

Research product repositories: strategies for data and metadata quality control

Luisa De Biagi*, Roberto Puccinelli, Massimiliano Saccone*,
Luciana Trufelli***

***National Research Council of Italy – Central Library**

****National Research Council of Italy – IT Office**

Open Access: strategies and actions

Cooperative strategies acting at different levels

- **Political level**: effective initiatives are needed at the national and international level to favor open access to research results achieved through public funding; those initiatives should address and harmonize the interests of the **different stakeholders of the R&D system**.
- **Institutional level**: academic and research institutions should define institutional and operational policies and carry out effective and widespread **advocacy actions** in their reference communities.

“For institutional record-keeping, research asset management, and performance-evaluation purposes, and in order to maximize the visibility, accessibility, usage and impact of our institution's research output”

- **Economic and legal level**: Open Access is not zero-cost. Economic strategies are needed to sustain open access to public research products, based on the “**author/institution pay**” model; on the legal side, the adoption of **Creative Commons (CC)** licenses should protect intellectual property rights while granting open access.
- **Technical-organizational level**: **standards** and commonly-agreed **guidelines** (based on a cooperative approach) are needed to certify **data and metadata quality**.
- **Technological level**: OA greatly benefits from the development and widespread adoption of **open standards and protocols** and from the development of modular, interoperable and open source-based **platforms for the management and diffusion of digital contents**.

OA and grey literature valorization

- Grey literature products plays a significant role in the context of the scientific production managed and diffused through Open Access archives, indexed and aggregated by the main service providers.
- Since the Seventh International Conference on Grey Literature at Nancy in 2006, GreyNet community started increasing its research activities relating to the OA effect on grey literature.
- The adoption of open standards and of OAI protocols by the international OpenGrey network favors the interoperability between the main OA repositories and the “System for Information on Grey Literature in Europe”. This is a first important step in the development of cooperative networks for data and metadata certification.
- The diffusion of the international Open Access initiative can facilitate the development and coordination of cooperative networks, which can help define and implement sustainable processes and guidelines for:
 - a better quality certification of grey literature products (open peer review, open peer commentary, etc.) and of the related metadata (adoption of common metadata standards and mappings, cooperative bibliographic and authority control, versioning, persistent identification systems, etc.),
 - a better intellectual property protection especially for multimedia materials, containing a significant percent on Education, Learning and Professional Training (Creative Commons License is still weak)
 - Better information to users about copyright constraints (When and in which terms could I use it?)
 - a wider access to research products, which can improve their visibility and impact

Strategic role of cooperation

- R&D stakeholders together with IT experts must keep on working on the organizational, technological and operational levels in order to assure the interoperability between the different information and knowledge management systems, thus allowing for an effective exchange and reuse of certified data and metadata.
- In our opinion, a tighter collaboration is needed, both at the national and international levels, between the actors of the Scientific Information and Communication sector (libraries, editors/aggregators, public and private service providers, evaluation bodies, policy makers) in order to develop a network of organizational and technological infrastructures able to provide an actual interoperability between the different systems for the production, management, identification, certification and diffusion of information resources and meta-resources.

Green road: institutional and disciplinary archives

- *Green road to Open Access: an alternative strategy for the scientific communication and information.*

Benefits

- Significant increase of visibility, usage and impact of open access products, including grey literature
- Accelerated and expanded research cycle: quicker result sharing process starting from a broader base of products
- Open and transparent peer review processes (carried out by the academic and scientific communities)
- Metadata certification process carried out by the involved institutions (certification of own production)
- Reuse of metadata for research management and evaluation
- Increased intellectual property protection for grey literature (Creative Commons Public License)

Criticalities

- Lack of mandatory policies in many academic and research institutions
- Lack of initiatives for the adoption of commonly agreed strategies and solutions (also in the field of economic sustainability - OA is not zero-cost!)
- Various copyright policies depending on publisher type (green, blue, yellow)
- Lack of cooperative networks for data and metadata certification → lack of interoperability between systems

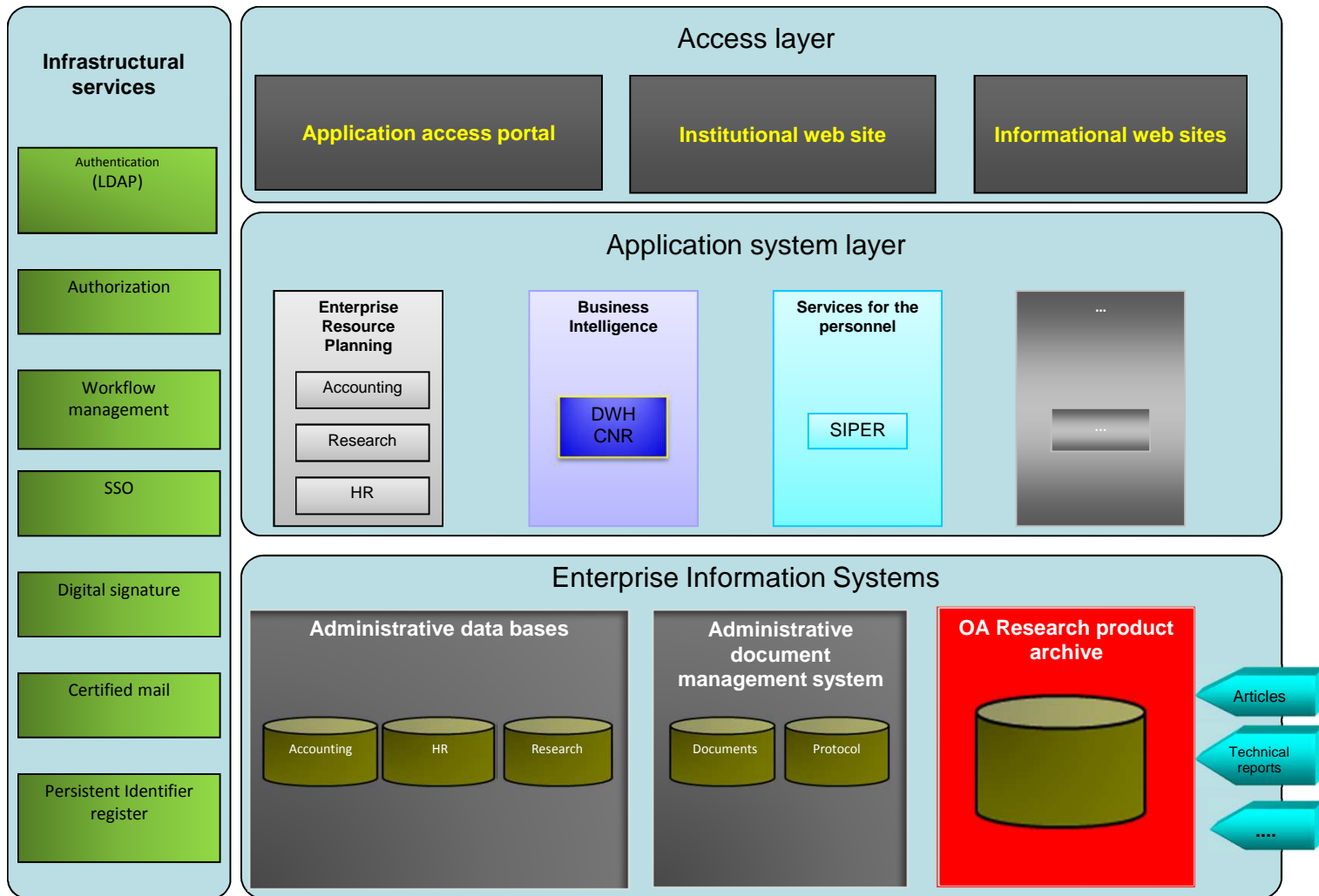
Green road interactions with other information channels

- Thanks to the stimulus provided by many institutions at the European and international level, many initiatives have been started for the creation of cooperative networks, which should design and implement systems and infrastructures for the integration of institutional and disciplinary archives in the context of international Information Systems (OpenGrey, OpenAire, etc.).
- Recently, some commercial publishers have established collaborations with institutions managing OA archives, in order to use those repositories as information sources for their data bases.
- Currently the Open Archive community is starting initiatives for the implementation of added value services, also leveraging the interoperability with other information systems (Current Research Information Systems) and bibliographic data bases, both public and commercial (e.g. OpenGrey, Web of Science, etc.)

CNR AI: a viable solution

- CNR:
 - Multidisciplinary research institution
 - 108 institutes distributed over the national territory and abroad
 - distributed hierarchical library system (Central library, research area libraries, institute libraries)
 - several digital archives but ...
 - Lack of an AI
- Viable solution:
 - Build on existing experiences and tools
 - Implementation of self archiving procedures
 - Implementation of multi-level data and metadata certification processes
 - Unique repository, several access points

IA integration with CNR IS



IA processes

- Self-archiving
- Different content reviewing procedures for conventional and unconventional research products
- Multi-level verification of metadata at the local (institute-area) and central (Central library) level
- Automated production of certified publication lists for each researcher (e.g. for internal career advancement procedures)

Quality control: methods and tools

- Bibliographic standards and authority control tools are not sufficient to assure data and metadata accuracy, completeness and consistency
- Quality management systems are needed to define processes for the production and management of data and metadata, which imply commonly agreed organizational models
- Only a shared effort can guarantee:
 - Quality certification of the main data and metadata production and management processes
 - Commonly agreed bibliographic and authority control tools for metadata certification
 - Highly customizable software solutions, based on open standards and platforms

Quality control: products, processes and strategies

- OA repositories keep a large variety of products and require complex management processes (production, validation, identification, indexing, preservation, diffusion, etc.)
- Products and metadata certification imply the adoption of strategies for the development of:
 - quality control policies, which allow data usage for statistical and evaluation purposes
 - ad hoc agreements with scientific and technological information stakeholders for the development of bibliographic and authority control systems.
 - configuration management systems for continuous control of content ingestion and management workflows.
 - open platforms (see also previous slide)
 - Technological support to innovative services: identification, versioning, right management, peer reviewing, etc.

Interoperability

OA repositories should be organized so as to guarantee:

- Interoperability with other national and international data bases (bibliographic, statistical, etc.);
- Metadata reuse, in order to support research product evaluation processes;
- Technical and organizational interoperability with national and international service providers.

OAI-PMH protocol limits interoperability to the unqualified Dublin Core schema, thus “flattening” research evaluation or increasing noise with an oversimplified metadata management process.

An effective interoperability can be guaranteed only through the adoption of organizational models based on the cooperation of all the R&D stakeholders.

Conclusions and future work

- Initiatives for research and academic institution evaluation are being established in Italy (ANVUR VQR 2004-2010)
- Within this context, OA archives are acquiring a great relevance thanks to their role of research product management systems and institutional data sources.
- In order to assure content reliability, a common effort is required for the development of cooperative certification systems
- This will be an important challenge for all the stakeholders of the Italian R&S system and, in our view, will increase the importance the OA archives within the scientific information and communication sector.