors, including both human subjects and external systems, are:

- authors, who generate intellectual contents
- publishers, including editorial-staff members
- end-users, the consumers (teachers, researchers, students, private companies members, professional experts, etc.) who look for publications about their scientific, research or personal interests
- brokers, service providers, public and academic libraries which support value-added services such as metadata, abstracts and TOCS creation, metadata and full-text indexing, document access
- national libraries that manage legal deposit services
- systems for the assignment and resolution of document unique identifiers (e.g. ISBN, ISSN, NBN, DOI)
- e-commerce services, such as banks and credit-cards networks, to support on-line payments and to manage the accounting of intellectual propriety rights
- Public Key Infrastructure (PKI) services, including Certification Authorities, to support cryptography, strong authentication and digital certificates
- delivery services, such as national postal systems and private couriers

2.2 Objects

In the academic context a wide range of digital objects can be organized in electronic archives or directly published on-line. Three reference axes have been explored for all documents: types, aggregation level, and formats.

Document types. On the basis of the Firenze University Press experience the following kinds of documents have to be considered:

- monographs and conference proceedings
- electronic journals
- pre-prints and post-prints of journal papers
- technical and research reports
- lecture notes, teaching notes and learning material
- multimedia publications, including audio-visual recordings
- collections of digitised documents
- theses and dissertations
- bibliographic and review collections about specialized thematic topics

Aggregation level. In the previous list many documents are basically monographic. They are standalone publications or can be collected in series depending on specific publishing policies. Multimedia publications can be organized, as monographic issues and their manifestation can be a file to be downloaded or a full hypermedia document to be navigated and read/played/seen on-line. Usually, bibliographic and review collections are managed by databases and a set of procedures which support a search and retrieve Web interfaces. On-line journals keep the same aggregation levels of their related paper version, that is the volume/number hierarchy. So the end-user navigates by year, then by single issue that contains the table of contents pointing to articles. These papers are described by a basic set of metadata, which contains the links to the full-text version, usually restricted by access control mechanisms such as user registration or IP address authentication. The contents of born-digital journals (e.g. D-Lib Magazine, Ariadne, IEEE Distributed Systems ONLINE) are more structured and their graphic layout is generally richer. In any case descriptive metadata must be associated to documents to represent their relationships. In particular, keywords and subject elements make up "virtual" aggregations by thematic areas based on selected subject and classification schemas. Relation elements must be used to express at least hierarchical relations like "is-part-of" and "has-part".

Formats. Basically DAFNE will support the commonly diffused document formats: HTML, Adobe PDF, Microsoft Word, RTF and TeX (used in mathematics and physics). The lack of consolidated e-books standards has suggested postponing these new kinds of publications for future investigation. In the advanced model the DAFNE production phase is two fold, both based on XML standards. The former deals with the generation of the information contents according to pre-defined XML DTD and schemas. The latter includes the layout formatting and the graphical rendering, even in different formats such as HTML and PDF, by means of a transformation process based on the Extensible Stylesheet Language (XLS) and XML Transformations (XLST). The first phase has the focus on the content and its structure, independent of proprietary formats. Descriptive metadata (e.g. Dublin Core elements) can be included to be automatically extracted in the next processing steps. The second phase concerns the possibility of generating different manifestations of the same intellectual work. In this step preservation and technical metadata should be added to properly manage the generated bit stream.

2.3 Services

Publishers like Addison Wesley, Elsevier or Springer are typically oriented to business support by offering virtual on-line shops to their consumers. Sometimes these Web sites are integrated with e-commerce sub-systems based on the shopping basket model. In other cases, consumers are redirected to booksellers and distributors sites such as Amazon.com and Fatbrain to buy on-line. The recent development of portals technology started a new generation of sites that support user profiling and monitor their activities. Thus the end-user operating environment can be dynamically tailored on the basis of her/his preferences and habits. Publishers are pushing consumers to join their portals as one-stop shop points for discovering, retrieving and accessing all the information they search for in the net.

DAFNE's focus is on the generation and diffusion of electronic publications in the scientific and academic context. For several years free consultation services have been exploited in this area such as ArXiv, NCSTRL and CogPrints. These e-print systems are based on the author/institution self-archiving model with dedicated Web interfaces for the direct submission of documents and the creation of metadata by registered authors. As regards papers the full-text visualization and print is free for the end-users.

Starting from this realm, since October 1999 the Open Archives Initiative (Van de Sompel & Lagoze 2001) has defined a general framework for open archives metadata interoperation based on the concepts of Data and Service Providers and the Protocol for Metadata Harvesting (PMH) now available in version 2 (OAI). The Electronic and Computer Science Department of the University of Southampton has implemented Eprints, a software for the creation and management of e-prints archives based on the self-archiving approach. Eprints version 2 supports the last PMH version and is free for download, compliant to GNU guidelines.

Experts like Steven Harnad (Harnad 1999) and Paul Ginsparg (Ginsparg 2001) have undertaken a true mission for promoting e-prints open archives in the scholarly publishing context as the solution for the rapid and free diffusion of scientific knowledge. Recently the Scholarly Publishing & Academic Resources Coalition (SPARC) has issued a paper to stimulate the discussion about scholarly communication and institutional repositories. The report contains two relevant positions: a) "institutional repositories can provide an immediate complement to the existing scholarly publishing model, while stimulating the emergence of a new disaggregated publishing model that will evolve over time"; and b) "Institutional repositories represent the logical convergence of faculty-driven self-archiving initiatives, library dissatisfaction with the monopolistic effects of the traditional and still-pervasive journal publishing system, and availability of digital networks and publishing technologies." (SPARC 2002, p. 29).

DAFNE, taking into account these two models, publishers "business shop" and academic free "knowledge diffusion", has defined a flexible framework to let both live together for pay publications and free e-prints archives (e.g. pre-prints, post-prints, lectures and learning material). In particular this approach is suitable for those University Presses which publish several kinds of documents and have two main categories of consumers: "internal" users (enrolled students, teachers, researchers) and "external" buyers (professional specialists, private companies, common citizens). The analysis phase has pointed out the following set of services to be supported by the system:

- document submission in digital format
- related metadata creation and upgrading
- peer-reviewing
- document digital signature, time stamping and encryption

- full-text indexing
- metadata indexing, search and retrieval
- authors registration and authentication
- end-users registration, profiling and authentication
- documents (e.g. full-text) access according to different kinds of policies (e.g. pay-per-view, print-on-demand, download, subscription)
- alerting, news, interest group subscription
- on-line payment (e.g. by credit cards)
- copyright management and accounting of royalties related to intellectual proprietary rights

3. The functional model

One of the main prerequisites of the project was to define a reference framework to manage several types of documents, to allow different organization/ aggregation models and to exploit flexible archiving systems. Dealing with the academic and scientific publishing context the model must support:

- a high level of autonomy for the authors who generate the intellectual contents
- local, national and international interoperability of document collections by thematic areas
- the integration with other information resources (primary and secondary) to support end-users uniform and simple access
- co-existence with business publishing

These goals have implied a clear distinction between two concepts: information (or intellectual) content and digital objects. The former is an abstract concept; the work generated by the intellectual activity of its authors, which is the core of the academic and scientific production, independent of formats, organizations and access modalities. The latter is the concrete representation of the information content within the system. According to this logical view, a digital object becomes an extension of the information content and includes the following components:

a) metatada which describes all the features of a digital object; at least four categories of metadata must be included: descriptive, representation and technical data, copyright and intellectual properties data, end-users' access rights;
b) the physical representation of the information content, formed by one or more bit-streams (i.e. texts, sounds, images, video sequences);
c) a persistent and unique identifier assigned when the digital object is created in the system.

In DAFNE the publishing pipeline has been partitioned in a sequence of relevant logical phases (Pasqui 2001). The submission of a new information content is the starting point A dedicated system module supports the direct immission of documents by pre-registered authors who also create a basic set of descriptive metadata. This activity can be executed by the editorial-staff (e.g. when a document is sent by e-mail). In any case, they perform formatting, graphical and layout restyling by using off-the-shelf authoring tools. Moreover they are in charge of the revision and full creation of descriptive metadada that feed the on-line catalogue to be searched/ browsed by end-users. At the end of this phase the information content becomes a digital object stored in a dedicated area of the publisher's repository, accessible only to the authors, the editorial staff and, if necessary, to the reviewers.

The peer-review is a well-known activity in the scientific publishing context. To manage and track all the interactions with the reviewers a "by hand" process can be used, using e-mails to exchange documents and review comments. Otherwise a dedicated system, such as Manuscript Central (ScholarOne), Xpress Track (XpressTrack), EdiKit (Berkeley Electronic Press) can be used. DAFNE's first prototype follows the first approach and the integration with an off-the-shelf tracking system has been planned in the advanced version.

When the information content is ready to be issued for publication the editorial-staff has to conclude the processing by performing several other operations. First, all the metadata related to the new publication has to be added: descriptive (including subject and relation entities), technical, copyright, end-user rights (including access modalities and related costs). Second, a persistent and unique identifier has to be assigned to the publication based on the Uniform Resource Name (URN) standard (IETF RFC 2141). URNs are persistent, location-independent, resource identifiers, a subset of the more general category of resource identifiers known as Uniform Resource Identifiers (URI) (IETF RFC 2396). In the publishing area exists a de-facto standard, the Digital Object Identifier (DOI) system, promoted by an international coalition of commercial publishers. Being a research project, DAFNE is investigating the possibility to develop a light identification system based on Light Directory Access Protocol (LDAP) technology. Third, a digital signature and a digital timestamp (based on PKI technology) can be generated to assure document authenticity and not repudiation by authors. In this phase the staff can submit the publication, including its digital signature and a subset of metadata, to the national service responsible for the legal deposit of electronic publications, which in Italy is the Biblioteca Nazionale Centrale di Firenze. Finally, the digital object enters the persistent storage area of the publisher's repository. This means that its related metadata becomes available for consultation and export. Now the publication is ready for access on the basis of the rights permission and control access mechanisms defined before.

In DAFNE publisher's catalogue consultation (i.e. end-users search and discovery) and other services such as alerting, news diffusion and even user registration and access control can be supported by the publisher directly or by a third party service provider, acting as a broker between the consumers and the publisher. In the first case the catalogue and the other services are integrated in the publisher repository, in the latter all the services have been delegated to a service provider and the publisher simply acts as a data provider which exports its metadata and objects.

Digital object access is the functional component, which deals with access control policies and business logic. Access control is based on the matching of two classes of properties: i) the access rights modalities associated to each digital object which are described by dedicated metadata; ii) the attributes which characterize each end-user (e.g. type, institution enrolment, subscriptions and registrations). Unless a publication is free, some kind of user identification and authentication should be implemented (e.g. IP based for institutional users and user/password for generic consumers). Business logic deals with all the aspects related to payments and royalties accounting. To this end the on-line catalogue Web interface must be integrated with an e-commerce subsystem which supports on-line payment transactions. DAFNE will exploit a synchronous approach that immediately notifies the seller's server about the successful completion of on-line payment. This allows the implementation of the pay-per-view access model without user subscription or pre-registration. Credit cards and mobile telephone payment models will be experimented. As far as royalties accounting is concerned DAFNE functional model asks for specific metadata to describe the fee (e.g. a fixed amount or a percentage of the cost) and to identify the subjects (persons, institutions or companies) entitled to that fee. These subjects must have a persistent and unique identifier. To avoid the maintenance of local repositories, a central (national) service for the registration of the subjects (e.g. authors) involved in copyright management should be implemented. Publishers and service providers use these metadata to compute the royalties resulting from end-users access to digital objects. The cumulated amounts will be periodically transferred to the central system to generate the payment transactions to be credited to the entitled subjects by a bank.

4. The architecture

The overall system architecture is distributed according to a three level logic depicted in Fig. 1 as a sequence of concentric rectangles. Data Providers are the core level, which includes basic services such as digital objects repositories. In DAFNE a publisher is a data provider. At the intermediate level there are the Service Providers, which manage value added services (e.g. resource cross-references). Moreover, they can support basic services delegated by publishers. The Support Services layer includes autonomous systems that provide specific services, shared by the whole publishing community on the basis of national and inter-institutional agreements.

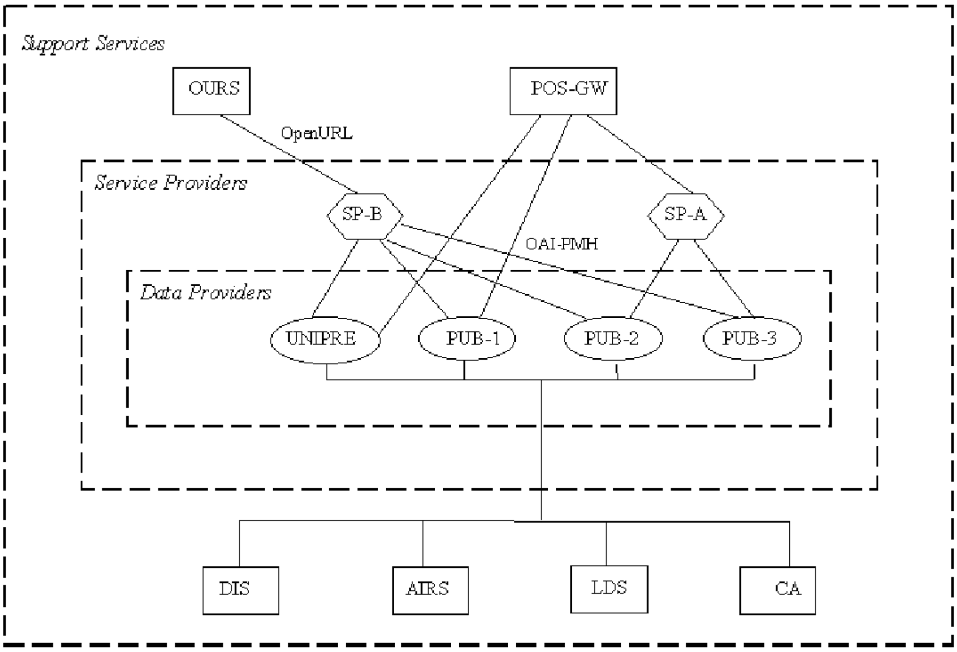

Figure 1. Three-level architecture and a possible deployment scenery

The design of Data and Service Providers subsystems is fully modular and their implementation should be component based to support the maximum flexibility and scalability.

Data Provider components are:

- *Storage* [+], the repository for the storage of documents and metadata.
- *Submission* [+], a Web interface for documents submission.
- *Authors Registration&Authentication* [+], the module responsible for authors registration and authentication.
- *PMH* [+], the interface that supports metadata harvesting from the repository according to PMH v.2 specification.
- *Document Identification Interface* [+], the module that supports the interaction with the service responsible for the assignment and resolution of persistent unique identifiers to digital objects.
- *Peer-review*, the tracking system to manage the reviewing activity.
- *End-Users Registration&Atuthentication*, the module responsible for authors registration and authentication.
- *End-Users Search Interface*, the Web interface to the catalogue which provides both simple and advanced search, main index browses, navigation by series and other documents aggregations.
- *End-Users Authorization*, the module that checks end-users access to digital objects.
- *E-commerce module*, the component for on-line payments management, interfaced to the POS Gateway subsystem.
- *Copyright module*, the component which implements rights management fees computation and accounting.

Only the basic components, marked with [+], must be deployed by a publisher (i.e. a data provider). The other modules can be omitted if their functions are delegated to an external service provider. Each Data Provider must implement the Protocol for Metadata Harvesting. This module supports any interaction with the external world, in particular with Service Providers, to export the metadata related to the digital objects hosted in the repository.

Fig. 2 is the schematic representation of a full-component Data Provider subsystem.

Service Provider components are:

- *Storage*, to host the metadata harvested from data providers' repositories.
- *Cross-reference linking*, to extend the items with links to other related resources (e.g. OPACs, bibliographic databases, abstracting services) based on the OpenURL standard (Van de Sompel & Beit-Arie 2001).
- *PMH*, the interface to harvest metadata from Data Providers repositories
- *End-Users Registration&Authentication*, the module responsible for authors registration and authentication.
- *End-Users Search Interface*, the Web interface to the catalogue that provides simple and advanced search functions, main index browses, navigation

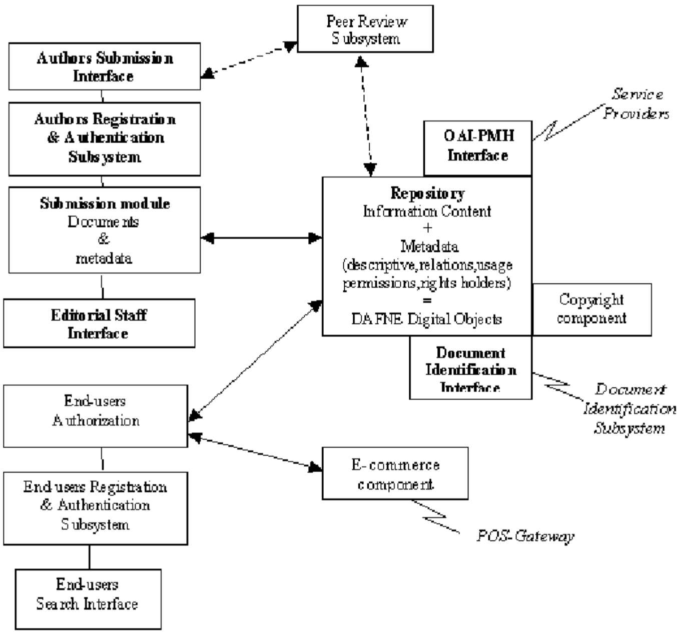
Peer Review Subsystem
Authors Submission Interface
Authors Registration & Authentication Subsystem
Service Providers
OAI-PMH Interface
Submission module Documents & metadata
Repository Information Content + Metadata (descriptive,relations,usage permissions,rights holders) = DAFNE Digital Objects
Editorial Staff Interface
Copyright component
Document Identification Interface
Document Identification Subsystem
End-users Authorization
E-commerce component
End-users Registration & Authentication Subsystem
POS-Gateway
End-users Search Interface

Figure 2. Full Data Provider components

by series and other document aggregations.

- *End-Users Authorization*, the module that checks end-users access to digital objects.
- *E-commerce module*, the component for on-line payments management, interfaced to the POS Gateway subsystem.
- *Copyright module*, the component that implements rights management, fees computation and accounting.

A Service Provider can support many publishers. For instance many independent catalogues can be implemented. In this case orders relating to items from different publishers catalogues can be submitted. The modules which interface the digital object identifier resolution subsystem, the POS Gateway and the copyright accounting system are unique.

The Support Services are:

- *Document Identification Subsystem*, which supports the assignment and resolution of persistent-unique identifiers for digital objects, conformant to URN syntax.
- *POS-Gateway* to credit card circuits, a system to manage on-line payment transactions, usually hosted by a bank.
- *OpenURL Resolution Service*, a centralized subsystem which implements the knowledge base and the resolution mechanisms to provide proper (i.e. context sensitive) links related to a catalogue item activated by an end-user.
- *Authors Identification&Registration Subsystems*, which support the registration and unique identification of the subjects, entitled to copyright royalties. The related database contains personal data, including the bank coordinates for the automatic generation of credit transactions.
- *Legal Deposit Service*, a centralized subsystem responsible for the legal deposit of electronic documents at the national level.
- *Certification Authority*, which issues, revokes and renews digital certificates to support digital signatures and timestamps.

In Fig. 1 a possible scenery showing the main interactions is depicted:

5. Metadata

As widely defined in literature, metadata is data about data and "includes information about the context of data and the content of data and the control of or over data" (Pasquinelli 1997). Metadata permeates the entire life cycle of digital publications. As the functional and technological models show, DAFNE is totally plunged into metadata, which supports the retrieval, management, long term archival, use and reuse of its digital objects.

One of the basic prerequisites of the project was to

UNIPRE Is a University Press, which publishes several kinds of documents included pre-prints for free consultation and learning material, restricted to campus network. Registered teachers and researchers can self-archive their papers. The other publications are subject to payment and subscription policies.
UNIPRE implements all the modules defined for a Data Provider except the review subsystem.

PUB-1 Is a commercial publisher whose catalogue contains digital journals and monographs. PUB-1 exploits a peer-review subsystem. IP address check is the control access policy for journals, whereas on-line payment is requested for subscriptions and pay-per-print.
PUB-1 implements all the modules defined for a Data Provider.
PUB-2 and PUB-3 Are two commercial publishers which support the direct submission of papers by authors. They have delegated their catalogue interface and on-line payment to a service provider.

PUB-2, PUB-3 implement only the basic Data Provider components.

SP-A Is a Service Provider which implements consultation, access control and on-line payment for PUB-2 and PUB-3.
Metadata is captured from the publishers by PMH.

SP-B Under the conditions of a special agreement, this Service Provider integrates the collections from UNIPRE, PUB-1, PUB-2 and PUB-3 by using PMH. Moreover, SP-B adds to the items an OpenURL link for cross-referencing with other resources, resolved by a dedicated subsystem (OURS).

OURS The OpenURL Resolution Service.

DIS The Document Identification Subsystem.

POS-GW The POS-Gateway.

AIRS The Authors Identification&Registration Subsystem.

LDS The Legal Deposit Service.

CA A Certification Authority.

identify a metadata set without creating anything new, but just by using metadata vocabularies already defined in related domains. In order to define an appropriate metadata profile, research and analysis of existing metadata sets have been conducted. As far as interoperability issues are concerned, this approach promotes the integration between the digital library and the electronic publishing districts. Moreover, we must not forget the great amount of resources, which have been invested in the definition of common metadata, sets in the digital objects realm. DCMI is part of this effort. Therefore it is from Dublin Core (DC) that DAFNE metadata analysis has begun to go on with other specifications.

The Open Archival Information System (OAIS 1999) has guided the design of the DAFNE information model, helping to select the metadata sets relevant in the digital publishing pipeline.

OAIS defines a conceptual framework for generic archival systems and an information model that provides a systematic view of metadata gathered in several categories. This model is "a high-level description of the types of information generated by and managed within the functional components of a complete archiving system" (OCLC/RLG 2001). The archive information package includes four types of information objects: Content Information, Preservation Description Information, Packaging Information and Descriptive Information. Content Information consists of the digital object and its associated representation information, that is the technical metadata, which supports the correct rendering and interpretation of the associated digital object (i.e. bit stream). The Preservation Description Information includes four types of metadata: Provenance (about the origin and preservation), Reference (includes the identifiers associated to a digital object), Fixity (data related to the authentication mechanisms) and Context Information (metadata about relations with other objects). Descriptive Information supports search and retrieve of the archive contents.

The first phase of DAFNE focuses on descriptive and digital rights management metadata. Technical, structural and long term preservation metadata will be analysed in depth in a further phase of the project. Concerning descriptive metadata, Publishing Requirements for Industry Standard Metadata (PRISM) is the reference metadata set in DAFNE, in particular for the design of data providers' repositories. PRISM is a specification promoted by a group of publishers and vendors, joined under the IDEAlliance initiative, which defines a set of elements useful for "interoperable content description, interchange and reuse in both traditional and electronic publishing contexts" (PRISM 2002, p. 1). This specification recommends the use of existing standards such as Dublin Core, XML, RDF and various ISO standards for location, languages and date/time format. PRISM descriptions are expressed as standalone XML documents or PRISM metadata can be embedded inline within the document. The XML representation is totally compatible with OAI-PMH and its capability to transport multiple metadata sets.

PRISM elements are gathered by six functional groups: General Purpose Descriptive Elements, Provenance, Timestamps, Subject Description, Rights and Permissions, Resource Relationships. The last group is proposed in place of "dc:relation" to provide more peculiar definitions about relations among the resources. A series of controlled vocabularies (PCV) enrich PRISM elements: Rights and Usage (prl:usage) to specify resource usages, Resource Type to define the style of presentation in the resource's content (dc:type), Resource Category to specify the intellectual genre (prism:category). PCVs support a further functional description for basic use of documents. The list of terms in Resource Type Vocabulary and Resource Category vocabularies derive from third parties thesauri such as Arts and Architecture Thesaurus (AAT), WORLDNET and NewsML (PRISM 2002, p. 52-55).

In the digital environment the creation of intellectual content is tightly linked to its management and commercial use and reuse, where "commercial" is used in its broadest sense. Commerce for DAFNE includes trade with the consumers and "cultural transactions" with public libraries or other not for profit institutions (e.g. universities, schools, etc.). In this realm, any intellectual content that transforms into a digital object can be related to many actors claiming rights on it. Indecs project (<indecs>), which deals with identifiers and metadata supporting intellectual property rights, asserts that "while an apple bought at a market stall is a single physical entity owned entirely by one person, a single digital audiovisual creation may contain hundreds or even thousands of separate pieces of intellectual property." (Rust & Bide 2000, p. 4). This clearly identifies how complex digital objects and their Intellectual Property Rights (IPR) management are.

As remarked by Iannella (Iannella 2001), Digital Rights Management (DRM) is the "digital management of the rights", be they the rights in a physical manifestation (e.g. a book) or be they the rights in a digital manifestation (e.g. an e-book) of a work. The first generation of DRM systems aimed to prevent unauthorized copies of digital objects by security controls and encryption techniques. Second generation DRMs include the description, identification, trading, monitoring and tracking of rights and the relations with the entitled subjects.

Waiting for the consolidation of one of the emerging metadata standards for DRM, such as the Xtensible rights Markup Language (XrML) and the Open Digital Rights Language (ODRL), PRISM adopted a pragmatic approach. The Rights and Permission functional group specifies "a small set of elements that would encode the most common rights information to serve as an interim measure for interoperable exchange of rights information" (PRISM 2002, p. 14). This set is too limited to support the basic level of DRM needed in DAFNE, so other relevant models have been investigated. Starting with <indecs> project, DAFNE has selected ODRL as the reference model for rights management metadata. Basically ODRL is a language and a vocabulary to express terms and conditions over assets. An asset is any digital or physical intellectual content, over which the author or other parties can claim rights. "ODRL complements existing analogue rights management standards by providing digital equivalent and supports an expandable range of new services that can be afforded by the digital nature of the assets in the Web environment." (Iannella 2001a, p. 1). ODRL defines a model and a set of semantics for managing the rights holders and the permissible usages of asset manifestations. Thanks to ODRL any digital asset can have its digital rights management information linked to it.

According to ODRL specified semantics, some of its data dictionary elements can integrate PRISM helping to define DAFNE metadata reference set, mainly:

- Usage Permission elements (*display*, *print*, *play*, *execute*)
- Requirement elements (*payment* which contains amount and currency and tax percent and code, *prepay*, *postpay*, *peruse*)
- Rights Holder elements (*percentage*, *fixedamount*)

Permissions are linked to parties and assets through an *agreement* element. Requirements are associated to permissions. Rights Holder elements are included within party elements, the subjects entitled to royalties. Both assets and parties must have a unique identifier, which is expressed by the *context* element and its sub-elements.

DAFNE Metadata Profile is on the way to be finalized. The Appendix outlines the core elements set under specification, derived from DC, PRISM and ODRL. The listed identifiers follow the XML Namespaces syntax where "dc" and "prism" are the namespaces that include Dublin Core and PRISM elements, whereas "dafne" should be the name of a new metadata vocabulary to be defined. DAFNE repository implementation (i.e. the Storage component) will put in relation Content Information, Metadata, Parties claiming rights on them, Permissions, Usage Constraints and Requirements. XML syntax is used to define simple and complex elements relations.

6. Conclusions

DAFNE is a research project that concerns scientific and scholarly production in the human and social sciences. Aiming to define the prototype of the national infrastructure for electronic publishing in Italy a full analysis of the publishing pipeline was performed. This has led to the definition of the organizational, legal, technical and business models. This paper has outlined the functional model, the reference architecture and the core set of metadata. DAFNE deals with the academic realms within which, its functional model tries to create a co-existence between the traditional business publishing approach and the institutional, e-print archive repositories, which are more and more diffused in the scholarly international context. Therefore the framework proposed by the Open Archives Initiative, based on Data and Services Providers and on the Protocol for Metadata Harvesting, has revealed itself to be very suitable to design the system architecture. An OpenURL based resolution system is the most suitable component to support resources cross-referencing in order to assure the integration with the digital libraries context.

Concerning metadata, DAFNE has made evident how much metadata permeates the entire publishing pipeline. Descriptive, relation and rights management are the functional groups of primarily required elements to implement the prototype. Technical, long term preservation and secure (i.e. digital digests and signatures) metadata will be added in the advanced version. The project has been investigating relevant existing metadata standards to define DAFNE metadata set by aggregation of and reference to these vocabularies. Dublin Core and PRISM are the main reference standards for descriptive and relational metadata. PRISM rights and permission metadata have been extended with some elements derived from ODRL model. They make up a minimal set for rights management in order to experiment on-line payment and rights computation functions. By the end of the year a full specification of DAFNE metadata set will be issued. Next year the DAFNE prototype exploitation will allow the consolidation of the final metadata specification.

References

ArXiv. <http://arXiv.org/>

Berkeley Electronic Press. <http://www.bpress.com>

CogPrints. Cognitive Sciences Eprint Archive.
< http://cogprints.soton.ac.uk/>

DOI. Digital Object Identifier. <http://www.doi.org/>

Eprints.org. University of Southampton.
<http://www.eprints.org>

Ginsparg, P., 2001. Creating a Global Knowledge Network. Conference on Electronic Publishing in Science. Paris, 20 February 2001.
<http://arXiv.org/blurb7pg01unesco.html>

Harnad, S., 1999. Free at Last: The Future of Peer-Reviewed Journals. D-Lib Magazine, 5(12).
<http://www.dlib.org/dlib/december99/12harnad.htm>

Iannella, R., 2001. Digital rights management (DRM) architectures.D-Lib Magazine, 7 (6).
<http://www.dlib.org/dlib/june01/iannella/06iannella.html>

Iannella, R., 2001a. Open Digital Rights Language (ODRL) Version 1.0.
<http://odrl.net/1.0/ODRL-10.pdf>

IETF RFC 2141. URN Syntax.
<http://www.ietf.org/rfc/rfc2141.txt>

IETF RFC 2396. Uniform Resource Identifiers (URI): Generic Syntax. <http://www.ietf.org/rfc/rfc2396.txt>

<indecs>, Interoperability of data in e-commerce systems. <http://www.indecs.org>

NCSTRL. Networked Computer Science Technical Reference Library.<http://www.ncstrl.org>

OAI. Open Archives Initiative.
<http://www.openarchives.org>

OAIS, 1999. Reference Model for an Open Archival Information System (OAIS): CCSDS 650.0-R-1. Red Book. Issue 1.
<http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html>

OCLC/RLG, 2001. Preservation Metadata for Digital Objects: a Review of the State of the ART. A White paper by the OCLC/RLG Working Group on Preservation Metadata. < http://www.oclc.org/digital preservation/presmeta_wp.pdf >

Pasqui, V., 2001. DAFNE Deliverable D12: Definizione dell'architettura logica a macroblocchi del sistema complessivo sulla base del flusso dei servizi e delle transazioni da sviluppare nella architettura tecnologica. Draft vers. 1. November 2001. Restricted report.

Pasquinelli, A., 1997. Information technology directions in libraries: a Sun Microsystems white paper: August 1997. <http://www.sun.com/products-n-solutions/edu/libraries/libtechdirection.html>

PRISM 2002, PRISM:Publishing Requirements for Industry Standard Metadata, Feb. 2002.
<http://www.prismstandard.org/techdev/primspec11.asp>

Rust, G. and Bide, M., 2000. The <indecs> metadata framework : Principles, models and data dictionary
<http://www.indecs.org>

SPARC, 2002. The Case for Institutional repositories: A SPARC Position paper. Prepared by Raym Crow. SPARC. Release 1.0. <http://www.arl.org/sparc/home>

ScholarOne. http://www.ScholarOne.com

Van de Sompel, H. and Beit-Arie, O. Open Linking in the Scholarly Information Environment Using the OpenURL Framework. D-Lib Magazine. 7(3).
<http://www.dlib.org/dlib/march01/vandesompel/03vandesompel.html>

Van de Sompel, H. and Lagoze, C., 2001. The Open Archives Initiative: Building a low-barrier interoperability framework. JCDL 2001.
<http://www.openarchives.org/documents/oai.pdf>

XpressTrack. http://xpresstrack.com

XrML. Extensible Rights Markup Language. ContentGuard, Inc. < http://www.xrml.org/>

Appendix - DAFNE Metadata Element Set (provisional)

The first six functional groups are derived from PRISM specification and the same purpose and meaning is reported for each element.

<u>*General Purpose Descriptive Elements*</u>

dc:identifier Identifier(s) for the resource.
dc:title The name by which the resource is known.
dc:creator The primary creator(s) of the intellectual content of the resource.
dc:contributor Additional contributors to the creation or publication of the resource.
dc:language The principal language of the resource.
dc:description A description of the resource.
dc:format The file format of the resource.
dc:type The style of presentation of the resource's content, such as image vs. sidebar.
prism:category The genre of the resource, such as election results vs. biographies.

<u>*Elements for Provenance Information*</u>

dc:publisher An identifier for the supplier of the resource.
prism:distributor An identifier for the distributor of the resource.
dc:source An identifier for source material for the resource.

<u>*Elements for Time and Date Information*</u>

prism:creationTime Date and time the identified resource was first created.

prism:modificationTime Date and time the resource was last modified.
prism:publicationTime Date and time when the resource is released to the public.
prism:releaseTime Earliest date and time when the resource may be distributed.
prism:receptionTime Date and time when the resource was received on current system.

Subject Descriptions

dc:coverage Indicates geographic locations or periods of time that are subjects of theresource.
dc:subject The subject of the resource.
dc:description Prose description of the content of the resource.
prism:event An event referred to in or described by the resource.
prism:location A location referred to in or described by the resource.
prism:person A person referred to in or described by the resource.
prism:organization An organization referred to in or described by the resource.

Resource Relationships

prism:isPartOf The described resource is a physical or logical part of the referenced resource.
prism:hasPart The described resource includes the referenced resource either physically or logically.
prism:isVersionOf The described resource is a version, edition, or adaptation of the referenced resource.
prism:hasVersion The described resource has a version, edition, or adaptation, namely, the referenced resource.
prism:isFormatOf The described resource is the same intellectual content of the referenced resource, but presented in another format.
prism:isTranslationOf The described resource is a human-language translation of the referenced resource.
prism:hasTranslation The described resource has been translated into an alternative human-language. The translated version is the referenced resource.

Rights and Permissions

dc:rights Container element for specific rights data
prism:copyright A copyright statement for this resource.
prism:expirationTime Time at which the right to reuse expires.
prism:releaseTime Time as which the right to reuse a resource begins, and the resource may be published.
prism:rightsAgent Name, and possibly contact information, for the agency in order to establish contacts and to determine reuse conditions if none specified in the description are applicable.
prl:geography Specifies geographic restrictions.
prl:industry Specifies restrictions on the industry in which the resource may be reused.
prl:usage Specifies ways that the resource may be reused.

DAFNE elements

dafne:usagePermission Specifies the available access modalities (e.g. display, print).
*dafne:payment C*ontains the following four sub-elements to express payment information.
dafne:amount Specifies the cost.
dafne:currency Specifies the currency (e.g. €).
dafne:taxpercent Specifies the tax percentage (between 0 and 100).
dafne:code the tax code (e.g. IVA or VAT).
dafne:paymentRequirements Specifies the payment modalities requested (e.g. prepay, post pay, peruse).
dafne:rightsHolders Contains the three following subelements to specify royalties accounting data.
dafne:holderId Specifies the unique identifier of a subject entitled to rights fees.
dafne:percentage Specifies the percentage of dafne:amount to be used to compute the fee due.
dafne:fixedamount Specifies a fixed amount due for rights.

If different modalities of access exist for the same publication *dafne:paymentRequirements*, *dafne: rightsHolders* and *dafne:payment* are related to different instances of *usagePermission*.