Cooperative Digital Asset Management in the Scientific Field: Strategies, Policies, Interoperability and Persistent Identifiers

Maurizio Lancia[1], Roberto Puccinelli[1], Massimiliano Saccone[2], Marco Spasiano[1], and Luciana Trufelli[2]

[1] Office for Information Systems, Italian National Research Council, Piazzale Aldo Moro, 7 – 00185 Rome, Italy
{maurizio.lancia,roberto.puccinelli, massimiliano.saccone,marco.spasiano}@cnr.it
[2] Central Library "G. Marconi", Italian National Research Council, Piazzale Aldo Moro, 7 – 00185 Rome, Italy
{luciana.trufelli}@cnr.it

Abstract. In this paper we present a series of activities carried out within the National Research Council of Italy (CNR) and aimed at the development of a unique, certified and open archive of CNR's digital research products. Starting from the description of CNR's distributed library system, we then briefly talk about CNR's involvement in OA initiatives and the role played by CNR's Information System Office in providing technological tools for digital asset management. Afterwards we try to point out some criticalities of OA archives. We then talk about the solution we propose for the development of a unique, certified and open archive using a cooperative approach that takes into account previous experiences, existing repositories, policy and organizational issues. We also present the processes we designed for content ingestion and validation and the strategies for persistent identification. We finally illustrate the technical solutions we have developed as prototype proposals for the community.

1 Introduction

The use of digital libraries in research institutions presents some peculiarities. Besides the necessity of effective digital resource management, this type of organizations need to design and deploy effective methods and tools to tackle complex problems of research product quality assessment and overall performance measurement.

Scientific production statistics and bibliometric indicators are widely leveraged in research evaluation processes. Although criticisms have been raised by the scientific community especially against the latter, the Italian legislation has recently confirmed their use [1].

In order to address the challenges of a wide and diffuse access to scientific production and to improve the reliability of quantitative and qualitative measures of

M. Agosti et al. (Eds.): IRCDL 2011, CCIS 249, pp. 144–159, 2011.

research production, this paper proposes an approach based on the cooperation of all the stakeholders working in the scientific environment. We give particular emphasis to the peculiarities of OA repositories, which, in our opinion, should be the back-bone of the scientific community knowledge base. We think that this approach could favor the innovation of the system of scientific libraries in Italy.

The increased accessibility of digital resources provided by the web technologies and the development of the information systems demand an increased effort for resource certification and contextualization (i.e. keeping the relations between digital resources and their respective contexts). The quantity and complexity of the problems posed by a multi-institutional environment, require organizational and political measures in first place. Technical solutions will follow, provided that all the adopted standards and technologies will have to be open.

Within CNR[1] Central Administration we are working to design and implement prototype solutions to the problems described above, in order to provide practical inputs to the discussion both at the National and International level. In order to achieve our goal, we look for cooperation and inputs from all CNR's scientific network and from external organizations playing key roles in the research field. The results of our work may represent a starting point for future cooperative developments and are made available to the scientific community in order to ensure the maximum consensus about the choices to be made.

2 CNR's Environment

2.1 The Organization of CNR's Libraries

CNR's library infrastructure reflects CNR's organization, featuring a Central Administration in Rome and a Scientific network made up of thematic institutes distributed all over the national territory. A significant percentage of CNR's institutes are hosted inside territorial Research Areas, which provide common services thus increasing efficiency.

CNR's library system features a hierarchical and distributed organization, which includes a Central Library (Biblioteca Centrale), Research Area Libraries (Biblioteche delle Aree di Ricerca), Institute Libraries (about 80). It can provide a wide range of services to the entire scientific community and has recently adopted new organizational measures in order to increase the coordination of its different branches and improve the quality of the services provided to the internal scientific community. This effort has already produced some results in terms of process rationalization and digital resource sharing. The medium term objective is to complete the integration between CNR's libraries and to provide new added value services both to the internal and external scientific community.

[1] The National Research Council of Italy is a public body whose mission is to carry out and promote research activities in the main sectors of knowledge and to disseminate the results and their applications in order to foster the scientific, technological, economic and social development of the Country (Legislative decree n. 127, 4th of June 2003).

2.2 CNR's Libraries and OAI Initiatives

In the recent years CNR's libraries have been involved in numerous Open Access initiatives, in the belief that these can help tackling some problems like the decrease of available economic resources; the continuous evolution of technological tools; the issues regarding long term digital preservation; the great amount of inadequately identified and certified digital resources; the unfavorable price policies of some commercial services.

Open Access initiatives are by necessity implemented through the tight cooperation between the CNR's libraries and the CNR's organizational units which provide ICT services. A first outcome of this cooperation is the development of some OA repositories for CNR's research products. Those repositories can contain several types of research products (publications, data sets, patents, grey literature, etc).

At present, CNR's libraries are contributing to the implementation of an integrated system connecting all CNR's OA research product archives and able to overcome some of the problems described in the previous sections (e.g. interoperability with the main information and knowledge management systems, transparent access to the entire CNR's scientific production, quality issues, persistent and unique identification of digital resources, unified digital preservation policies, etc.).

2.3 The Role of CNR's Information Systems

The management of digital archives within CNR's Central Administration is carried out through a tight cooperation between the Central Library and the Information System Office. This collaboration has been recently boosted in order to achieve a higher integration between the digital libraries and the other components of CNR's information system, which include the ERP system, the business intelligence platform, web portals and informational web sites.

Figure 1 shows the new integrated architecture of CNR's information systems. At the bottom there are the so called Enterprise Information Systems (EIS), including the ERP data bases, administrative document bases and the scientific digital libraries. At the upper level we find the applications which access and manage the data kept at the EIS level. At the top of this architecture there is the level of access portals and informational web sites, which provide different interfaces to the different types of users. Orthogonal to this layered architecture is a set of infrastructural services for unified authentication and authorization, digital signature, certified e-mail, persistent identifier assignment and resolution, etc.

The integration of scientific digital libraries with the other information systems can provide many benefits in terms of increased accessibility and can also allow for a more effective evaluation process. Actually, the possibility of linking the research products with administrative information regarding personnel, projects and financial accounting is a pre-requisite for evaluating the performance at different levels (single researchers, research groups, institutes, departments, organization).

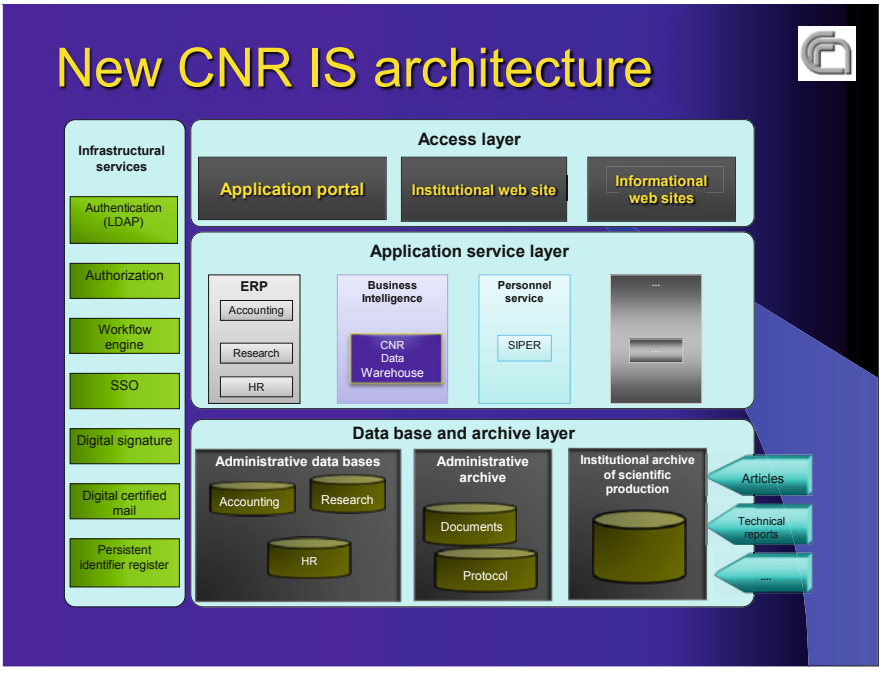

Fig. 1. The new CNR's Information System Architecture

In order to achieve the desired integration of digital libraries within CNR's information system it is crucial to take into account the existing situation, which features multiple and distributed repositories. We think that the best approach is to federate the archives by introducing a series of common policies which define metadata sets, authority files, data interchange protocols, document ingestion and management processes. Obviously this change have to be introduced by gathering inputs and consensus from CNR's scientific network. Needless to say that all the technical solutions have to be based on open standards and open source platforms. With respect to this point, it is worth mentioning that CNR's information systems have started more than ten years ago a process of migration towards open source solutions, shifting the investments from commercial products to highly skilled personnel. A software factory has been established some years ago inside the Information System Office, which develops and maintain the majority of the components of CNR's information system. This, in our opinion, helps in tailoring solutions to the specific needs of our organization.

3 Criticalities of OA Archives

In recent years the main scientific institutions and the stakeholders of the scientific publishing sector have spent a significant effort in enhancing and improving

the quality of OA initiatives and of the annexed services. Among the many initiatives in this field we would like to mention: the development of aggregated services for OA repositories, managed by numerous institutional and disciplinary service providers; the significant stimulus provided at the international level by the European Commission[2] and by the European Research Council (ERC) [2,3] and, at the Italian national level, by the Italian Conference of University Rectors (Conferenza dei Rettori delle Universit Italiane – CRUI) [4] to the deployment of organizational and technological infrastructures for the collection of and access to research papers funded by public institutions and for their deposit in institutional and/or disciplinary archives[3].

Although much effort has been spent to overcome the criticalities of OA archives, some issues still need to be addressed. Among these we would like to mention:

- interoperability at the policy, organizational and technological level, especially with respect to national and international administrative data management systems and to some of the most important public and commercial bibliographic data bases (e.g. OECD, European Commision data banks, Anagrafe Nazionale dei Prodotti della Ricerca, Web of science, Scopus, etc);
- extensive coverage of the institutional scientific production;
- quality control as regards the scientific contents of the deposited items[4];
- unique and persistent identification of works and authors;
- quality control and certification of the descriptive metadata (bibliographic and authority control).
- version history tracking.

In general, some of the main problems affecting OA repositories are caused by or connected to the fragmentation of OA initiatives, which, at the Italian national level, is combined with the lack of common strategies and policies. This sometimes leads to duplicated efforts [5]. All those problems hinder the use of

[2] European Commission, Open Access Pilot in FP7, European Communities Publication Office, August 2008. The 12th of December 2008 the European Commission launched the "Open Access Pilot in FP7", with the aim of defining innovative strategies and tools that could guarantee open access to data and publications coming from EC funded research projects. The EC provided recommendations to the R&D institutions about the adoption of mandatory policies regarding the deposit in disciplinary or institutional OA archives. *A Digital Agenda for Europe – Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions*, Brussels, 26.8.2010 - COM(2010) 245 final/2. `http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52010DC0245(01):EN:NOT`

[3] It is also worth mentioning, among the EU projects in which CNR is involved: DRIVER – Digital Repository Infrastructure Vision for European Research – `http://www.driver-repository.eu/`; OpenAIRE – Open Access Infrastructure Research for Europe – `http://www.openaire.eu/`

[4] The institutional repositories are not always able to certify the quality of unpublished works, neither are they designed for easy integration with open peer-review services.

institutional repositories as reliable and certified tools for statistics collection and research evaluation [6].

In our opinion, the main causes of the problems listed above are:

- insufficient allocation of economic and human resources, which raise sustainability issues;
- lack of mandatary policies for the deposit of the scientific production in OA institutional repositories;
- problems in carrying out an effective advocacy of the scientific and technical communities of reference;
- lack of cooperative systems for metadata control and certification; many OA repositories, even among those playing the role of institutional repositories, do not pursue the objective of providing bibliographic and authority metadata controls; this hinders or makes more difficult the retrieval and aggregation of research products and metadata (e.g. it is very difficult to retrieve the works of a single researcher if his/her name is recorded in different ways for different articles; in this case an author unique identifier and an author authority file would be of help). At present neither metadata standards[5] nor commonly used digital library software favor the adoption of shared rules and conventions which would improve the quality of OA repositories [7,8,9].
- Lack of peer-review procedures for unpublished works [7].
- Lack of agreement between the sytakeholders about policies and strategies for the effective design, implementation and management of OA repositories.
- Lack of added value services like: helpdesks tailored on the final users needs, effective DRM strategies.

Despite the criticalities described above, many information professionals acknowledge the relevance of the Institutional Archives (IA), among all OA initiatives, as a strategic channel for an open, certified and immediate access to the products of universities and research institutions.

According to the definition provided by Clifford A. Lynch [10], the Institutional Archive is as the set of services that a research organization offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members. The types of managed digital materials must be indicated in an Open Access mandatory management policy, issued by the organization itself.

The Institutional Archive is primarily an organizational commitment aimed at assuring:

- the digital preservation of the materials mentioned above and of the related metadata;
- the certified access to digital resources and metadata.

The responsibilities regarding the management of this types of services may be assigned to different organizational units within an institution, provided that all

[5] The adoption of commonly agreed metadata sets (Dublin Core, MARC, etc.) and the definition of highly refined mapping systems are not "per se" sufficient to guarantee the certified exchange and reuse of data and metadata.

those units cooperate within a coordinated and unified context, characterized by the collaboration between the stakeholders which take part in the decision processes: librarians, ICT professionals, archivists, user communities and policy makers.

4 An Open Archive for Research Products

4.1 Institutional Objectives of CNR's OA Repository and Metadata Quality Definition

Within the framework of the new integrated architecture of CNR's information systems, a preliminary study have been carried out in order to verify the feasibility and the implications of an OA repository, which should collect, archive, catalogue, identify, certify and deliver CNR's scientific production, with the aim of providing: a strategic support to CNR's institutional mission, an increased visibility of CNR's production, a quality certification of data and metadata, a certification tool able to support research evaluation processes [11].

In particular, great attention has been paid to descriptive and administrative metadata quality certification in order to guarantee:

- unique and persistent bibliographic identification of the digital items and of their different versions as well as of the intellectual rights and authors' affiliations;
- visibility and impact maximization of the deposited works;
- interoperability with internal catalogues and administrative data bases, bibliographic data bases, national and international statistics;
- the reuse of metadata to support research evaluation processes;
- organizational-technical interoperability with Italian and international service providers.

The chosen approach is to define metadata quality by specifying their functions and usage contexts (functional approach): "Quality is very difficult to define. A definition that can be used in the context of metadata is: 'high quality metadata supports the functional requirements of the information and documentation system it is designed to support', which can be summarized as 'quality is about fitness for purpose" [8]. "Quality metadata reflect the degree to which the metadata in question perform the core bibliographic and management functions of discovery, use, provenance, currency, authentication, and administration. The functional perspective is closely tied to the criteria and measurements used for assessing metadata quality." In this case, "accuracy, completeness, and consistency are the basic criteria used in measuring metadata quality" [9].

In order to achieve accuracy, completeness and consistency of descriptive and administrative metadata, a content and metadata management system has been designed based on the OAIS conceptual model [12] and on a hierarchical and

distributed paradigm, which involves all CNR library system and the technical units of the Central Administration that provide knowledge management services. The goal is to tackle the issues described above, regarding accurate metadata quality control, persistent and unique identification, versioning, metadata contextualization and linkage with other CNR's administrative data bases (personnel, projects, etc.), interoperability, etc.

4.2 Workflows for Digital Asset Ingestion and Validation

In this section we describe the process we designed for ingesting and verifying digital assets in our repository, which is summarized in Figure 2. In our opinion, this process suits large and distributed organizations with a distributed library system. The process includes three main workflows: archiving (which can be triggered by the author himself), first level verification (performed by institute and research area libraries – "local libraries" in what follows) and second level verification (which involves both the local libraries and the Central Library) [13].

In the self-archiving phase authors upload their documents along with a signed declaration in which they state their authorship rights and fill in a form for descriptive metadata. They can also ask for a peer review of the contents, which will be carried out by appointed reviewers. The same steps can be performed by the local libraries on behalf of the authors.

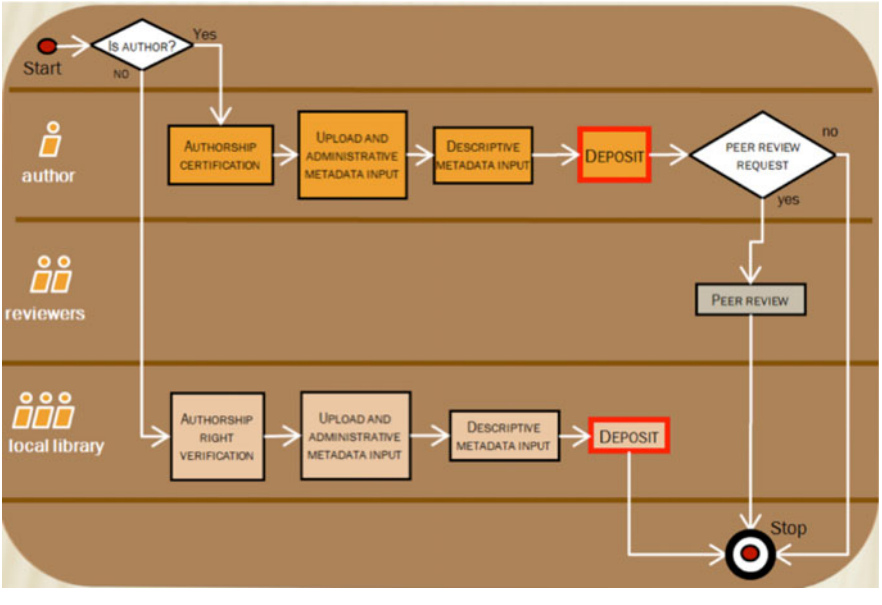

Fig. 2. Archiving Workflow

In the first level verification phase (Figure 3) the local libraries check metadata for completeness and correctness. They can interact with authors in case modifications are needed.

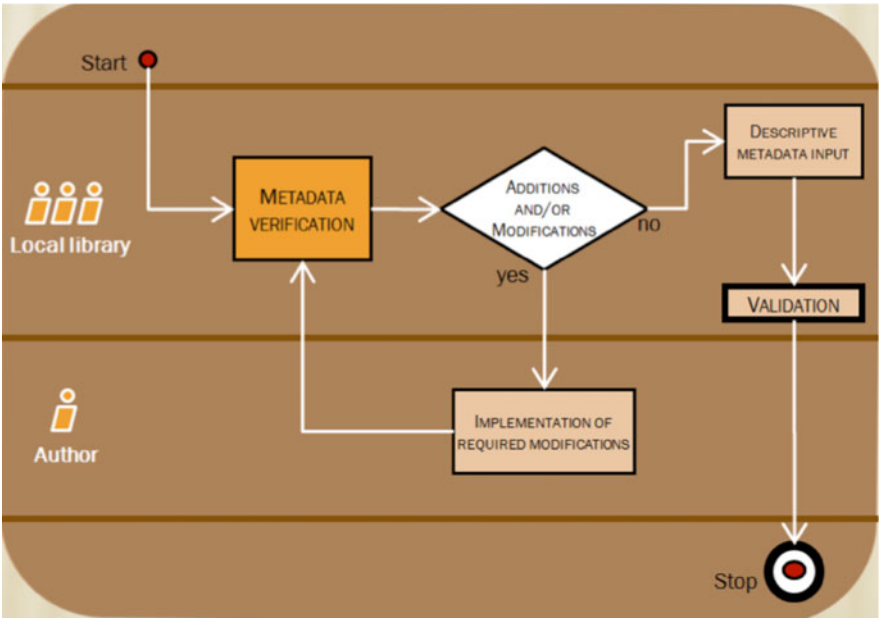

Fig. 3. First level verification

In the second level verification phase (Figure 4) the Central Library performs a second check on metadata, interacting with the authors and the local library for any required clarification/modification. The final part of this workflow includes filling in some metadata fields that are specific to the Central Library and performing the final steps for validation, certification and publication.

4.3 Metadata Certification: Policies, Organization and Tools

As regards authentication, ingestion, metadata verification/integration, final validation and publishing phases, the system allows for a constant feed-back between users and system managers, by means of an helpdesk service, which support the user in all the steps of the deposit procedure.

The system features three different workflows:

a) deposit accomplished by one of the authors – i.e. self archiving;
b) deposit accomplished by the local library of the institute to which the authors belong;

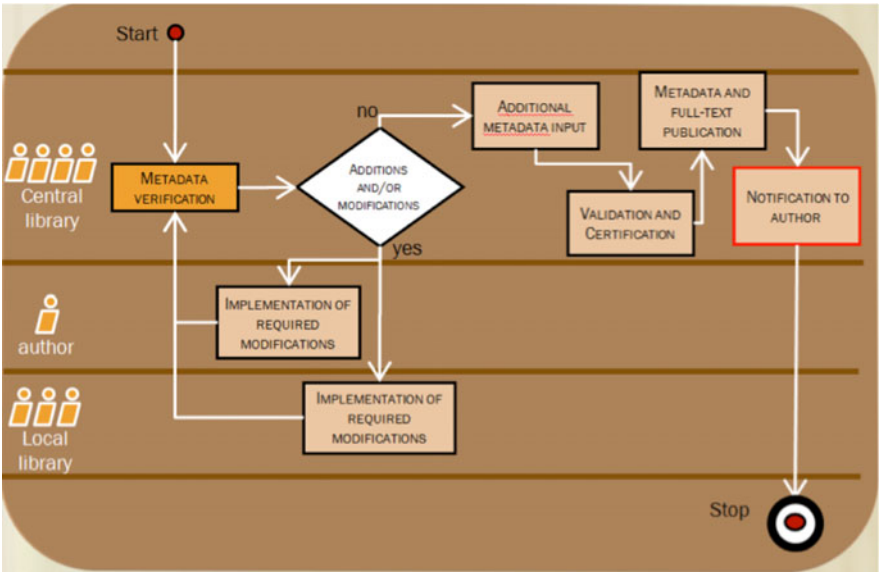

Fig. 4. Second level verification

c) deposit accomplished by the Central Library (generally for the Organizational Units of the Central Administration or for those without a local library).

In all three cases the digital material undergo a thorough review and validation process (bibliographic control) that can be performed at various levels (Institute, Research Area, Central Library).

We have also planned to trigger a cooperative development of authority control tools as regards, for instance, intellectual responsibility (authors, editors, etc.), periodical journals and series titles (linked to the ISSN International Centre data base), etc.

4.4 Interoperability with External Information and Knowledge Management Systems: A Cooperative Approach

From a technical view point, though OAI-PMH protocol allows – as a container – to embed different types of metadata formats, it limits interoperability to the elements of the unqualified Dublin Core schema. If this on one hand simplifies metadata management on the provider's side, on the other hand it increases the "noise" in research products evaluation [7, 9].

In order to guarantee an actual interoperability at the service provider level, it would be necessary to develop cooperative bibliographic and authority control systems for the main shared bibliographic elements (author, contributor, title,

subject, type, relation, etc.) and adopt, in a cooperative fashion, persistent identification systems for intellectual responsibilities, types of resources and different versions of the same objects.

In our opinion, in order to pursue metadata quality, the cooperative approach should take into account functional requisites of the national and international information systems to which they provide support, thus assuring:

- persistent and unique identification of the deposited works and of their different versions as well as of the authors' affiliations;
- visibility and impact maximization for the deposited items;
- interoperability between bibliographic, statistical and administrative data bases, both at the national and international level;
- technical and organizational interoperability between Italian service providers.

4.5 Persistent Identification

Persistent identifiers (PI) are alphanumeric codes that allow to uniquely and persistently identify digital resources. Besides, they are also a pre-requisite for the certification of authenticity, provenance and managing rights.

There are currently several PI standards, which present different features and suit the needs of different communities. It is highly probable that the situation will not change in the future, i.e. there won't be a unique PI technology or only one central digital resource registry for the entire world [13]. Within this section we provide a short list of the most diffused PI standards.

The **Document Object Identifier** system (DOI [14]) is a business-oriented solution widely adopted by the publishing industry, which provides administrative tools and a Digital Right Management System (DRM).

Archival Resource Key (ARK [15,16]) is an URL-based persistent identification standard, which provides peculiar functionalities that are not featured by the other PI schemata, e.g., the capability of separating the univocal identifier assigned to a resource from the potentially multiple addresses that may act as a proxy to the final resource.

The **Handle System** [17, 18, 19] is a technology specification for assigning, managing, and resolving persistent identifiers for digital objects and other resources on the Internet. The protocols specified enable a distributed computer system to store identifiers (names, or handles) of digital resources and resolve those handles into the information necessary to locate, access, and otherwise make use of the resources.

Finally, the **Persistent URL** (PURL [20]) is simply a redirect-table of URLs and it's up to the system-manager to implement policies for authenticity, rights, trustability, while the **Library of Congress Control Number** (LCCN [21]) is a persistent identifier system with an associated permanent URL service (the LCCN permanent service), which is similar to PURL but with a reliable policy regarding identifier trustability and stability.

In the Library domain the National Bibliography Number has been defined and is currently promoted by the CENL. IETF RFC 3188 [22, 23] describes the

use of URNs to represent NBN persistent identifiers. The NBN standard assumes that the national libraries are responsible for the national name registers. The first implementations of NBN registers in Europe are available at the German and Swedish national libraries.

We have chosen to use NBN codes to identify the resources used in our testing activities, because this is an open standard, whose URN-based format well maps a distributed and hierarchical organization of digital libraries like CNR's one. To this end in the recent years we have participated to some initiatives regarding persistent identification, in order to identify and implement appropriate solutions for the research environment.

The NBN syntax is quite straightforward. According to RFC 3188 an URN-NBN PI which uses ISO 3166 codes to identify countries has the general form:

URN:NBN:<ISO 3166 country code>:
{<sub-namespace code>}–<assigned NBN string>

where the assigned NBN string can be any sequence of alphanumeric characters. This syntax allows to split the national domains in any number of sub-domains, hierarchically organized.

The effectiveness of any type of PI is more a policy issue than a technical one. In the case of NBN, policies are mainly established at the national level and require the cooperation and the agreement of all the main actors in the field of digital preservation [24]. In Italy an initiative for establishing a national infrastructure for NBN assignment and resolution is being carried out, involving the National Library of Florence (BNCF), the National Library of Rome (BNCR), the Central Institute for Unique Catalogue (ICCU), the National Research Council of Italy (CNR), Fondazione Rinascimento Digitale (FRD) and University of Milano (UniMi). The proposed model for the Italian infrastructure is based on a hierarchical and distributed approach [13].

At the highest level there is a root node, which is responsible for the top-level domain (IT in our case). The root node delegates the responsibility for the different second-level domains (e.g.: IT:UR, IT:FRD, etc.) to second-level naming authorities. Sub-domain responsibility can be further delegated using a virtually unlimited number of sub-levels (eg.: IT:UR:CNR, IT:UR:UNIMI, etc.). At the bottom of this hierarchy there are the leaf nodes, which are the only ones that harvest publication metadata from the actual repositories and assign unique identifiers to digital objects.

Each agency adheres to the policy defined by the parent node and consistently defines the policies its child nodes must adhere to [13].

Within this architecture each node harvests PI information from its child nodes and is able to directly resolve all identifiers belonging to its domain and sub-domains. Besides, it can query other nodes to resolve NBN identifiers not belonging to its domain.

A first testbed of the Italian infrastructure has been deployed and tests have been carried out for massive NBN assignment and for resolution of single identifiers.

4.6 PID Interoperability Issues

There are cases in which it may be useful to assign an NBN to a digital resource already having an assigned PI of a different type. For example, if a resource has an assigned DOI, it may still be useful to assign an NBN to it, because the latter type of PI has a syntax that facilitates accounting tasks (e.g. grouping and counting resources by sub-domain). In this cases it is useful to make provisions for easy retrieval of the previously assigned identifier. In our opinion there are two main ways for achieving this result: the first is to include the previously assigned PI in the final part of the NBN, the second is to include it in the metadata set. Both solutions have advantages and drawbacks. The first allow for easy retrieval of the alternate PI but may violate local policies for NBN generation and can be used for only one additional and pre-existing PI. The second makes it less immediate to retrieve the alternate PI but does not interfere in the NBN generation process. Also In this case, a common agreement is needed, which takes into account practical issues and existing situations [24].

4.7 Duplicate Detection

The problem of duplicate detection within large and distributed document bases is not a trivial one but effective solutions in this field can have a major impact on the ability to certify contents, especially if combined with the use of persistent identifiers. A simple solution is to associate an MD5 (or equivalent) hash code to any resource. This code is calculated from the binary content of the file itself. Comparison of hash codes allow detecting bitwise equality but if two files differ for a single bit, they are considered different entities. More robust solutions can be devised using other techniques. We are currently exploring the possibility of using I-Match [25] or equivalent algorithms, which enable near duplicates detection. We are currently working to assess the scalability of this approach in large and distributed environments

5 Technical Solutions

A series of activities have been carried out and collaborations have been established in order to identify and implement appropriate technical solutions for the problems described in the previous sections. In particular, we have developed a first prototype of document repository called JDIAM which satisfy the following requirements:

- Customizable workflows
- Customizable metadata sets
- Authority file management
- Document versioning
- Document classification
- Information retrieval

- User profiling
- Support to a hierarchical organization of digital libraries
- Support to OAI-PMH

The application is based on open standards and open source applications. In particular, it has been developed on Java Enterprise platform and uses an Alfresco document repository. The interaction with the repository is performed via CMIS protocol. This means that Alfresco can be substituted by an CMIS compliant repository (e.g. Nuxeo).

Future developments include: integration with the PI register for easy PI assignment, integration with automated classification tools for fast categorization of existing large document bases; enhancements of IR capabilities, DRM management, interoperability with external document bases through web service interface

JDIAM will be interoperable with JNBN, a software for the management of the nodes of a distributed and hierarchical infrastructure for NBN assignment and resolution. JNBN has been developed by the National Research Council of Italy with contributions from Fondazione Rinascimento Digitale and is Powered by CNR's JADA Development Framework. It is based on open technologies and standards and is released under EUPL license.

6 Conclusions

In this paper we have discussed the main issues that, in our opinion, should be addressed in order to improve the quality and reliability of the OA repositories and foster their use as dissemination tools, able to support research management and evaluation.

OA repositories maybe leveraged both at the national and international level as relevant data sources for statistical surveys and bibliometric indicator construction [26]. In order to achieve this goal it is crucial to resolve all the problems of data certification, stability and reliability. In our opinion this goal can be achieved with a cooperative and flexible approach applied both on the policy and technical levels and involving all the relevant stakeholders of the scientific environment.

Within CNR, experts in library and information science and in ICT are carrying out a series of joint studies and research activities aimed at designing technical and organizational solutions that can be exported to broader distributed and cooperative environments. To this end CNR is also involved in related activities with other relevant actors of the information and knowledge management community.

The results of our work may represent a starting point for future cooperative developments and are made available to the scientific community in order to ensure the maximum consensus about the choices to be made.

References

1. MIUR: Ministero dellIstruzione, dellUniversit e della Ricerca: Decreto Ministeriale 28 luglio 2009, n. 89: valutazione dei titoli e delle pubblicazioni scientifiche (2009)
2. ERC: European Research Council: Relaunching the European Research Area, ERA (2007)
3. ERC: European Research Council: ERC Scientific Council Guidelines for Open Access (2007)
4. CRUI, Roma: CRUI. Gruppo Open Access: Linee guida per gli archivi istituzionali (2009)
5. Guédon, J.C.: Open Access and the divide between "mainstream" and "peripheral" science (2008)
6. Organization for Economic Co-operation and Development, Paris: Organization for Economic Co-operation and Development (OECD): Frascati Manual 2002: proposed standard practice for surveys on research and experimental development: the measurement of scientific and technological activities (2002)
7. Guerrini, M.: Gli archivi istituzionali: open access, valutazione della ricerca e diritto dautore. Bibliografica, 33–60 (2010)
8. Guy, M., Powell, A., Day, M.: Improving the quality of metadata in eprint archives. Ariadne 38 (2004)
9. Park, J.R.: Metadata quality in digital repositories: a survey of the current state of the art. Cataloging & Classification Quarterly 47, 213–228 (2009)
10. Lynch, C.: Institutional Repositories: essential infrastructure for scholarship in the digital age. ARL Bimonthly Report 226, 1–7 (2003)
11. White, W.: Institutional repositories: contributing to institutional knowledge management and the global research commons. In: Proceedings of the 4th International Open Repositories Conference. Georgia Institute of Technology, Atlanta (2009)
12. ISO: International Organization for Standardization – ISO TC 20/SC 13: Space data and information transfer systems: open archival information system: reference model: ISO 14721:2003 (2003)
13. Bellini, E., Cirinná, C., Lancia, M., Lunghi, M., Puccinelli, R., Saccone, M., Sebastiani, B., Spasiano, M.: Persistent identifier distributed system for digital libraries. International Cataloguing and Bibliographic Control Journal (ICBC) 39, 30–36 (2010)
14. Paskin, N.: Digital Object Identifiers for scientific data. Data Science Journal 4, 1–20 (2005)
15. ARK: Archival Resource Key (2011)
16. Kunze, J., Rodgers, R.: The ARK identifier scheme. Technical report, Network Preservation Group (2008)
17. Sun, S., Lannom, L., Boesch, B.: Handle system overview: RFC 3650. Technical report, The Internet Society (ISOC) – IETF (2003)
18. Sun, S., Reilly, S., Lannom, L.: Handle system namespace and service definition: RFC 3651. Technical report, The Internet Society (ISOC) – IETF (2003)
19. Sun, S., Reilly, S., Lannom, L., Petrone, J.: Handle system protocol (ver 2.1) specification: RFC 3652. Technical report, The Internet Society (ISOC) – IETF (2003)
20. Library of Congress – National Digital Library Program: Library of Congress: Relationship Between URNs, Handles, and PURLs (1997)
21. Library of Congress: Library of Congress – Network Development MARC Standards Office: Library of Congress Control Number (LCCN): restructuring to accommodate century change (1999)

22. Hakala, J., Hones, A.: Using national bibliography numbers as uniform resource names: draft IETF urnbis RFC 3188bis-nbn-urn-00. Technical report, IETF (2010)
23. Hakala, J.: Using national bibliography numbers as uniform resource names: RFC 3188. Technical report, IETF (2001)
24. Lancia, M., Sebastiani, B., Puccinelli, R., Spasiano, M., Saccone, M., Trufelli, L., Bellini, E., Cirinná, C., Lunghi, M.: Towards a european global resolver service of persistent identifiers. In: Cirinná, C., Lunghi, M. (eds.) Cultural Heritage Online Empowering Users: an Active Role for Users Communities, Florence, pp. 137–142 (2009)
25. Chowdury, A., Frieder, O., Grossmann, D., McCabe, M.C.: Collection statistics for fast duplicate document detection. ACM Transactions on Information Systems (TOIS) 20, 171–191 (2002)
26. Harnad, S.: Open access scientometrics and the UK research assessment exercise. Scientometrics 79, 147–156 (2009)