Des archives aux données de la recherche : projets internationaux

Francis ANDRE
DIST, CNRS

Ethnographic archives workshop
Nanterre, Laboratoire d'ethnologie et de sociologie comparative, MAE 24 – 25 janvier 2013

Credits : EC, GFII, RDA, EUDAT, Tony Hey

Content

The context : the 4th paradigm of science

A few keywords : the Open Family

European Commission policy framework

Publishers initiatives

What’s next ? Research Data Alliance

Emergence of a Fourth Research Paradigm

1. Thousand years ago – **Experimental Science**
   - Description of natural phenomena
2. Last few hundred years – **Theoretical Science**
   - Newton's Laws, Maxwell's Equations…
3. Last few decades – **Computational Science**
   - Simulation of complex phenomena
4. Today – **Data-Intensive Science**
   - Scientists overwhelmed with data sets from many different sources
     - Data captured by instruments
     - Data generated by simulations
     - Data generated by sensor networks
   - ➢ eScience is the set of tools and technologies to support data federation and collaboration
     - For analysis and data mining
     - For data visualization and exploration
     - For scholarly communication and dissemination

(With thanks to Jim Gray)

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - \mathrm{K}\frac{c^2}{a^2}$$

Content

The context : the 4th paradigm of science

A few keywords : the Open Family

European Commission policy framework

Publishers initiatives

What’s next ? Research Data Alliance

Open Science

Open Science means optimal sharing of research results and tools such as publications, research data, software, educational resources and infrastructures across institutional, disciplinary and national boundaries.

Open Content

Publications, data and educational resources resulting from publicly funded research should be openly made available and preserved for reuse.

To ensure the highest quality publications and datasets should be reviewed by peers taking advantage of the new opportunities offered by information and communication infrastructures.

Open Infrastructures

Research activities need reliable, high-performance and economically efficient infrastructures that are able to manage the expected scale of future data resources.

The Internet that changed our lives should keep its ability to provide global, open high-speed connectivity and should mitigate knowledge divides.

Data, computational resources and software resulting from publicly funded research should be made openly available and preserved, with appropriate open source licensing, for re-use in research and education activities.

Open Culture

Scientists and their organisations should apply open sharing principles.

In an open science world the academic and research career systems should support and reward scholars and students who participate in the culture of sharing.

Open science should enable adequate education and preparation of citizens to benefit from the abundance of technical tools and scientific information.

The young should be inspired by an ambition for new discoveries and the creation of new business and industries, and join the ranks of scientists, engineers and entrepreneurs in far greater numbers.

Those using or reusing scientific information should comply with codes of conduct and integrity in his/her discipline.

Content

The context : the 4th paradigm of science

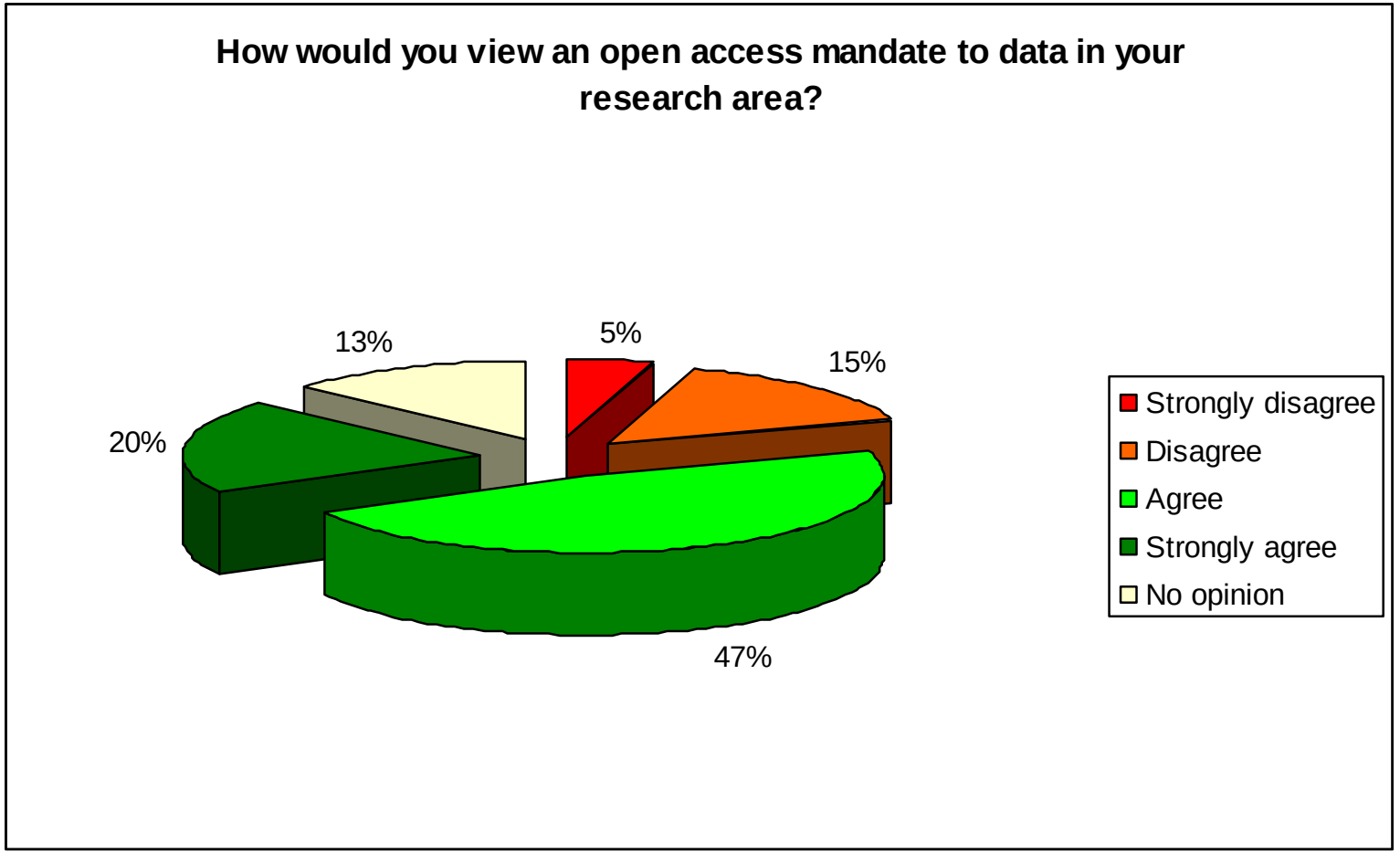
A few keywords : the Open Family

European Commission policy framework

Publishers initiatives

What’s next ? Research Data Alliance

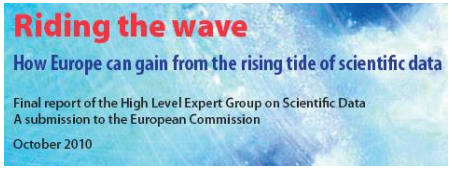
Open access to data?



Three quarters of respondents with opinion: agree or strongly agree

Source: EC Survey (July-August 2011)

e-infrastructure for data



"Our vision is a scientific e-infrastructure that supports seamless access, use, re-use of data."

from Riding the Wave report, High-Level Group on Data

Supporting 21st Century scientific data factories

Basis for global multi disciplinary research

Work on global collaborations (G8+5 working group on data)

Opportunities: Horizon 2020, ERA

Make open access to publications the general principle

Start addressing consistently scientific data infrastructure

Coordinate with EU Member States on open access and preservation policies

Scientific information package

- *Two texts (17.7.2012):*
  - **Commission Communication (COM(2012)401): "Towards better access to scientific information: Boosting the benefits of public investments in research"**

    → announces EC actions in Horizon 2020
  - **Commission Recommendation C(2012)4890) on access to and preservation of scientific information**

    → suggests actions for the MS
- *Adopted together with ERA Communication "A Reinforced European Research Area Partnership for Excellence and Growth"*

Open access to research data

3. Define clear policies for the dissemination of and open access to research data resulting from publicly funded research.

These policies should provide for:

- concrete objectives and indicators to measure progress;
- implementation plans, including the allocation of responsibilities (including appropriate licensing);
- associated financial planning.

Open access to research data

Ensure that, as a result of these policies:

- research data that result from publicly funded research become publicly accessible, usable and re-usable through **digital e-infrastructures**. Concerns in particular in relation to privacy, trade secrets, national security, legitimate commercial interests and to intellectual property rights shall be duly taken into account. Any data, know-how and/or information whatever their form or nature which are held by private parties in a joint public/private partnership prior to the research action and have been identified as such shall not fall under such an obligation;
- datasets are made **easily identifiable** and can be **linked** to other datasets and publications through appropriate mechanisms, and additional information is provided to enable their proper evaluation and use;
- institutions responsible for managing **public research funding** and academic institutions that are publicly funded assist in implementing national policy by putting in place mechanisms enabling and rewarding the **sharing of research data**;
- advanced-degree programmes of **new professional profiles** in the area of data-handling technologies are promoted and/or implemented.

Pilot on open access to data

- **To be developed during 2013**
- **Will take into account privacy, security issues, commercial interests**
- **Issues to consider:**
  - What data? (underlying, raw, data collections)
  - What areas? (ERC, FET, societal challenges)
  - Where should data go?
  - Use and re-use / ownership
  - More than just access: data management & curation
- **Process: internal and external consultations**

Recommendation to the MS

- **Member States to define policies for and implement:**
  - OA to publications
  - OA to research data – taking into account data-specific concerns
  - preservation and re-use of sientific information
  - e-infrastructures
- **Consistency between EU H2020 policy and MS policy**
- **Structured co-ordination of MS at EU-level and reporting**
- **Multi-stakeholder dialogue to be set up**

Content

The context : the 4th paradigm of science

A few keywords : the Open Family

European Commission policy framework

Publishers initiatives

What’s next ? Research Data Alliance

Les mégarevues

Le modèle des mégarevues marque une rupture importante par rapport aux critères classiques de l'édition scientifique :

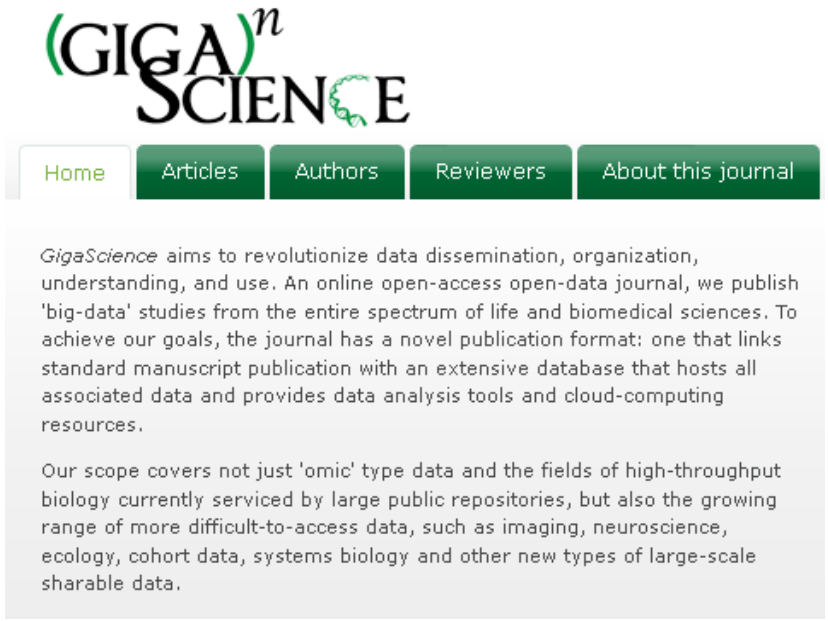
- Libre accès aux publications sans équivalent papier : diffusion on line uniquement.
- Périmètre résolument multidisciplinaire : l'ancrage dans une discipline, élément structurant de l'édition scientifique traditionnelle, s'efface.
- *Peer-reviewing* « allégé » : l'évaluation se limite à l'examen de la cohérence scientifique des travaux, voire intervient même dans certains cas après publication (« *post peer-reviewing* »), les résultats de recherche négatifs sont parfois acceptés.
- Contenus à périmètre large : séquences multimédias, *datasets*, macros sont adjointes aux articles ; les formes documentaires traditionnelles convergent et s'effacent parfois dans de vastes bases de données (*Big Data*).

Springer lance « SpringerPlus »

« Si les contenus de SpringerPlus seront bien « peer reviewed », cette validation par les pairs se limitera à l'examen de la solidité scientifique des travaux et de l'intérêt des résultats rapportés dans l'article. (…) Il n'y aura pas de limitation de la longueur des articles ni du nombre d'illustrations. (…) Tout l'enjeu pour Springer et les éditeurs qui suivront cette voie sera de développer pour leur « mégarevue » une notoriété suffisante pour s'imposer dans le paysage déjà chargé de la publication scientifique. »
http://www.springer.com/about+springer/media/pressreleases?SGWID=0-11002-6-1329722-0 (12/01/2012)

BioMedCentral lance « GigaScience »

BioMedCentral, filiale de Springer, a lancé en juillet 2012 la mégarevue « GigaScience », positionnée sur les « Big Data » dans le domaine de la recherche sur le génome humain, en partenariat avec le BGI (Beijing Genome Institute). La revue, qui diffuse des ressources textuelles et multimédia très diverses (articles scientifiques, rapports d'expériences, graphes et schémas) est adossée à une base de données (GigaDB) pour l'hébergement et la publication des jeux de données.
http://www.knowledgespeak.com/newsArchieveviewdtl.asp?pickUpID=15945&pickUpBatch=1992#15945 (11/07/2012)



Wiley-Blackwell lance « GeoScience »

Presqu'au même moment, Wiley Blackwell a lancé, en partenariat avec la Royal Meteorological Society (UK) « GeoScience Data Journal » : une mégarevue embrassant toutes les sciences du vivant (climatologie, géologie, océanographie, …). La revue s'inscrit dans une logique de « data curation », et ne publie que des articles factuels (description des jeux, méthodologie pour la collecte, macros et logiciels pour la manipulation des données). Les données, structurées et identifiées par un DOI sont stockées dans un entrepôt, et liées aux articles.

http://eu.wiley.com/WileyCDA/PressRelease/pressReleaseId-104139.html (16/07/12)

Geoscience Data Journal

© 2013 John Wiley & Sons Ltd.

Each article is made available under the terms of the Creative Commons Attribution License

Edited By: Dr Rob Allan, Met Office, UK

Online ISSN: 2049-6060

Associated Title(s): Atmospheric Science Letters, International Journal of Climatology, Meteorological Applications, Quarterly Journal of the Royal Meteorological Society, Weather

Meet the Editor | Rob Allan

Dr Rob Allan joined the Met Office Hadley Centre in the UK in mid-2000 after some 10 years working in the Climate Impact Group for CSIRO's Atmospheric Research Division in Melbourne, Australia. His research foci originally centred on the El Niño Southern Oscillation phenomenon, interannual to secular scale climatic variability and climate change. At the Met Office Hadley Centre, he was responsible for the development of historical gridded global monthly mean sea level pressure (MSLP) and daily North Atlantic-European MSLP data sets back to 1850. This involved the rescue, recovery, quality control and archiving of historical global MSLP data from both terrestrial and marine sources.

Now accepting papers

Provides an open access platform where scientific data can be formally published, in a way that includes scientific peer-review.

An online-only journal, GDJ publishes short data papers cross-linked to, and citing, datasets that have been deposited in approved data centres and awarded DOIs.

A data article describes a dataset, giving details of its collection, processing, software, file formats etc, without the requirement of novel analyses or ground breaking conclusions. It allows the reader to understand the when, how and why data was collected and what the data-product is.

F1000 lance « F1000 Research »

Le service de veille scientifique sur abonnement Faculty of 1000 (F1000) a lancé *F1000 Research* : une plateforme orientée « Big Data » pour la publication immédiate de contenus scientifiques en libre accès intégral à spectre large (les jeux de données incomplets sont acceptés). La validité scientifique des travaux fait l'objet d'une évaluation collaborative post-publication.
http://f1000.com/resources/F1000Research_PressRelease_WebFINAL.pdf

The right to read is the right to mine

- besoin des communautés scientifiques d'accéder en masse aux « données brutes » de la recherche en mutualisant l'interprétation des résultats.
- besoin de plus en plus fort des outils du « Big Data » (*text mining, data mining*, visualisation, outils de découverte, …), afin.
- Dans cette optique, la mise en libre accès des corpus est une condition indispensable à l'interconnexion généralisée entre les données et à leur réutilisation.

Sur ce sujet, lire l'article du Guardian : http://www.guardian.co.uk/science/2012/may/23/text-mining-research-tool-forbidden (23/05/12)

L'Open Knowledge Fondation (OKF), qui réalise un important travail de lobbying auprès des pouvoirs publics afin d'encourager les démarches Open Data et Open Science, a publié une déclaration sur ce sujet : *« the right to read is the right to mine »* en mai : http://blog.okfn.org/2012/06/01/the-right-to-read-is-the-right-to-mine/ (01/05/12)

En 2010, l'Open Knowledge Fondation avait, à travers son groupe de travail Open Data in Science, participé à la rédaction des Principes de Panton (Panton Principles), un texte de référence pour le mouvement Open Science : http://pantonprinciples.org/
Traduction française : http://www.commissionrecherche.com/2012/09/panton-principles-for-open-data-in.html

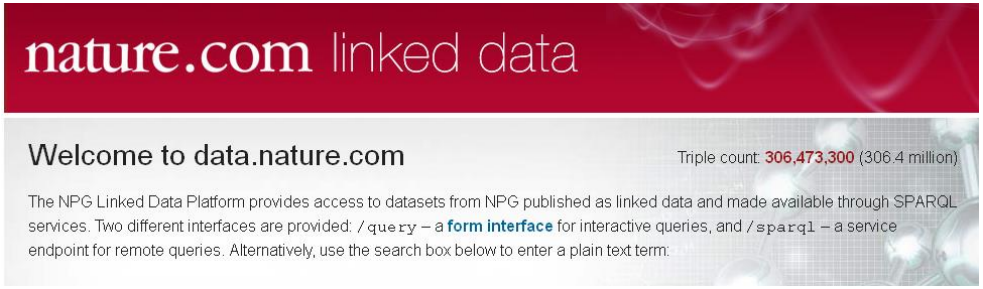
Selon Phillip Campbell, directeur des publications de NPG, le libre accès à la recherche est inévitable car les chercheurs auront de plus en plus besoins de « fouiller » les corpus pour donner du sens aux grands ensembles de données :
http://www.guardian.co.uk/science/2012/jun/08/open-access-research-inevitable-nature-editor (08/06/12).

La fourniture de services de découverte et de *« linking »* vers les jeux de données constitue désormais un positionnement possible pour les éditeurs afin de valoriser leurs contenus. Le ***Data Citation Index***, récemment lancé par Thomson Reuters, sera vraisemblablement amené à jour un rôle de modèle pour l'offre commerciale et Open Access.
http://thomsonreuters.com/content/press_room/science/730914 (16/10/12).

NPG ouvre sa plateforme de « données liées »

NPG a lancé sa plateforme de « données liées » en avril 2012 : http://data.nature.com. La plateforme est alimentée en temps réel par l'ensemble des métadonnées des publications diffusées sur la plateforme commerciale de l'éditeur. Un communiqué de juillet indique que la plateforme héberge désormais plus de 270 millions d'objets RDF (soit 10 fois plus qu'à son lancement en avril), organisés en 12 jeux de données interrogeables (langage SPARQL). Ceux-ci comprennent les métadonnées bibliographiques, de citation et d'ontologies utilisées pour toutes les publications de la maison d'édition. Les jeux sont publiés dans le respect des normes et standards du web sémantique (Dublin Core, FOAF, PRISM, BIBO, OWL, …). Leur antériorité remonte à 1945 pour les articles publiés dans la revue *Scientific American*. Les données sont diffusées sous le régime des licences Creative Commons Zero (CC0), pour une liberté de réutilisation maximale.

http://www.nature.com/press_releases/ldp.html (20/07/12)



Partenariat entre l'ALPSP et l'association P-D-R

« L'ALPSP (Association of Learned and Professionnal Society Publishers, www. alpsp.org) ; STM, l'association internationale des éditeurs scientifiques, techniques et médicaux, et l'association P-D-R (Pharma Documentation Ring), qui regroupe les documentalistes et professionnels de l'information au sein des grand groupes pharmaceutiques, ont annoncé le 17 septembre un accord modifiant la licence type qui lie les éditeurs scientifiques aux membres du « club » P-D-R. Une clause nouvelle permettra aux membres du P-D-R d'utiliser les corpus d'articles scientifiques issus des revues auxquelles ils sont abonnés pour des usages de text et de data-mining.. Elle illustre le fait que les éditeurs tentent de passer par la voie contractuelle pour encadrer le développement des usages du text mining. »

http://www.p-d-r.com/content/press_releases/archive/2012/

BiomedCentral : adapter les licences CC au text-mining

BioMed Central a annoncé son souhait de modifier la licence attachée aux données de ses articles pour les placer sous licence CC0, autorisant les réutilisations commerciales et le text mining. Par cet acte les métadonnées seraient considérées comme des « données publiques » relevant du bien commun, notamment les références bibliographiques, et les données de recherche attachées (*« tabular data », « supplementary material »*). A ce stade, BioMed Central attend le retour des auteurs et des éditeurs sur un projet de draft de la licence destiné à formaliser le cadre juridique de ces usages :

http://blogs.biomedcentral.com/bmcblog/2012/09/10/put-the-open-in-open-data/ (10/09/12)

Rapport de l'ESF sur les enjeux de l'OA pour la recherche biomédicale:

La European Science Fondation (ESF) regroupe 72 organismes de recherche européens et travaille en étroite collaboration avec Science Europe pour l'unification de l'Espace de Recherche Européen (ERA) pour lequel elle considère que le libre accès sera un levier. La Fondation vient de publier un rapport sur les enjeux de la mise en libre accès des résultats de la recherche biomédicale: *"open access in biomedical research across Europe".* Au-delà des aspects moraux (accélérer la diffusion de la recherche médicale pour améliorer la santé publique), le rapport souligne les enjeux liés à l'ouverture des données aux différentes formes de réutilisation, notamment la fouille et le *text-mining* pour accélérer la découverte scientifique.

http://www.esf.org/media-centre/ext-single-news/article/new-report-assesses-open-access-in-biomedical-research-across-europe-850.html (22/10/12)

Projet SHARE / De Gruyter

SHARE est un projet multidisciplinaire financé par la Commission européenne, visant à créer une vaste base de données sur les statistiques européennes en rapport avec le vieillissement de la population : gros volumes de microdonnées sur un panel de plus de 80 000 individus, résidant dans plus de 20 pays, âgés de 50 ans et plus. La base est alimentée en collaboration par les différents observatoires et instituts de recherche publics impliqués dans la démarche, en charge des campagnes d'enquête au niveau national. L'initiative sert de modèle dans le cadre de l'unification de l'ERA (European Research Area). Elle répond au besoin de l'UE de pouvoir observer sur le temps long l'évolution de la pyramide démographique des états-membres en bénéficiant de données mutualisées et homogènes. Les données étaient déjà ouvertes gratuitement à la réutilisation par les chercheurs (le projet est entre autre, piloté par l'Institut Max Planck de Berlin, à l'initiative de la Déclaration de Berlin).

http://www.share-project.org/fileadmin/press_information/PI_De_Gruyter_Share_ENG-1.pdf (09/10/12)

http://www.nature.com/press_releases/linkeddata.html (04/04/1

BioMed Central : fournir des environnements « Open Science » aux auteurs

BioMed Central, la filiale intégralement Open Access de Springer, se positionne sur la fourniture d'environnements de travail « Open Science » pour les chercheurs, avec pour enjeu l'intégration des publications dans l'écosystème des données de la recherche. Dans cette optique, l'éditeur a interfacé le logiciel de publication et d'archivage des données de la recherche « *LabArchives Electronic Laboratory Notebook »* (ELN) sur sa plateforme. Les auteurs déposant leur article sur BioMed Central peuvent désormais charger *datasets* et métadonnées associées à leurs travaux dans un espace de stockage dédié (100Mb gratuits). Ils peuvent également attribuer gratuitement un DOI (Digital Object Identifier) aux jeux pour en faciliter la citation, l'exploration et garantir la pérennité des liens vers les objets.

http://www.biomedcentral.com/presscenter/pressreleases/20120404 (04/04/12).

BiomedCentral : adapter les licences CC au text-mining

BioMed Central a annoncé son souhait de modifier la licence attachée aux données de ses articles pour les placer sous licence CC0, autorisant les réutilisations commerciales et le text mining. Par cet acte les métadonnées seraient considérées comme des « données publiques » relevant du bien commun, notamment les références bibliographiques, et les données de recherche attachées (*« tabular data », « supplementary material »*). A ce stade, BioMed Central attend le retour des auteurs et des éditeurs sur un projet de draft de la licence destiné à formaliser le cadre juridique de ces usages :

http://blogs.biomedcentral.com/bmcblog/2012/09/10/put-the-open-in-open-data/ (10/09/12)

BioMed Central
The Open Access Publisher

BioMed Central blog

Help put the open in Open Data and Open Bibliography

Iain Hrynaszkiewicz on September 10, 2012 at 3:08 pm - 9 Comments

Legal restrictions and uncertainties surrounding scientific data are a barrier to efficient data sharing and reuse, and ultimately the pace of research. Copyright in particular is problematic for data. It is often unclear if data are protected by copyright and the law differs greatly internationally.

To try and more universally clarify the legal status of data published in our open access journals – maximizing the potential for secondary data uses such as text mining – we have been working towards a solution: public domain dedication of data under the Creative Commons CC0 waiver.

BioMed Central : fournir des environnements « Open Science » aux auteurs

BioMed Central, la filiale intégralement Open Access de Springer, se positionne sur la fourniture d'environnements de travail « Open Science » pour les chercheurs, avec pour enjeu l'intégration des publications dans l'écosystème des données de la recherche. Dans cette optique, l'éditeur a interfacé le logiciel de publication et d'archivage des données de la recherche « *LabArchives Electronic Laboratory Notebook* » (ELN) sur sa plateforme. Les auteurs déposant leur article sur BioMed Central peuvent désormais charger *datasets* et métadonnées associées à leurs travaux dans un espace de stockage dédié (100Mb gratuits). Ils peuvent également attribuer gratuitement un DOI (Digital Object Identifier) aux jeux pour en faciliter la citation, l'exploration et garantir la pérennité des liens vers les objets.

http://www.biomedcentral.com/presscenter/pressreleases/20120404 (04/04/12).

Linking and lightening: New partnership connects and reveals dark data

04 Apr 2012

• BioMed Central and LabArchives partnership ensures that valuable datasets are made publically available and linked to journal articles

• Authors granted enhanced access to Electronic Laboratory Notebook software

Sharing and reuse of data has become a vital part of modern scientific research. Having access to datasets ensures that the pace of scientific discovery is not unnecessarily hindered by data being kept under lock and key or hidden away in lab drawers.

BioMed Central, a long supporter of data sharing and the open data movement has now partnered with LabArchives, LLC to shed further light on science's "dark data". Through this new collaboration, authors submitting articles to selected BioMed Central journals will be provided with complimentary subscriptions to an enhanced version of the popular LabArchives Electronic Laboratory Notebook (ELN) software.

LabArchives provides a powerful yet simple-to-use platform which scientists can utilize to store and publish their laboratory data. The special BioMed Central version of LabArchives will provide authors with 100Mb of complimentary storage and optional integrated submission to BioMed Central's journals. Published datasets can be assigned a Digital Object Identifier (DOI) which serves as a permanent identifier for a data set, making data more discoverable and citable, helping to enable scientists to get credit for publishing their data and permanently link journal articles to supporting data.

Content

The context : the 4th paradigm of science

A few keywords : the Open Family

European Commission policy framework

Publishers initiatives

What’s next ? Research Data Alliance

Accelerating Data Sharing and Exchange Across the Globe → the Research Data Alliance

- The Research Data Alliance (RDA) is an emerging organization that will facilitate **specific, short-term efforts that accelerate the sharing and exchange of research data**
- **Working groups serve as accelerants** to data sharing practice and infrastructure. Work products / deliverables include
  - Adopted standards
  - Deployed infrastructure
  - Adopted policy
  - Implemented best practice, etc.

Research Data Alliance Approach

Goal: Focused efforts, tangible progress:

- **Create and adopt** focused pieces of code, policy, infrastructure, standards, or best practices that enable data to be shared and exchanged
- **"Harvestable" efforts** for which roughly a year of work can eliminate a roadblock for a substantial community
- **Efforts that have substantive applicability** to segments of the data community, but may not apply to everyone
- **Efforts for which working scientists and researchers can start today** while more long-term or far-reaching solutions are being appropriately discussed through other vehicles

"Case Statements" being used to focus development of outcome-oriented Working Groups

Creating a new organization:

- *Organization "jumpstarted" through international R&D agency discussions*
- *Development support and participation by US (NSF, NIST), EU (iCORDI), AU (ANDS)*
- *Activities, organization being developed by international non-governmental Steering Group*
- *Global Data Meeting held 10/1-3 in Washington D.C. involved 120+ international participants from 3 continents*
- ***"Official" launch March 18-20, 2013 in Sweden***

RDA Organizational Structure – Work in Progress …

Working Groups

- Carry out work of RDA
- Outcome-oriented, consensus-driven
- Approval based on vetted "Case Statement" defining group charter, deliverables, and processes
- Open to all; commitment expected

RDA Community

- Open registration through rd-alliance.org, commitment to participation expected / encouraged
- Participate in Interest Groups (TBD), Working Groups, other component groups – multiple entry points
- Attend RDA Plenary (Forum for working groups, interest groups, organizational members, partner affiliates, etc.)

Technical Expert Group(s) (TBD)

- Advise WGs and Council
- Provide expertise to enable working groups and interest groups / BOFs

Online Open Interaction Fora

- use for many activities, open to all RDA members

Admistration and Management Team (Secretariat and NGS)

- Implements strategic direction set by Council
- Supports RDA component groups including Council, Working Groups, etc.
- Supports the activities of the RDA
  - Arranges plenary meetings
  - Runs on-line forums
  - Manages documents, etc.

Council

- Confirms consensus, manages strategic direction
- Provides general oversight of RDA
- Serves as a point of reference
- Drives economic sustainability for RDA
- Decision-making responsibilities for WGs, etc.

How to Participate

- **Suggest a Working Group** (focused12-18 month efforts with measurable outcomes that accelerate data sharing and exchange)
  - Candidate Working Groups creating **"Case Statements"** on their topic for community and RDA review
- **"Join" RDA and Participate in the Forums**
  - **Community members** will join by registering at rd-alliance.org. (Discussion Forum at **forum.rd-alliance.org**).
  - **Organizational members** contribute funding or time for a 2+ years to support core RDA structure
- **Provide feedback** through **rd-alliance.org** and to the organizing committee as the organization develops
  - Community feedback welcome/needed to "debug" the RDA.
  - Help us make this an impactful, effective, and functional organization that achieves its goals.
- **Come to the first RDA Plenary**
  - **March 18-20, 2013 in Gothenburg, Sweden**. More information will be available at rd-alliance.org

RDA - early thematic groups

- Foundation of Data Objects	- basic data model and reference terminology
- Data Objects	- diversity of formats, structures, semantics
- Collection Properties	- specifying properties for automatic checking
- Metadata Principles	- architecture basics of metadata systems
- Policy Rules	- harmonization of (formal) policies and rules
- Content Interoperability	- "simple" architectures for semantic interop
- PID Information Types	- harmonization on info associated with PIDs
- Data Type Registry	- a registry of types with actionable links
- Center Registry	- a worldwide registry of repositories
- Overcoming Data Friction	- improvements towards services on data
- Data as a Service	- architecture basics for services on data
- Registry Interchange	- discovery across discipline boundaries
- Certification	- principles of certification of repositories
- Data-Publication Linking	- principles/architectures for D-P-Linking
- IPR Issues	- harmonizing and simplifying IPR/license issues
- Marine Data Management	- domain specific harmonization
- Urban Sciences Data	- domain specific harmonization

Non Government Structures (NGS) Funded to support RDA

US:

- Fran Berman – RPI
- Bill Michener – DataOne
- Beth Plale – Indiana
- Sayeed Choudhury – Johns Hopkins

Australia:

- Ross Wilkinson – ANDS
- Andrew Treolar - ANDS

Europe:

- Leif Laaksonen (iCORDI/CSC)
- Peter Wittenburg (iCORDI/Max Planck Institute)
- Juan Bicarregui (iCORDI/STFC)

iCORDI key facts

Project Name	iCORDI – Global Data Interoperability
Start date	1st September 2012
Duration	24 months
Budget	3,3 M€ (including 2,3 M€ from the EC) and 240 PMs (CSA)
EC call	Call 10 (INFRA-2012-3.2): International cooperation with the USA on common e-Infrastructure for scientific data (11.2011)
Participants	14 partners from 8 countries (national data centers, technology providers and research communities)
Objectives	"iCORDI's prime objective is to establish a ***coordination platform*** between Europe and the USA to discuss and improve the interoperability of today's and tomorrow's scientific data infrastructures of both continents and to extend this to the global levels."

Who's Involved in iCORDI

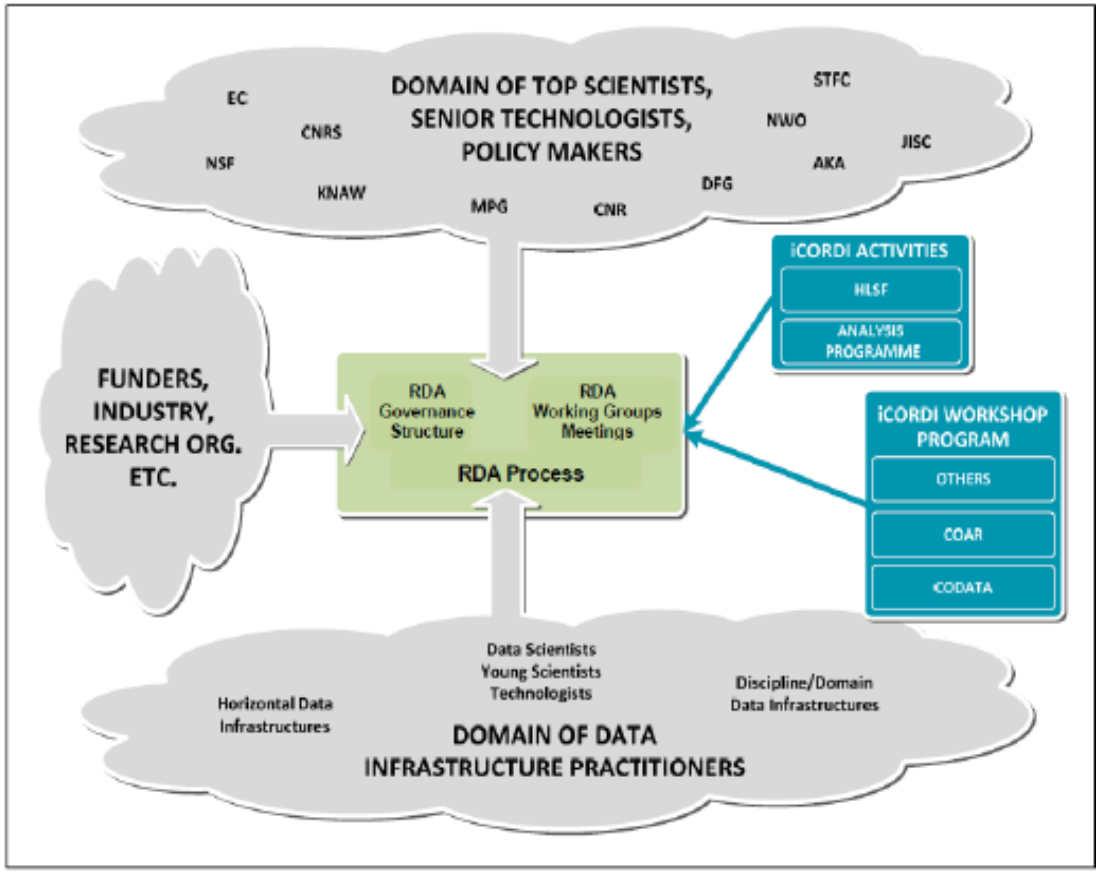
icordi-info@postit.csc.fi | www.icordi.eu

iCORDI goals

- Facilitate and support the creation of an IETF (Internet Engineering Task Force) like organisation for the global data interoperability and data exchange – **Research Data Alliance** (RDA)
- **Support** the work of **European research communities** in their effort to **collaborate** with their US counter parts through **a coordinated platform** and to extend this to **the global levels**

High Level Strategy Forum (HLSF) 1/2

- **Data-oriented scientists** mean leading scientists who are working at the cutting edge of data-driven science in various disciplines and who can help determining the needs with respect to the integration of data to create interoperable domains.
- **Data architects** mean leading IT experts who are working on concrete data organizations and data solutions that aim to overcome the many technical barriers currently faced.



RESEARCH DATA ALLIANCE

Researchers around the world sharing and using research data without barriers

- About
- Organisation
- Working Groups +
- News & Events +
- First Plenary -
  - Invitation
  - Programme
  - Venue and Facilities

Please Join Us for the Research Data Alliance Launch and First Plenary

March 18-20, 2013, Gothenburg, Sweden

Open Access Research Data without Barriers

As "big data" emerges as an international priority, the Research Data Alliance (RDA) is a newly formed organization whose goal is to accelerate data-driven innovation world-wide through research data sharing and exchange.

At its first Plenary, the RDA will be launched by sponsors from the European Commission, the U. S. Government and the Australian Government and leaders in the data community. The Plenary will be a working meeting to accelerate discussion, Working and Interest Group interaction, and data community development.

Research Data Community Gathering and Working Meeting

Join us for 3 days of outstanding speakers, working sessions, active contributions, and the launch of the Research Data Alliance. We invite you to participate in the RDA and join international colleagues working to increase the data sharing and exchange needed to drive global innovation and success. We look forward to seeing you in Gothenburg!

Registration is free of charge, please register **here**. *Please note that the number of participants is limited, so please register now to avoid disappointment.*

Links

- G8+05 working group on Data Infrastructures

http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/g8.pdf

- Research Data Alliance

http://rd-alliance.org/

- iCordi

https://www.icordi.eu/

- French website on research data

http://www.donneesdelarecherche.fr/