The way to open resources

Laurent Romary
CNRS

Two aspects of scientific communication

- Research papers
  - All types (Conferences, journals, grey literature etc.)
  - Classical media vs. Online publications
  - *Strong institutional needs*
- Research data/digital resources
  - Databases (e.g. genomics, chemistry, etc.)
  - Semi-structures documents (e.g. corpora and lexica in the humanities)
  - *Required for faster scientific progress*

Research papers at CNRS

- Institutional repository as a way towards open access
  - Win-win situation between the researcher and the institution
    - Researchers: bring the scientific content
    - Institution: brings the infrastructure (money!) and validation (documentary expertise)
  - Overall strategy
    - Added value through quality and services
    - Foster open access within the institutional repository

Quality

- Quality of information
  - Systematic check of metadata descriptions by a librarian
  - Institutional acknowledgement
  - Wide coverage
    - Cf. annual CNRS production: ~ 20 000 papers
- Quality of infrastructure provision
  - Long-term archiving environments
  - Wide dissemination

Services

- To our researchers
  - Quality, accessibility, longstanding archives, reporting aid (annual assessment, lab assessment), legal support
- To the institution
  - Quality, wide coverage of lab production, better evaluation, prospective tools
- To the research (and tax payer) community
  - Quality, wide accessibility of the French (multidisciplinary) research production

Basic roles in the workflow

- Researcher
  - Provides basic metadata information
  - Provides the actual content (file)
  - Expresses a will with regards OA
- Librarian
  - Checks and improve metadata
  - Validates the researcher’s will (legal constraints)
- Research manager
  - Approves the inclusion of the publication in the institutional repository

Rich metadata

- Multiplicity of MD configurations, from researcher to open archives
- Impossibility to standardize one single format
  - Solution: implements an ISO 11179 compliant MD registry ensuring semantic interoperability across standards (DC, HAL DTD, TEI, RDF, …)
- Additional services:
  - Diary of researchers, laboratories (multi-institutional)
  - Typology of scientific domains (in-house?)
  - Multidisciplinary terminological database

Implementation

- The HAL platform (CCSD; http://ccsd.cnrs.fr)
  - Coupled to ArXiv
  - Mechanisms allowing the creation of views (stamping) and collections
  - Already widely used in specific communities (Physics); PhD theses
- Several experiments
  - Large laboratory (350) with librarian in Nancy
  - Cluster of institutions (INRIA, CNRS, Univ. J. Fourier) with three documentation centers
  - Humanities laboratory with “remote” librarian at INIST

Digital resources

- Wide variety of resource types
  - Community specific approach (e.g. standards)
- High technical level required
  - Specific creation and maintenance methods
- Less copyright constraints
  - Replaced by privacy issues (humanities) and specific maintenance and distribution models
- Policy — win-win strategy again
  - Include resource production and dissemination in academic evaluation criteria
  - Support the development of resources through a network of competence centres

Summary

- Institutional repositories to leverage open access
  - One single infrastructure - Several views
  - Two policy levels
    - Mandatory contribution to IR
    - Encourage contribution to OA
- Institutions should put emphasis on quality and services
  - What is good for the institution is good for open access
- Beyond OAI
  - More standardization efforts on data description and representation