



Quelles évolutions pour l'IST et ses métiers ?



ou



francis.andre@cnrs-dir.fr

contenu

- Des changements...
 - Les pratiques de recherche
 - Les défis sociétaux
 - Open Science
 - Évolutions techniques
 - Évolutions juridiques
 - Nouveaux services
- Impact sur les fonctions d'appui
 - IST : des objets, des métiers
 - Des opportunités à saisir
- Les freins au changement
 - Les éditeurs
 - Changer les mentalités
 - Des infrastructures à créer, à piloter
- En résumé

Crédits: Commission
Européenne, Jim Gray,
The Royal Society, Fran
Berman, OCDE, L.Schmitt

Conseils de lecture...

<https://www.innovationpolicyplatform.org/content/open-science>

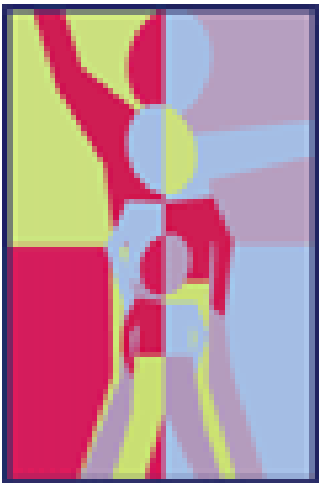


<http://ec.europa.eu/digital-agenda/en/news/final-report-science-20-public-consultation>

Research today

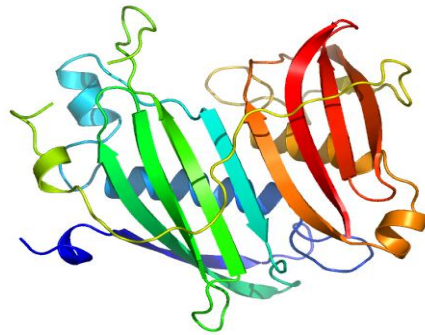
Which has the greatest impact – nature or nurture?

PSID: longitudinal data on 8000 families over 40 years



How does disease spread?

PDB: World wide reference collection of protein structure information



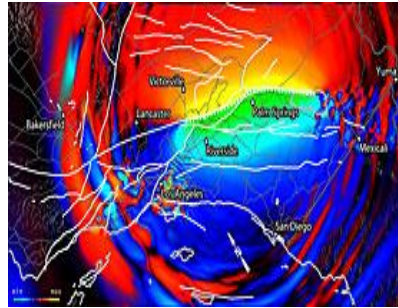
Are current stresses on this bridge dangerous?

Terabridge data set: Structure sensor data for real-time data mining, event detection, decision support and alert dissemination



What is the impact of a large-scale earthquake on the Southern San Andreas Fault?

Digital data from Southern California Earthquake Center simulations used for disaster planning and building requirements



Where are the brown dwarfs?

NVO: Data from 50+ astronomical sky surveys and large-scale telescopes.

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



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



(With thanks to Jim Gray)

Evolution des pratiques de science

Science du 21^{ème} siècle : plus...

- Numérique
- Collaborative
- Interdisciplinaire
- Réactive
- Citoyenne
- Partagée

**Open
Research**

Science 2.0

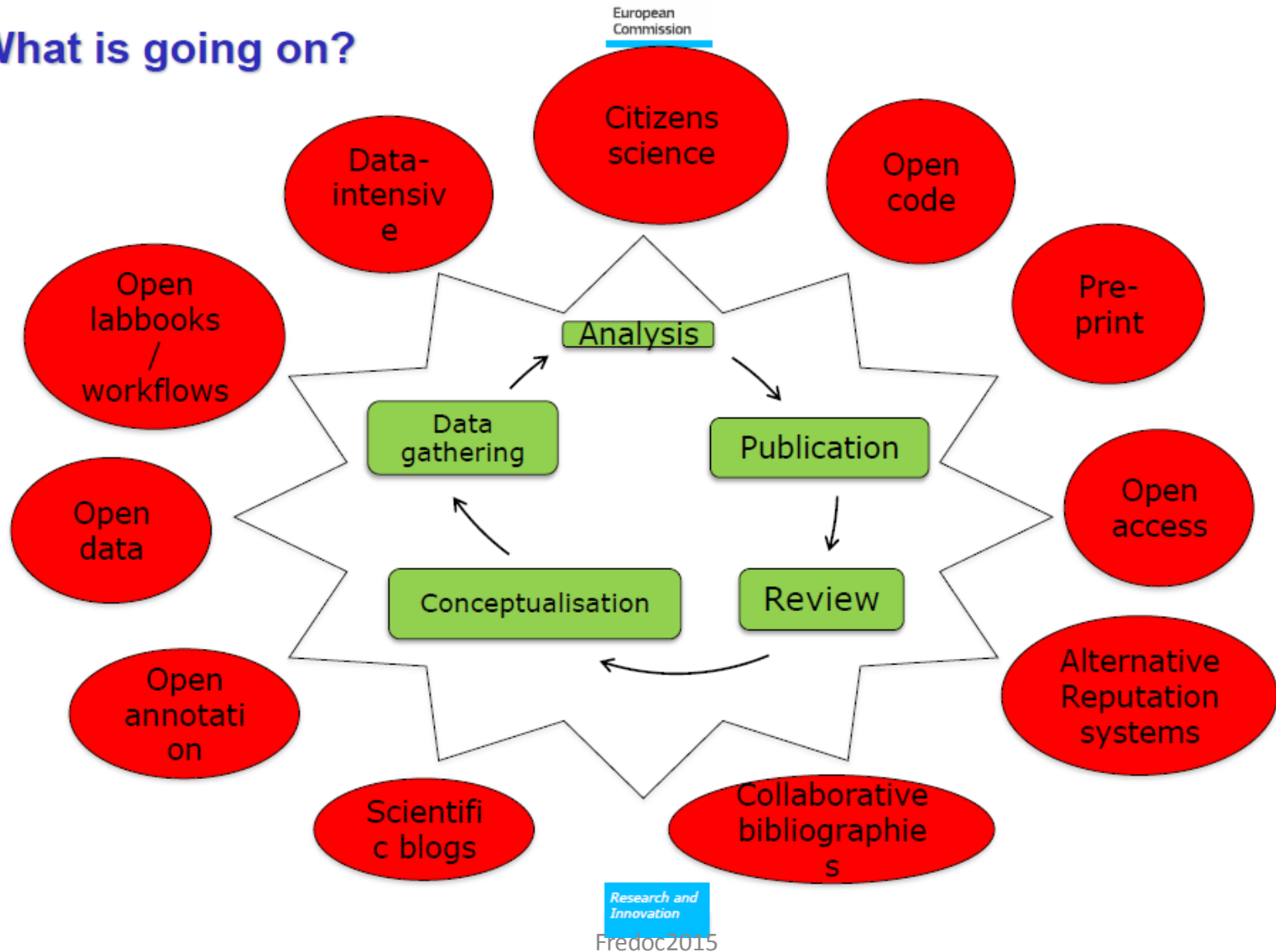
OPEN SCIENCE

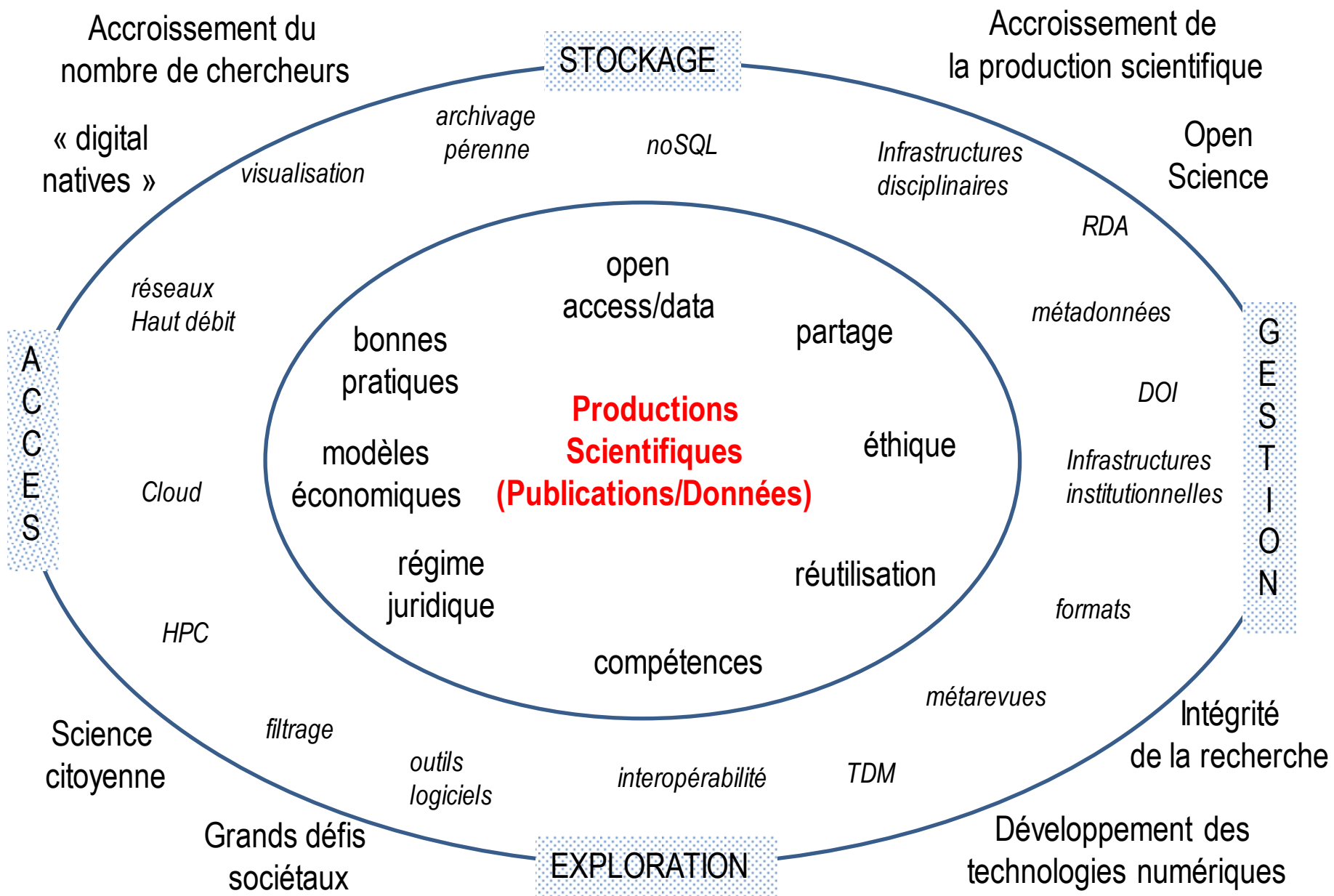
eScience

Tendances : accroissement de la production scientifique, du nombre de chercheurs, nouvelle façon de faire de la science, guidée par les données massives, importance des défis sociétaux

Science 2.0

2. What is going on?





Growing at different speed

Trend	Status	Data
Pre-print	Mature	916.000 articles in arXiv
Open access	Fast growing	Exponential growth of OA journals. Up to 50% of scientific publications are OA (depending on criteria: gold or green)
Data intensive	Fast growing	Helix-Nebula: the ScienceCloud Big science (CERN, EMBL, ESA) teams up with big business to provide cloud computing for big data
Citizen science	Medium growth	1 Mio. Zooniverse users
Open research data	Medium growth	29% of a sample of research funding organizations mandate open data archiving and 10% encourage it
Reference sharing	Medium-fast growth	2.7 Mio. users of Mendeley (reference-sharing tools), 500 million docs uploaded
Open code	Sketchy growth	21% of JASA articles make code available 7% journals require code
Open Notebook	Sketchy growth	Isolated projects

Tendances

Output Time	Bibliography	Data	First Analysis, working notes	Draft paper	Article	Comment on other people's work
Traditionally	Not public	Not Public	Not Public	Not public	Public	Internal, public only through articles
Emerging trend	Public	Public	Public	Public	Public	Public by all means and at all stages of work



Examples:

Mendeleey

DataNet

Science
blogs

Arxiv

OpenAire

Openannotation.org

Research and
Innovation

Impact sur l'IST

Vers un nouveau paradigme en IST:

nouveaux acteurs dominants, nouveaux modèles économiques, nouvelles technologies

- Papier, sociétés savantes, abonnements, catalogues
- Bases de données, édition numérique, discovery tools, concentration des éditeurs, big deals
- Mouvement Open Access, modèles OA en concurrence, data deluge, Web x .0 - Open science

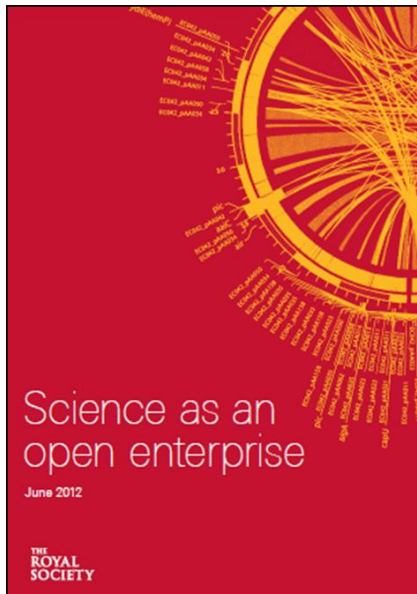


IST : un périmètre élargi

- De l'accès à la documentation extérieure
 - Discovery tools, textes intégraux, numérique, sélection, base de données, analyse de contenus, synthèses
- A la valorisation de la production interne
 - Archives ouvertes, référencement, enrichissement sémantique, plateformes éditoriales
- Pour activités de recherche, mais aussi
 - Administration de la recherche : CRIS, évaluation,
 - Politique scientifique : bibliométrie, scientométrie

IST : un périmètre élargi

- De l'accès à l'IST extérieure
 - Discovery tools, textes intégraux, numérique, sélection, base de données, analyse de contenus, synthèses, jeux de données, corpus,
- à la valorisation de la production interne
 - Archives ouvertes, référencement, enrichissement sémantique, plateformes éditoriales, plateformes d'annotation, curation de données, base de données factuelles,...
- Pour activités de recherche, mais aussi
 - Administration de la recherche : AO, évaluation,
 - Politique scientifique : bibliométrie, scientométrie
 - Plateformes coopératives : gérer, partager, réutiliser



“A particular dilemma for universities is to determine the role of their science libraries in a digital age.

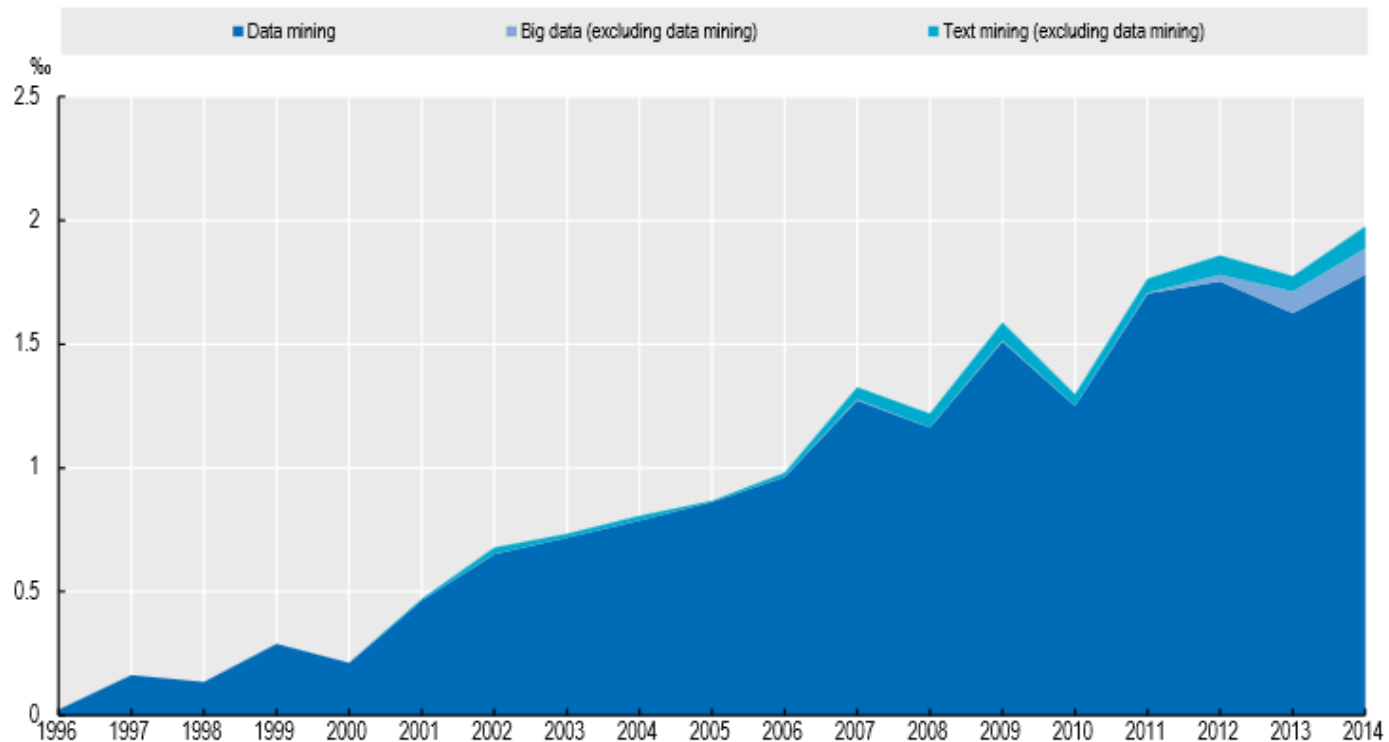
The traditional role of the library has been as a repository of data, information and knowledge and a source of expertise in helping scholars access them.

That role remains, but in a digital age, the processes and the skills that are required to fulfil the same function are fundamentally different. ”

Science as an Open Enterprise Report, Royal Society, UK

Articles traitant de TDM

1995-2014, per thousand article



Source: OECD (2014), *Measuring the Digital Economy: A New Perspective*, OECD Publishing, Paris.

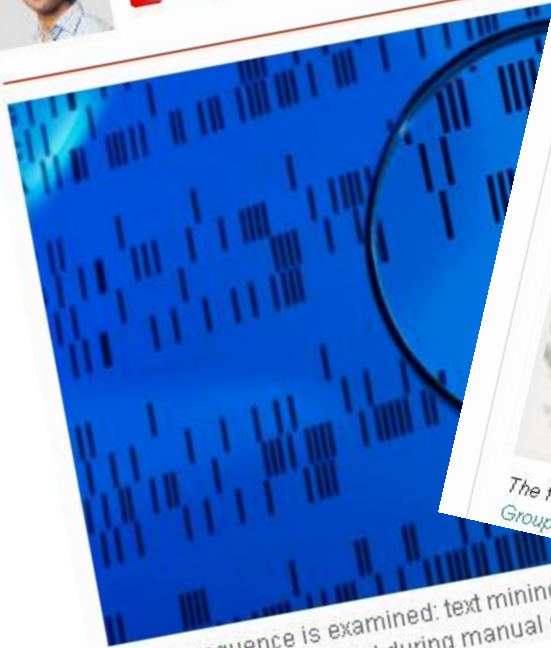
Text mining: what do publishers have against this hi-tech research tool?

Researchers push for end to publishers' default ban on computer scanning of tens of thousands of papers to find links between genes and diseases



Alok Jha, Science correspondent
The Guardian, Wednesday 23 May 2012 16

Jump to comments (62)



A DNA sequence is examined: text mining allows links that
never have been noticed during manual searches. Photograph
Belcher/Alamy

La fouille de textes

The Right to Read Is the Right to Mine

June 1, 2012 in Bibliographic, OKF Projects, Open Access, Open Content, Open Data, Open Science, Texts, WG Open
Bibliographic Data, Working Groups



The following is a draft content mining declaration developed by the Open Knowledge Foundation's Working
Group on Open Access

Services de base :

- Reconstruction de références citées
- Indexation automatique
- Extraction d'entités nommées
- Extraction terminologique

Services à valeur ajoutée :

- Enrichissement
- Fouille de texte (TDM)
- Jeu de données induites
- Nano publication
- Schéma de collaboration
- Schéma d'influence
- Analyse sémantique
- Analyse d'impact politique
- Document automatique

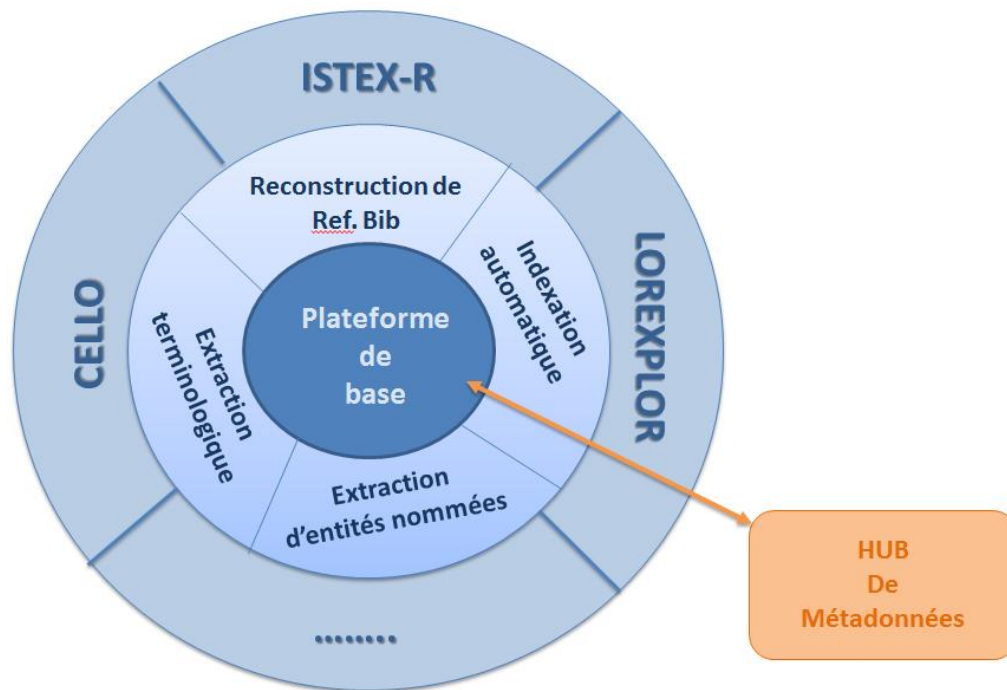
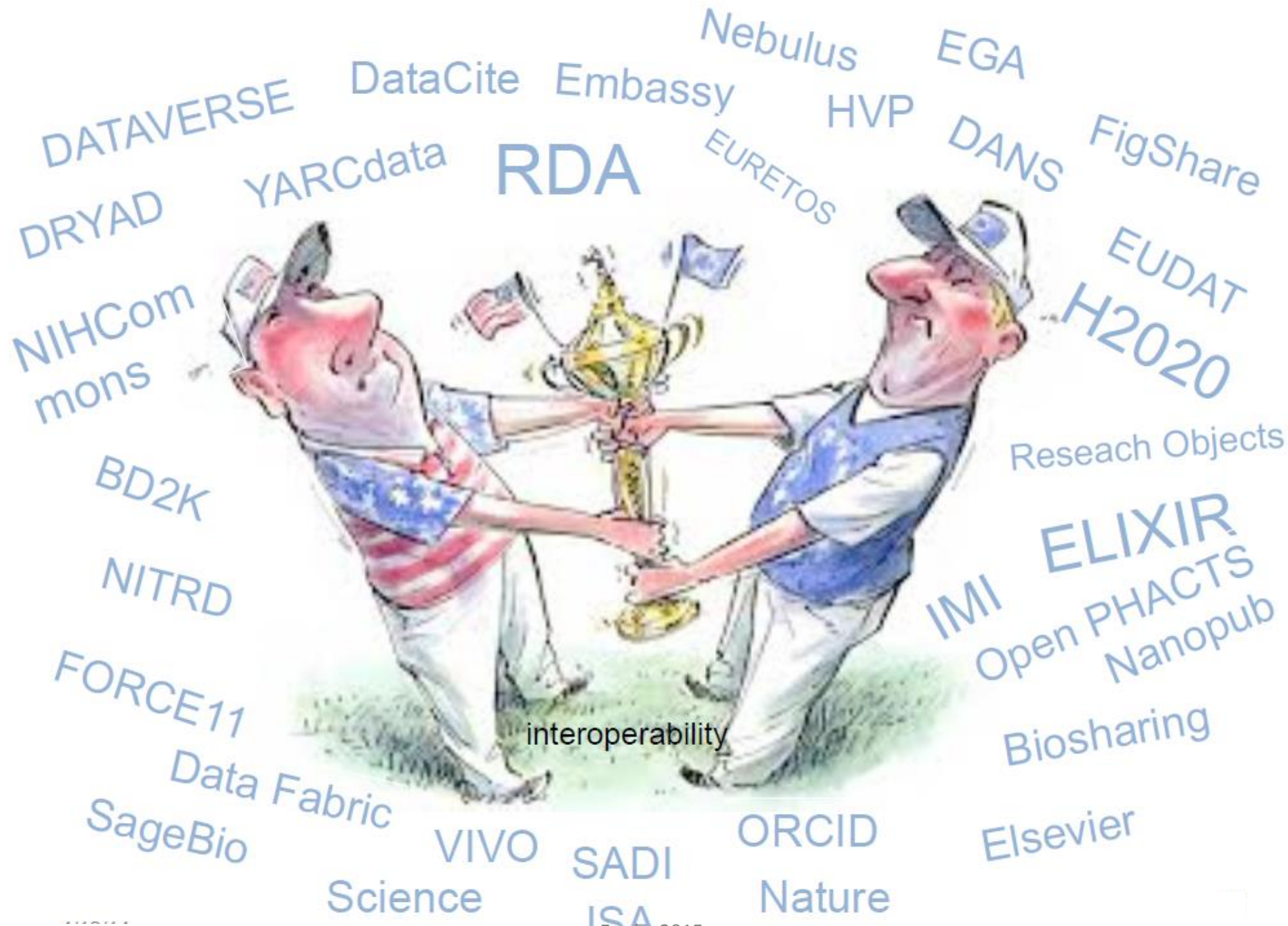


Schéma L.SCHMITT, CNRS/INIST
« Plateforme ISTEX – intégration et complémentarité »
Journées de l'ABES 21 mai 2014 Montpellier



RDA

interoperability

RDA : Research Data Alliance



De l'hiver suédois en 2013
à l'automne 2015 à Paris



Collaboration internationale, 50+ groupes de travail, 3500+ membres

Des freins

<http://ec.europa.eu/licences-for-europe-dialogue/node/7>



The screenshot shows the 'Licences for Europe' website header with the European Commission logo and the title 'LICENCES FOR EUROPE Structured stakeholder dialogue 2013'. Below the header is a navigation menu with tabs for 'Home', 'Working Groups -', '1st Plenary', 'Mid-term review', and 'Final Plenary Meeting'. The 'Working Groups -' tab is selected. Underneath, there is a 'Related content' section with a grid of links to various documents and presentations. The 'Text and Data Mining Working Group (WG4)' section is highlighted, containing a sub-section 'Text and Data Mining for Scientific Research Purposes (WG4)' with a paragraph of text.

European Commission

LICENCES FOR EUROPE
Structured stakeholder dialogue 2013

Home Working Groups - 1st Plenary Mid-term review Final Plenary Meeting

Related content

- WG4 Presentations, 6th meeting-14... Customer-Industrial Chemist-Perspective.pdfPublishers-Perspective-Initiatives.pdf
- WG4 Presentations, 3rd meeting-22... COM-open access.ppt ELRA-HumanLanguageTechnologies.pdf UK-planned exception.ppt CrossRef.pdf
- List of participants WG4 Working group 2 - List of participants (last updated: 13/12/2013)
- WG4 Conclusions from meetings conclusions-6th-meeting 14 October.docReport-4th-meeting.docReport-3rd-meeting.docConclusions-2nd-meeting.docConclusions-1st-meeting.doc
- WG4 Presentations, 4th meeting-29... STM-sample clause.ppt
- Intro-to-TDM-SAS.pptTDM-CCC.pptTDM-PLS.pptTDM-researcher.pdf

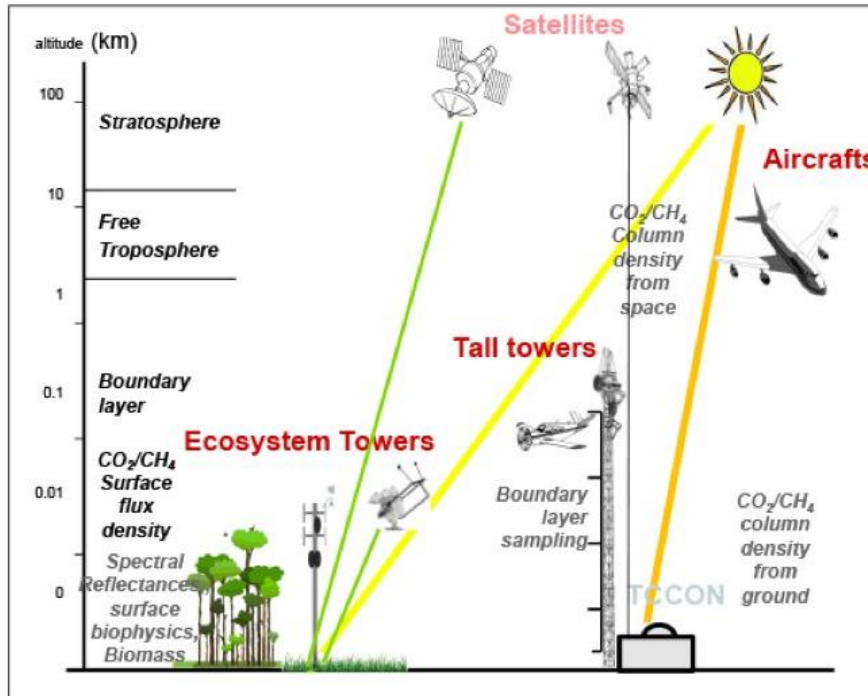
Text and Data Mining Working Group (WG4)

Text and Data Mining for Scientific Research Purposes (WG4)

The Commission's objective is to promote the efficient use of text and data mining (TDM) for scientific research purposes. TDM currently requires contractual agreements between users (e.g. typically research institutions) and rights holders (e.g. publishers of scientific journals) to establish the modalities for technical access to the relevant data sets. The Group should explore solutions such as standard licensing models as well as technology platforms to facilitate access.

Un simulacre de dialogue : les représentants des chercheurs et des bibliothèques ont quitté la discussion...

ICOS



Communiqué de presse..

Présentée comme une "infrastructure scientifique de classe mondiale" pour lutter contre le changement climatique, Icos mobilisera, à plein régime, 500 chercheurs, ingénieurs et techniciens en 2017, pour une **production de 1 000 articles scientifiques par an dès 2020**, espèrent ses promoteurs.

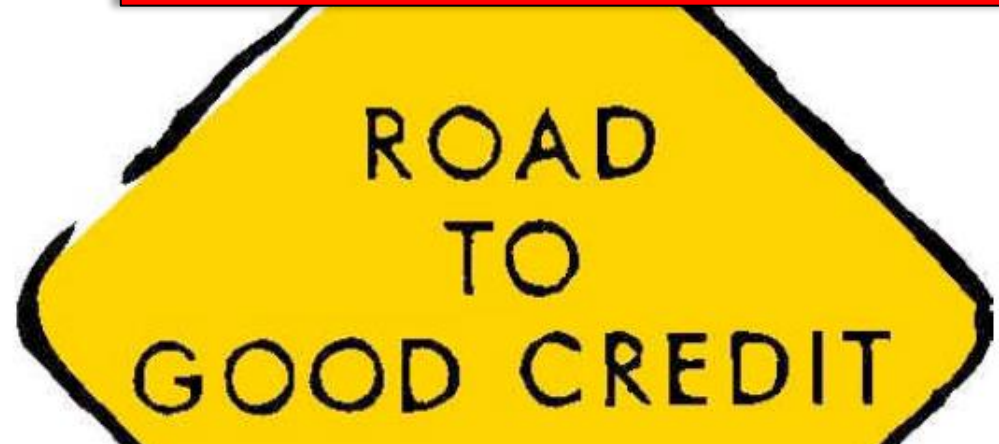
Le concept ICOS : une infrastructure d'observation intégrée à long terme des flux et concentrations des gaz à effet de serre d'origine anthropique

Et pourquoi ne pas plutôt mettre en avant l'effet démultiplicateur du partage des données ?

Rewards

- Publications
- Publications
- Publications
- Publications
- Publications
- Publications
- Grants
- Awards and honours
- Teaching
- Service
- Data

Manque de processus de rewarding liés aux productions de connaissance autres que les publications...



« To ensure successful adoption by scientific communities, open scientific research data principles will need to be underpinned by an appropriate policy environment, including **recognition of researchers fulfilling these principles** »

G8 Science Ministers Statement, 2013

Trust, Trust, Trust

- Certifications des publications
 - Peer reviewed journals
 - Les archives institutionnelles : preprints/postprints
 - Open peer reviewing
 - Le couplage publication/données
- Certifications des réservoirs : workflow, format, procédures qualité,...



En résumé

- IST acteurs
 - signalement des productions scientifiques
 - exploration de contenus textuels
 - Gestion/valorisation des données
 - développement des métriques alternatives
 - architecture des SI des labos



Garants des **processus de certification** des résultats scientifiques produits dans les laboratoires (publications ET données, blogs, carnets de labos,...)

Acteurs de la confiance, une légitimité à conquérir ?

