TRE

Comunicazione scientifica

Nuove funzioni

Web

Open data

Open Access

Nuove riviste

Disgregazione/riaggregazione

Nanopublications

Enhanced publications

Altmetrics

Workflows online

Data-driven Science

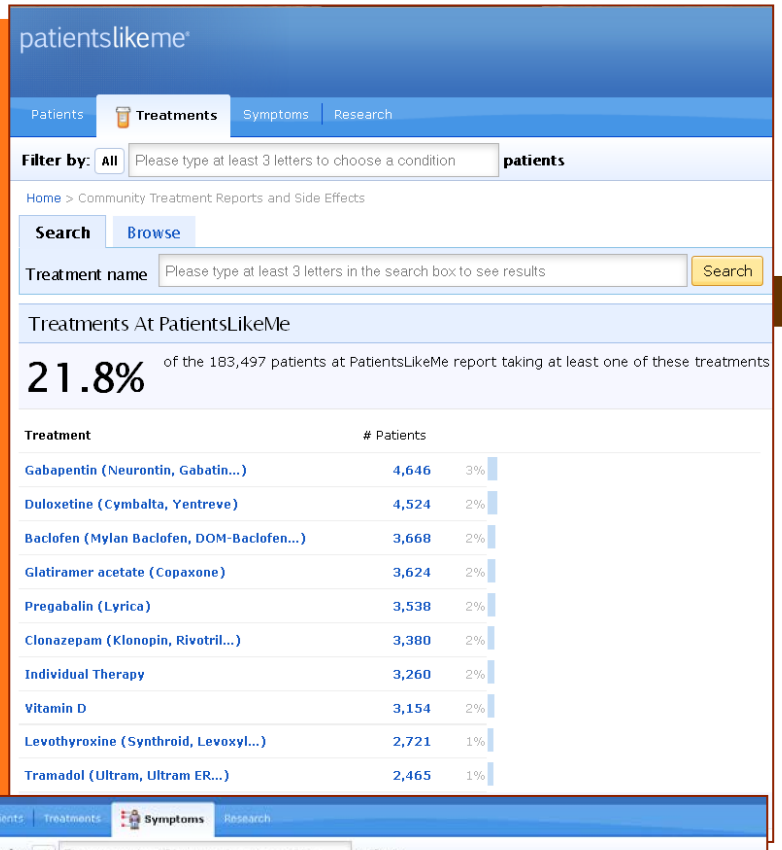
Cosa sono gli Open Data?

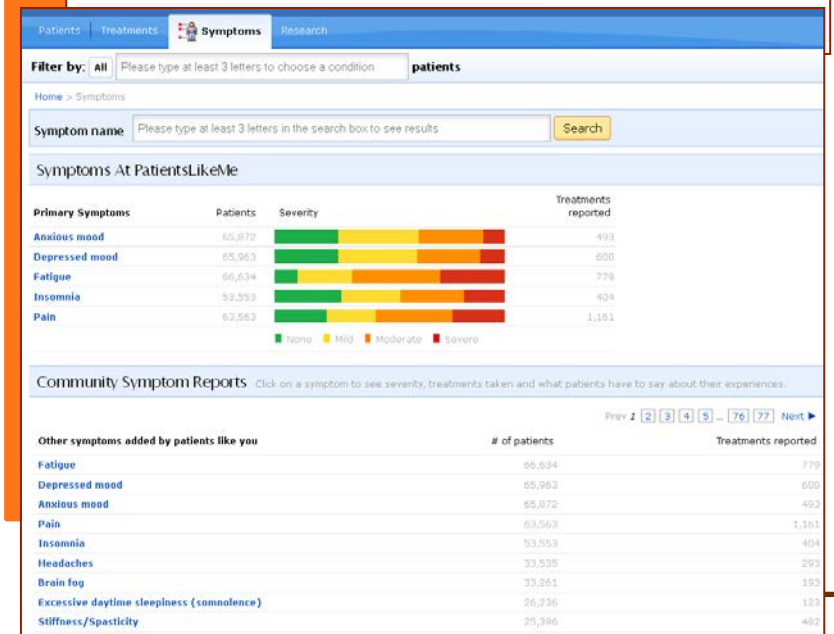
http://goo.gl/2jQB

Open Data al lavoro

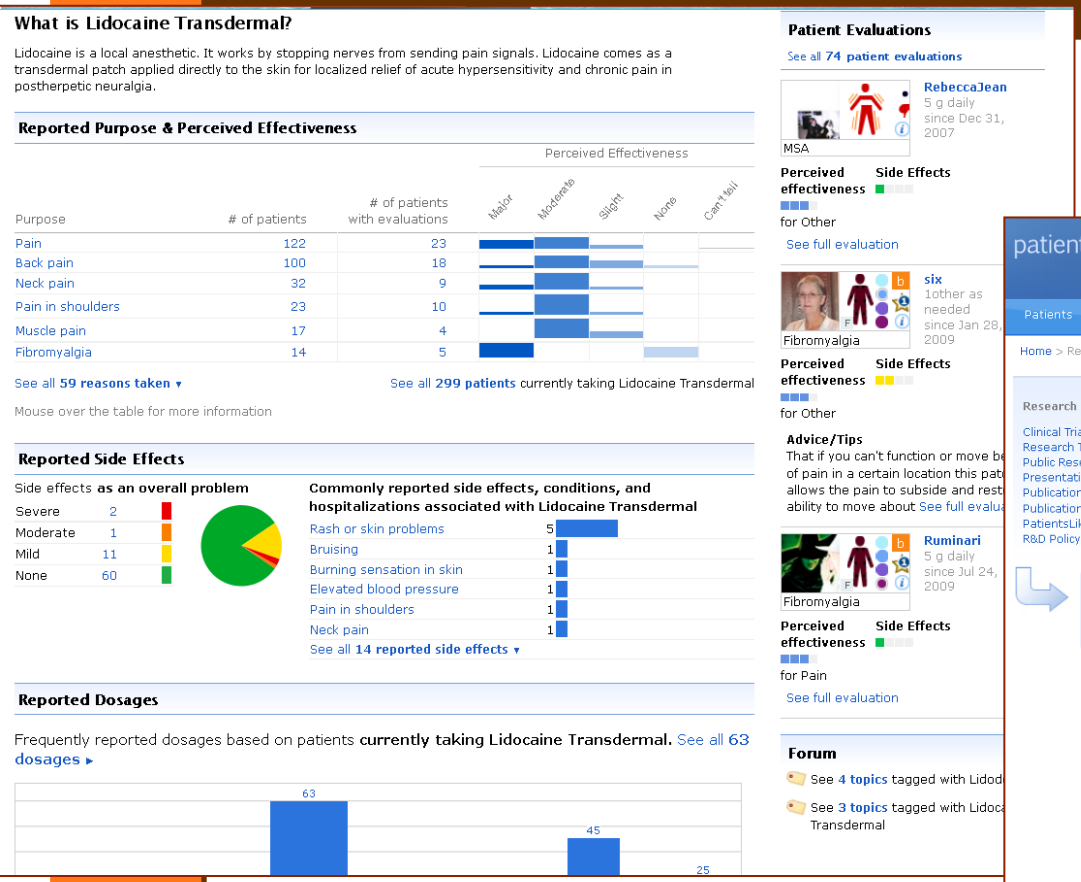
http://www.patientslikeme.com/

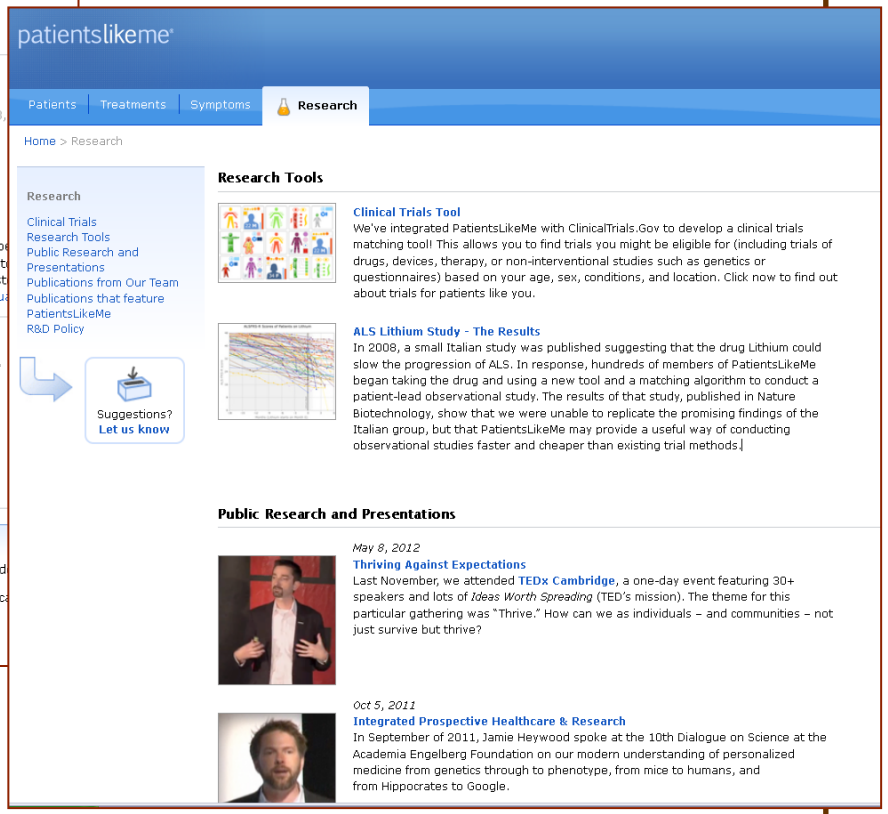
Open Data al lavoro





Open Data al lavoro





The value of openness

patientslikeme® | The Value of Openness Blog

Home | Archives | What is our Openness Philosophy?

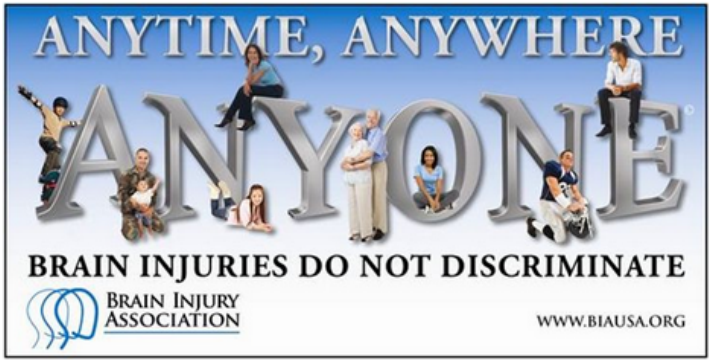
About

At PatientsLikeMe, people with every type of condition are coming together to share their health experiences, find patients like them and learn how to take control of their health. The result is improved care for patients as well as an acceleration of real-world medical research.

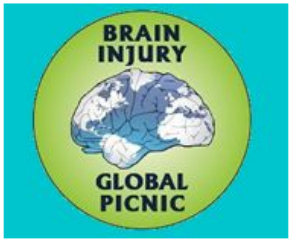
A Traumatic Brain Injury Can Happen to Anyone

Posted by admin | March 18, 2013

Did you know that 1.7 million Americans sustain a traumatic brain injury each year? Or that the term includes any type of blow, bump or jolt to the head or penetrating head injury that disrupts normal brain function?

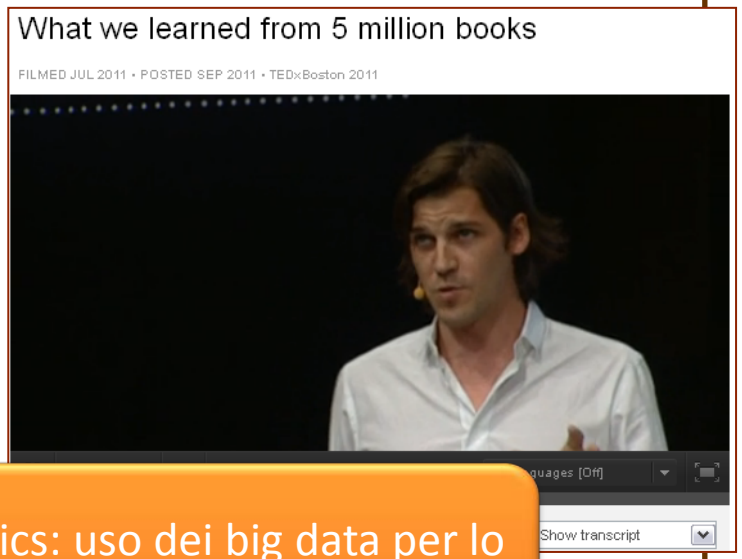


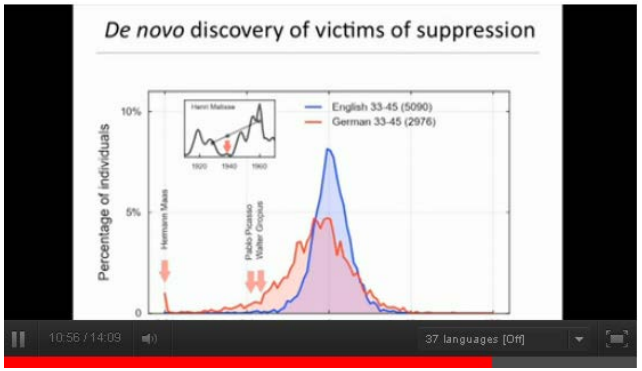
March is Brain Injury Awareness Month, a time for reinforcing the seriousness of head injuries, given that traumatic brain injuries are a contributing factor in a third of all injury-related deaths in the US. In addition, 3.1 million individuals are living with life-long disability as a result of a traumatic brain injury. Some common causes of these injuries are falls, car accidents, workplace accidents and assaults, but the effects can vary greatly from person to person. No two brain injuries are alike.



Are you living with a traumatic brain injury? Share your symptom and treatment histories with the 500+ members of PatientsLikeMe's traumatic brain injury community and discuss your experiences in our Injuries and Traumas Forum. In addition, consider raising awareness in your area by getting involved with the Brain Injury Global Picnic, which takes place September 21, 2013. The goal is to organize 1,000 picnics around the world – thus setting the Guinness World Record for the most people picnicking in a 24-hour period – to promote awareness, education and change.

Culturomics: big data and humanities

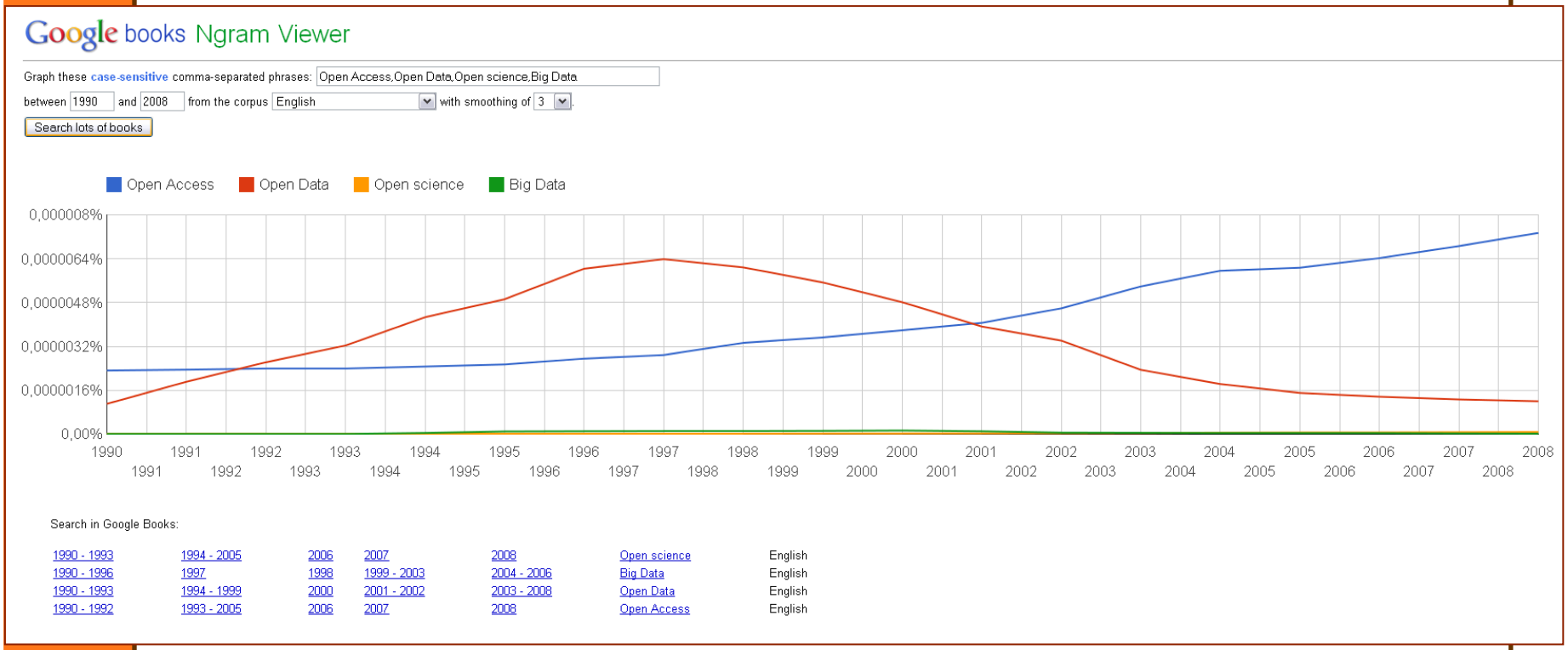




Culturomics: uso dei big data per lo studio della cultura umana

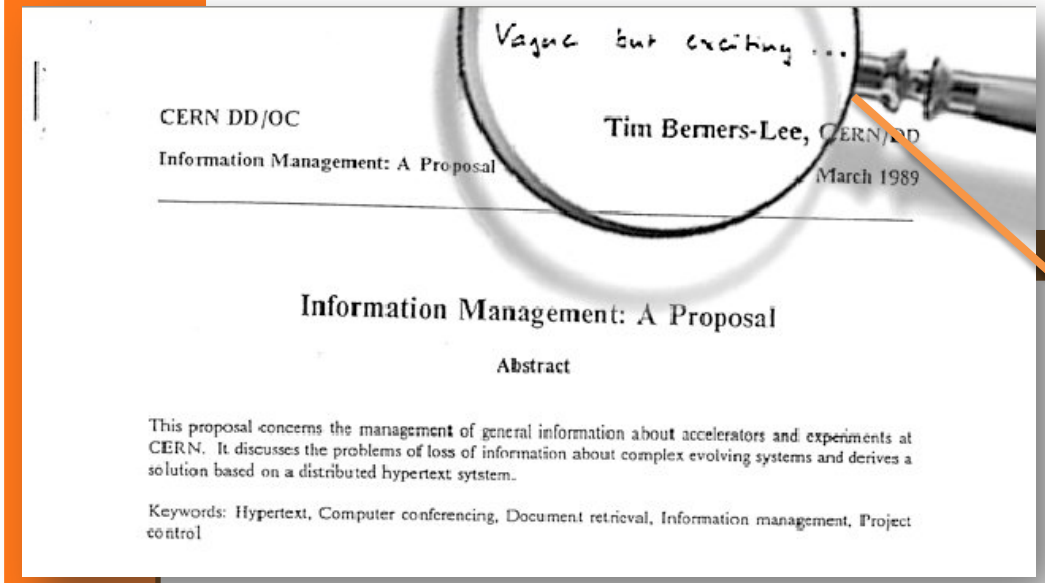
http://goo.gl/MaUIq

[Google Books Ngram Viewer]



http://books.google.com/ngrams

e se…?


Vague but exciting ...

CERN DD/OC

Tim Berners-Lee, CERN/DD

Information Management: A Proposal

March 1989

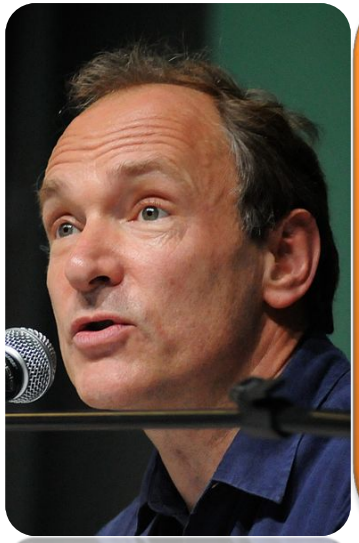
Information Management: A Proposal

Abstract

This proposal concerns the management of general information about accelerators and experiments at CERN. It discusses the problems of loss of information about complex evolving systems and derives a solution based on a distributed hypertext sytstem.

Keywords: Hypertext, Computer conferencing, Document retrieval, Information management, Project control

Vague, but exciting…

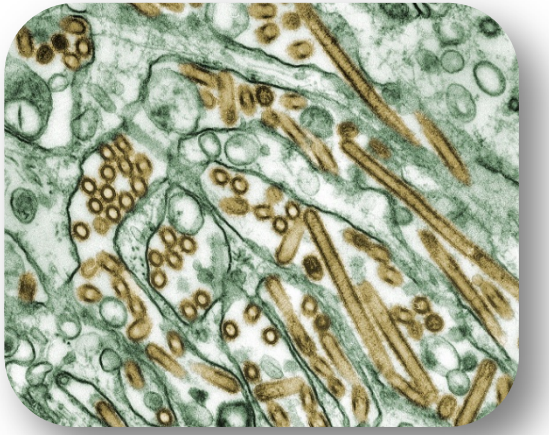


…nel 1989 Tim Berners Lee
o il CERN avessero deciso di
tenere **CHIUSO**
il protocollo HTTP e **i dati**
per il suo sviluppo?

e se…?

… nel 2006 Ilaria Capua non avesse deciso di depositare la sequanza del virus N5H1 (influenza aviaria) in GenBank?

…l'OMS non avrebbe adottato il suo approccio trans-disciplinare per le strategie pre-epidemiche, che riguarda la salute di tutti noi

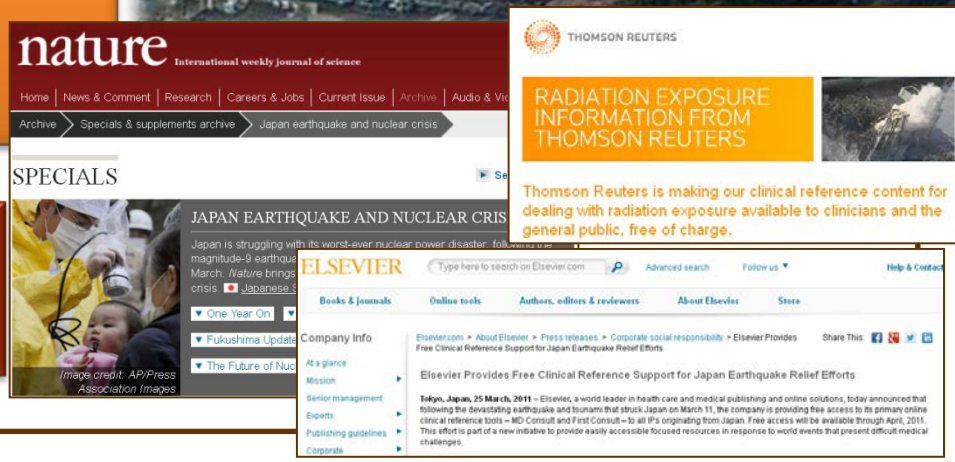


… e se?

11 marzo 2011:
tsunami e
incidente nucleare
a Fukushima

pochi giorni dopo,
gli editori scientifici
rendono disponibili
i dati e le ricerche
sull'inquinamento da
radiazioni

…significa che fino a quel momento erano CHIUSI…



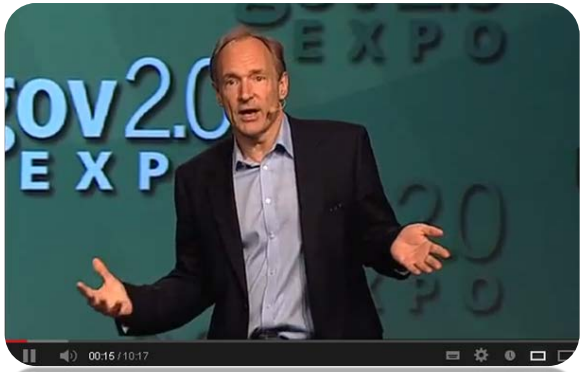
Open research data: chi

T.Berners-Lee, Raw data now, TED talks, 2009

«Put your data on the Web! Let's unlock our data and reframe the way we use it together»

«Data combination gives much better answers than data itself»

T.Berners Lee, The year Open data went worldwide, TED talks, 2010



«The data work only open and connected...»

Tim Berners-Lee, Open, Linked Data for a Global Community, Gov 2.0, 2010

Open research data: chi

«**Data** are the **new oil**»

«You can **give away your data now** – and **generate revenue and jobs**, and even **save money** from the better information and decisions that will flow»

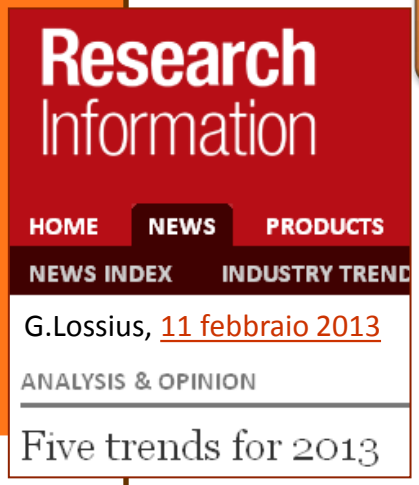
«**Big data means big opportunities**»

«Data become an infrastructure that scientists can use on their way to new frontiers»

Neelie Kroes, Blog

5 trends per il 2013

- Impact Factor: riposi in pace?
- Consolidamento del panorama editoriale
- Il momento degli e-textbook
- La crescita degli strumenti mobili
- IL FUTURO SONO I DATI



Data deluge

International innovation, Dec 2012

Financial Times Magazine, 1 febb 2013

IEEE Spectrum, Feb 2011

I dati creano ponti…

• dati sempre associati a termini negativi come “diluvio”, “eccesso”
• NON DOBBIAMO OPPORRE RESISTENZA MA “ABBRACCIARLI”
• facilità di gestione di dati fa girare melgio il commercio globale
• **I DATI CREANO PONTI FRA LE DISCIPLINE**
• **I DATI DISSOLVONO LE BARRIERE,**
APRONO NUOVI CANALI DI COMUNICAZIONE, LINEE DI RICERCA E OPPORTUNITÀ COMMERCIALI

WORLD ECONOMIC FORUM

It’s a funny thing about data: If you look up “data” in the literature, it invariably comes laden with negative baggage – words such as “overload” and “deluge”. With terms such as these, no wonder scientists and scholars alike shy away! But instead of resisting large data sets, we should be embracing them. Multinational companies are learning how to work with large data sets, and they’re getting better at it all the time. But mid-size and smaller businesses don’t have that expertise. Yet the largest companies will benefit greatly if smaller businesses, many of them key to their supply, production and workforce, can, too, handle massive data. Such ease of data handling can grease the wheels of global commerce.

To realize this vision, data and data awareness needs to be integrated into the educational curriculum. We need to teach students how to use massive amounts of data, in higher-education curriculums and even at secondary-school levels. Our educational focus must change, or our children won’t be prepared to handle the vast amounts of data that they invariably will encounter every day.

Data creates a bridge between traditional disciplines, spawning discovery and innovation from the humanities to the hard sciences. Data dissolves barriers, opening up new channels of communication, lines of research, and commercial opportunities. Data will be the engine, the spark to create a better world for all.

Jan Hesthaven, professor of applied mathematics at Brown University.

http://goo.gl/ExaGW

Bridging the gap

nei primi 60 secondi in cui leggete questo articolo, 1 miliardo di Gigabytes sarà transitato sulla rete

equivale a un decimo delle informazioni contenute nella Library of Congress

nei prossimi sette anni questi numeri cresceranno di dieci volte

I dati crescono più rapidamente di quanto facciano gli strumenti adatti ad archiviarli e scambiarli

è un gap da colmare, quindi un'opportunità...

Forbes · New Posts · Most Popular 29 Youngest Billionaires · Lists The World's Billion

CIO Network
INSIGHTS AND IDEAS FOR TECHNOLOGY LEADERS.
+ Follow (495)

TECH | 8/23/2012 @ 6:15PM | 2,462 views

Bridging The Data Deluge Gap

Eric Savitz, Forbes Staff

1 comments, 1 called-out + Comment Now + Follow Comments

Guest post written by Abhi Talwalker

Abhi Talwalkar is CEO of LSI Corp.


Abhi Talwalkar

In the first 60 seconds of reading this article, 1 billion gigabytes of information will flow across mobile networks around the world. That's the equivalent of a tenth of all the information contained in the Library of Co ss ve flow of ill IDC. drives or servers – also is expected to expand at an incredibly rapid ra s individuals and businesses, we are all dealing with the impact of this da deluge.

A. Talkwalker, Bridging the data deluge gap, Forbes 23 ago 2012

Science is an open enterprise

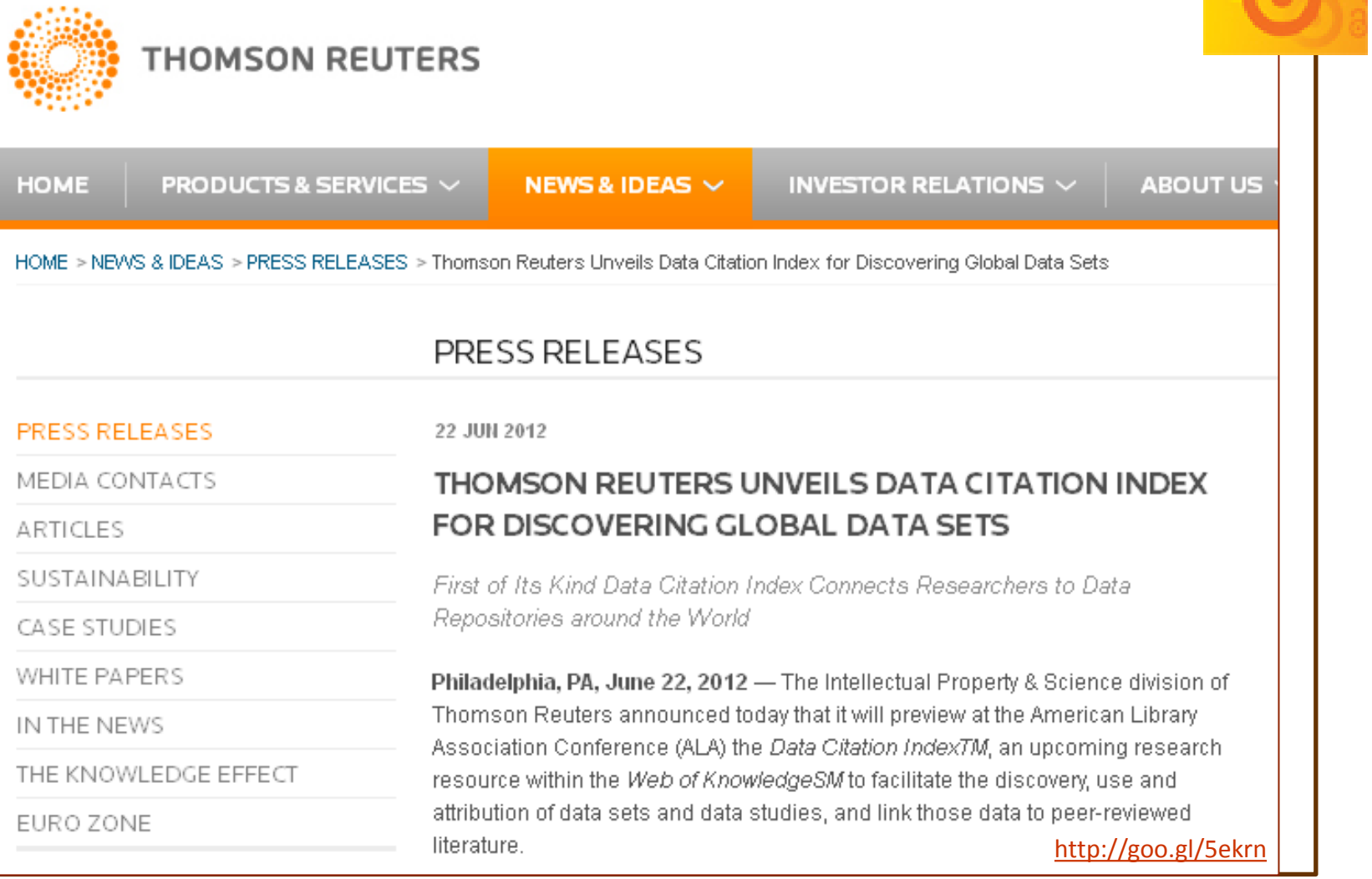
- preservare il principio di “apertura”
- SFRUTTARE I DATI POTENZIALMENTE PUÒ FAVORIRE UNA SECONDA RIVOLUZIONE “OPEN SCIENCE”

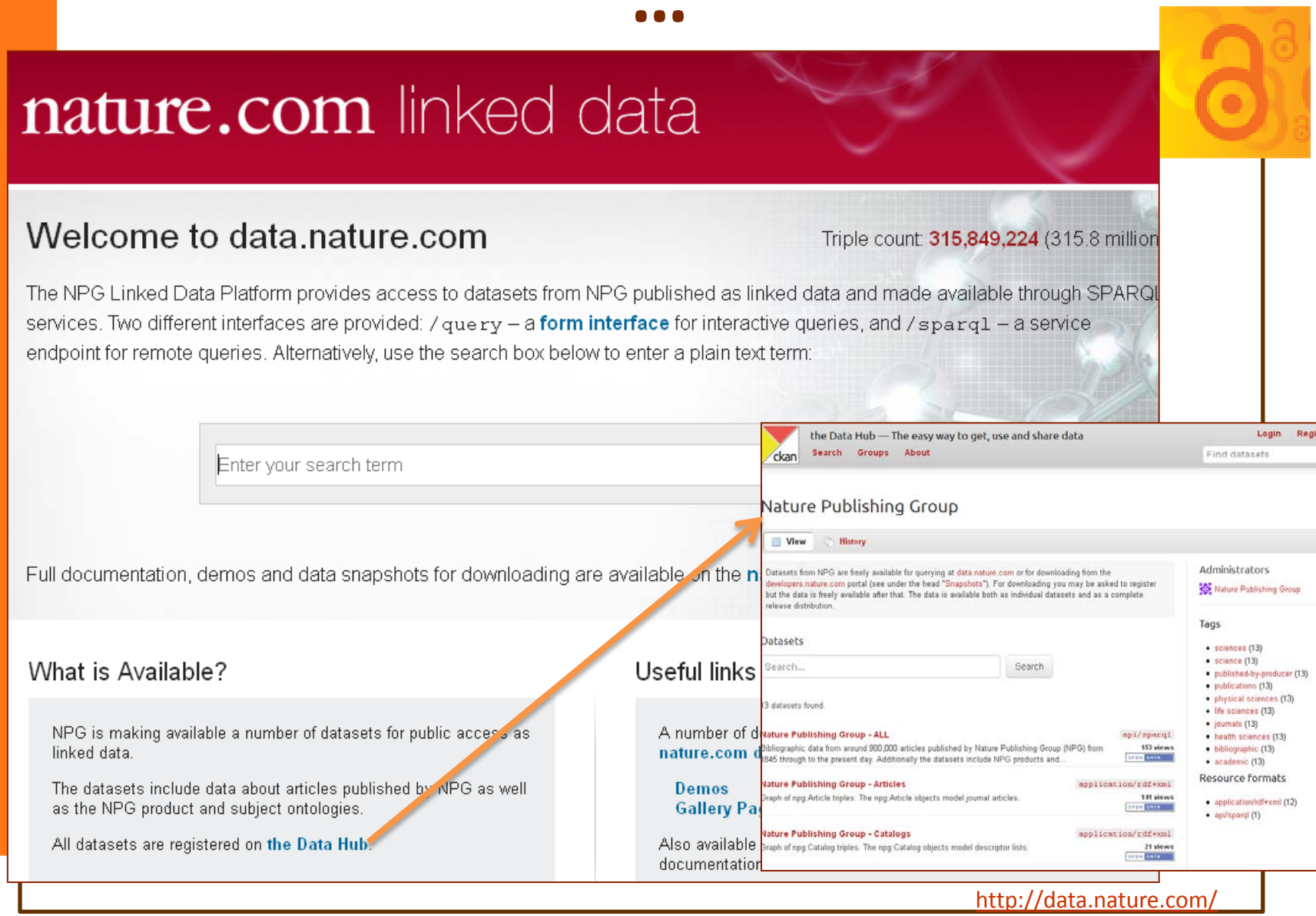
Areas for action

Six key areas for action are highlighted in the report:

- Scientists need to be more open among themselves and with the public and media
- Greater recognition needs to be given to the value of data gathering, analysis and communication
- Common standards for sharing information are required to make it widely usable
- Publishing data in a reusable form to support findings must be mandatory
- More experts in managing and supporting the use of digital data are required
- New software tools need to be developed to analyse the growing amount of data being gathered

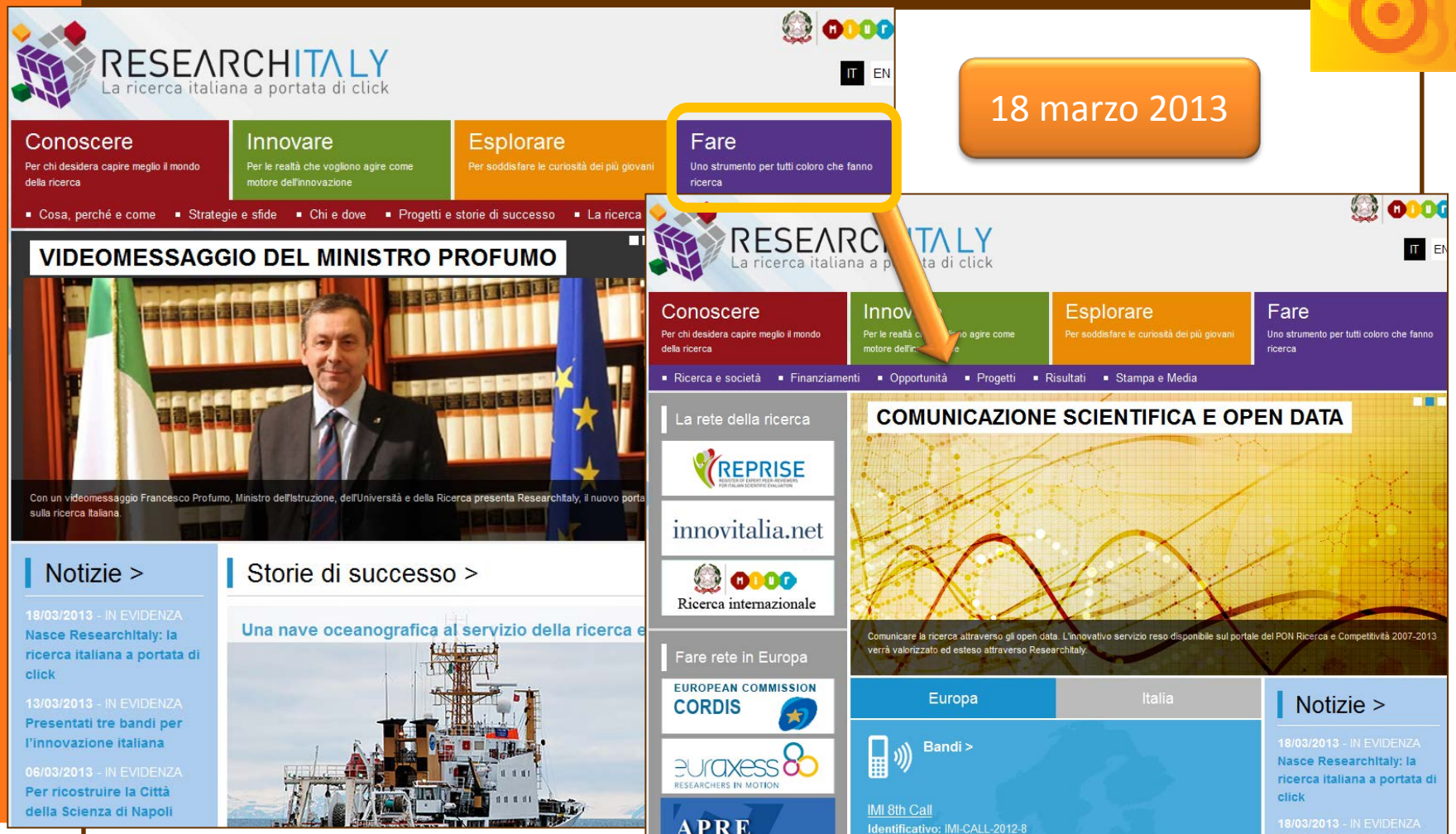
Royal Society, Science is an open enterprise, Report, June 2012


THOMSON REUTERS
HOME
PRODUCTS & SERVICES
NEWS & IDEAS
INVESTOR RELATIONS
ABOUT US
HOME > NEWS & IDEAS > PRESS RELEASES > Thomson Reuters Unveils Data Citation Index for Discovering Global Data Sets
PRESS RELEASES
PRESS RELEASES
MEDIA CONTACTS
ARTICLES
SUSTAINABILITY
CASE STUDIES
WHITE PAPERS
IN THE NEWS
THE KNOWLEDGE EFFECT
EURO ZONE
22 JUN 2012
THOMSON REUTERS UNVEILS DATA CITATION INDEX FOR DISCOVERING GLOBAL DATA SETS
First of Its Kind Data Citation Index Connects Researchers to Data Repositories around the World
Philadelphia, PA, June 22, 2012 — The Intellectual Property & Science division of Thomson Reuters announced today that it will preview at the American Library Association Conference (ALA) the Data Citation IndexTM, an upcoming research resource within the Web of KnowledgeSM to facilitate the discovery, use and attribution of data sets and data studies, and link those data to peer-reviewed literature.
http://goo.gl/5ekrn



http://data.nature.com/

18 marzo 2013



https://www.researchitaly.it/conoscere/

Opening up science

non ci può essere scienza moderna senza condivisione

nuovi strumenti ci possono schiudere l'era della Open science

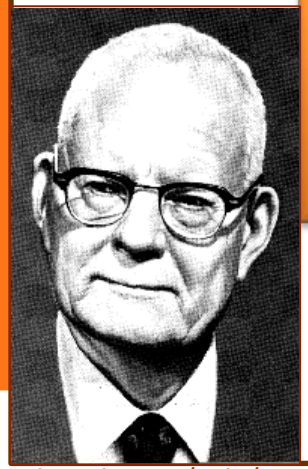
EU sta supportando Open Science

…certo richiede investimenti e anche in tempo di crisi abbiamo deciso di aumentarli, per investire sul futuro

…ma i soldi sono dei taxpayers, che quindi devono trarne beneficio

diretto
(Open Access/ Open Data)

indiretto
scienza migliora la vita



deve essere una rivoluzione condotta dal basso (ricercatori)

deve avere dimensione globale

In God we trust, all others bring data

http://en.wikipedia.org/wiki/W._Edwards_Deming

N. Kroes, 18 marzo 2013 http://goo.gl/RDTi8

Il quarto paradigma: la scienza dei dati

Emergence of a Fourth Research Paradigm

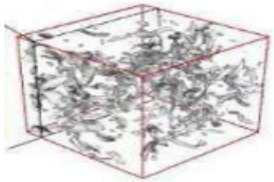
Thousand years ago – **Experimental Science**
- Description of natural phenomena

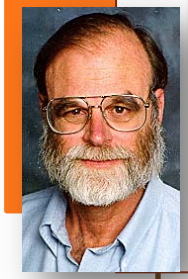
Last few hundred years – **Theoretical Science**
- Newton's Laws, Maxwell's Equations...

Last few decades – **Computational Science**
- Simulation of complex phenomena

Today – **Data-Intensive Science**
- Scientists overwhelmed with data sets from many different sources
  - Captured by instruments
  - Generated by simulations
  - Generated by sensor networks

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



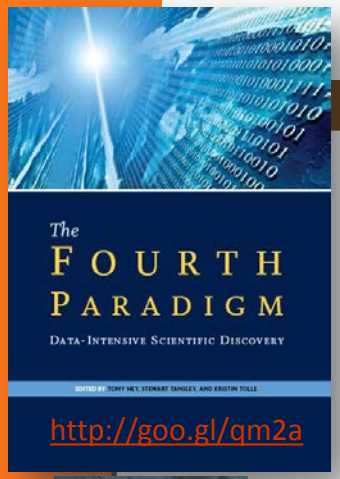


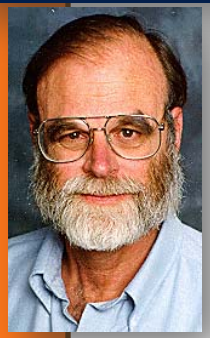
eScience is the set of tools and technologies to support data federation and collaboration
- For analysis and data mining
- For data visualization and exploration
- For scholarly communication and dissemination

Thanks to Jim Gray

T. Hey, Open Access, Open Data e Open Science, Copenhagen, June 2012

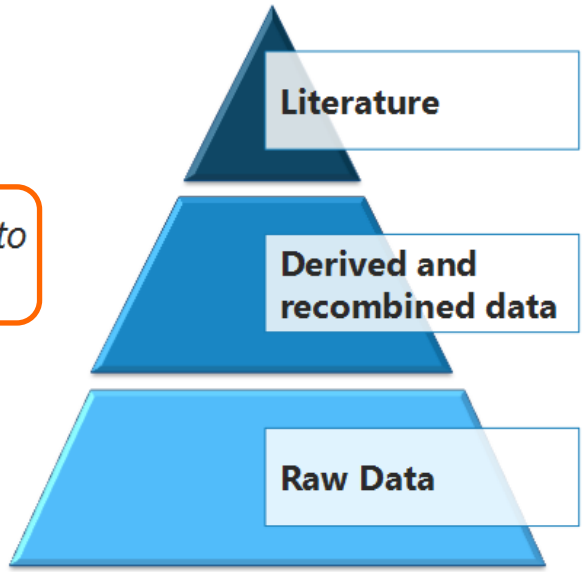
Il quarto paradigma





All Scientific Data Online

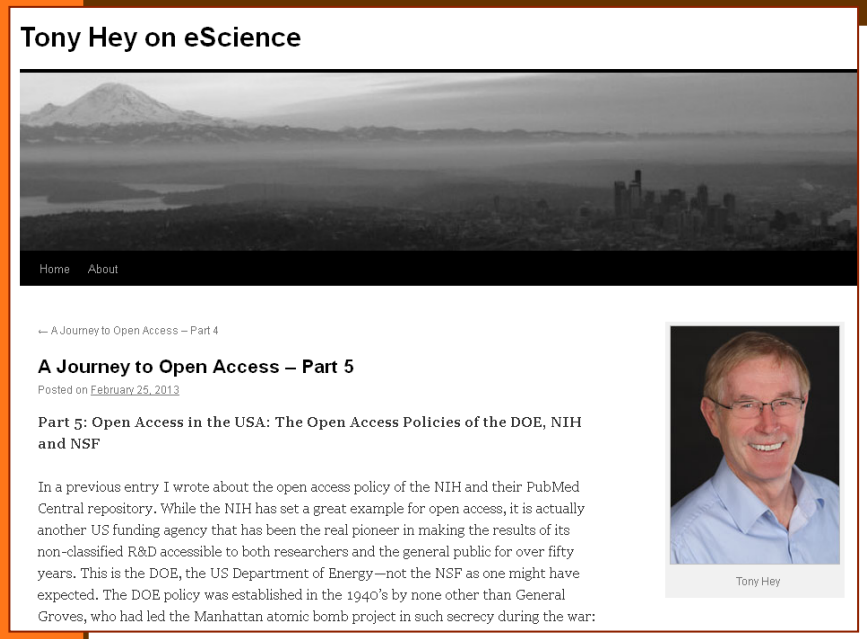
- Many disciplines overlap and use data from other sciences.
- Internet can unify all literature and data
- Go from literature *to* computation *to* data *back to* literature.
- Information at your fingertips – For everyone, everywhere
- Increase Scientific Information Velocity
- Huge increase in Science Productivity



(From Jim Gray's last talk)

Microsoft Research Connections

[Open Access e Open Data]


Tony Hey on eScience

Home About

← A Journey to Open Access – Part 4

A Journey to Open Access – Part 5

Posted on February 25, 2013

Part 5: Open Access in the USA: The Open Access Policies of the DOE, NIH and NSF

In a previous entry I wrote about the open access policy of the NIH and their PubMed Central repository. While the NIH has set a great example for open access, it is actually another US funding agency that has been the real pioneer in making the results of its non-classified R&D accessible to both researchers and the general public for over fifty years. This is the DOE, the US Department of Energy—not the NSF as one might have expected. The DOE policy was established in the 1940's by none other than General Groves, who had led the Manhattan atomic bomb project in such secrecy during the war:

Tony Hey

in questo scenario,
giocano un ruolo primario
• gli archivi Open Access
• le politiche degli enti di finanziamento

vedi l'ultimo discorso di Jim Gray:
due rivoluzioni,
• il quarto paradigma
• la futura rivoluzione nella comunicazione scientifica ovvero le politiche di Open Access dell'NIH

vision: an open access world of full text publications and data, a global digital library that can truly accelerate the progress of science

T.Hey, A journey to Open Access

Il quarto paradigma

Condizioni:

- I ricercatori devono cooperare su standard per la provenienza dei dati, la loro gestione e conservazione
- la ricerca scientifica deve orientarsi verso la condivisione dei dati come condizione irrinunciabile
- I processi di pubblicazione devono essere più flessibili, in tempo reale, e collaborativi
- tutti gli enti di ricerca pubblici e privati devono condividere le spese per creare e mantenere l'infrastruttura tecnologica per pubblicare e rendere accessibili I datasets
- l'infrastruttura e i servizi per i dati devono essere riconosciuti come beni fondamentali per la scienza

Il quarto paradigma

considerare e offrire strumenti per l'intero CICLO DI VITA DEI DATI

allo studio un add-in per MS Excel per data curation

Acquisition and modelling

Archiving and preserving

Collaboration and visualisation

Data

Disseminating and sharing

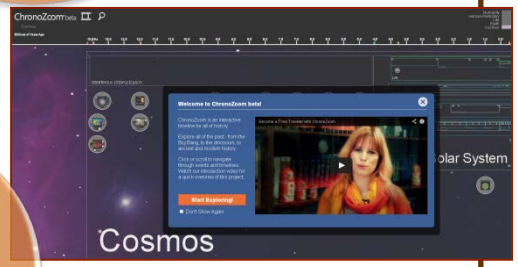
Analysis and data mining

- For organizing their data
  - DCXL , EZID
- To keep their data safe
  - Merritt, Micro-services
- To help them get grants
  - Data Management Plan tool
- To help get their work noticed
  - EZID, Data Papers
- To help them find other data
  - EZID, Data Papers

servizi offerti dalla Univ. of California

http://www.planets-project.eu/

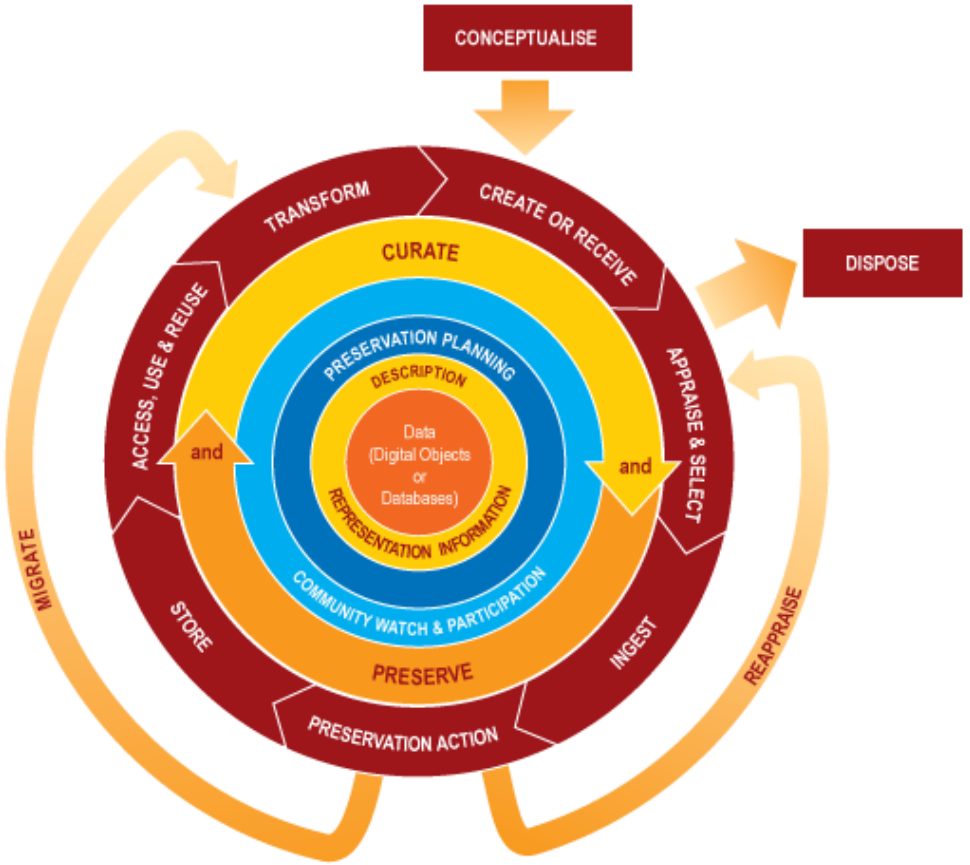
archivio distribuito, http://www.dataone.org/



http://www.chronozoomproject.org/

T.Hey, Open Access, Open Data e Open Science, Copenhagen, June 2012

Data life cycle



Key elements of the DCC Curation Lifecycle Model

http://www.dcc.ac.uk/resources/curation-lifecycle-model

Openness as infrastructure / 1

available to us. The papers have become finely tuned objects where some of the text is used to show the author understands the existing paradigm of the field, some of the text is used to describe the methods and results, and some is used to describe the implications. Each of these sections needed to be terse, as paper was expensive to print and ship.

This hid the fact that science was, in fact, actually much more like a wiki. Every topic in science is open for back and forth, and new discoveries spark rounds of editing and re-editing, and the print equivalent of flame wars in biting letters to the editor. But it was a wiki that was camouflaged as physical media. And in an era of increasingly computerized science, with automated and massively parallel lab equipment pumping data into massively parallel processing power, we're starting to see an absolutely overwhelming increase in the number of digital papers. Leaving behind the irony of digital paper, there is a strong

Science is drowning in its own outputs, and a lot of those outputs are turning out to be either non-reproducible [2] or downright false [3].

What we need is a full-scale revolution in the way that we publish knowledge, and there are many claimaints to carry the standard of that revolution. Some are from the radical incrementalism school(Endnote 2)-into which I would put Open Access, a movement that puts literature

validity in the peer review process [5]. Others go farther, arguing for the abandonment of the article as the core unit of knowledge transfer, for nano-publication of individual assertions [6], for the publication of figures or data rather than articles [7], for the rise of wiki science and the end of peer review entirely [8].

per questa rivoluzione,
in uno scenario della scienza sempre più data-intensive,
per cui non valgono gli stessi parametri di un documento di testo, mancano tre livelli di infrastruttura…

Openness as infrastructure / 2

Infrastruttura per collaborazioni scientifiche [virtual laboratories, distributed computing…]

Infrastruttura per classificazione formale dei dati [per il riuso]

Infrastruttura legale per l'APERTURA DEI DATI [licenze adatte]

…tutto questo consente la riproducibilità…

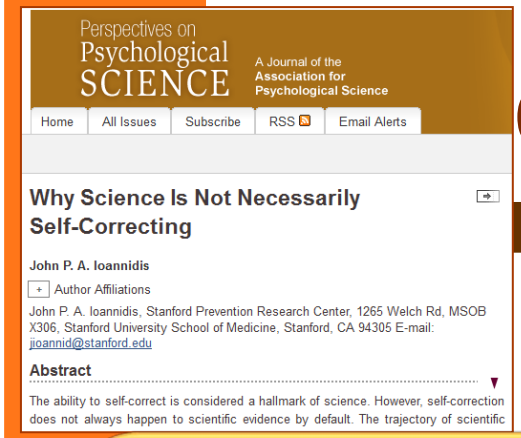
Taken together, these three skeins of collaboration, classification, and openness draw us inevitably towards the long-claimed, but rarely-achieved, goal of the scientific method: to make claims that are reproducible under similar circumstances by someone other than the claimant, to be reproducible.

Openness as infrastructure / 3

But open data will in the end win out, just as open systems have won out for networking, for document sharing, for software, and are beginning to win for culture and education. It is, in the end, the better way to do science, one in which there is less duplication of effort, less fraud, more reproducibility, more return on investment, and faster times to market of knowledge. It is, moreover, one that returns scientific data to its most natural state, one that is a pure public good, that gains more value as more people possess it.

J. Wilbanks, Openness as infrastructure, Journal of Cheminformatics 2011, **3**:36

Open data = riproducibilità


Perspectives on Psychological SCIENCE
A Journal of the Association for Psychological Science

Home | All Issues | Subscribe | RSS | Email Alerts

Why Science Is Not Necessarily Self-Correcting

John P. A. Ioannidis

Author Affiliations

John P. A. Ioannidis, Stanford Prevention Research Center, 1265 Welch Rd, MSOB X306, Stanford University School of Medicine, Stanford, CA 94305 E-mail: jioannid@stanford.edu

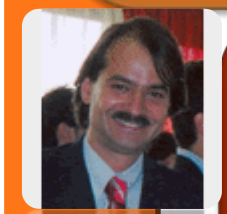
Abstract

The ability to self-correct is considered a hallmark of science. However, self-correction does not always happen to scientific evidence by default. The trajectory of scientific

scienza è approssimarsi alla verità,
NON pubblicare risultati eclatanti (e magari falsi)

oggi **riproducibilità è tendente a zero…**
MA riproducibiltà è ciò che dà fiducia nella scienza

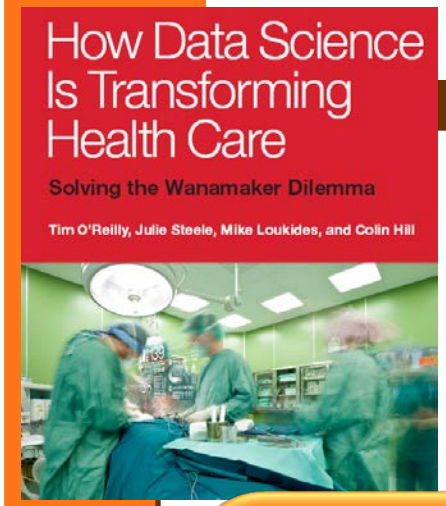
possibili strumenti utili:
- registro degli esperimenti, suddivisi fra ricerca esplorativa e di conferma
- peer review basata su solidità scientifica e non su "pubblicabilità"
- metriche per ciò che si può replicare
- contare gli sforzi di replica
- ASSOCIARE DATI GREZZI E WORKFLOW



la medicina rischia di distruggere la civiltà se le spese per una sanità inefficiente basata su nessuna o limitata o gonfiata prova scientifica continuano a salire esponenzialmente…
è essenziale che ci siano esperimenti sulla reale efficiacia e sui benefici reali

J. Ioannidis, Why Science Is Not Necessarily Self-Correcting, Perspectives on Psychological Scienc nov. 2012, 7 (6):645-654

[in medicina…]



«I know that half of my advertising doesn't work. The problem is that I don't know which half»
John Wanamaker

Google ha risolto il problema proponendo pubblicità basate sui click (i dati) dei singoli utenti…

In medicina è lo stesso:
ci sono cure standard per pazienti "medi",
che però non funzionano su tutti i pazienti.
Solo che finora non sapevamo su quali no.
Ora invece si possono raccogliere dati per capire le differenze, e quindi prevedere su quali pazienti può funzionare e su quali no

USA: 600 miliardi di $ in cure inefficaci

raccogliendo ed elaborando i dati e confrontando i clinical trial con le osservazioni reali sui pazienti quotidiani si può capire
quali cure sono efficaci e pagare solo per quelle!!!

outcome-based payment model

http://goo.gl/CkPIM

Riding the wave… the rising tide of data

Vision per il 2030

✓ il pubblico ha accesso ai dati e ne fa un uso creativo

✓ tutti gli attori coinvolti saranno consapevoli dell'importanza di CONSERVARE e CONDIVIDERE i dati prodotti durante le ricerche

✓ tutti i ricercatori e i professionisti saranno in grado di TROVARE, ACCEDERE E PROCESSARE i dati di cui hanno bisogno

http://goo.gl/bEwYH

✓ rete internazionale di archivi di dati, interoprabili, con standard condivisi, per L'AFFIDABILITÀ E L'INTEGRITÀ

✓ i finanziamenti pubblici cresceranno, dopo aver visto i ritorni sugli investimenti derivanti dal maggiore USO E RIUSO DEI DATI prodotti con fondi pubblici

✓ scambio di dati fra settore pubblico e privato porterà beneficio a entrambi

✓ politici potranno prendere decisioni più informate basate su solide evidenze

Riding the wave…

✓ necessario un quadro internazionale a supporto di una infrastruttura collaborativa per la gestione dei dati
✓necessario finanziamento (ingente)
✓necessari sistemi di incentivo e premio per chi mette a disposizione i dati

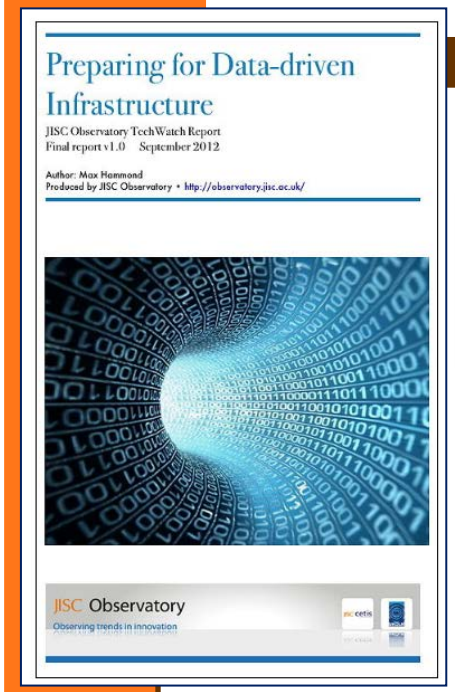
http://goo.gl/bEwYH

We believe that we all benefit from a far-seeing, collaborative and open approach to science and the e-infrastructure to support it. We urge action now.

Who benefits from scientific e-infrastructure

Beneficiaries	Benefits
Citizens	Appreciate the results and benefits arising from research and feel more confident in how their tax money is spent
	Find their own answers to important questions, based on real evidence
	Pass on knowledge and experience to others, and make a contribution to the knowledge society beyond their immediate circle and life-spans
Funders and Policy Makers	Make evidence-based decisions
	Eliminate unnecessary duplication of work
	Get greater return on investment
Researchers	Have all data and tools easily available, increasing productivity
	Cross disciplinary boundaries, gaining new insights and producing new solutions
	'Stand on the shoulders of giants'
Enterprise and Industry	Use the best available information for R&D, increasing productivity
	Create new knowledge, markets and job opportunities
	Provide a strong industrial and economic base for European prosperity
	Increase opportunities for mobility and knowledge exchange

Data driven infrastructure



tutte le organizzazioni hanno bisogno di condividere i dati, anche solo internamente

sempre più forte tendenza a condividere i dati scientifici in modo aperto

sempre maggiori passi verso logiche data-centric

nuovi database: **il valore risiede nei dati**, non più nei servizi, che servono solo a esporre i dati

individuare modelli coerenti di gestione dei dati

sfruttare il cloud computing

l'informazione è prioritaria

M. Hammond, Preparing for data driven infrastructure, JISC report, 2012

Quali dati?

Data are facts, observations or experiences on which an argument, theory or test is based.
Data may be numerical, descriptive or visual.
Data may be raw or analysed, experimental or observational.
Data includes: laboratory notebooks; field notebooks; primary research data (including research data in hardcopy or in computer readable form); questionnaires; audiotapes; videotapes; models; photographs; films; test responses.
Research collections may include slides; artefacts; specimens; samples.

http://goo.gl/6QgtP

- **Raw or primary data:** Information recorded as notes, images, video footage, paper surveys, computer files, etc., pertaining to a specific research project
- **Processed data:** Analyses, descriptions, and conclusions prepared as reports or papers
- **Published data:** Information distributed to people beyond those involved in data acquisition and administration

http://goo.gl/m0TSz

dati di base della ricerca finanziata con i fondi pubblici

i **dati** di cui già si parla **nell'articolo scientifico** pubblicato su rivista

sulla carta: solo riassunto della ricerca (l'articolo)

i dati:
•oggetti complessi
•generati in momenti diversi della ricerca
•con metodi e fini diversi per disciplina

dati grezzi (raw data)

sul **web** possiamo **integrare** con **l'intero set di dati**

Quali dati?

- **NON** i dati della **ricerca applicata**
- **NON** i dati correlati ai **brevetti**
- **NON** dati **personali**
- **NON** dati **confidenziali**
- **NON segreti industriali**
- **NON** dati di **produzione**

… solo i dati di base della ricerca…

… che sono alla base dei prodotti innovativi…

Novità?

molti **ricercatori** già **scambiano** i dati (Virtual Observatory, CERN…)

alcune **discipline** (Genomics) **funzionano solo** in virtù della **condivisione** dei dati

• molte **riviste** (Nature Genetics) già **richiedono la pubblicazione dei dati** insieme all'articolo
• molte riviste Open Access lo consigliano

nature genetics

PLOS ONE

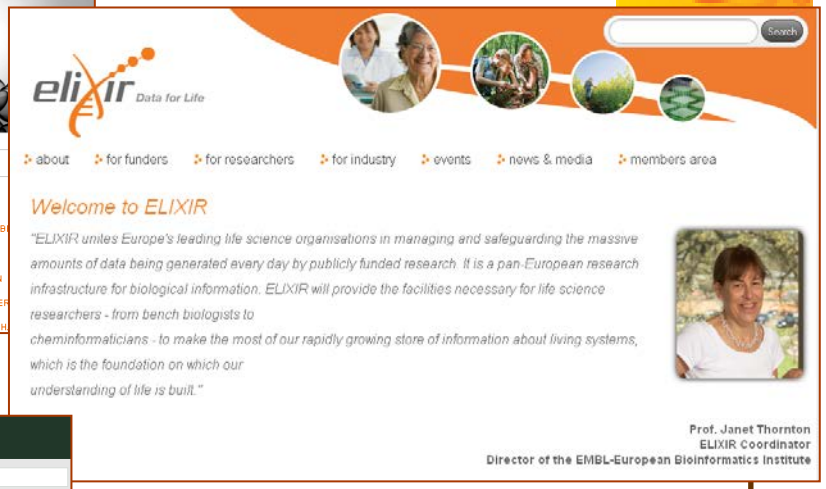
ci sono già database aperti

		Policy of Required Deposition for Types of Data				Policy of Provision of Materials and Methods			Full data deposited
Journal	Impact Factor	Microarray	Nucleic Acid	Protein	Macromolecular	Materials upon request	Protocols upon request	Conditions of publication	% of papers
New England Journal of Medicine	52.589	■							0
Cell	29.887	■	■	■	■	■	■	■	1
Nature	28.751	■	■	■	■	■	■	■	0
Lancet	28.638	■							0
Nature Medicine	26.382	■	■	■	■	■	■	■	0
Science	26.372	■	■	■	■	■	■		1
Nature Immunology	26.218	■	■	■	■	■	■	■	9
Nature Genetics	25.556	■	■	■	■	■	■	■	0
JAMA	25.547		■	■					1
Nature Biotechnology	22.848	■	■	■	■	■	■	■	5
Nature Materials	19.782	■	■	■	■	■	■	■	0
Immunity	19.266	■	■	■	■	■	■	■	0
Nature Cell Biology	17.623	■	■	■	■	■	■	■	0
Journal of Clinical Investigation	16.915	■	■	■		■	■	■	0
Archives of General Psychiatry	15.976		■	■					0
Journal of the National Cancer Institute	15.678	■	■	■		■	■		0
Nature Neuroscience	15.664	■	■	■	■	■	■	■	1
Journal of Experimental Medicine	15.612	■	■	■	■	■	■		0
Annals of Internet Medicine	15.516								0
Journal of Clinical Oncology	15.484								0
Nature Methods	15.478	■	■	■	■	■	■	■	6
Genes and Development	14.795		■	■	■	■			3
Nature Physics	14.677	■	■	■	■	■	■	■	0
PLoS Biology	13.501	■	■	■	■				2
Neuron	13.41	■	■	■	■	■			0
Molecular Cell	13.156	■	■	■	■	■	■	■	0
Circulation	12.755	■	■	■					0
PLoS Medicine	12.601	■	■	■	■				0
Development Cell	12.436	■	■	■	■	■	■	■	0
Gastroenterology	11.673	■	■	■		■			0
Genome Research	11.224	■	■	■	■	■		■	6
American Journal of Human Genetics	11.092	■	■	■	■	■	■	■	3
Nature Structural and Molecular Biology	11.085	■	■	■	■	■	■	■	0
Journal of the American College of Cardiology	11.054								0
Blood	10.896	■	■		■				0
Hepatology	10.374		■						0
Current Biology	10.539	■	■	■	■	■	■	■	0
Gut	10.015								0
British Medical Journal	9.723								0
Circulation Research	9.721	■	■		■	■	■		1
Plant Cell	9.653				■			■	0
Nano Letters	9.627								0
Journal of Cell Biology	9.598	■	■	■	■				0
PNAS	9.598	■	■	■	■	■			1
Molecular and Cellular Proteomics	9.425		■	■					7
PLoS Pathogens	9.336	■	■	■	■	■	■	■	0
American Journal of Psychiatry	9.127								0
American Journal of Respiratory and Critical Care Medicine	9.074	■	■						0
Annals of Neurology	8.813								0
PLoS Genetics	8.721	■	■	■	■	■	■	■	0

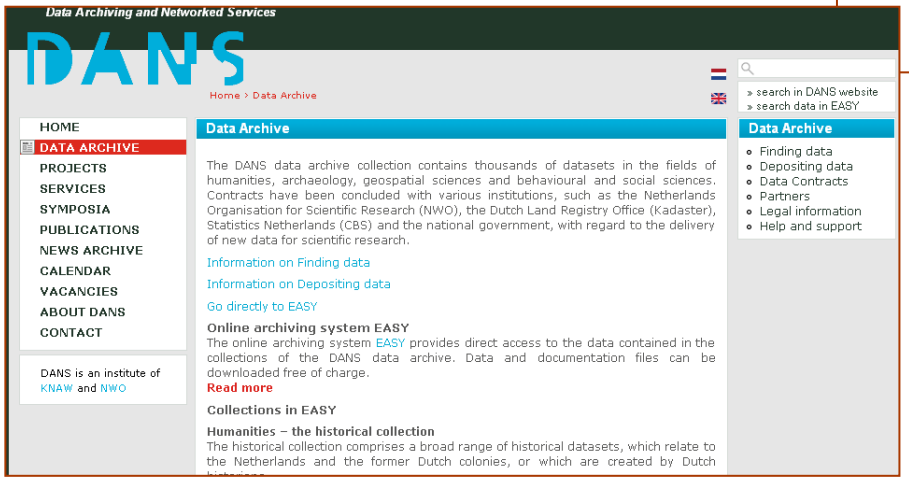
Royal Society, Science is an open enterprise, Report, June 2012

Novità?

http://www.openphacts.org/



http://www.elixir-europe.org/



http://www.dans.knaw.nl/en/content/data-archive

http://echo.mpiwg-berlin.mpg.de/home

Pro e contro

Researchers' Incentives and Disincentives to Share Data

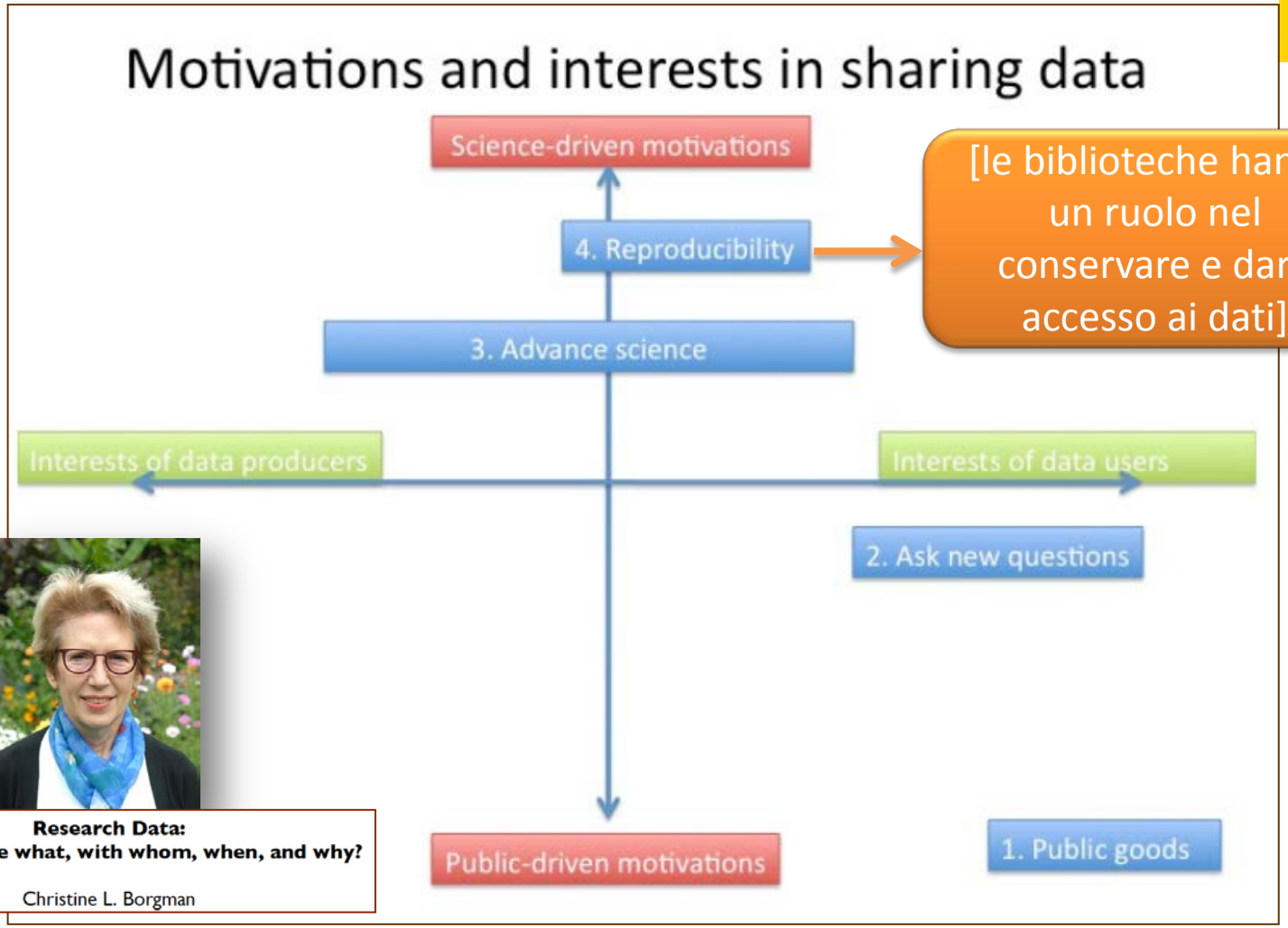
Incentives for researchers to share their data include the ethos of open science and peer review; the value of collaborating with others, for which data may be the "glue;" benefits to reputation; and reciprocity. Depositing one's data may be a condition of gaining access to the data of others, and of access to useful tools for analysis and management. Coercion may also play a role: some funding agencies or individual grant contracts may require data contribution as a condition for funding.

incentivi

Researchers also have multiple incentives *not* to share their data. In most fields, the rewards come from publication, not from data management. Scholars are hired and promoted based on their publication record rather than on the quality of their metadata. Secondly, documenting data is a labor-intensive process even for local use. Documenting methods, instrumentation, and software, and producing metadata at a level that the data are interpretable by others, can require much more labor than documentation for use by oneself or one's team. Thirdly, researchers are concerned about establishing the priority of their claims on research findings in the face of competition. Embargo periods, where they exist, protect the investigator by providing a period of time to analyze data and publish results prior to the public release of their data. Lastly is the set of concerns for intellectual property, both the ability to control one's own resources and the ability to gain access to resources controlled by others

http://goo.gl/40MHd

Perché condividere?



[le biblioteche hanno un ruolo nel conservare e dare accesso ai dati]

Research Data:
Who will share what, with whom, when, and why?

Christine L. Borgman

http://goo.gl/40MHd

Perché non condividere?

What might prevent you from sharing the data you have produced or intend to produce for this project?

Steinhart et al., Journal of eScience Librarianship, 2012; 1(2)

G. Steinhart et al. Prepared to Plan? A Snapshot of Researcher Readiness to Address Data Management Planning Requirements, Journal of eScience Librarianship, 2012, 1(2): Article 1

Condividere o non condividere?

uso dei dati

	Agree Strongly	Agree Somewhat	Neither Agree Nor Disagree	Disagree Somewhat	Disagree Strongly
Lack of access to data generated by other researchers or institutions is a major impediment to progress in science.	353 (27.2%)	520 (40%)	230 (17.7%)	149 (11.5%)	48 (3.7%)
Lack of access to data generated by other researchers or institutions has restricted my ability to answer scientific questions.	228 (17.6%)	422 (32.5%)	297 (22.9%)	238 (18.4%)	112 (8.6%)
Data may be misinterpreted due to complexity of the data.	383 (29.6%)	590 (45.6%)	217 (16.8%)	77 (6%)	26 (2%)
Data may be misinterpreted due to poor quality of the data.	379 (29.4%)	540 (41.8%)	232 (18%)	107 (8.3%)	33 (2.6)
Data may be used in other ways than intended.	410 (31.8%)	539 (41.8%)	249 (19.3%)	68 (5.3%)	23 (1.8%)

	Others can access my data easily	
	Agree strongly	Agree somewhat
social sciences	11(5.4%)	36(17.8%)
computer science/engineering	12(10.3%)	29(24.8%)
physical sciences	17(11.3%)	41(27.3%)
environmental sciences & ecology	56(12.0%)	124(26.5%)
atmospheric science	12(23.5%)	13(25.5%)
biology	28(15.6%)	50(27.9%)
medicine	2(6.5%)	2(6.5%)
other	12(13.0%)	21(22.8%)

C.Tenopir et al. Data Sharing by Scientists: Practices and Perceptions, PLoS ONE 2011, 6(6): e21101.

Condividere o non condividere?

	Agree Strongly	Agree Somewhat	Neither Agree Nor Disagree	Disagree Somewhat	Disagree Strongly
I share my data with others.	418 (32.3%)	551 (42.6%)	199 (15.4%)	95 (7.3%)	30 (2.3%)
Others can access my data easily.	150 (11.6%)	317 (24.6%)	310 (24%)	307 (23.8%)	207 (16%)

	Agree Strongly	Agree Somewhat	Neither Agree Nor Disagree	Disagree Somewhat	Disagree Strongly
I would use other researchers' datasets if their datasets were easily accessible.	561 (43.2%)	524 (40.3%)	136 (10.5%)	62 (4.8%)	16 (1.2%)
I would be willing to place at least some of my data into a central data repository with no restrictions.	539 (41.6%)	472 (36.4%)	141 (10.9%)	104 (8%)	39 (3%)
I would be willing to place all of my data into a central data repository with no restrictions.	191 (14.9%)	338 (26.3%)	234 (18.2%)	318 (24.7%)	205 (15.9%)
I would be more likely to make my data available if I could place conditions on access.	317 (24.8%)	506 (39.6%)	279 (21.8%)	107 (8.4%)	68 (5.3%)
I am satisfied with my ability to integrate data from disparate sources to address research questions.	156 (12.2%)	419 (32.7%)	363 (28.3%)	275 (21.5%)	69 (5.4%)
I would be willing to share data across a broad group of researchers who use data in different ways.	476 (37%)	565 (43.9%)	185 (14.4%)	48 (3.7%)	13 (1%)
It is important that my data are cited when used by other researchers.	885 (68.6%)	298 (23.1%)	87 (6.7%)	14 (1.1%)	7 (0.5%)
It is appropriate to create new datasets from shared data.	505 (38.9%)	475 (36.6%)	261 (20.1%)	36 (2.8%)	20 (1.5%)

	For others to use my data		To use other people's data	
	Yes	No	Yes	No
Co-authorship on publications resulting from use of the data	751 (59.7%)	506 (40.3%)	750 (61.2%)	476 (38.8%)
Formal acknowledgement of the data providers and/or funding agencies in all disseminated work making use of the data	1168 (93%)	88 (7%)	1147 (93.3%)	83 (6.7%)
Formal citation of the data providers and/or funding agencies in all disseminated work making use of the data	1166 (94.5%)	68 (5.5%)	1152 (95.1%)	59 (4.9%)
The opportunity to collaborate on the project (including, for example, consultation on analytic methods, interpretation of results, dissemination of research results, etc.)	991 (80.6%)	239 (19.4%)	980 (81.2%)	227 (18.8%)
Results based (at least in part) on the data could not be disseminated in any format without the data provider's approval.	585 (47.7%)	642 (52.3%)	594 (48.9%)	620 (51.1%)
At least part of the costs of data acquisition, retrieval or provision must be recovered.	364 (30%)	851 (70%)	374 (31.2%)	826 (68.8%)
Results based (at least in part) on the data could not be disseminated without the data provider having the opportunity to review the results and make suggestions or comments, but approval not required.	746 (61.7%)	464 (38.3%)	750 (62.7%)	447 (37.3%)
Reprints of articles that make use of the data must be provided to the data provider.	860 (70.1%)	367 (29.9%)	850 (70.4%)	357 (29.6%)
The data provider is given a complete list of all products that make use of the data, including articles, presentations, educational materials, etc.	846 (69.3%)	375 (30.7%)	831 (69.1%)	372 (30.9%)
Legal permission for data use is obtained.	545 (44.8%)	672 (55.2%)	552 (45.8%)	652 (54.2%)
Mutual agreement on reciprocal sharing of data	880 (72.2%)	339 (27.8%)	865 (71.9%)	338 (28.1%)
The data provider is given and agrees to a statement of uses to which the data will be put.	810 (66.8%)	403 (33.2%)	799 (67%)	394 (33%)

Le politiche (vedi National Science Foundation)
possono giocare un ruolo determinante

Open Research Data: i vantaggi / 1

una scienza più solida

meglio basarsi sui DATI che sull'INTERPRETAZIONE degli stessi contenuta negli articoli…

per **confrontare/ dibattere** con i propri dati

per **creare nuova conoscenza** aggiungendo i propri dati

per **favorire un approccio inter-disciplinare o trans-disciplinare**

utile per risolvere le sfide globali (es. clima)

Open Research Data: i vantaggi / 2

una **scienza più efficiente/efficace**

i dati già raccolti possono essere **riusati** senza ulteriori spese, da **differenti prospettive/differenti approcci**

se i dati **circolano rapidamente e globalmente**, la **creazione di conoscenza accelera**

dati
- archiviati
- catalogati
- ricercabili
- accessibili

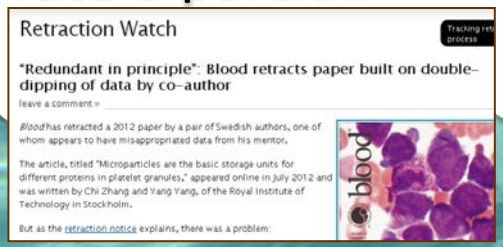
infrastruttura adeguata

interoperabilità/standard

Open Research Data: i vantaggi / 3

una **scienza più trasparente**

i dati possono essere **riusati per scoprire frodi/falsificazioni /interpretazioni errate** o parziali


Retraction Watch

Tracking retr process

"Redundant in principle": Blood retracts paper built on double-dipping of data by co-author

leave a comment »

Blood has retracted a 2012 paper by a pair of Swedish authors, one of whom appears to have misappropriated data from his mentor.

The article, titled "Microparticles are the basic storage units for different proteins in platelet granules," appeared online in July 2012 and was written by Chi Zhang and Yang Yang, of the Royal Institute of Technology in Stockholm.

But as the retraction notice explains, there was a problem:

i dati possono essere usati per **replicare l'esperimento**

Open Research Data: i vantaggi / 4

una **società più competitiva**

le piccole-medie imprese non hanno accesso ai dati della ricerca accademica, chiusa in riviste con abbonamenti esosi

se avessero accesso ai dati, immetterebbero sul mercato i loro prodotti DUE ANNI prima

Access to Research and Technical Information in Denmark, Report, May 2011

Open Research Data: i vantaggi / 5

una società più efficiente

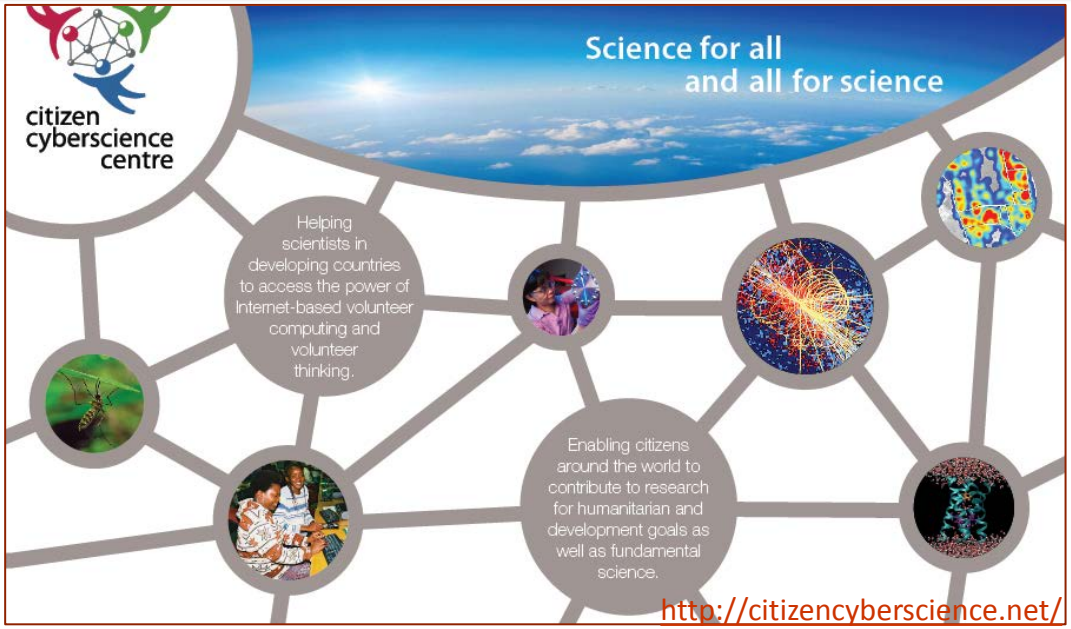
con nuove tecniche di data-mining si possono creare nuovi dati, con valore aggiunto per l'intera società

USA: 300bn $ nel settore sanità

Big data: The next frontier for innovation, competition, and productivity, 2011

Open Research Data: i vantaggi / 6

una **società più partecipativa**

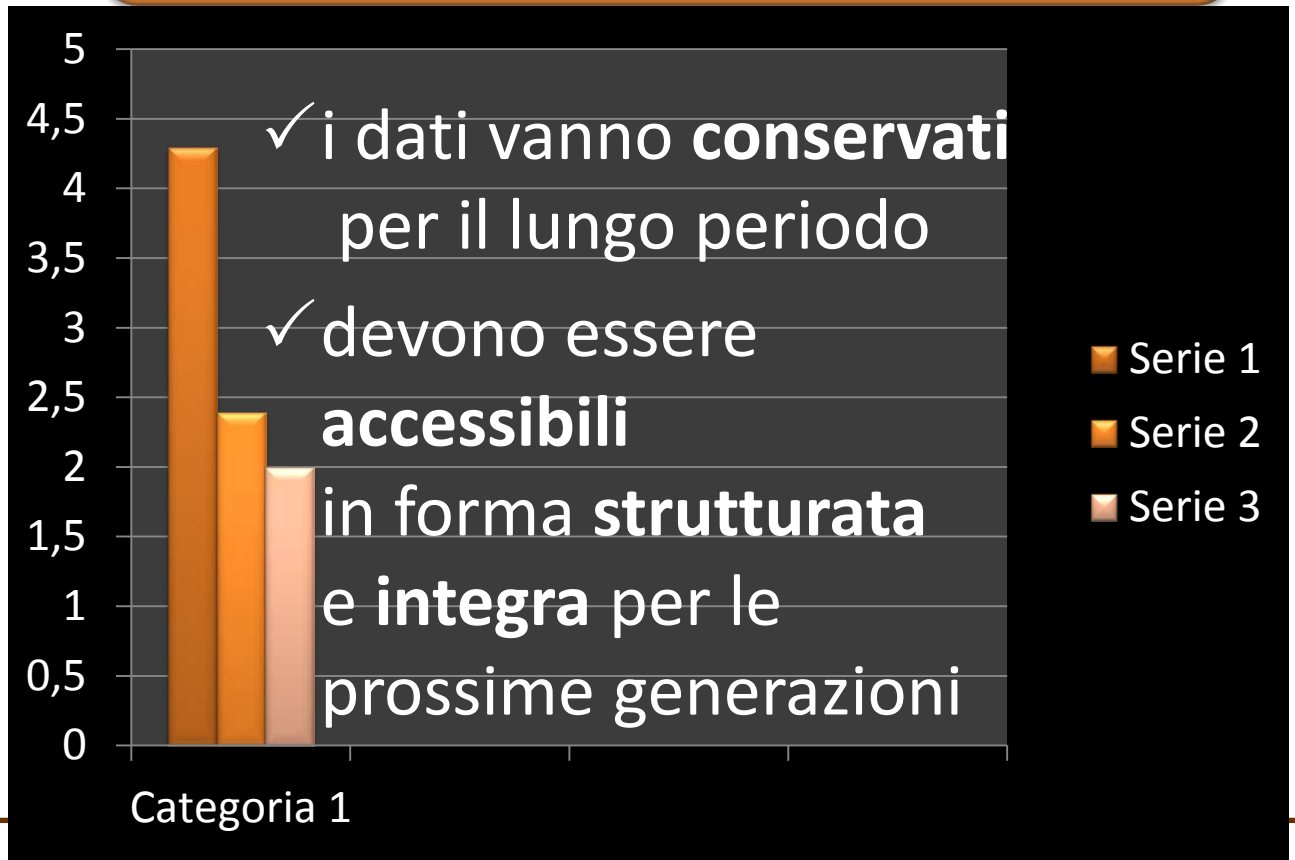


scienza aperta significa anche citizen cyberscience:
i cittadini comuni partecipano alla raccolta di dati

F.Grey, The rise of citizen cyberscience and its impact on professional research, OAI 7 Geneva, 2011

Open Research Data: i vantaggi / 7

per la società del futuro



Open Research Data: i vantaggi / 8

nuove opportunità
per una società che cresce

la cura, la gestione e la conservazione dei data sets richiede nuove professionalità

- data manager
- data curator

si aprono prospettive di **lavoro** e **formazione**

- nuove competenze

Open Research Data: i vantaggi / 9

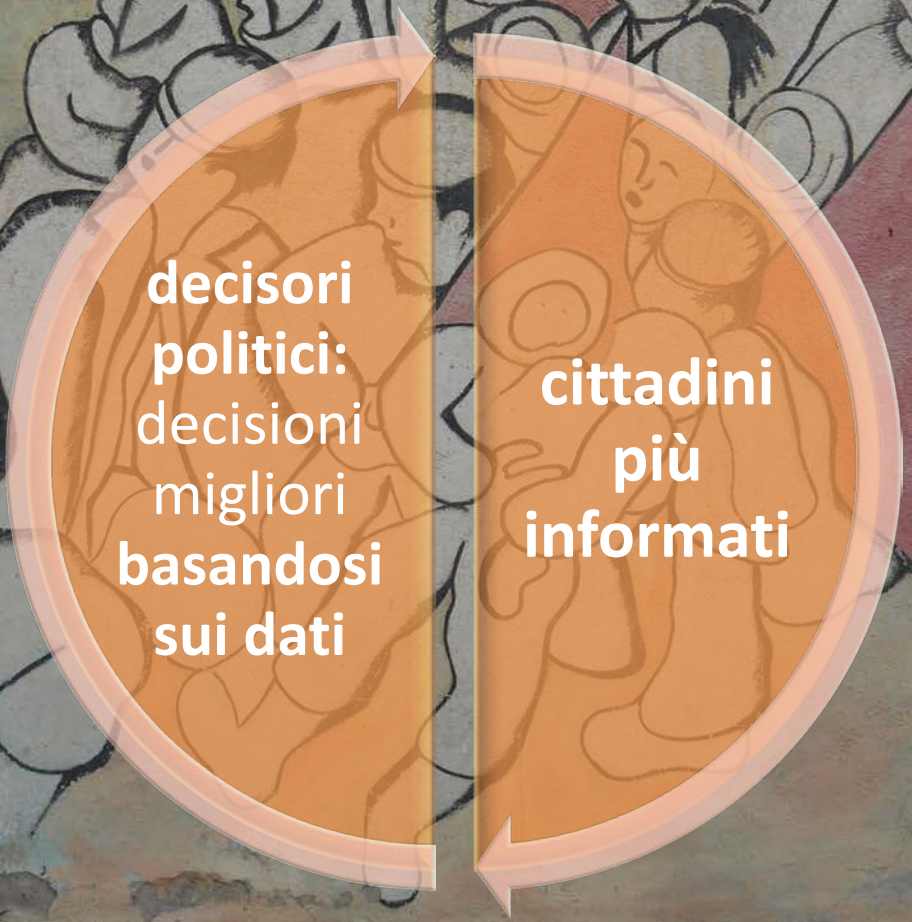
	Percent increase in citation count (95% confidence interval)	p-value
Publish in a journal with twice the impact factor	84% (59 to 109%)	<0.001
Increase the publication date by a month	−3% (−5 to −2%)	<0.001
Include a US author	38% (1 to 89%)	0.049
Make data publicly available	**69% (18 to 143%)**	**0.006**

We examined the citation history of 85 cancer microarray clinical trial publications with respect to the availability of their data. The 48% of trials with publicly available microarray data received 85% of the aggregate citations. Publicly available data was significantly ($p = 0.006$) associated with a 69% increase in citations, independently of journal impact factor, date of publication, and author country of origin using linear regression.

HA Piwowar et al. Sharing Detailed Research Data Is Associated with Increased Citation Rate. PLoS ONE 2007 2(3): e308

Open Data (PS): i vantaggi / 1

una **società della conoscenza**



Open Data (PS) : i vantaggi / 2

riuso per remix (API, Apps)

«The market for mobile apps has outgrown the information and communication technologies market over the past two years (2009-2010), and its growth will accelerate in the future to reach $ US 35 billion in 2015. It is to be one of the fastest growing segments in the information technology market»

POPSIS report, Oct. 2011

«Public Sector Information generates over 30 billion euros per year in economic activity, with services from geo-location services to weather forecasts. By opening up this resource fully, we could more than doubly the value of this activity, around 70 billion euros»

Neelie Kroes, Blog, Dec. 12, 2011

Box 6 Benefits of curating and sharing publicly funded research data

Open access to, and sharing of, data from publicly funded research offer many research and educational advantages over a closed, proprietary system that places high barriers to both access and subsequent re-use. Open access to such data:

- Reinforces open scientific and scholarly inquiry;
- Encourages diversity of analysis and opinion;
- Promotes new research and new types of research;
- Enables the application of automated knowledge discovery tools online;
- Allows the verification of previous results;
- Makes possible the testing of new or alternative hypotheses and methods of analysis;
- Establishes a broader base set of data than any one researcher can hope to collect, thereby providing a greater baseline of factual information for the research community;
- Supports studies on data collection methods and measurement;
- Facilitates the education of new researchers;
- Enables the exploration of topics not envisioned by the initial investigators;
- Permits the creation of new data sets, information, and knowledge when data from multiple sources are combined;
- Helps transfer factual information to, and promote development and capacity building in developing countries;
- Promotes interdisciplinary, inter-sectoral, inter-institutional, and international research; and
- Helps to maximize the research potential of new digital technologies and networks, thereby providing greater returns from the public investment in data collection and research.

Source: GEO (2010) *GEOSS Data Sharing Action Plan*. GEO-VII Plenary.

http://www.ands.org.au/resource/houghton-cost-benefit-study.pdf pag. 49

Dati : costi e benefici

VICTORIA UNIVERSITY
MELBOURNE AUSTRALIA

Costs and Benefits of Data Provision

Report to the Australian National Data Service

By

John Houghton

The evidence from previous studies suggests that individual cases vary greatly, making generalisation extremely difficult and of limited value. Perhaps, what could more usefully be generalised are the methods of analysis. For example, it would be useful to combine the frameworks and models into a tool that could be applied in assessing the costs and benefits of research data curation and sharing, and to further develop the framework for estimating cost-benefits outlined in this study to produce a tool tailored to the analysis of the costs and benefits

Studies of the costs and benefits associated with open access to publications and data arising from publicly funded research suggest that the benefits can be significant and the costs relatively small. The return on investment in curation and open accessibility is all the greater when the data are long-lived, as the returns are recurring during the useful life of the data, although the time lag between investment and return (cost and benefit) can be substantial. What this study demonstrates is that the direct and measurable benefits of making PSI available freely and unrestrictedly typically outweigh the costs. When one adds the longer-term benefits that we cannot fully measure, cannot even foresee, the case for open access appears to be strong.

Settembre 2011, http://www.ands.org.au/resource/houghton-cost-benefit-study.pdf

Dati: questioni aperte / 1

Extracting, Transforming and Archiving Scientific Data

Daniel Lemire[1] and Andre Vellino[2]

[1] Université du Québec à Montréal,
100 Sherbrooke West
Montreal, QC, H2X 3P2 Canada
lemire@gmail.com
[2] CISTI - National Research Council of Canada
Building M-55, 1200 Montreal Road
Ottawa, ON K1A 0S2 Canada
andre.vellino@nrc.ca

Abstract. It is becoming common to archive research datasets that are not only large but also numerous. In addition, their corresponding metadata and the software required to analyse or display them need to be archived. Yet the manual curation of research data can be difficult and expensive, particularly in very large digital repositories, hence the importance of models and tools for automating digital curation tasks. The automation of these tasks faces three major challenges: (1) research data and data sources are highly heterogeneous, (2) future research needs are difficult to anticipate, (3) data is hard to index. To address these problems, we propose the Extract, Transform and Archive (ETA) model for managing and mechanizing the curation of research data. Specifically, we propose a scalable strategy for addressing the research-data problem,

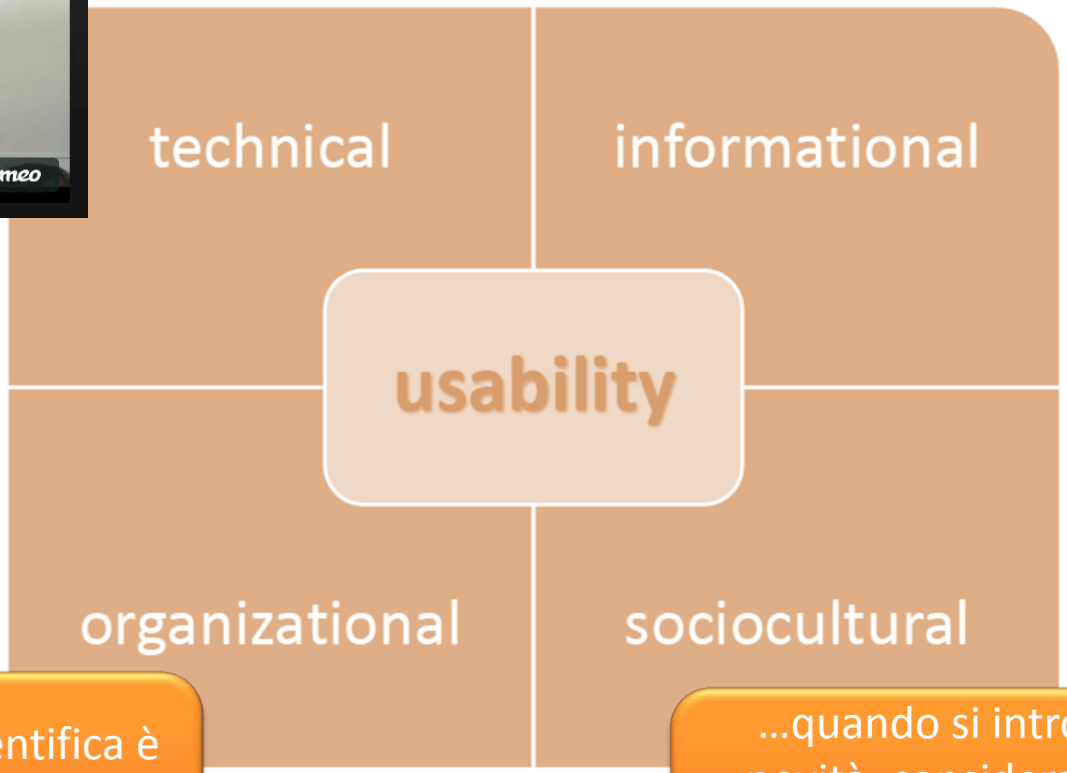
1) i dati sono eterogenei, sono eterogenee le loro fonti

2) difficile anticipare le esigenze future della ricerca

3) i dati sono difficili da indicizzare

http://arxiv.org/abs/1108.4041

Open data: questioni aperte / 2



…comunicazione scientifica è un SISTEMA in cui tutto si tiene…

…quando si introduce una novità, considerare sempre effetti e ripercussioni sull'intero sistema e le sue componenti…

Oya Rieger, Cornell University , The research data landscape– an overview, OPENAire Workshop, Copenhagen, June 2012

Questioni tecniche

technical infrastructure

- scalable and flexible systems to store, discover, access, and archive content
- interoperability standards to link related information objects and various types of repositories
- metadata standards to facilitate discovery, access, archiving, and repurposing

trovare la quantità di metadati adatta a descrivere SENZA scoraggiare il deposito…

Oya Rieger, Cornell University , The research data landscape– an overview, OPENAire Workshop, Copenhagen, June 2012

Questioni socio-culturali

sociocultural issues

- community-based standards for deposit, use, and maintenance of data
- different access provisions in support of academic, and entrepreneurial requirements
- incentives and rewards for scientists to share the outputs of their research endeavors

Oya Rieger, Cornell University , The research data landscape– an overview, OPENAire Workshop, Copenhagen, June 2012

Questioni di diritti

information policies

- information policies to support:
  - IPR
  - privacy
  - confidentiality
  - institutional ownership
  - security
  - access limitations
  - retention and deaccessioning

Oya Rieger, Cornell University , The research data landscape– an overview, OPENAire Workshop, Copenhagen, June 2012

Questioni organizzative

organizational infrastructure

- business and sustainability plans
- governance models
- recognition and engagement of stakeholders
- collaboration strategies
- communication and marketing strategies

Oya Rieger, Cornell University , The research data landscape– an overview, OPENAire Workshop, Copenhagen, June 2012

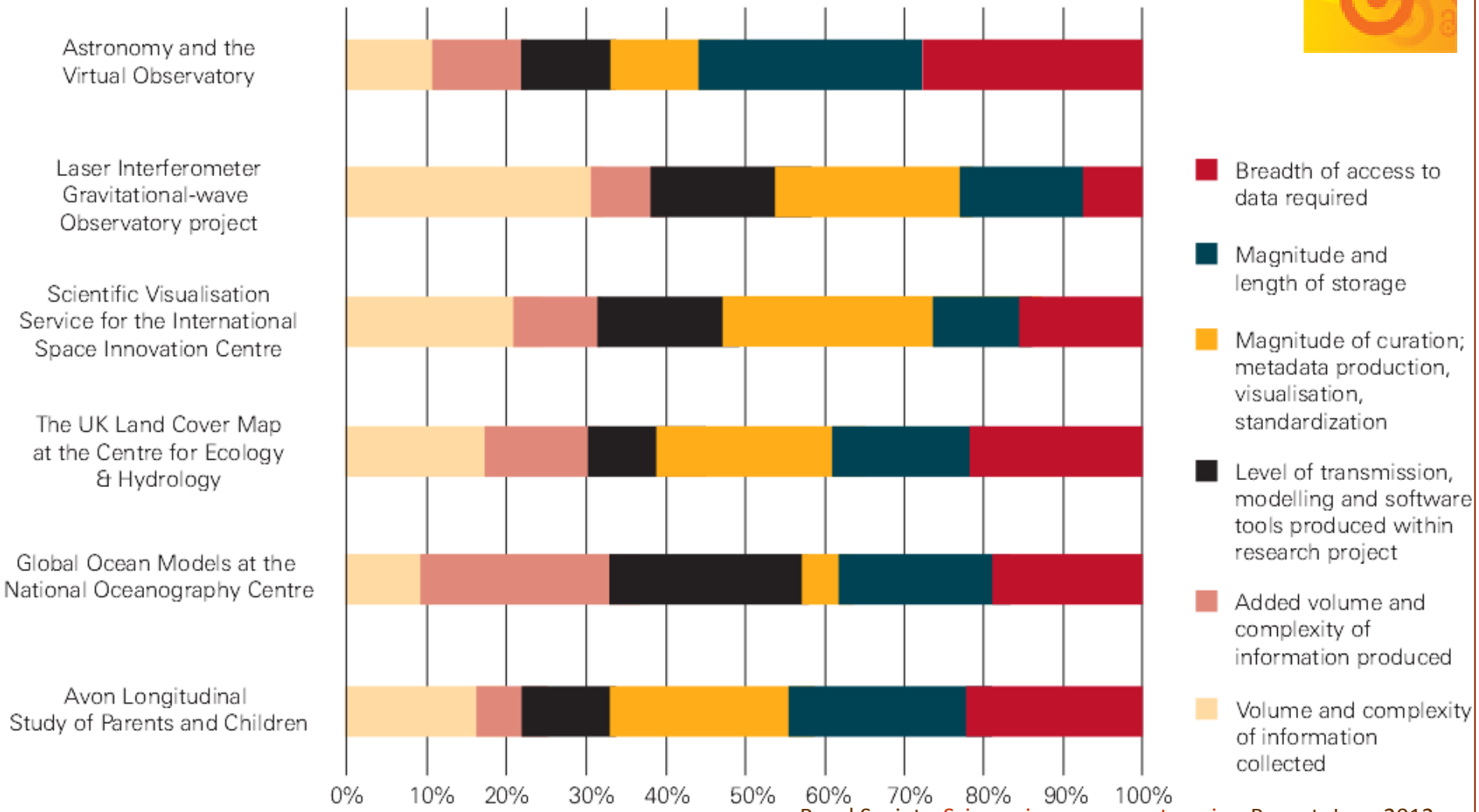
…al centro resta l'usabilità

usability

- data quality standards
- ease of deposit to encourage end-users
- tools to support analytics, mining, integration, and visualization
- digital identifiers to persistently locate
- citation standards to reference resources
- metrics to track and communicate impact

Oya Rieger, Cornell University , The research data landscape– an overview, OPENAire Workshop, Copenhagen, June 2012

Esempi pratici



Royal Society, Science is an open enterprise, Report, June 2012

Open Data: questioni aperte / 3

Beyond Sharing and Re-using:

Toward Global Data Networking

With ideas provided by (in alphabetical order)

Fred Friend, Jean-Claude Guédon, and Herbert Van de Sompel

1. Monitoring the growth of data and their provenance;
2. Monitoring the formats in which data are produced;
3. Surveying the tools to ensure interoperability of data, and their use across disciplines: curation;
4. Classifying kinds of data: metadata formats
5. Creating harvestable, distributed databases of metadata to support the creation of cross-database services, e.g. discovery;
6. Tackling the data preservation challenge

The suggested approach

1. Given what precedes, a top-down approach to achieve global sharing of data is not likely to succeed; if perchance it should, it will initially meet with a great deal of resistance and this will delay the whole process, especially in the absence of a global governance structure with real power of implementation;
2. Because the present situation reveals a complex landscape of fragmented and disparate data sets, working toward an ever denser set of linkages between them appears best. Such a set of linkages will grow best if they grow from actual needs, in particular those of various research communities. At the same time, research communities should easily find their way to sets of evolving guidelines that would ensure a growing degree of coherence across all boundaries.
3. The experience of the Internet and its evolution suggests that an approach based on layered thinking is best. This source of inspiration is all the more compelling that, like the Internet, data is part of a vast communication system, that of science, and that it is presently expressed and processed in the form of digital bits.

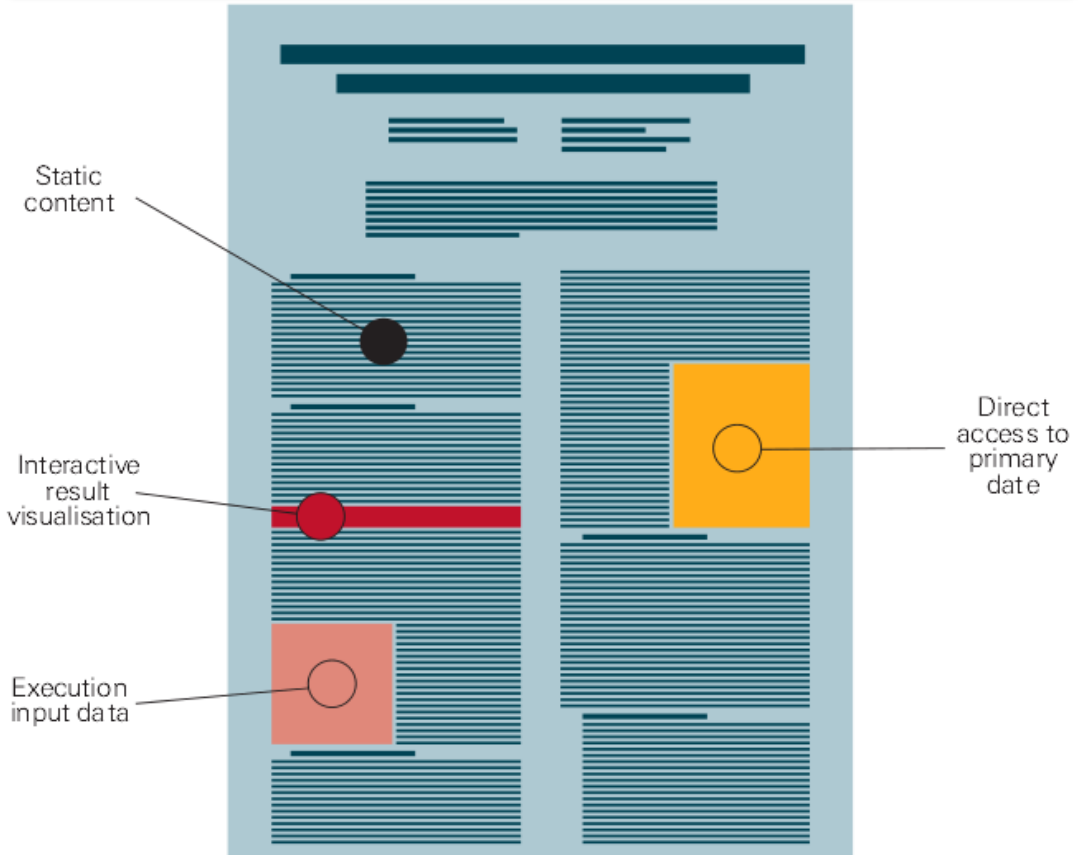
Ottobre 2011, http://goo.gl/BD5nQ

Open Data: questioni aperte / 4

4. The experience of the Internet also suggests that no grand Cartesian architecture will prevail in such an environment. The data landscape, very much like the network landscapes at the end of the sixties and the early seventies, is much too fluid and much beyond the full control of any one entity to fit in any Procrustean bed anyone would want to design for it. It is better to adopt an iterative approach based on modest gains that gradually add up. Some working principles of the Internet deserve being recalled in this context: **rough consensus and working code are to be preferred to formal negotiations**; **releasing code soon and often is preferable to trying to come up with a fully worked out solution from the outset**; **favouring implementation over standardization in order to achieve the latter has already been mentioned;**
5. The impetus must come from the research communities themselves. Consequently, the best starting point for the European Community is:
   a. To begin work with some of the research communities they directly support and use this lever to nudge them into the desired direction. In effect, they would be the seed groups that emerged with GATS before the IETF in the case of the Internet.
   b. Relevant interlocutors from other regions or countries would be invited to do the same.
   c. Small, by invitation only, workshops in various countries and regions (e.g. The E.C.) would ignite the process aiming at testing the various possibilities available and begin working towards **networking of data** across communities.

Open Data: scenario

articolo eseguibile



Royal Society, Science is an open enterprise, Report, June 2012

Un po' di definizioni

Data type	Definition
Big Data	Data that requires massive computing power to process.
Broad Data	Structured big data, so that it is freely available through the web to everyone, eg on websites like www.data.gov
Data	Qualitative or quantitative statements or numbers that are (or assumed to be) factual. Data may be raw or primary data (eg direct from measurement), or derivative of primary data, but are not yet the product of analysis or interpretation other than calculation.
Data-gap	When data becomes detached from the published conclusions
Data-intensive science	Science that involves large or even massive datasets
Data-led approach	Where hypotheses are constructed after identifying relationships in the dataset.
Data-led science	The use of massive datasets to find patterns as the basis of research.
Dataset	A collection of factual information held in electronic form where all or most of the information has been collected for the purpose of provision of a service by the authority or carrying out of any other function of the authority. Datasets contain factual information which is not the product of analysis or interpretation other than calculation, is not an official statistic, and is unaltered and un-adapted since recording.
Linked Data	Linked data is described by a unique identifier naming and locating it in order to facilitate access. It contains identifiers for other relevant data, allowing links to be made between data that would not otherwise be connected, increasing discoverability of related data.
Metadata	Metadata "data about data", contains information about a dataset. This may be state why and how it was generated, who created it and when. It may also be technical, describing its structure, licensing terms, and standards it conforms to.
Open Data	Open data is data that meets the criteria of intelligent openness. Data must be accessible, useable, assessable and intelligible.
Semantic Data	Data that are tagged with particular metadata - metadata that can be used to derive relationships between data.

Royal Society, Science is an open enterprise, Report, June 2012

Open Data / 1

What is Open Data?

Share 90 Tweet 120 ShareThis 660

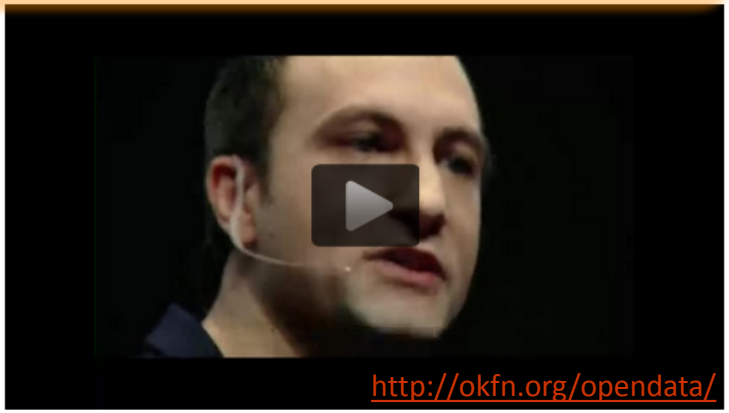
"Open data is data that can be freely used, reused and redistributed by anyone – subject only, at most, to the requirement to attribute and sharealike." OpenDefinition.org

The Open Definition sets out in detail the requirements for 'openness' in relation to content and data. The key features are:

Availability and Access: the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.

Reuse and Redistribution: the data must be provided under terms that permit reuse and redistribution including the intermixing with other datasets.

Universal Participation: everyone must be able to use, reuse and redistribute – there should be no discrimination against fields of endeavour or against persons or groups. For example, 'non-commercial' restrictions that would prevent 'commercial' use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.



Open Data: How We Got Here, and Where We're Going.
From the LIFT 2012 conference.

Open Data / 2

Open Definition

The Open Definition | Conformant Licenses | Licensing Guide

Definizione di Conoscenza Aperta

Version: 1.0

Terminologia

Con il termine **conoscenza** si intende:

1. Contenuti come musica, film, libri;
2. Dati, siano essi scientifici, storici, geografici o di altro tipo;
3. Informazione del settore pubblico.

Nonostante la sua evidente importanza, il software è escluso, poiché è già stato trattato in maniera appropriata in altre sedi.

Il termine **opera** sarà utilizzato per indicare l'oggetto o l'elemento di conoscenza che viene trasferito.

Il termine **pacchetto** può essere utilizzato anche per indicare una raccolta di opere. Naturalmente un tale pacchetto può essere considerato un'opera di per sé.

Il termine **licenza** si riferisce alla licenza in base alla quale il lavoro è reso disponibile. Nel caso in cui non sia stata concessa alcuna licenza, ci si riferisce alle normali condizioni giuridiche a cui il l'opera è soggetta (per esempio le norme sul diritto d'autore).

Open Data / 3

Definizione

Un'opera è aperta se la relativa modalità di distribuzione soddisfa le seguenti condizioni:

1. Accesso

L'opera deve essere disponibile nella sua interezza ed a un costo di riproduzione ragionevole, preferibilmente tramite il download gratuito via Internet. L'opera deve inoltre essere disponibile in un formato comodo e modificabile.

2. Ridistribuzione

La licenza non deve imporre alcuna limitazione alla vendita o all'offerta gratuita dell'opera singolarmente considerata o come parte di un pacchetto composto da opere provenienti da fonti diverse. La licenza non deve richiedere alcuna "royalty" o altra forma di pagamento per tale vendita o distribuzione.

3. Riutilizzo

La licenza deve consentire la realizzazione di modifiche e di opere derivate e deve consentire la loro distribuzione agli stessi termini dell'opera originaria.

4. Assenza di restrizioni tecnologiche

L'opera deve essere fornita in un formato che non ponga ostacoli tecnologici allo svolgimento delle attività sopraelencate. Ciò può essere conseguito mediante la messa a disposizione dell'opera in un formato aperto, vale a dire un formato le cui specifiche siano pubblicamente e liberamente disponibili e che non imponga nessuna restrizione economica o di altro tipo al suo utilizzo.

5. Attribuzione

La licenza può richiedere di citare i vari contributori e creatori dell'opera come condizione per la ridistribuzione ed il riutilizzo di quest'ultima. Se imposta, questa condizione non deve essere onerosa. Per esempio, se viene richiesta la citazione, un elenco di coloro che devono essere citati deve accompagnare l'opera.

6. Integrità

La licenza può richiedere, come condizione perché l'opera venga distribuita in forma modificata, che l'opera derivata abbia un nome o un numero di versione diverso dall'opera originaria.

7. Nessuna discriminazione di persone o gruppi

La licenza non deve discriminare alcuna persona o gruppo di persone.

8. Nessuna discriminazione nei settori d'attività

La licenza non deve impedire a nessuno di utilizzare l'opera in un determinato settore d'attività. Per esempio, la licenza non può impedire che l'opera sia utilizzata da un'azienda, o che venga utilizzata ai fini di ricerca genetica.

9. Distribuzione della licenza

I diritti relativi all'opera devono valere per tutte le persone a cui il programma viene ridistribuito senza che sia per loro necessario accettare o sottostare ad alcuna licenza aggiuntiva.

10. La licenza non deve essere specifica per un pacchetto

I diritti relativi all'opera non devono dipendere dal fatto che l'opera sia parte di un particolare pacchetto. Se l'opera viene estratta da quel pacchetto e usata o distribuita in conformità con i termini della licenza dell'opera, tutte le persone a cui il lavoro viene ridistribuito devono avere gli stessi diritti concessi in congiunzione con il pacchetto originario.

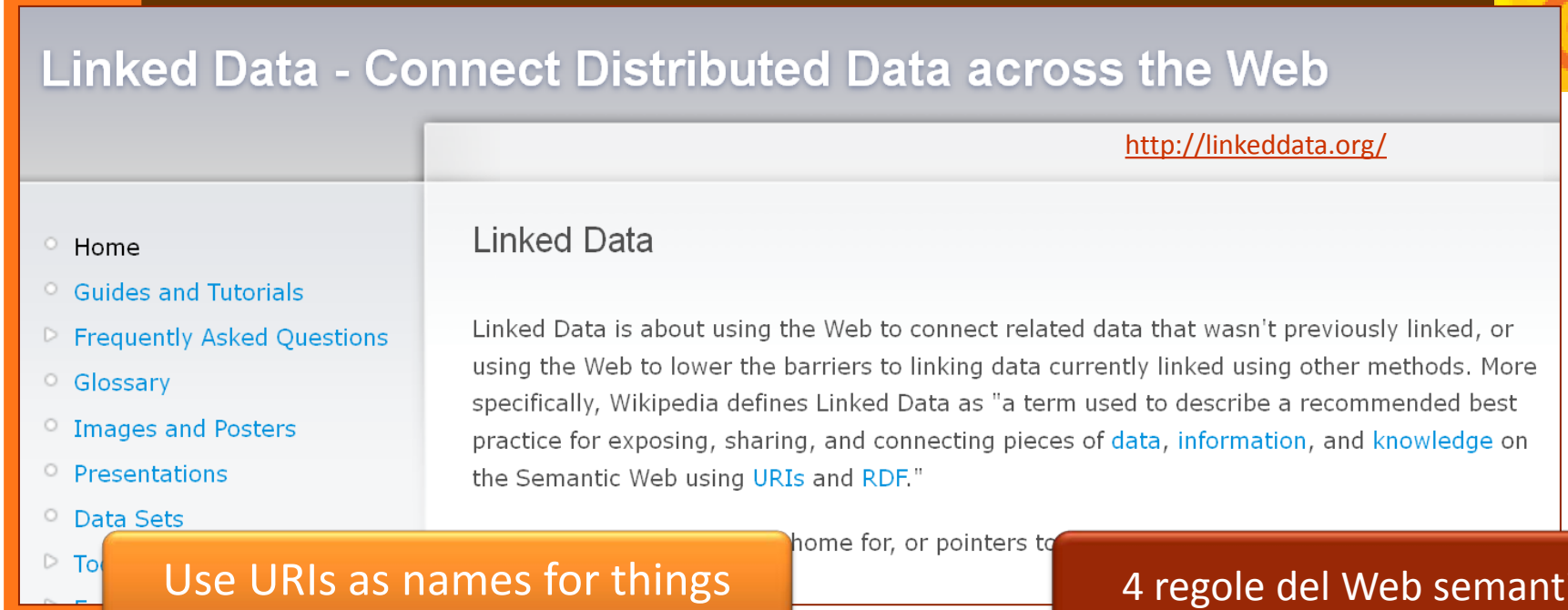
11. La licenza non deve limitare la distribuzione di altre opere

La licenza non deve imporre restrizioni su altre opere distribuite insieme all'opera licenziata. Per esempio, la licenza non deve insistere sul fatto che tutte le altre opere distribuite sullo stesso supporto siano aperte.

Translated by Primavera De Filippi, Andrea Glorioso and Juan Carlos De Martin at the NEXA Center for Internet & Society, Politecnico di Torino.

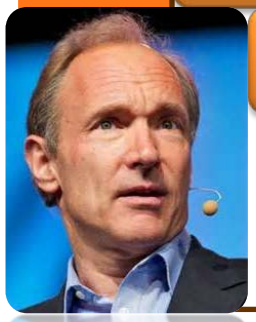
http://opendefinition.org/okd/italiano/

Linked Data



4 regole del Web semantico

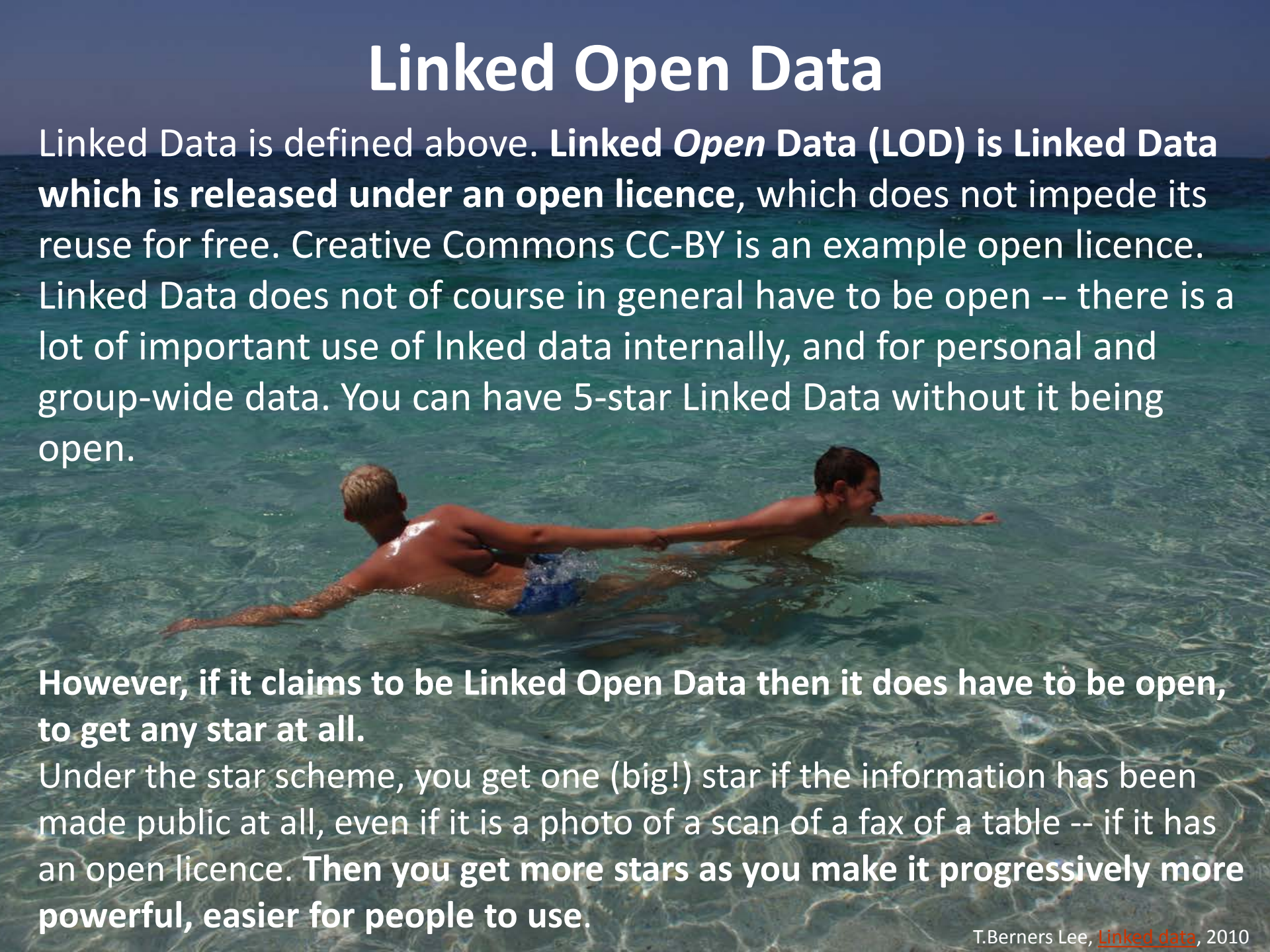
- Use URIs as names for things
- Use HTTP URIs so that people can look up those names
- When someone looks up a URI, provide useful information, using the standards (RDF*,SPARQL)
- Include links to other URIs. so that they can discover more things



T.Berners Lee, Linked data, 2006-2009

Linked Open Data

Linked Data is defined above. **Linked *Open* Data (LOD) is Linked Data which is released under an open licence**, which does not impede its reuse for free. Creative Commons CC-BY is an example open licence. Linked Data does not of course in general have to be open -- there is a lot of important use of lnked data internally, and for personal and group-wide data. You can have 5-star Linked Data without it being open.



However, if it claims to be Linked Open Data then it does have to be open, to get any star at all.

Under the star scheme, you get one (big!) star if the information has been made public at all, even if it is a photo of a scan of a fax of a table -- if it has an open licence. **Then you get more stars as you make it progressively more powerful, easier for people to use**.

T.Berners Lee, Linked data, 2010

Linked Open Data

★ Available on the web (whatever format) *but with an open licence, to be Open Data*

★★ Available as machine-readable structured data (e.g. excel instead of image scan of a table)

★★★ as (2) plus non-proprietary format (e.g. CSV instead of excel)

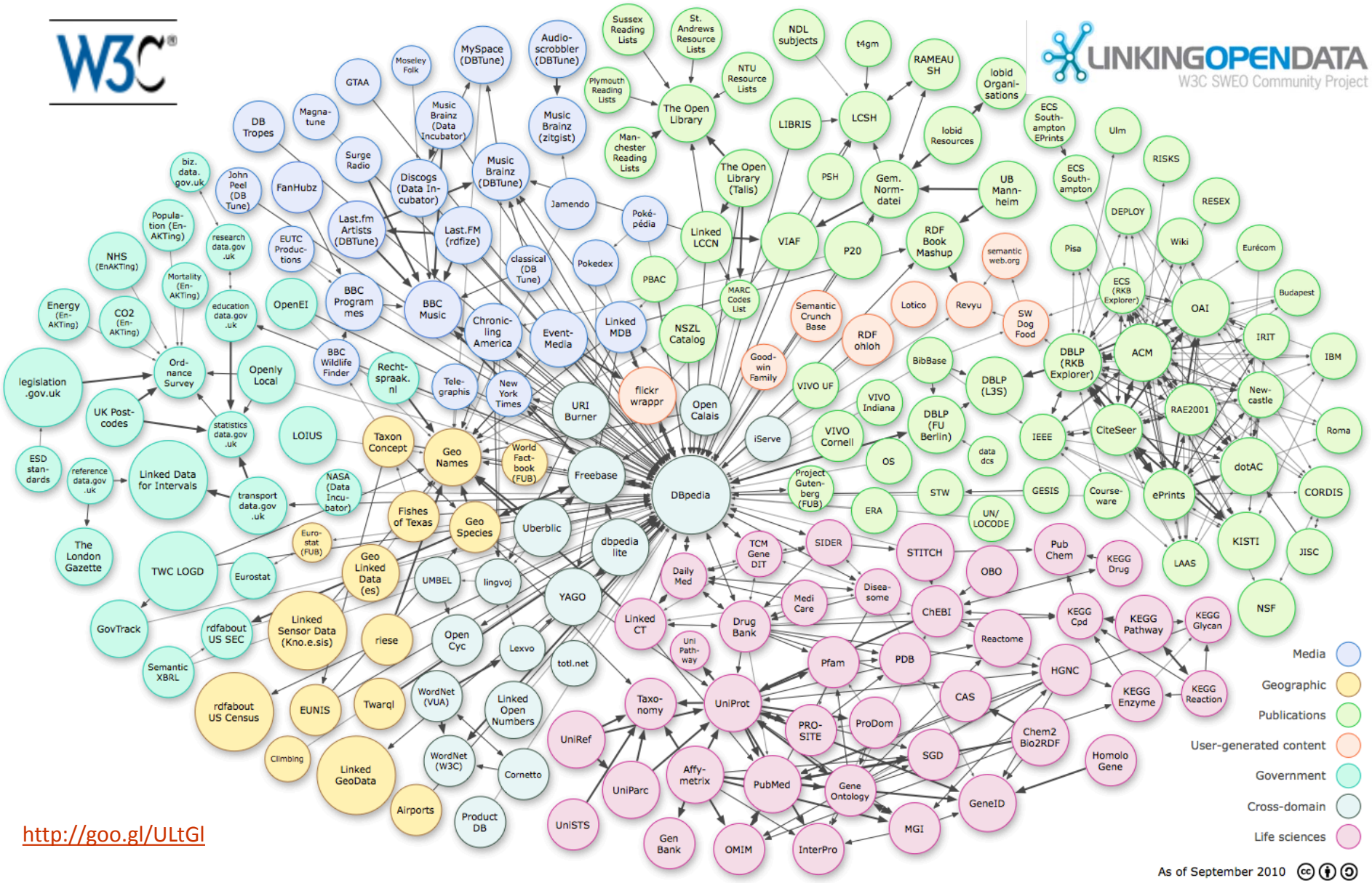
★★★★ All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff

★★★★★ All the above, plus: Link your data to other people's data to provide context

T.Berners Lee, Linked data, 2010

Linked Open Data

Video da Europeana



http://goo.gl/ULtGl

SOMMARIO: *1. Introduzione – 2. Interoperabilità tra i dati – 3. Tecnologie per favorire l'interoperabilità: RDF e SPARQL – 3.1. Resource Description Framework (RDF) – 3.2. SPARQL Protocol and RDF Query Language (SPARQL) – 4. Linee guida per la pubblicazione dei dati – 5. Pubblicare i Linked Data – 5.1. Pubblicazione di dati strutturati – 5.2. Pubblicazione di dati non strutturati: testo – 6. Open Data incontra il Linked Data – 6.1. Open Data e Open Licenses – 7. Creare i link verso archivi esterni – 8. Conclusioni*

Linked Open Data

differenti risorse. Da qui l'esigenza di poter descrivere i dati in maniera strutturata sfruttando formati in grado di gestire la relazione tra gli stessi. Per rispondere a questa esigenza, il *Web* sta evolvendo da spazio di raccolta e collegamento di documenti ad uno in cui i dati grezzi contenuti nelle risorse sono collegati o collegabili facilmente dalle macchine. Questa evoluzione è conosciuta con il nome di *Linked Data*. Il *Linked Data* è una metodologia che permette di aggregare e collezionare dati provenienti da fonti distribuite[2]. Per rendere completamente accessibili questi dati al mondo del Web, i dati stessi devono essere pubblicati sotto condizioni d'uso "aperte" (o "libere"), che ne consentano consultazione, navigazione (con qualsiasi mezzo e anche tramite *deep-linking*[3]) e aggregazione.

L'*Open Data*, letteralmente "dati aperti", è la corrente di pensiero (e il relativo "movimento") che cerca di rispondere all'esigenza di poter disporre di dati legalmente "aperti"[4], ovvero liberamente (ri-)usabili da parte del fruitore, per qualsiasi scopo[5]. L'obiettivo dell'Open Data può essere raggiunto per legge, come negli USA dove l'informazione generata dal settore pubblico federale è in pubblico dominio[6], oppure per scelta dei detentori dei diritti, tramite opportune

Linked Open Data

possiamo usare una comparazione del tipo: l'*Open Data* sta al *Linked Data*, come la rete *Internet* sta al *Web*. L'*Open Data*, quindi, è l'infrastruttura (o la "piattaforma") di cui il *Linked Data* ha bisogno per poter creare la rete di inferenze tra i vari dati sparsi nel Web. Il *Linked Data*, in altre parole, è una tecnologia ormai abbastanza matura e con grandi potenzialità, ma ha bisogno di grandi masse di dati tra loro collegati, ossia "linkati", per diventare concretamente utile. Questo, in parte, è già stato ottenuto ed è in corso di miglioramento, grazie a progetti come *DBpedia*[8] o *FreeBase*[9]. In parallelo ai contributi delle community online, un altro tassello importante – una sorta di "bulk upload" molto prezioso – potrebbe essere dato dalla disponibilità di grosse masse di dati pubblici, idealmente anche già linkati dalle istituzioni stesse – o comunque messi a disposizione in modo strutturato – che aiutino a raggiungere una "massa critica" di dati collegati. A partire dal substrato, rappresentato dalla disponibilità di fatto dei dati e dalla loro piena riutilizzabilità (in modo legale), il *Linked Data* può offrire una potente rappresentazione degli stessi, in termini di relazioni (collegamenti): in questo senso, *Linked Data* ed *Open Data* convergono e raggiungono la loro piena realizzazione nell'approccio *Linked Open Data*.

Rizzo –Morando-De Martin, Open Data: la piattaforma di dati aperti per il Linked Data, Informatica e diritto, 2011, 20, 493-511

Panton Principles

Principles for Open Data in Science

Endorse About Comment FAQ Translations Di

Science is based on building on, reusing and openly criticising the published body of scientific knowledge.

For science to effectively function, and for society to reap the full benefits from scientific endeavours, it is crucial that science data be made **open**.

By open data in science we mean that it is freely available on the public internet permitting any user to download, copy, analyse, re-process, pass them to software or use them for any other purpose without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. **To this end data related to published science should be explicitly placed in the public domain.**

http://pantonprinciples.org/

Panton principles

Formally, we recommend adopting and acting on the following principles:

1. Where data or collections of data are published it is critical that they be published with a clear and explicit statement of the wishes and expectations of the publishers with respect to re-use and re-purposing of individual data elements, the whole data collection, and subsets of the collection. This statement should be precise, irrevocable, and based on an appropriate and recognized legal statement in the form of a waiver or license.

   When publishing data make an explicit and robust statement of your wishes.

2. Many widely recognized licenses are not intended for, and are not appropriate for, data or collections of data. A variety of waivers and licenses that are designed for and appropriate for the treatment of data are described **here**. Creative Commons licenses (apart from CCZero), GFDL, GPL, BSD, etc are NOT appropriate for data and their use is STRONGLY discouraged.

   Use a recognized waiver or license that is appropriate for data.

Panton principles

3. The use of licenses which limit commercial re-use or limit the production of derivative works by excluding use for particular purposes or by specific persons or organizations is STRONGLY discouraged. These licenses make it impossible to effectively integrate and re-purpose datasets and prevent commercial activities that could be used to support data preservation.

   If you want your data to be effectively used and added to by others it should be open as defined by the ***Open Knowledge/Data Definition*** *– in particular non-commercial and other restrictive clauses should not be used.*

4. Furthermore, in science it is STRONGLY recommended that data, especially where publicly funded, be explicitly placed in the public domain via the use of the Public Domain Dedication and Licence or Creative Commons Zero Waiver. This is in keeping with the public funding of much scientific research and the general ethos of sharing and re-use within the scientific community.

   Explicit dedication of data underlying published science into the public domain via PDDL or CCZero is strongly recommended and ensures compliance with both the Science Commons ***Protocol for Implementing Open Access Data*** *and the* ***Open Knowledge/Data Definition***.

http://pantonprinciples.org/

Denton Declaration / 1

The Denton Declaration

An Open Data Manifesto

Introduction

On May 22, 2012 at the University of North Texas, a group of technologists and librarians, scholars and researchers, university administrators, and other stakeholders gathered to discuss and articulate best practices and emerging trends in research data management. This declaration bridges the converging interests of these stakeholders and promotes collaboration, transparency, and accountability across organizational and disciplinary boundaries.

Declarations

Open access to research data is critical for advancing science, scholarship, and society.

Research data, when repurposed, has an accretive value.

Publicly funded research should be publicly available for public good.

Transparency in research is essential to sustain the public trust.

The validation of research data by the peer community is an essential function of the responsible conduct of research.

Managing research data is the responsibility of a broad community of stakeholders including researchers, funders, institutions, libraries, archivists, and the public.

http://openaccess.unt.edu/denton_declaration

Denton Declaration / 2

Principles

1. Open access to research data benefits society, and facilitates decision making for public policy.
2. Publicly available research data helps promote a more cost-effective and efficient research environment by reducing redundancy of efforts.
3. Access to research data ensures transparency in the deployment of public funds for research and helps safeguard public good will toward research.
4. Open access to research data facilitates validation of research results, allows data to be improved by identifying errors, and enables the reuse and analysis of legacy data using new techniques developed through advances and changing perceptions.
5. Funding entities should support reliable long-term access to research data as a component of research grants due to the benefits that accrue from the availability of research data.
6. Data preservation should involve sufficient identifying characteristics and descriptive information so that others besides the data producer can use and analyze the data.
7. Data should be made available in a timely manner; neither too soon to ensure that researchers to benefit from their labor, nor too late to allow for verification of the results.

http://openaccess.unt.edu/denton_declaration

Denton Declaration / 3

8. A reasonable plan for the disposition of research data should be established as part of data management planning, rather than arbitrarily claiming the need for preservation in perpetuity.

9. Open access to research data should be a central goal of the lifecycle approach to data management, with consideration given at each stage of the data lifecycle to what metadata, data architecture, and infrastructure will be necessary to support data discoverability, accessibility, and long-term stewardship.

10. The costs of cyberinfrastructure should be distributed among the stakeholders – including researchers, agencies, and institutions – in a way that supports a long-term strategy for research data acquisition, collection, preservation, and access.

11. The academy should adapt existing frameworks for tenure and promotion, and merit-based incentiv to account for alternative forms of publication and research output including data papers, public data sets, and digital products. Value inheres in data as a standalone research output.

12. The principles of open access should not be in conflict with the intellectual property rights of researchers, and a culture of citation and acknowledgement should be cultivated rigorously and conscientiously among all practitioners.

13. Open access should not compromise the confidentiality of research subjects, and will comply with principles of data security defined by HIPAA, FERPA, and other privacy guidelines.

http://openaccess.unt.edu/denton_declaration

Denton Declaration / 4

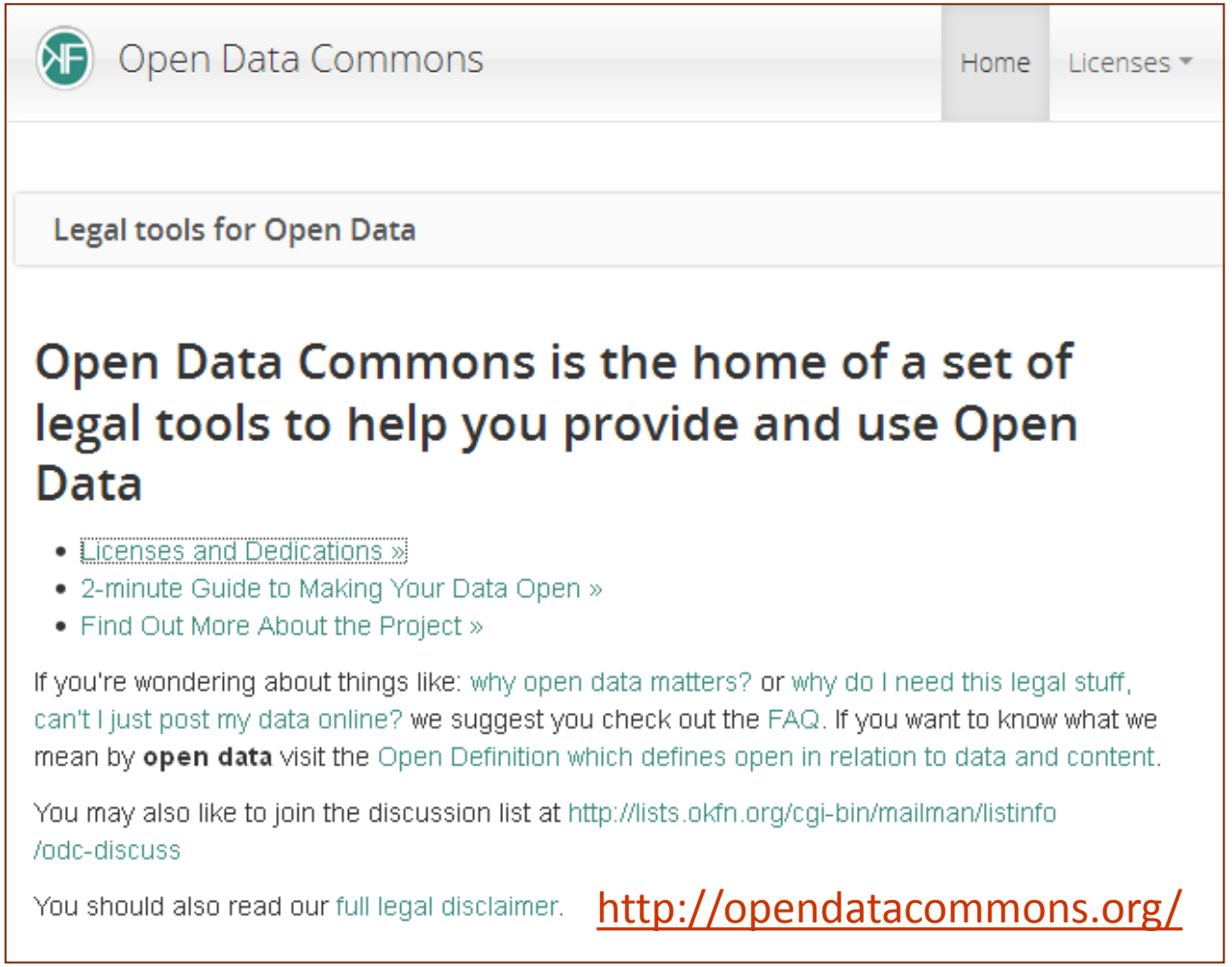
Intentions

In our professional interactions at meetings, on review panels, conferences, teaching, etc. we will advocate the following positions:

1. A culture of openness in research.
2. A federated model of archiving data to enable discoverability, transparency, and open access.
3. A robust and sustainable funding regime for research data management infrastructure (technical, policy, and human resources).
4. The development and adoption of metadata standards for research data.
5. Long-term access to data that supports published research outputs.
6. Support for researchers in negotiations with publishers to allow open access to research in repositories.
7. Recognition of researchers' intellectual property in data and scholarly research outputs.

http://openaccess.unt.edu/denton_declaration

Open Data: strumenti



http://opendatacommons.org/

Open Data: strumenti / 2

Open Data Handbook

Il manuale degli Open Data

Questo manuale affronta gli aspetti giuridici, sociali e tecnici degli open data (dati aperti). Il manuale può essere utilizzato da chiunque, ed è stato appositamente studiato per coloro che intendono **aprire** i dati. Il manuale discute del **perché, cosa e come** degli open data – quindi perché percorrere la strada dell'apertura, cosa si intende con 'aperto' (open), e come si fa 'open data' ("aprire/liberare" i dati).

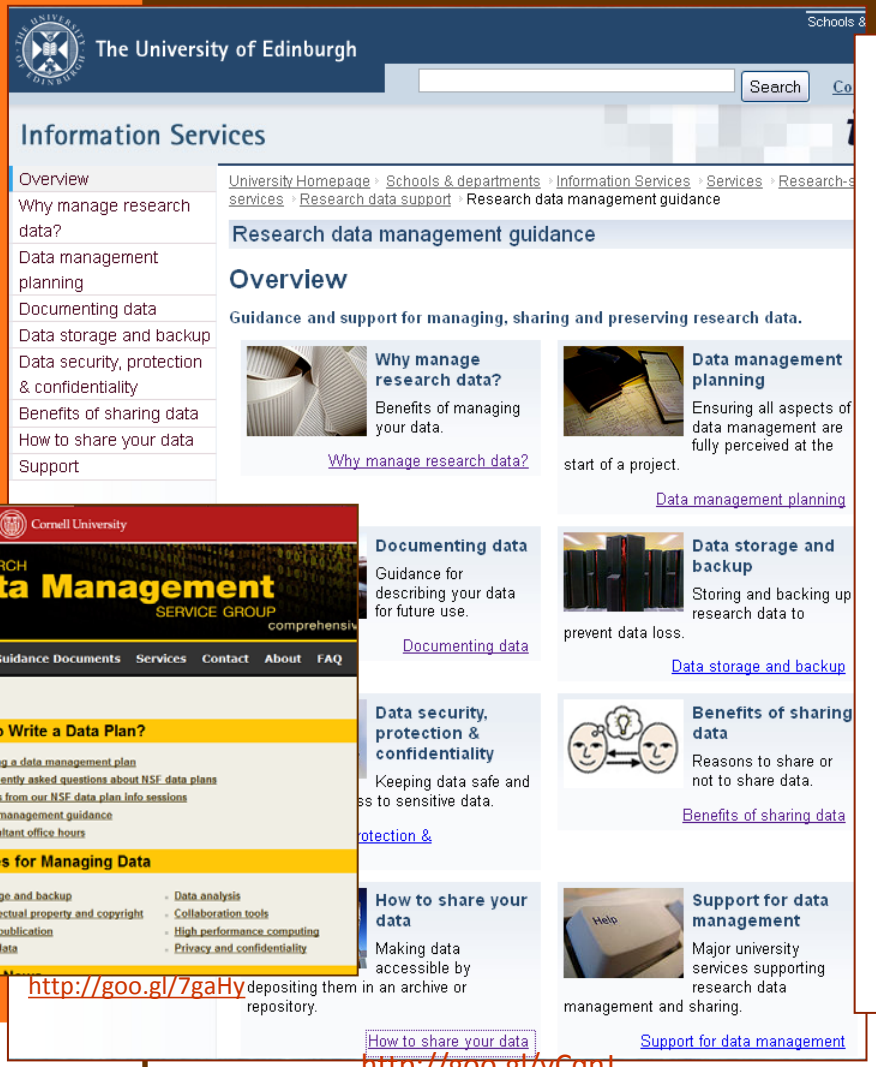
Per iniziare, puoi dare un'occhiata all'**introduzione**. Puoi navigare nel manuale usando l'indice dei contenuti (vd. barra laterale o sotto)

Accogliamo calorosamente ogni commento sul testo e incorporiamo i feedback di volta in volta. Siamo anche felici di coinvolgere chiunque voglia contribuire o suggerire nuove sezioni o aree da esaminare.

Indice dei Contenuti

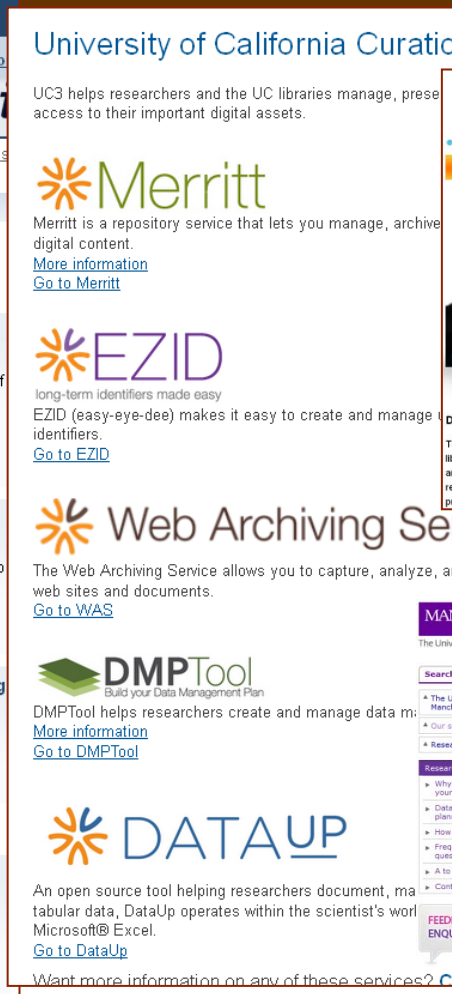
http://opendatahandbook.org/it/

Open Data: strumenti / 3

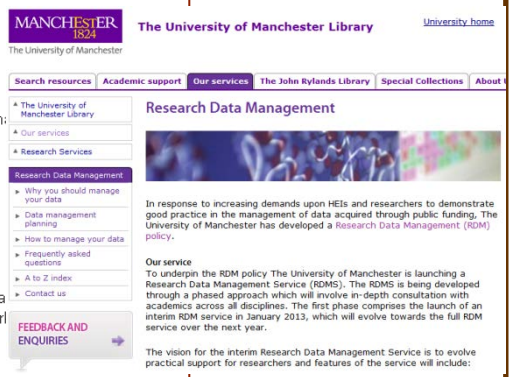


http://goo.gl/7gaHy

http://goo.gl/yCqnJ



http://d2c2.lib.purdue.edu/



http://www.cdlib.org/services/uc3/

Open Data: strumenti / 4

Implementing an Open Data Policy

A Primer for Research Funders

Prepared by Greg Tananbaum

On behalf of the Scholarly Publishing and Academic Resources Coalition (SPARC)

20 marzo 2013, http://goo.gl/9F5Zm

Open Data: strumenti / 5

gli enti di ricerca iniziano a richiedere
DATA MANAGEMENT PLAN

http://www.nsf.gov/bfa/dias/policy/dmp.jsp

Open Data: strumenti / 6

DCC because good research needs good data

DMP*online*

The DCC Data Management Planning Tool

Front Page

Plans

Data mangement plan

Welcome

Funding bodies increasingly require their grant-holders to produce and maintain Data Management Plans (DMPs), both at the bid-preparation stage and after funding has been secured.

DMP Online has been developed by the Digital Curation Centre to enable you to build and edit DMPs according to the requirements stipulated by the major UK funders.

The tool also contains helpful guidance and links

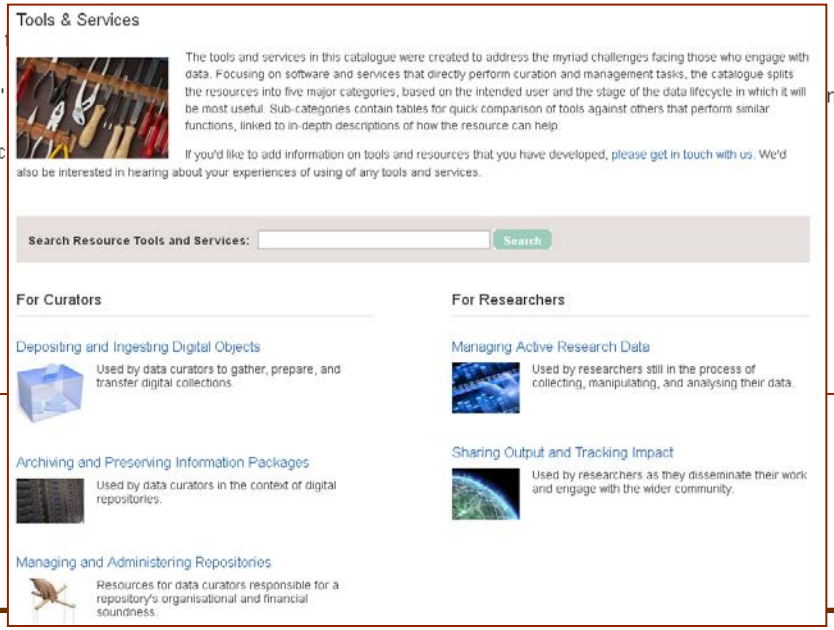
To get started, click on 'Create account' or 'Sign in'

n get going by clicking 'Start a new plan', below.

If you already have plans stored in the system, clic

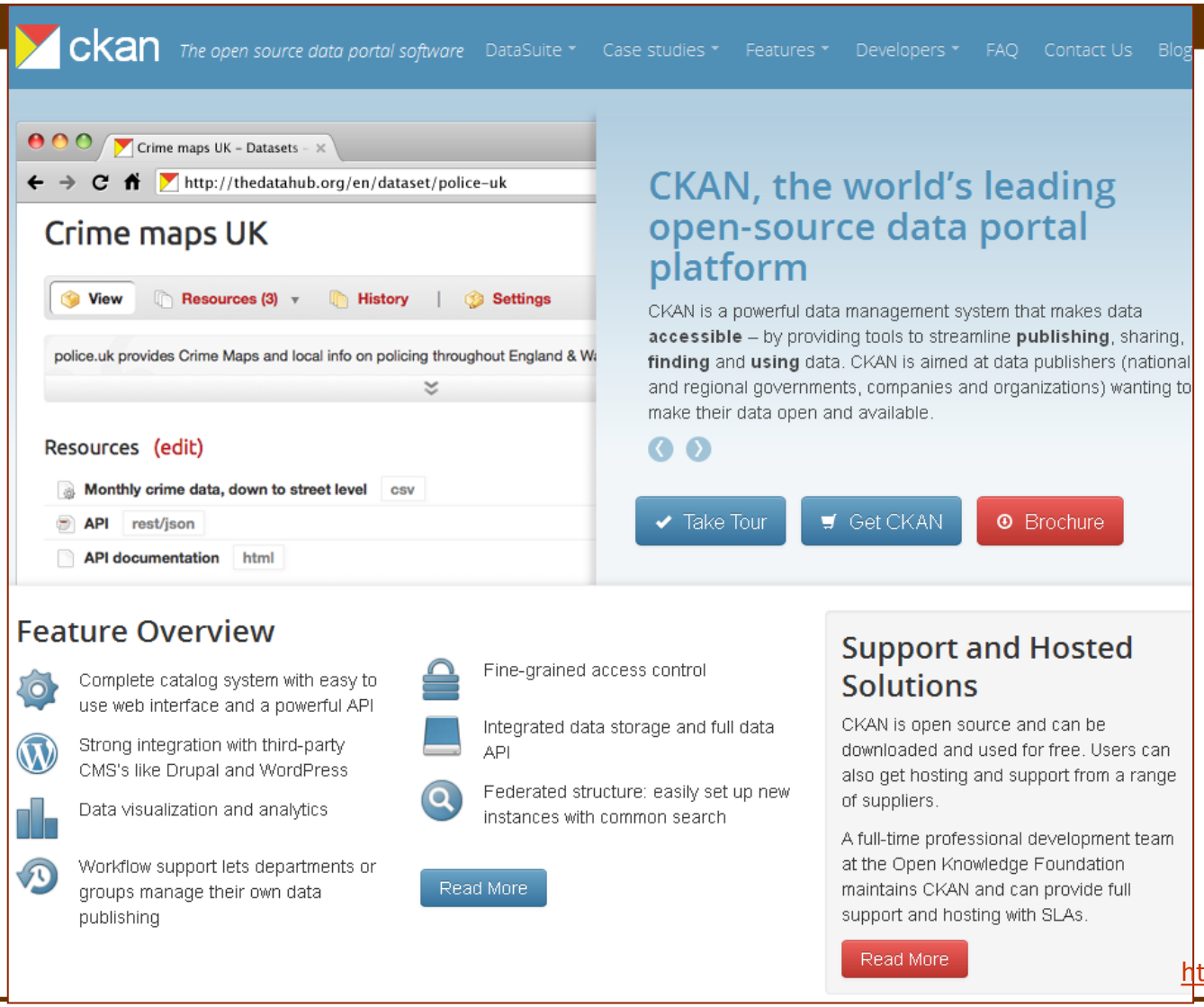
Start a new plan

Return to a saved plan



https://dmponline.dcc.ac.uk/

Open Data: strumenti / 7



http://ckan.org/

Licenze

1. Before you begin a database project, convene a meeting of all of the stakeholders. Expose all of the expectations of the group and decide if your goals are primarily scientific, commercial, or mixed. If mixed, take a stern look at the actual commercial potential of the project. Invite technology transfer offices to join you—they have greate experience in the realities of commercialization.
2. If your project is scientific in nature, and not commercial, explore the benefits of open licensing and drawbacks of enclosure. Go through the various definitions and find the most common ground possible, always placing the burden of proof on those who want more control and not less. This will create less "default enclosure" but allow for those increasingly rare situations in which "open" is not appropriate. Attempt to hew as closely as possible to the admittedly rigorous open definitions and standards, and do not write your own intellectual property licenses—instead, use existing and well deployed ones.
3. Develop simple explanations of your terms of use, and make them easy to find for users. Make sure that your licensing, expectations for attribution, terms of use, and more are linked in many ways to your data and database. Do not expect your users to read the legal text of your terms and conditions and licenses; instead, create simple summaries with linkages to the detailed text for users to access. Whenever possible, use metadata to indicate the licensing terms explicitly—the Creative Commons Rights Expression Language [30] is a good tool for this.
4. Don't ever lock up metadata. A significant swath of data will be incompatible with an open regime, whether it's to protect trade secrets or patient privacy. But the metadata that describes closed data, and how to access closed data, can be almost as valuable. If you can't make the data public domain, make the metadata public domain.

le licenze diventeranno sempre più importanti

servono a salire sulle spalle dei giganti e a evitare battaglie legali di Davide contro Golia

Williams AJ, Wilbanks J, Ekins S, Why Open Drug Discovery Needs Four Simple Rules for Licensing Data and Models. PLoS Comput Biol 2012, 8(9): e1002706

Open Data in dominio pubblico

Open Data Commons Home Licenses

Open Data Commons Public Domain Dedication and License (PDDL)

Open Data Commons Legal tools for Open Data Home

ODC Public Domain Dedication and License Summary

This is a human-readable summary of the Public Domain Dedication and License 1.0. Please see the disclaimer below.

You are free:

To Share: To copy, distribute and use the database.

To Create: To produce works from the database.

To Adapt: To modify, transform and build upon the database.

As long as you:

Blank: This section is intentionally left blank. The PDDL imposes no restrictions on your use of the PDDL licensed database.

Disclaimer

http://opendatacommons.org/licenses/pddl/

This is not a license. It is simply a handy reference for understanding the PDDL 1.0 — it is a human-readable expression of some of its key terms. This document has no legal value, and its contents do not appear in the actual license. Read the full PDDL 1.0 license text for the exact terms that apply.

creative commons About Licenses Public Domai

CC Wiki Home Page Discussion View source Histo

What links here Related changes Recent Changes Sp

Go

Log in / create account (OpenID)

Data

This page supersedes the 2006 document Databases and Creative Commons.

Much of the potential value of data is to society at large — more data has the potential to facilitate enhanced scientific collaboration and reproducibility, more efficient markets, increased government and corporate transparency, and overall to speed discovery and understanding of solutions to planetary and societal needs.

A big part of the potential value of data, in particular its society-wide value, is realized by use across organizational boundaries. How does this occur (legally)? Many sites give narrow permission to use data via terms of service. Much ad hoc data sharing occurs among researchers. And increasingly, open data is facilitated by sharing under public terms to manage copyright restrictions that might otherwise limit dissemination or reuse of data, e.g. CC licenses or the CC0 public domain dedication.

Many organizations, institutions, and governments are using CC tools for data. For case studies about how these tools are applied, see:

Uses of CC Licenses with Data and Databases

Uses of CC0 with Data and Databases

You can also read more about Creative Commons' most up-to-date thinking on data and databases, and what you can do to contribute.

Frequently asked questions about data

Can databases be released under CC licenses?

Yes, CC licenses can be used on any copyrighted work, including a copyrighted database. A CC license may be applied to any or all copyrighted aspects of a database and its contents. See below for more information regarding how to provide clear notice of what is licensed. Any use of the licensed database or its contents that is restricted by copyright law requires compliance with the relevant license conditions (BY, SA, NC, ND). In their current version (3.0) CC licenses do not require compliance with the license conditions when only sui generis database rights (and not copyright) are implicated. Additionally, the international and "ported" version 3.0 licenses, excluding EU jurisdiction ports, do not grant any permissions where sui generis database rights are implicated. Please see below for more detail.

CC0, the public domain dedication, can also be used on databases. The effect is to waive all copyright and related rights in the database, placing it as close as possible into the worldwide public domain. In certain domains, such as science and government, there are important reasons to consider using tools like CC0. Waiving copyright and related rights eliminates all uncertainty for potential users, encouraging maximal reuse and sharing of information. Where waiver is not a viable option and some conditions on reuse are necessary, rights holders should consider using CC licenses that give the public more freedom to reuse and remix the content.

http://wiki.creativecommons.org/Data

Which components of a database are protected by copyright?

Contents

[hide]

Creative Commons 0

CC0 1.0 Universal (CC0 1.0)
Public Domain Dedication

This is a human-readable summary of the **Legal Code (read the full text)**.

Disclaimer

No Copyright

The person who associated a work with this deed has **dedicated** the work to the public domain by waiving all of his or her rights to the work worldwide under copyright law, including all related and neighboring rights, to the extent allowed by law.

You can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission. See **Other Information** below.

Other Information

- In no way are the patent or trademark rights of any person affected by CC0, nor are the rights that other persons may have in the work or in how the work is used, such as **publicity or privacy** rights.
- Unless expressly stated otherwise, the person who associated a work with this deed makes no warranties about the work, and disclaims liability for all uses of the work, to the fullest extent permitted by applicable law.
- When using or citing the work, you should not imply **endorsement** by the author or the affirmer.

http://creativecommons.org/publicdomain/zero/1.0/

Europeana Data Exchange Agreement

dal 1 luglio 2012 richiede obbligatoriamente Licenze Creative Commons 0 [pubblico dominio]

Dati e licenze

Licences for Europe

text e data mining possono portare ingenti risparmi

possono ridurre il tempo di lettura dell'80%

per essere competitivi con USA e Asia (in cui sui dati c'e' eccezione al copyright, anche per prodotti europei, se risiedono su server locali) è necessario adottare soluzioni che facilitino text e data mining

l'incontro dello scorso 4 febb aveva come scopo una ulteriore negoziazione di licenze e pagamenti su materiale che invece sarebbe già legalmente accessibile

è un paradosso che mentre una parte della Commissione spinga per apertura massima in H2020, altre iniziative innalzino barriere

non è provato che alzare più barriere porti più innovazione e ricerca e beneficio, anzi…

l'uso dei dati non è in concorrenza con lo sfruttamento del lavoro originale di ricerca, anzi ne accresce il valore

http://goo.gl/yUX56

Dati e licenze

le nuove tecnologie venivano viste come una minaccia anche dai mercati. L'approccio corretto anche per i diritti è all'opposto: ADATTARE LE PRATICHE ALLE NUOVE ESIGENZE DEL DIGITALE

al momento sono necessari permessi e permessi per sfruttare i dati… questo impedisce ai ricercatori di competere e ai cittadini di trarre beneficio dalla scienza

l'iniziativa si chiama Licences for Europe MA non sempre nuove licenze sono la soluzione… spesso sono meglio soluzioni agili e pragmatiche

dobbiamo dimostrare che copyright e tecnologia possono andare d'accordo

N. Kroes 4 feb 2013

Portable legal consent

ABOUT US

The goal of Consent to Research is to play a part in the transformation of health from something we experience passively to something we experience actively. Our health is more than the medical visits we make, or the genomes we carry inside us, or our environments. We can now measure an enormous set of data about an individual, and if there were an enormous pool of publicly available health data, it will be far easier to apply powerful modeling techniques and begin to develop new kinds of hypotheses about the connections between our health, our DNA, and our choices.

Our strategy to achieve this goal is to create a massive pool of openly available, user-contributed data about health and disease. Right now data about people's health is expensive, complex to access, and the sample sizes are tiny.

The problem is that right now, it's not easy to donate your data to health research.

Thus, our tactics are to create standard, free tools for managing data donation and personal data, promote the use of open standards, and embed our tools into clinical study. These tools help people not only gather data about themselves and their health, but get that data into the hands of data-driven research scientists. Everything we do is free.

To help potential data donors, we make systems that let data donation happen: informed consent processes, institutional review board protocols, tools to extract data from the systems where it normally lives, and a software system that can receive data and get it to health researchers.

PORTABLE LEGAL CONSENT

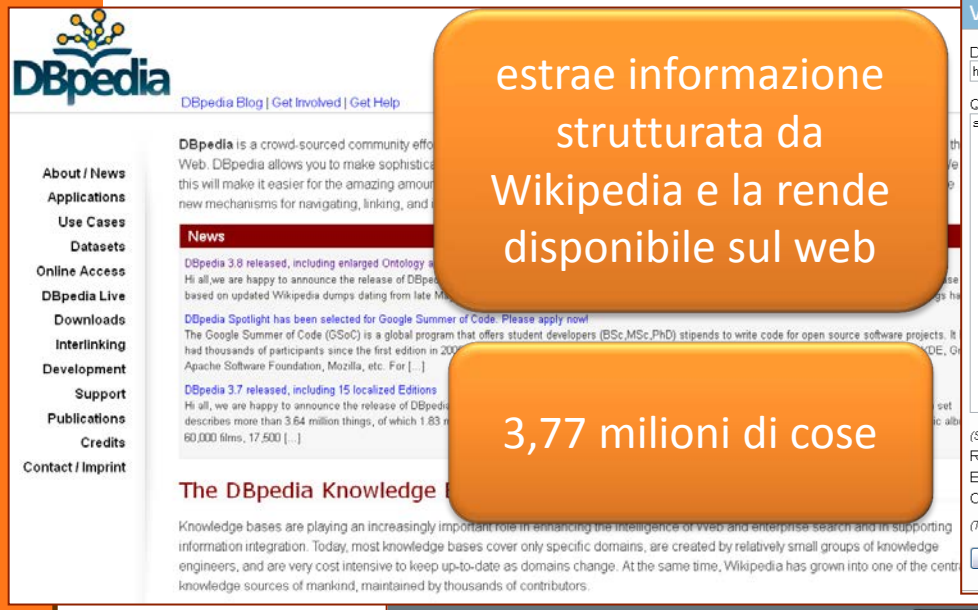
The legal system is Portable Legal Consent, so named because it allows data donors to carry their consent with them, and to attach that consent to the data they donate – it's portable.

PLC is at heart designed as an open standard, inspired by the way that technology protocols such as HTTP and TCP/IP work. That means that we're building a methodology by which multiple "open consent" projects might use their own consent forms and protocols, but have all the data be legally interoperable and usable as a common pool of information. We're starting by writing a specific implementation being used in the Self Contributed Cohort for Common Genomics Research study, or SCC-CGR, which has been positively received by the research and patient advocacy communities. SCC-CGR is a study run by Sage Bionetworks and you are welcome to enroll.

http://weconsent.us/

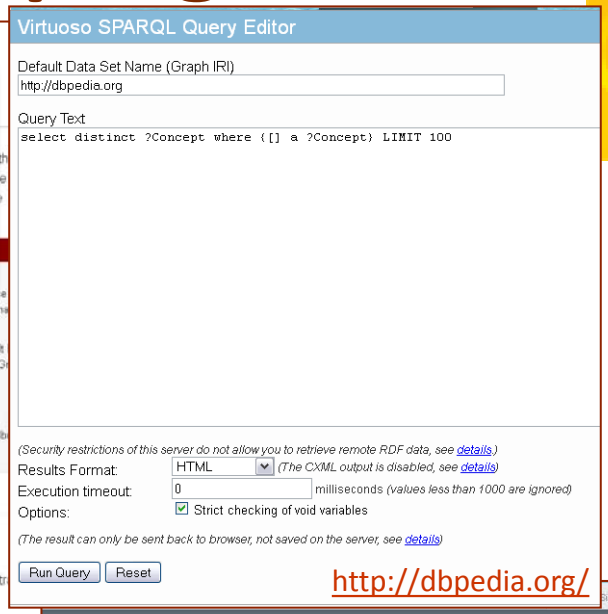
http://weconsent.us/portable-legal-consent-ted-talk-online/

Open Data: progetti

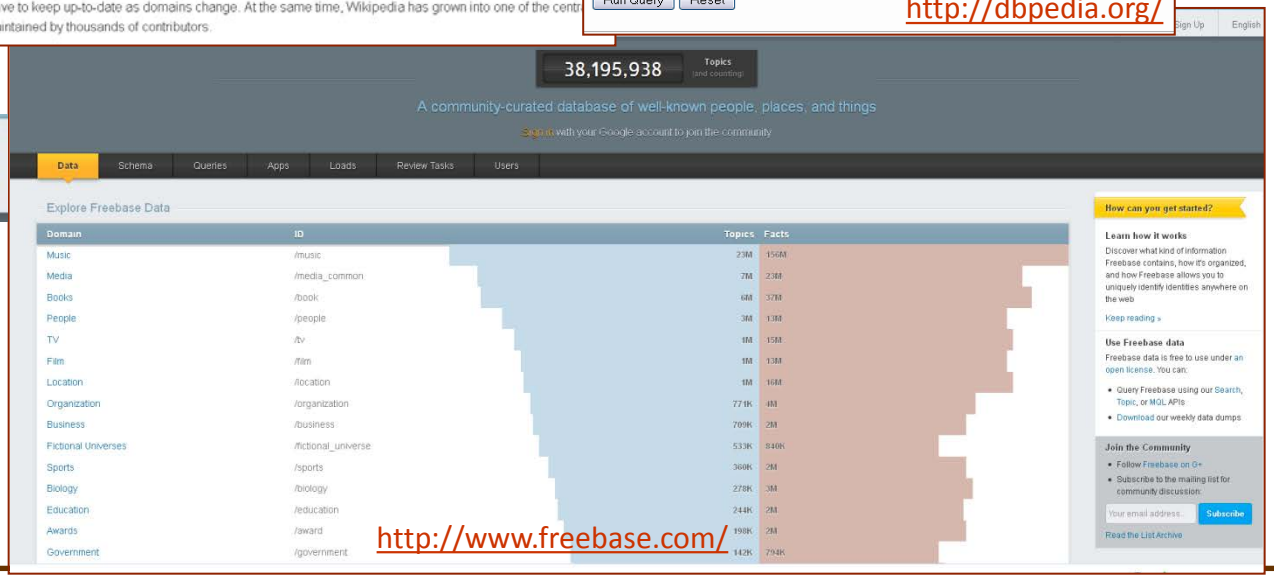


estrae informazione strutturata da Wikipedia e la rende disponibile sul web

3,77 milioni di cose

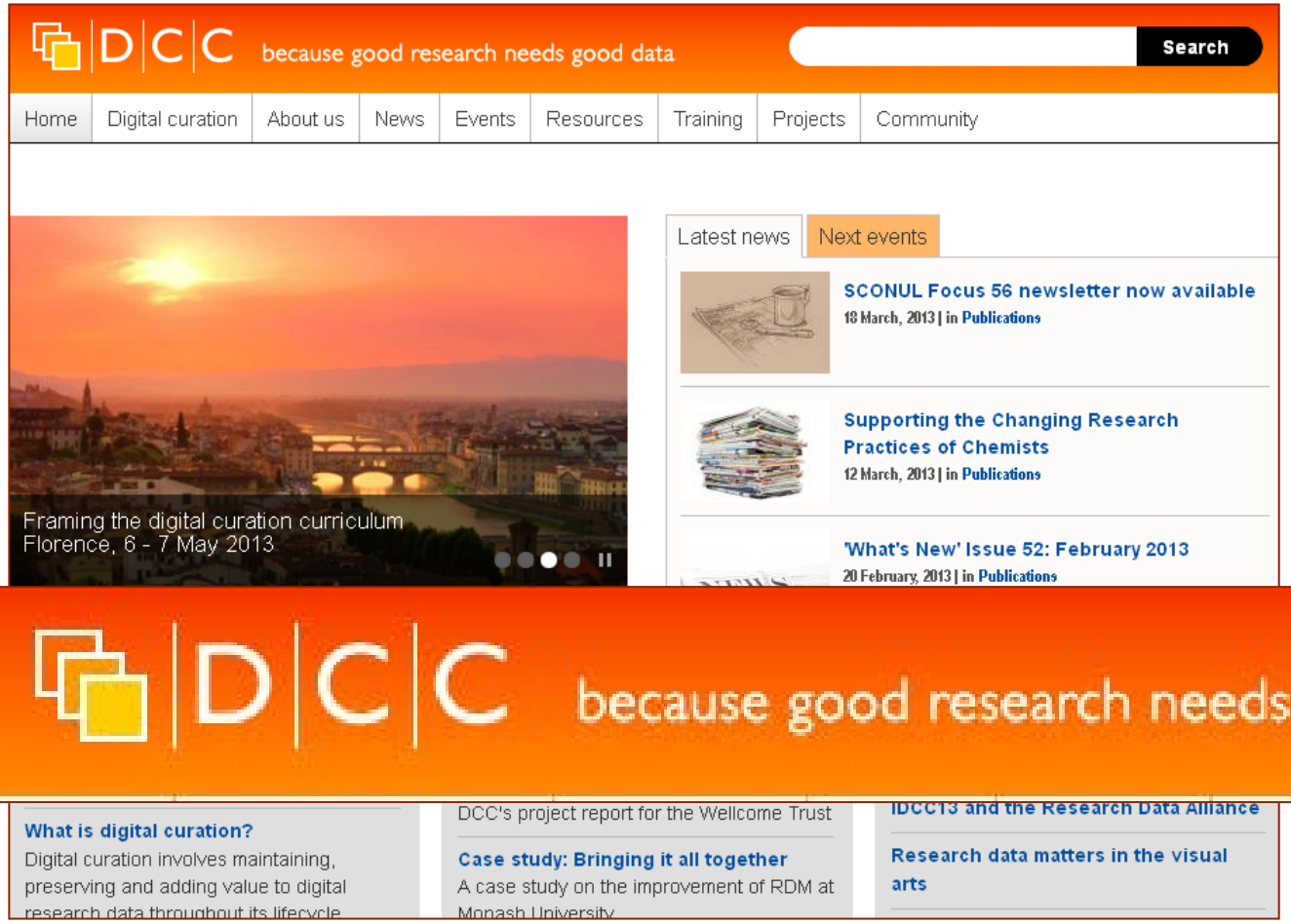


http://dbpedia.org/



http://www.freebase.com/

Progetti / Digital Curation Centre



http://www.dcc.ac.uk/

Progetti: Research data alliance

RESEARCH DATA ALLIANCE

Researchers around the world sharing and using research data without barriers

- About
- Organisation
- Working Groups +
- Participate
- News & Events +
- First Plenary +
- Documents & Presentations +

Research Data Alliance

The Research Data Alliance is a organisation that aims to accelerate and facilitate research data sharing and exchange. The work of the Research Data Alliance will primarily be undertaken through its **working groups**. Participation in working groups, starting new working groups, and attendance at the twice-yearly plenary meetings is open to all.

The Research Data Alliance Launch and First Plenary will be live streamed on the iCORDI web channel .

Launch

The first plenary meeting will be held in Gothenburg from March 18-20, 2013. Please mark your diaries now and plan to attend!

See the **RDA Launch Press Release**.

Progetto USA, EU, Australia

Participate

Participation in the RDA is open to anyone. There are several ways to participate in the RDA. Please see the overview **here**.

Purpose

http://rd-alliance.org/

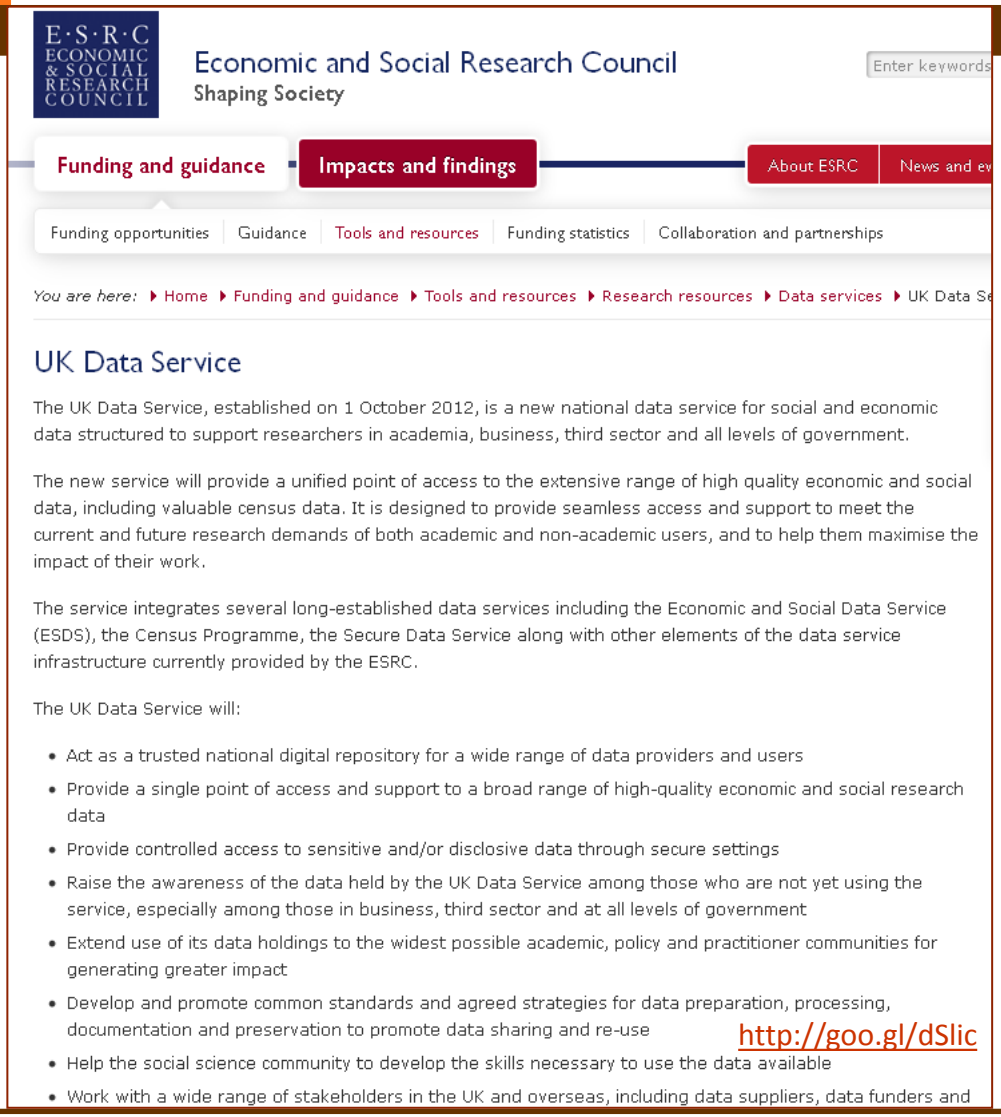
Progetti: OpenAIRE Plus

http://www.openaire.eu/

Linking Open Access publications to data, Workshop, Copenhagen 11 giugno 2012

OpenAIRE ha appena ricevuto il premio Europeo Better society (13 marzo 2013)

Progetti: UK data service



per dati sociali ed economici

unico punto di accesso

per ricerche accademiche e non

Progetti: Australian National Data Service

ANDS Home | Contact Us | Guides

Find Research Data: Search **Search our site:** Google™ Custom Search Go

Australian National Data Service

Our Vision: More Australian researchers reusing research data more often

ANDS is enabling the transformation of:

Data that are:	to	Structured Collections that are:
Unmanaged	→	Managed
Disconnected	→	Connected
Invisible	→	Findable
Single-use	→	Reusable

More>>

Australian Research Data Commons

ANDS is building the Australian Research Data Commons: a cohesive collection of research resources from all research institutions, to make better use of Australia's research data outputs.

Research Data Australia

Research Data Australia, ANDS' flagship service, provides a comprehensive window into the Australian Research Data Commons. It is an Internet-based discovery service designed to provide rich connections between data, projects, researchers and institutions, and promote visibility of Australian research data collections in search engines.

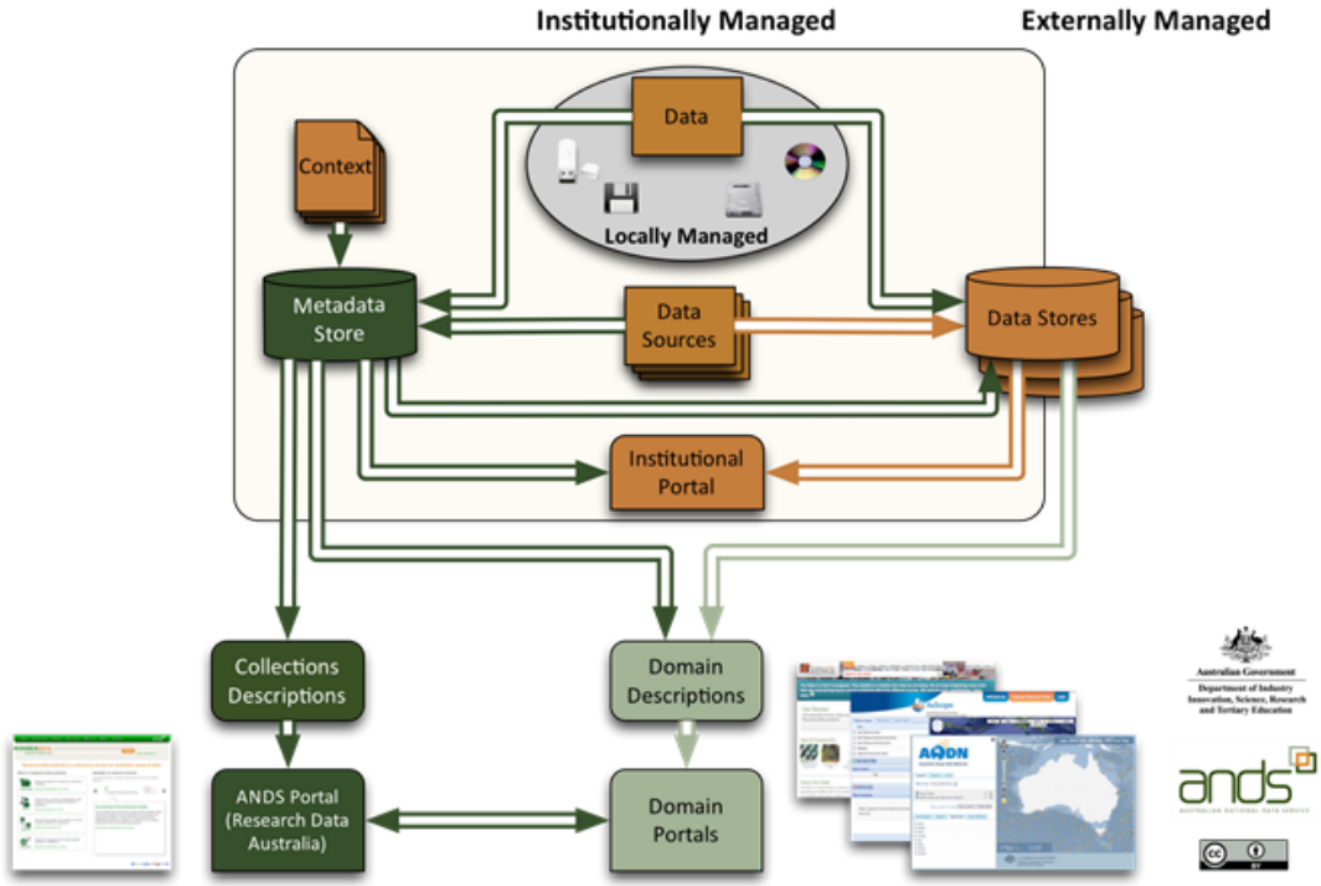
Search for research data on researchdata.ands.org.au

http://ands.org.au/

Progetti: Data Commons

ANDS is creating the Australian Research Data Commons—a meeting place for researchers and data, to provide:

- a set of data collections that are shareable
- descriptions of those collections
- an infrastructure that enables populating and exploiting the commons
- connections between the data, researchers, research, instruments and institutions



http://www.ands.org.au/about/approach.html#ardc

Progetti : MAPPA

mappa
metodologie applicate alla predittività del potenziale archeologico

- Il progetto
- Chi siamo
- Pubblicazioni
- Biblioteca
- News
- Multimedia
- MAPPAGIS
- MAPPAopenDATA
- Servizi
- Newsletter
- Opening the Past 2013
- Sondaggio OD

MAPPAopenDATA

Conservare, disseminare, collaborare. L'archeologia libera e aperta è la nuova frontiera della ricerca.

Il MOD (MAPPA Open Data) è l'archivio digitale archeologico pensato all'interno del gruppo di lavoro del Progetto MAPPA per conservare e disseminare la documentazione archeografica (Dataset) e la letteratura grigia (Relazioni) prodotta nel corso di una qualsiasi indagine archeologica.

Il Progetto MAPPA attraverso il MOD vuole tutelare chi ha prodotto la documentazione in tutte le sue forme sia a livello di dato grezzo archeografico (Dataset), sia a livello di letteratura grigia (Relazioni), per questo ha individuato nell'apposizione di un DOI ad ogni Dataset e/o Relazione e nell'utilizzo della licenza CC BY-SA 3.0, il modo per consentire contemporaneamente la corretta circolazione dei dati e la tutela della paternità intellettuale.

Il MOD nasce dall'esperienza di Pisa, ma con la collaborazione di tutti ambisce a diventare una piattaforma condivisa per tutto il territorio nazionale. Chiunque voglia collaborare aggiungendo nuovi Dati all'archivio può contattarci all'indirizzo **info@mappaproject.org**

L'inserimento della documentazione (dataset/relazioni) di un intervento all'interno del MOD è fatta su base volontaria dai detentori della paternità intellettuale dei dati (vedi Disclaimer). L'immissione di tali dati nel sistema è curata dal Progetto MAPPA.

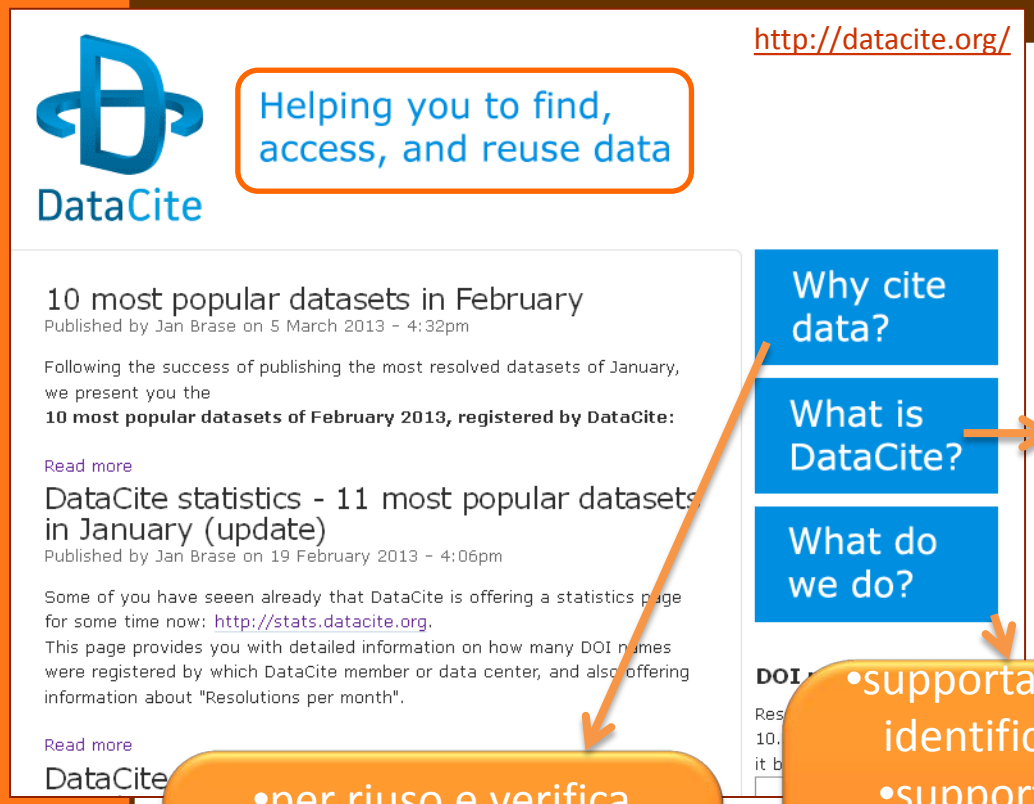
L'accesso al MOD è gratuito.

L'accesso all'applicazione presuppone l'accettazione delle **Condizioni di Servizio** e del **Disclaimer**.

>> ACCEDI AL MOD

http://goo.gl/FQkel

Progetti: DataCite



scopi:
- stabilire accesso ai dati sul web
- accrescere l'accettazione dei dati come parte legittima e citabile del record scientifico
- supportare l'archiviazione per il riuso futuro

- per riuso e verifica
- per tracciare l'impatto
- per consentire il riconoscimento agli autori

- supportare i ricercatori aiutandoli a cercare, identificare e citare i datasets con fiducia
- supportare i data centres con standard e identificatori persistenti per datasets e workflows
- supportare le riviste nella pratica di collegare gli articoli con i datasets

DataCite metadata schema

DataCite Metadata Schema Repository

Metadata Store

For our Metadata store all versions hosted on this site are valid although only one version is considered *preferred* at any given moment

Each incoming to Metadata store XML document should indicate schema location in the root element.

Deprecation notification

DataCite Metadata Schema 2.0 will be retired on 1st Aug 2011. After this date MDS will not accept documents based on this version.

Available versions

- DataCite Metadata Schema 2.2 (released 2011-07-01; preferred)
- DataCite Metadata Schema 2.1 (released 2011-03-28)
- DataCite Metadata Schema 2.0 (released 2011-01-24; deprecated)

citazione tipo:

Creator (PublicationYear): Title. Publisher. Identifier

ID	Property
1	Identifier (with type attribute)
2	Creator (with name identifier attributes)
3	Title (with optional type attribute)
4	Publisher
5	PublicationYear

Table 2: DataCite Optional Properties

ID	Property
6	Subject (with schema attribute)
7	Contributor (with type and name identifier attributes)
8	Date (with type attribute)
9	Language
10	ResourceType (with description attribute)
11	AlternateIdentifier (with type attribute)
12	RelatedIdentifier (with type and relation type attributes)
13	Size

Progetti: ODIN

ODIN | Our mission | Events | Project outputs | Workplan | Partners | Contact | Internal

What is ODIN?

ODIN – *ORCID and DataCite Interoperability Network* - is a two-year project which started in September 2012, funded by the European Commission's 'Coordination and Support Action' under the FP7 programme.

Partners in ODIN are innovators in science, information science and the publishing industry: CERN, the British Library, ORCID, DataCite, Dryad, arXiv and the Australian National Data Service (see **Partners**).

http://odin-project.eu/

The ODIN mission

ODIN will build on the ORCID and DataCite initiatives to uniquely identify scientists and data sets and connect this information across multiple services and infrastructures for scholarly communication. It will address some of the critical open questions in the area:

- Referencing a data object
- Tracking of use and re-use
- Links between a data object, subsets, articles, rights statements and every person involved in its life-cycle.

ODIN
ORCID and DataCite Interoperability Network

Recent Posts

- Save the date: ORCID Codefest and other events in Oxford May 23-24 March 12, 2013
- Dates set for ODIN codesprint and 1st year conference February 19, 2013
- ANDS publishes conference paper on ODIN February 7, 2013
- Report from October 2012 Kickoff Meeting now available February 4, 2013

RSS - Posts

twitter

olyerickson: Hearing about #DataCite #DéjàVu #RDALaunch

7 hours ago from TweetDeck

olyerickson: Now hearing about

Progetti: ODIN

The ODIN mission

'*Data as infrastructure*' is a critical concept for a fully-integrated European Research Area (ERA) to drive innovation forward as envisaged by the Digital Agenda for Europe. The lack of data availability hinders this vision. In academic publishing, peer review and citation have long been recognised as mechanisms for endorsing the trustworthiness of research outputs and incentivizing researchers to contribute. Trustworthy research data will only be widely available if the same principles are applied. Key, participative, initiatives have emerged to address this challenge.

The DataCite consortium aims to establish easier access to research data and increase acceptance of research data a legitimate contributions in the scholarly record, and to support data archiving to permit results to be verified and re-purposed for future study. DataCite has assigned over 1m digital object identifier (DOI) names in the last few years to make research data citable, true to emergence of the '4th paradigm', Jim Gray's vision of "data-intensive scientific discovery".

The Open Researcher & Contributor ID Initiative (ORCID) is an open, non-profit, community-based effort to provide a registry of unique researcher identifiers and a transparent method of linking research activities and outputs to these identifiers. ORCID is unique in its ability to reach across disciplines, research sectors, and national boundaries and its cooperation with other identifier systems.

ODIN aims to build on the success of DataCite and ORCID by designing an '*awareness layer*' for persistent author and object identifiers, thereby reducing technical, cultural and logistical barriers to the accessibility, attribution and trust of data. Identifier awareness will make it possible to stabilise: (1) References to a data object; (2) tracking of use and re-use; (3) Links between a data object, subsets, articles, rights statements and every person involved in its life-cycle (creator, editor, reviewer, aggregator, etc.).

Given the importance of these functions as we approach HORIZON 2020, we aim to prove the feasibility of author, data and rights identification, promote trust building towards open scientific data e-Infrastructures and lay the foundation necessary to promote future interoperability (technical, semantic, reference architecture, etc.) in the scientific data doma in Europe and globally.

dati come infrastruttura sono un concetto cruciale per ERA

trovare meccanismi degni di fiducia per i dati

1. fornire riferimento a un dato
2. tracciare uso e riuso
3. collegare dati, articoli, persone

[Den Haag manifesto]

identificatori persistenti e Linked Open Data

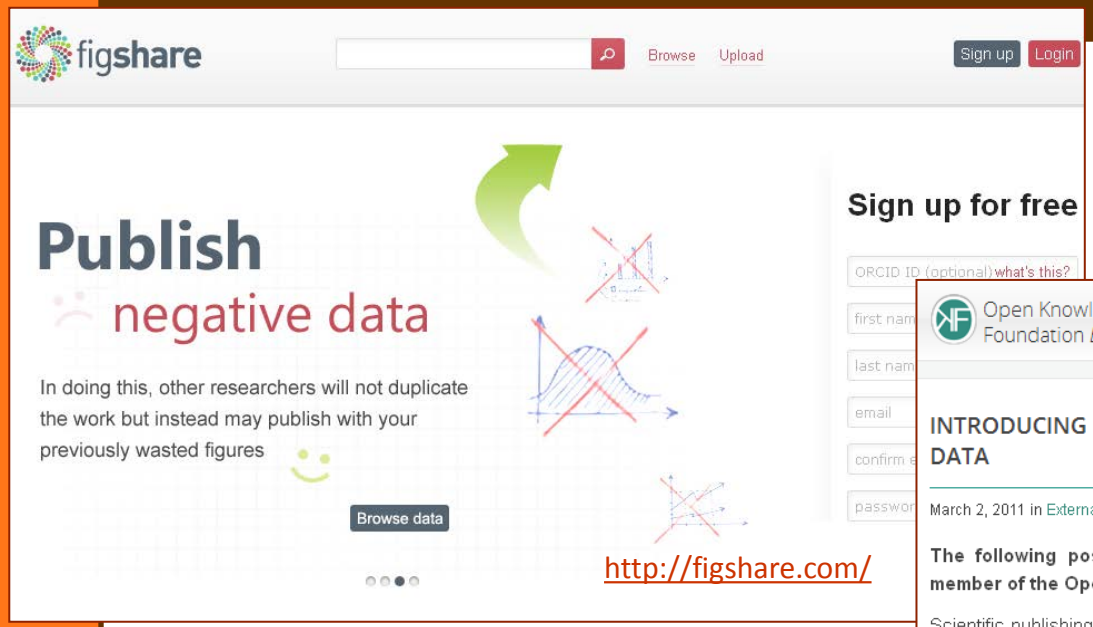
http://www.knowledge-exchange.info/Default.aspx?ID=462

Progetti: Databib

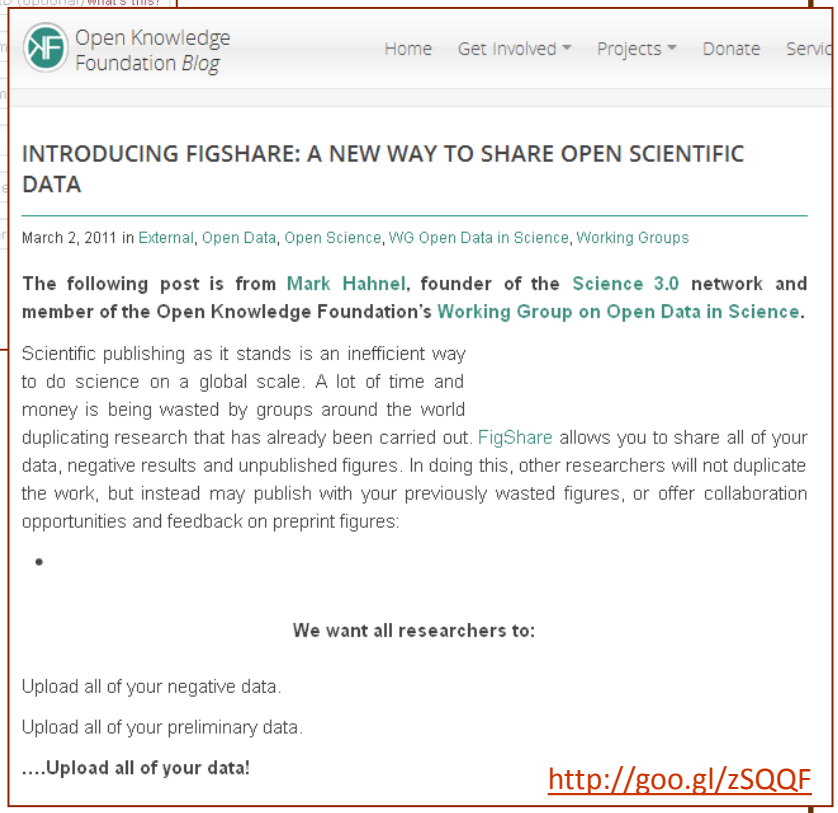
519 data repositories

http://databib.org/

Open Data: progetti / Figshare



http://figshare.com/



http://goo.gl/zSQQF

Progetti: DRYAD

DRYAD

Submit Data Now!

See how to submit

My Account

About Dryad

Dryad is both an international repository of data underlying peer-reviewed articles in the basi
related to published, or accepted, scholarly publications.

Dryad's objectives are to serve as a repository for tables, spreadsheets, and all other kinds

- validate published findings,
- explore new analysis methodologies,
- repurpose data for research questions unanticipated by the original authors,
- perform synthetic studies, and
- utilize data for educational purposes.

Data sharing makes our shoulders broader

DRYAD

- validare i risultati
- esplorare nuove metodologie
- porsi nuove domande, diverse da quelle dell'autore
- fare sintesi
- usare i dati a scopo didattico

Progetti: DRYAD

What should you deposit?

It is impossible to predict precisely which data will be most useful to researchers in the future. Dryad recommends depositing as much data as possible, in accordance with the following guidelines:

- Dryad collects data underlying peer-reviewed articles in the basic and applied biosciences.
- All data files must be associated with a publication.
- Authors should ideally submit sufficient data and metadata (i.e., documentation or descriptive information about the data) such that another researcher would be able to evaluate and reproduce the findings described in the publication.
- Data should be archived at the stage where it is rea
- Archived data files may include, but are not limited
  - spreadsheets or other tables
  - images or maps
  - alignments
  - character matrices
- You may also submit data that was originally collec is referenced by the current publication.

• solo dati riferiti a una pubblicazione peer reviewed
• con documentazione adeguata a riprodurre i risultati
• depositati in uno stadio in cui siano pronti per analisi statistica (non grezzi)

Progetti: DRYAD

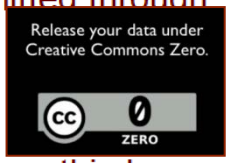
•integrazione con l'articolo
•LICENZA DI RIUSO
•embargo fino alla pubblicazione
•a disposizione dei referees

Why should I choose Dryad for my data?

Dryad aims to make data archiving as **simple** and as

Simple

- Dryad welcomes **data files associated with any published article in the biosciences,** as well as **software scripts and other files** important to the article.
- There is no restriction regarding data formats.
- Dryad works with journals to **integrate article and data submission**, streamlining the deposit process. Once the files are prepared, deposition typically takes **less than 15 minutes** (2-minute video here).
- Data destined for **more specialized repositories** can, in some cases, be submitted through Dryad, reducing the time and complexity of data submission yet further.
- Dryad provides a single clear and best-practice option for terms of reuse.
- A **curator** will check your files for technical problems before they are released.
- By default, data are **embargoed** until journal article publication. Dryad makes sure this happens so you do not need to.
- If it is supported by the policy of the journal, you may, during the submission process, select a **'no-questions-asked' embargo** on data downloads for one year post-publication. Dryad will support a longer embargo if directed by a journal editor.
- You are free to provide additional **keywords** that make the data easier to discover and additional **documentation** (in the form of ReadMe files) to help ensure proper data reuse.
- You have the ability to add **new versions** of data files in order to make updates or corrections.
- Dryad can make data **securely available for peer review** at the request of the journal.



Progetti: DRYAD

Rewarding

- Dryad works to ensure that you get **credit for reuse of your data** by promoting adopti best-practice data citation policy and the trackability of data citations.
- Data files receive **persistent, resolvable** Digital Object Identifiers (DOIs) that can be used in a citation as well as listed on your CV.
- Dryad's terms of reuse for data **facilitate the maximum impact** for your work.
- Data in Dryad are **independently discoverable**, providing a new route by which others may learn about your work.
- Discovery is supported through the **indexing** of Dryad's contents by services such as Google Scholar, Web of Science, and others.
- **Usage statistics** are available for you to highlight when your datasets are frequently downloaded.
- Since Dryad does not reject data for being of the wrong type or in the wrong format, **all the data files associated with an article can be archived together.**
- Dryad can host files that are **larger** than those accepted by most journal websites (up to 1 GB per file and 10 GB per package).
- Your data are **preserved** and made available for the long-term, even beyond the lifespan of Dryad, through continuous backup and replication services.
- Dryad is **community-led**, with priorities and policies shaped by the members of the Dryad Consortium, including scientific societies, publishers, and other stakeholder organizations.
- Dryad is a nonprofit, but takes **sustainability** seriously, ensuring that funds are available for long-term preservation.
- Dryad is an active participant in organizations developing **best-practices for data management** such as Biosharing, DataCite and DataONE. You as a researcher benefit from, and contribute to, the work of these organizations by depositing and using Dryad.

Progetti: DRYAD

DRYAD NEWS AND VIEWS

Dryad updates and goings on

Feeds: Posts Comments

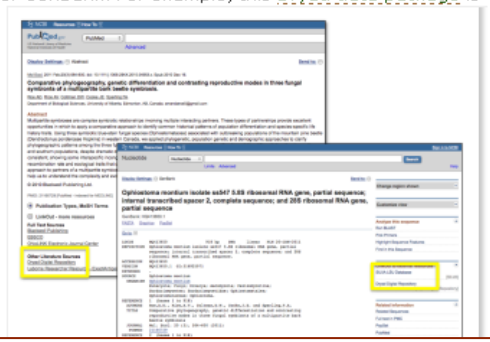
« Hope and change for research data in the US

Linking from PubMed and GenBank to data in Dryad

2013/03/06 by Hilmar

PubMed and GenBank, from the National Center for Biotechnology Information (NCBI), are hugely popular resources for searching and retrieving article abstracts and nucleotide sequence data, respectively. PubMed indexes the vast majority of the biomedical literature, and deposition of nucleotide sequences in GenBank or one of the other INSDC databases is a near universal requirement for publication in a scientific journal.

Thanks to NCBI's "LinkOut" feature, it is now easy to find associated data in Dryad from either PubMed or GenBank. For example, this Dryad data package is linked from:



- the article's abstract in PubMed. "LinkOut" is at the bottom of the page; expand "+" to see the links to Dryad and other resources.
- nucleotide data associated with the same publication in GenBank. "LinkOut" is

DRYAD

DRYAD TWEETS

RT @openscience: 'EU Opens Up Access to Scientific Research' ow.ly/jcNhL #OpenScience in the @WSJ 3 hours ago

Beyond the PDF 2 will be televised live - like any self-respecting revolution, starting Tuesday shrd.by/5GHsLP #btpdf2 1 day ago

Exciting session on open science at the Ecological Soc Amer meetings in Minneapolis this August bit.ly/141rIQx 2 days ago

Linking from PubMed and GenBank to data in Dryad

Progetti: Dataverse

The Dataverse Network™ Project

A Web Application for Sharing, Citing, Analyzing and Preserving Research Data

IQSS
The Institute for Quantitative Social Science
HARVARD UNIVERSITY

ABOUT | SOFTWARE | DATA CITATION | DATA MANAGEMENT | GUIDES

It enables data archiving and preservation through re-formatting, standards and exchange protocols.

It provides control and recognition for researchers through data management, branding and formal data citation.

Fact Sheets for:

- Scholars
- Research Project
- Journals
- Dissertation

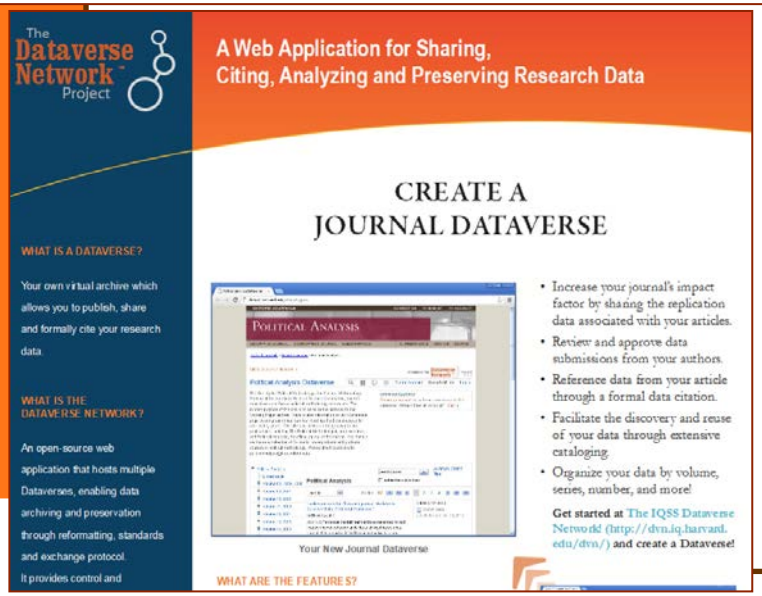
http://thedata.org/

Progetti: Dataverse

ABOUT THE PROJECT

The Dataverse Network is an open source application to publish, share, reference, extract and analyze research data. It facilitates making data available to others, and allows you to replicate others work. Researchers, data authors, publishers, data distributors, and affiliated institutions all receive appropriate credit.

A Dataverse Network hosts multiple dataverses. Each dataverse contains studies or collections of studies, and each study contains cataloging information that describes the data *plus* the actual data and complementary files.



IQSS
The Institute for Quantitative Social Science
HARVARD UNIVERSITY

Harvard's *home for social science research.*
Creating, sharing, and preserving scientific knowledge about human societies.

POWERED BY THE Dataverse Network™ PROJECT v. 3.3

IQSS Dataverse Network

Create Account Log In

Access the world's largest collection of social science research data here by searching across or browsing through one of the virtual data archives - called "dataverses" - listed below. You may also create a dataverse of your own to share your social science data and get a formal persistent citation. Your dataverse may easily be branded as your web page, or embedded in your web site. Learn more about the Dataverse Network Project at http://thedata.org.

CREATE A DATAVERSE
Create a Dataverse to upload your own data sets and create collections of data.

Show Released Dataverses

Filter by Organization
- Association, Society
- Business, Industry
- Government
- School (Harvard)
- School (Other)

Filter by Type
- Institutions, Depts, Centers
- Journal Replication Archive
- Large Collections
- Research Projects
- Scholars

Released Dataverses

Search Studies Go Advanced Search Tips

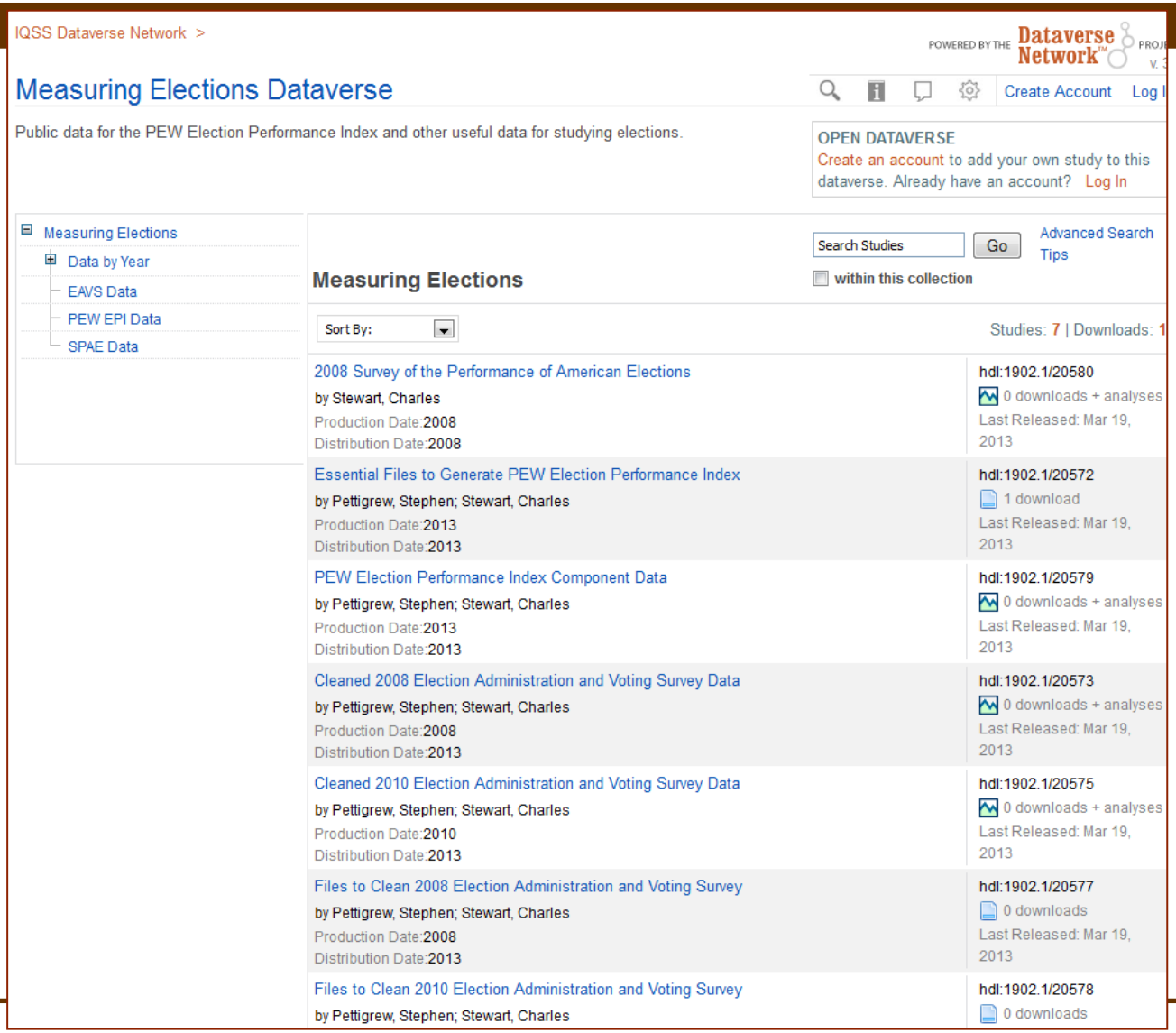
ALL # A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Dataverses: 513 | Studies: 51,771 | Files: 720,732

1 2 3 4

Name	Affiliation	Released	Activity
Measuring Elections View Info [+]	MIT	Mar 19, 2013	
IPod Reading Eye tracking experiment View Info [+]	Harvard-Smithsonian Center for Astrophysics	Mar 10, 2013	
Kathryn J. Alexander View Info [+]	Duke University	Mar 7, 2013	
Andrew Ballard's Dataverse View Info [+]	Duke University	Mar 7, 2013	
Humidtropics, a CGIAR Research Program on View Info [+]	Humidtropics	Mar 3, 2013	
Josh Lerner View Info [+]	Duke University	Feb 27, 2013	
Stephen Morgan View Info [+]	Duke University	Feb 27, 2013	
Replication of Pollard and Morgan (2002) View Info [+]	Duke University	Feb 27, 2013	
Carl Klarner View Info [+]	Indiana State University	Feb 27, 2013	
Sentinel Landscape Nicaragua Honduras - GIS View Info [+]	Research Project	Feb 18, 2013	

Progetti: Dataverse



IQSS Dataverse Network >

POWERED BY THE Dataverse Network™

Measuring Elections Dataverse

Create Account Log I

Public data for the PEW Election Performance Index and other useful data for studying elections.

OPEN DATAVERSE
Create an account to add your own study to this dataverse. Already have an account? Log In

- Measuring Elections
  - Data by Year
  - EAVS Data
  - PEW EPI Data
  - SPAE Data

Search Studies Go Advanced Search Tips

within this collection

Measuring Elections

Sort By:

Studies: 7 | Downloads: 1

2008 Survey of the Performance of American Elections
by Stewart, Charles
Production Date:2008
Distribution Date:2008
hdl:1902.1/20580
0 downloads + analyses
Last Released: Mar 19, 2013

Essential Files to Generate PEW Election Performance Index
by Pettigrew, Stephen; Stewart, Charles
Production Date:2013
Distribution Date:2013
hdl:1902.1/20572
1 download
Last Released: Mar 19, 2013

PEW Election Performance Index Component Data
by Pettigrew, Stephen; Stewart, Charles
Production Date:2013
Distribution Date:2013
hdl:1902.1/20579
0 downloads + analyses
Last Released: Mar 19, 2013

Cleaned 2008 Election Administration and Voting Survey Data
by Pettigrew, Stephen; Stewart, Charles
Production Date:2008
Distribution Date:2013
hdl:1902.1/20573
0 downloads + analyses
Last Released: Mar 19, 2013

Cleaned 2010 Election Administration and Voting Survey Data
by Pettigrew, Stephen; Stewart, Charles
Production Date:2010
Distribution Date:2013
hdl:1902.1/20575
0 downloads + analyses
Last Released: Mar 19, 2013

Files to Clean 2008 Election Administration and Voting Survey
by Pettigrew, Stephen; Stewart, Charles
Production Date:2008
Distribution Date:2013
hdl:1902.1/20577
0 downloads
Last Released: Mar 19, 2013

Files to Clean 2010 Election Administration and Voting Survey
by Pettigrew, Stephen; Stewart, Charles
hdl:1902.1/20578
0 downloads

Progetti: DataONE

DataONE
Data Observation Network for Earth

Search ONEMercury For Go Connect

About Participate Resources Education Data

New DataONE Education Resources

DataONE provides CC0 access to a suite of ten training modules that you can download and include in your data management teaching. Modules cover all aspects of the Data Life... read more

Tweets

DataONE @DataONEorg 18 Mar
RT @trishacruse: Chris Greer: Biggest threat to success is losing track of the mission #RDALaunch

DataONE @DataONEorg 18 Mar
Peter Fox: If it's not just about the data, then what is it about? the alliance #RDALaunch

DataONE @DataONEorg 18 Mar
Peter Fox: Work as if you've succeeded #RDALaunch

Expand

Compose new Tweet...

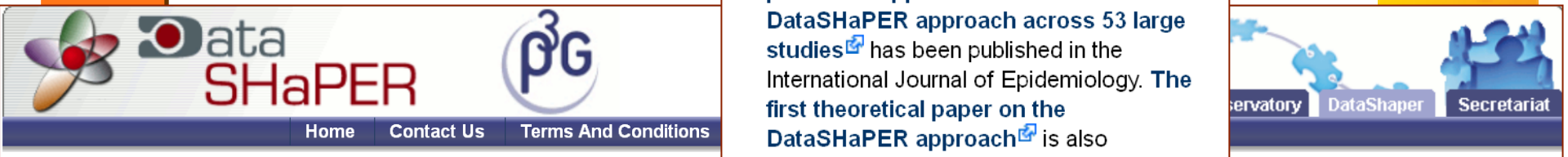
Latest News

Data Management Guide for Citizen Scientists released
Release of DataONE R Client
DMPTool Partners Awarded Funding for Enhancement
View All News

Find it Fast

ONE*Mercury*
Data Management Planning
Best Practices
Software Tools

Progetti: Datashaper



What's New ?

- **Is rigorous retrospective harmonization possible? Application of the DataSHaPER approach across 53 large studies** has been published in the International Journal of Epidemiology. **The first theoretical paper on the DataSHaPER approach** is also available on the IJE website
- The **Renal DataSchema** is now online. This DataSchema addresses Chronic Kidney Disease (CKD) and was developed in collaboration with **GENECURE**;

WHAT IS THE DATASHAPER

GENERAL DESCRIPTION

The DataSHaPER (DataSchema and Harmonization Platform for Epidemiological Research) is both a scientific approach and a suite of practical tools. Its primary aims are to facilitate the prospective harmonization of emerging biobanks, provide a template for retrospective synthesis and support the development of questionnaires and information-collection devices, even when pooling of data with other biobanks is not foreseen.

Its basic structure reflects a four step approach to harmonization:

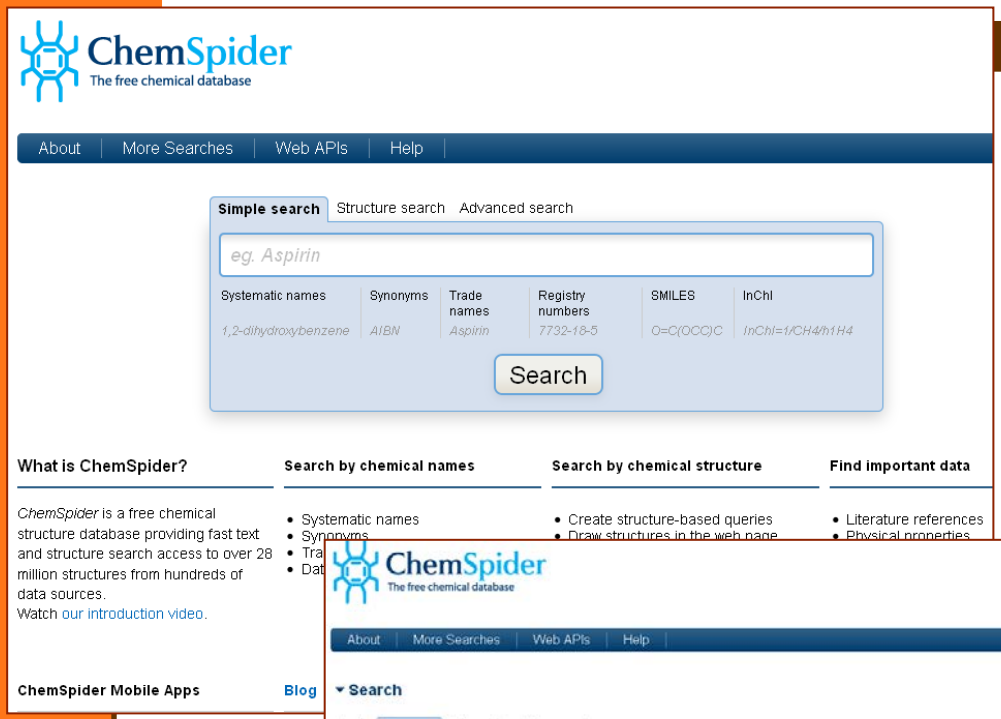
- Identify and document the set of core variables to be shared;
- Formally assess the potential to share each variable between participating studies;
- Define appropriate data processing algorithms;
- Process and synthetize real data.

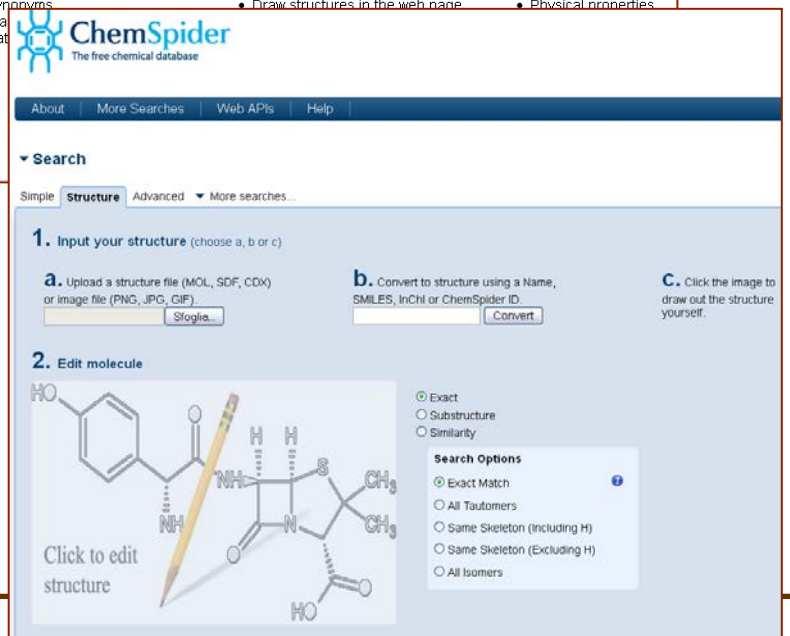
In the context of the DataSHaPER, the term "variables" refers to the primary units of interest in a statistical analysis (e.g. current smoker [yes/no], or body mass index as a quantitative trait). An important distinction is drawn between such variables and the specific "assessment items" that are collected by a particular study (e.g. questions in a questionnaire or physical measures collected by a study). Crucially, it is variables that are harmonized between studies and it is this that provides for flexible yet robust harmonization, because a given variable may potentially be built using different assessment items in different studies.

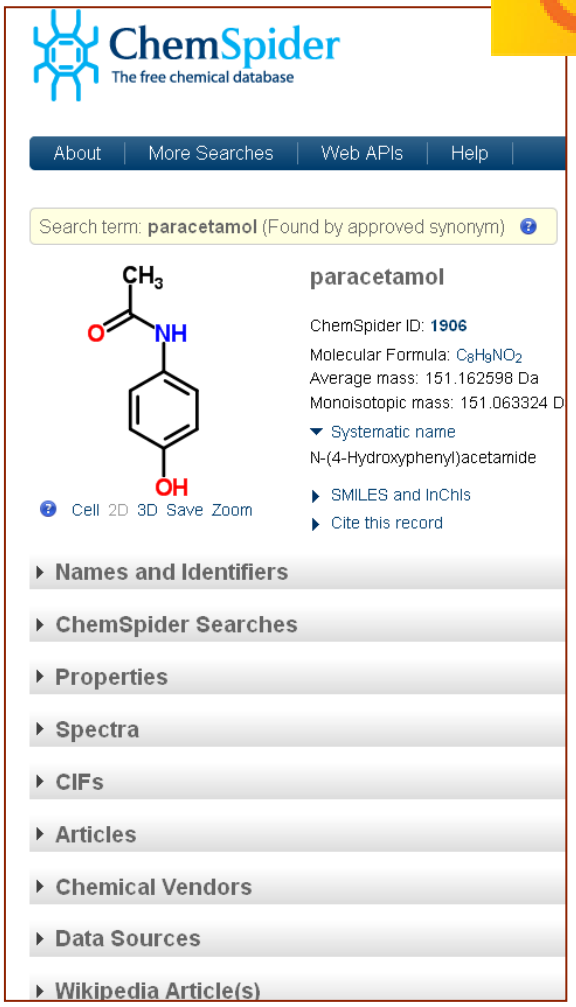
Structurally, the DataSHaPER is a dynamically evolving entity with two primary components: the **DataSchema Platform** and the **Harmonization Platform**.

http://www.datashaper.org/

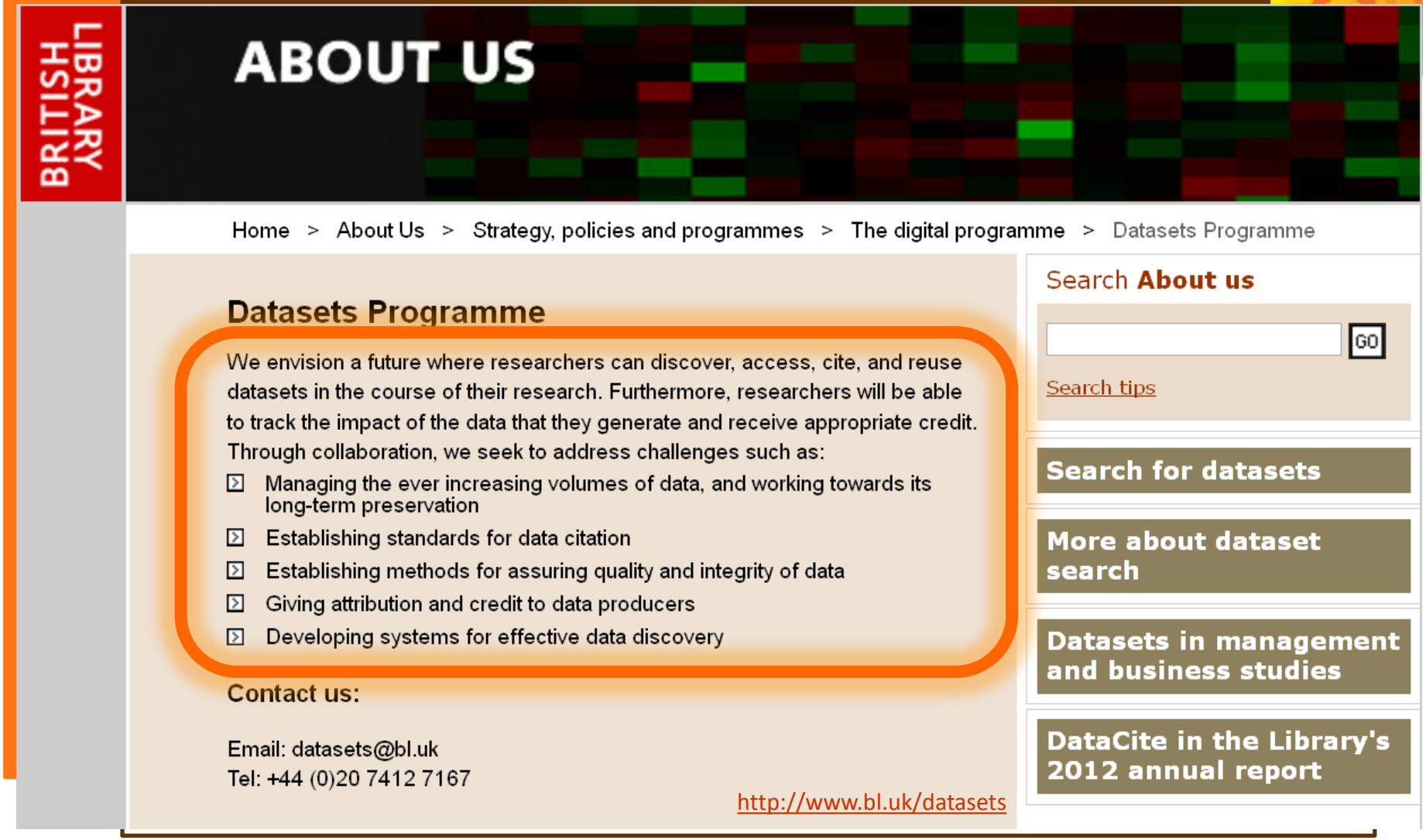
Progetti: ChemSpider







Open data: vision



Open Metadata principles

discovery A metadata ecology for UK education & research

Home | Vision | The Business Case | Developer Zone | Projects | News & Events

- Principles
- Principles Signatories
- Related Reading
- Videos

Discovery Open Metadata Principles

Discovery, the metadata ecology for UK education and research, invites stakeholders to join us in adopting a set of principles to enhance the impact of our knowledge resources for the furtherance of scholarship and innovation.

1. **Open metadata** creates the opportunity for enhancing impact through the release of descriptive data about library, archival and museum resources. It allows such data to be made freely available and innovatively reused to serve researchers, teachers, students, service providers and the wider community in the UK and internationally.
2. **Libraries, archives and museums** exist in an environment where they, their stakeholders and partners share an interest in exposing and repurposing metadata.
3. **Government and the wider community** are increasingly embracing a commitment to openness, sharing and reuse of metadata, especially where resources have been collected using public funds.

Further information

Licensing Open Data: A Practical Guide (PDF)

The Technical Principles

http://goo.gl/AVkxQ

Open Metadata principles

4. **We therefore recognise** the importance of promoting a clear endorsement of open metadata with practical guidance about licensing.
5. **We recommend** that institutions and agencies should proceed on the presumption that their metadata is by default made freely available for use and reuse, unless explicitly precluded by third party rights or licences.
6. **We strongly advocate** that all metadata releases require licensing, for which institutions and agencies should adopt a standard open licensing framework that is suited to their purposes.
7. **Reference to permissible usage** under the terms of a standard open licence will promote confident and appropriate use. When licensing open metadata in the majority of circumstances, the standard Open Data Commons Public Domain Dedication & Licence (ODC-PDDL), the broadly similar Creative Commons CC0 licence or the UK Open Government Licence (OGL) will be appropriate.
8. **Avoidance of variations** to such standard licences will make it easier to combine data from different resources and will reduce repeated requirement for legal advice.

http://goo.gl/AVkxQ

Library Linked Data


W3C Incubator Group Report

Library Linked Data Incubator Group Final Report

W3C Incubator Group Report 25 October 2011

Key recommendations of the report are:

- That **library leaders** identify sets of data as possible candidates for early exposure as Linked Data and foster a discussion about Open Data and rights;
- That **library standards bodies** increase library participation in Semantic Web standardization, develop library data standards that are compatible with Linked Data, and disseminate best-practice design patterns tailored to library Linked Data;
- That **data and systems designers** design enhanced user services based on Linked Data capabilities, create URIs for the items in library datasets, develop policies for managing RDF vocabularies and their URIs, and express library data by re-using or mapping to existing Linked Data vocabularies;
- That **librarians and archivists** preserve Linked Data element sets and value vocabularies and apply library experience in curation and long-term preservation to Linked Data datasets.

2 Benefits of the Linked Data approach

The Linked Data approach offers significant advantages over current practices for creating and delivering library data while providing a natural extension to the collaborative sharing models historically employed by libraries. Linked Data and especially Linked Open Data is **sharable, extensible,** and easily **re-usable.** It supports multilingual functionality for data and user services, such as the labeling of concepts identified by language-agnostic URIs. These characteristics are inherent in the Linked Data standards and are supported by the use of Web-friendly identifiers for data and concepts. Resources can be described in collaboration with other libraries and linked to data contributed by other communities or even by

W3C, Library Linked Data, ott. 2011

Library Linked Data

notion is fully compatible with the long-term mandate of libraries. Libraries, and memory institutions generally, are in a unique position to provide trusted metadata for resources of long-term cultural importance as data on the Web.

principles. Links between libraries and non-library services such as Wikipedia, GeoNames, MusicBrainz, the BBC, and The New York Times will connect local collections into the larger universe of information on the Web.

Linked Data is not about creating a different Web, but rather about enhancing the Web through the addition of structured data. This structured data, expressed using technologies such as RDF in Attributes (RDFa) and microdata, plays a role in the crawling and relevancy algorithms of search engines and social networks, and will provide a way for libraries to enhance their visibility through search engine optimization (SEO). Structured data

W3C, Library Linked Data, ott. 2011

Bibliographic framework transition initiative

http://www.loc.gov/marc/transition/

Bibframe (Library of Congress)

BIBFRAME.ORG New Bibliographic Framework

Overview | Vocabulary | Demos | Tools | Contribute

Home

BIBFRAME Model Overview

The Library of Congress has launched the Bibliographic Framework Transition Initiative to better accommodate future needs of the library community. A major focus of the initiative will be to determine a transition path for the MARC 21 exchange format to more Web based, Linked Data standards. Zepheira and The Library of Congress are working together to develop a Linked Data model, vocabulary and enabling tools / services for supporting this Initiative.

BIBFRAME.ORG provides an overview of this effort. This site offers:

- A demonstration area where the model is being tested and displayed using different test suites of data
- View of the BIBFRAME vocabulary along with supporting formal specifications
- Educational resource on how the BIBFRAME vocabulary relates to MARC21 fields and subfields
- Code and supporting services.

Additional background on this initiative can be found here:

- LC's Bibliographic Framework Initiative: An Update (November 29, 2012)
- Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services (November 21, 2012)
- Bibliographic Framework Initiative Approach for MARC Data as Linked Data (September 2012)

Linked Data in biblioteca

JLIS.it
Italian Journal of Library and Information Science
Rivista italiana di biblioteconomia, archivistica e scienza dell'informazione

Vol 4, N° 1 (2013): Global Interoperability and Linked Data in Libraries: Special issue

Editor: Mauro Guerrini
Proceedings by Gianfranco Crupi and Ginevra Peruginelli

Il fascicolo raccoglie gli atti del Seminar Global Interoperability and Linked Data in Libraries tenuto all'Università di Firenze il 18 e 19 giugno 2012. Tuttavia, rispetto al programma del convegno, il fascicolo non comprende alcune relazioni (Paola Mazzucchi, Federico Morando, Roberto Moriondo e Romano Nanni) e include invece tre ulteriori contributi (Barbara Tillett, Mauro Guerrini e Tiziana Possemato, Gianfranco Crupi). Inoltre, i saggi sono raccolti secondo un ordine e una struttura diversi dal programma originario.

University of Florence, Aula Magna, June 18-19, 2012
seminar
GLOBAL INTEROPERABILITY AND LINKED DATA IN LIBRARIES

JLiS, 2013, 4(1)

Introduction

Linked Data as a new Paradigm of Data Interconnection

Publishing Value Vocabularies and Standard as Linked Data

Linked Data in biblioteca

sentire la sua voce. Il problema, invece, è che le biblioteche sono lontane dalla principale fonte odierna di informazioni, cioè dal web. La spinta a muovere le biblioteche verso i linked data non risponde semplicemente al desiderio di modernizzare i cataloghi, ma rappresenta la necessità di trasformare i cataloghi delle biblioteche da database chiuso e separato in un sistema integrato con la tecnologia che la gente utilizza per la ricerca e per la creazione di nuove idee. Insomma, il catalogo della biblioteca, per essere visibile all'utente di oggi, deve cessare di essere un'entità staccata, un database separato e divenire un insieme di dati, milioni di dati tra gli innumerevoli dati del web. Il nostro compito oggi, come bibliotecari e informatici, non è di tradurre i dati bibliografici di oggi in linked data; il nostro compito è quello di creare un nuovo sistema di accesso e d'uso dei dati bibliografici che sia compatibile con il funzionamento della rete.

Linked Data: an Evolution

Karen Coyle

K.Coyle, Linked Data: an evolution

Due sono gli aspetti di questa evoluzione. Il primo è rendere utilizzabili i dati bibliografici nel web. Ogni persona che fa ricerca, che studia, che scrive e che cita, ha bisogno di dati bibliografici, alcuni dei quali possono essere forniti dai cataloghi delle biblioteche. E con la disponibilità dei dati bibliografici sul web, ogni persona connessa in rete diviene potenzialmente un utente della biblioteca. Il secondo aspetto riguarda l'uso di dati presenti in rete per migliorare i servizi della biblioteca. Collegando dati bibliografici e risorse web si può, ad esempio, collocare un libro nel suo contesto storico o dimostrare l'influenza di un autore sulla cultura del tempo. A che punto siamo

Metadata e copyright

LIBRARYJOURNAL

McFarla

LATEST STORIES | SPECIAL FEATURES | MANAGING LIBRARIES

SUBSCRIBE TO LJ | SUBSCRIBE TO LIBRARY HOTLINE | PATRON PROFI

You are here: Home / Opinion / Peer to Peer Review / Metadata and Copyright | Peer

Metadata and Copyright | Peer to Peer Re

By Karen Coyle on February 28, 2013 1 Comment

Most of us are aware of the basics of U.S. copyright law, i copyrightable and non-copyrightable works. Some materials are explicitly exempted from copyright in this country, a key example being U.S. Federal documents. (Although if that sounds to you like a clearly distinguishable category, you should ask your local government documents librarian to fill you in on the complexities of defining "U.S. Federal document.") Another exempted category is that of facts and compilations of facts that have no creative component. This was determined in the famous Supreme Court ruling of Feist v. Rural Telephone, in which the Court interpreted the constitutional wording of "to promote the Progress of Science and useful Arts" as implying some level of creativity.

> "The constitutional requirement necessitates independent creation plus a modicum of creativity. Since facts do not owe their origin to an act of authorship, they are not origina and thus are not copyrightable." (Feist, p. 1)

As you might imagine, "modicum of creativity" is itself very difficult to define, much like the determination of Fair Use based on "substantiality of the portion used" and the effect on the n the work.

Recently, the Digital Public Library of America (DPLA) addressed copyright in metadata with its draft *Policy Statement on Metadata*, which was discussed at the board meeting of Feb. 14. The first statement of the draft policy document is to my mind the most controversial:

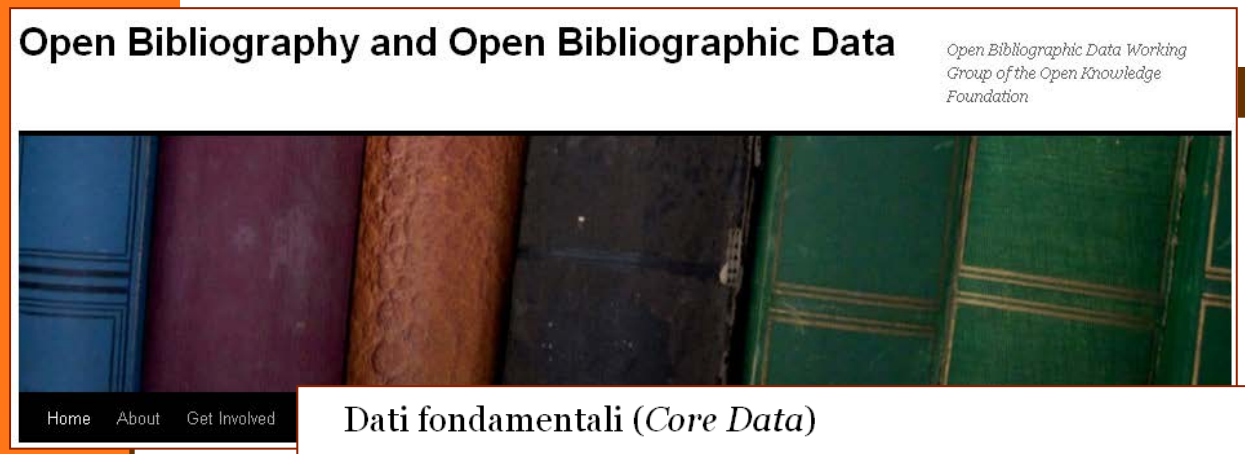
> **01. The Vast Majority of Metadata is Not Subject to Copyright Restrictions.**
>
> The DPLA believes that the vast majority of metadata is not subject to copyright, because it either expresses only objective facts (which are not original) or constitutes expression so limited by the number of ways the underlying ideas can be expressed that such expression has merged with those ideas.

e la creatività dei bibliotecari???

I can imagine that declaring library catalog data to be void of creativity will grate on the self-esteem of many catalogers. While some library metadata consists of the rote recording of facts about an object, it would be hard to explain the necessity of a 1,500 page set of cataloging rules to produce that outcome. Catalogers that I meet are proud of their ability to interpret those rules for the more complex cases that come into their hands. That different catalogers make different decisions (much to the consternation of downstream users of the catalog data for their own cataloging) is evidence that at least some of the cataloging process is not purely factual.

K.Coyle, Metadata and copyright, Library Journal, 28 feb 2013

Open Bibliographic Data



Open Knowledge Foundation

Total Petition Signatures

- OpenBiblio Principles: 182

Dati fondamentali (*Core Data*)

I dati bibliografici consistono di descrizioni bibliografiche. Una descrizione bibliografica descrive una risorsa bibliografica (articolo, monografia, etc. – sia essa cartacea o elettronica) allo scopo di:

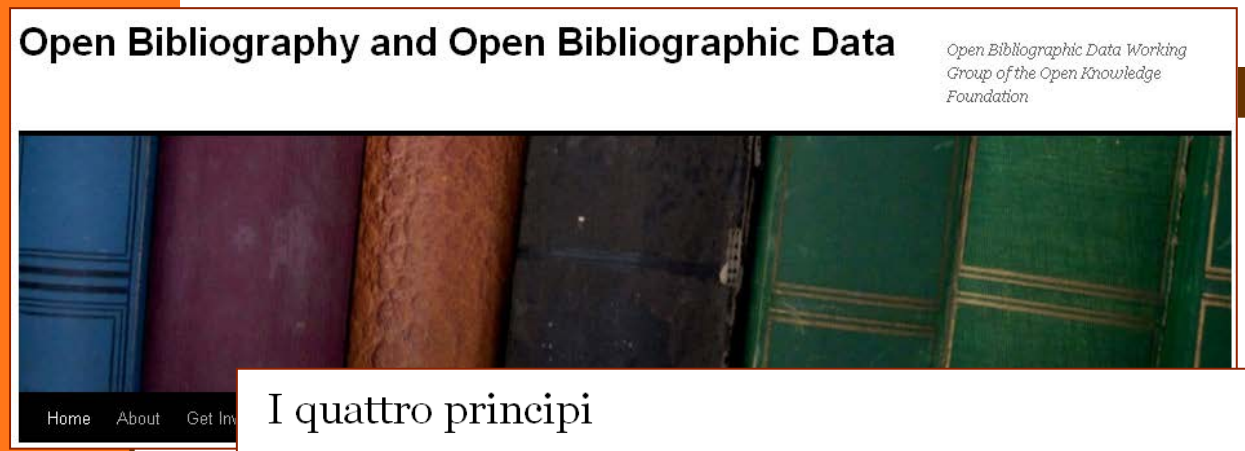
1. identificare la risorsa descritta, cioè indicare un'unica risorsa tra tutte quelle esistenti e
2. ubicare la risorsa descritta, cioè indicare come e dove trovarla.

Tradizionalmente, era sufficiente un'unica descrizione per entrambi gli scopi, dal momento che essa forniva informazioni su: autore/i, curatore/i, titolo, editore, luogo e data di pubblicazione, identificazione dell'opera "madre" (ad es. rivista), numeri di pagina.

Sul Web per identificare una risorsa si usano gli Uniform Resource Identifiers (URI) come URN, DOI, etc., mentre è possibile localizzarla tramite Uniform Resource Locators (URL), che sono URI accompagnati dal prefisso HTTP. Perciò, tutti gli URI di risorse bibliografiche rientrano in questo concetto ristretto di dati bibliografici.

http://openbiblio.net/

Open Bibliographic Data



Open Knowledge Foundation

Total Petition Signatures

- OpenBiblio Principles: 182

I quattro principi

Raccomandiamo formalmente di adottare i seguenti principi e di agire sulla base di essi:

1. Quando si pubblicano dati bibliografici e collezioni di dati bibliografici è cruciale che siano pubblicati con una chiara ed esplicita dichiarazione sulla volontà e le aspettative dell'editore riguardo al riuso e alla riproposizione di descrizioni bibliografiche individuali, di intere collezioni o di loro sottoinsiemi. Tali asserzioni devono essere precise, irrevocabili e basate su un'asserzione appropriata e legalmente riconosciuta, espressa in forma di deroga o di licenza.
   Quando pubblicate dati bibliografici, usate una licenza esplicita e robusta.
2. Molte licenze ampiamente riconosciute non sono pensate per i dati bibliografici o collezioni di dati e non sono a essi appropriate. Esiste una varietà di licenze progettate e appropriate per il trattamento di dati che sono descritte qui: (http://opendefinition.org/licenses/#Data). Le licenze Creative Commons (ad esclusione della CC0), GFDL, GPL, BSD, etc. NON sono adatte ai dati e il loro uso è FORTEMENTE scoraggiato.
   Usate una licenza o una deroga appropriata ai dati.

http://openbiblio.net/

Open Bibliographic Data

3. L'uso di licenze che limitano il riuso commerciale o la produzione di opere derivate escludendo l'uso per scopi particolari o per persone o organizzazioni specifiche è FORTEMENTE scoraggiato. Queste licenze rendono impossibile integrare efficacemente i dataset e riproporli. Inoltre, impediscono lo sviluppo di servizi commerciali che aggiungano valore ai dati bibliografici o di attività commerciali che possano servire come supporto alla preservazione dei dati.
   Se volete che i vostri dati siano realmente usati e arricchiti da altri, devono essere aperti così come definito dalla Open Definition (http://opendefinition.org) – in particolare, non devono essere usate la clausola non-commerciale e altre clausole restrittive.
4. Inoltre, si raccomanda che i dati bibliografici o le collezioni di essi, specialmente se sono finanziati con denaro pubblico, devono essere esplicitamente posti nel pubblico dominio tramite la licenza Public Domain Dedication (negli Stati Uniti) o la deroga Creative Commons Zero. Questo favorisce il massimo riuso possibile dei dati, in linea con l'etica generale di condivisione nell'ambito dei beni culturali finanziati dal pubblico.
   Ove possibile, ponete esplicitamente i dati bibliografici nel Pubblico Dominio tramite PDDL o CCo.

http://openbiblio.net/

Harvard Library catalog in Open data

THE HARVARD LIBRARY

Millions of Harvard Library Catalog Records Publicly Available

Harvard Library Releases Nearly 100% of Its Records

April 24, 2012 – The Harvard Library announced it is making more than 12 million catalog records from Harvard's 73 libraries publicly available.

The records contain bibliographic information about books, videos, audio recordings, images, manuscripts, maps, and more. The Harvard Library is making these records available in accordance with its Open Metadata Policy and under a Creative Commons 0 (CC0) public domain license. In addition, the Harvard Library announced its open distribution of metadata from its Digital Access to Scholarship at Harvard (DASH) scholarly article repository under a similar CC0 license.

"The Harvard Library is committed to collaboration and open access. We hope this contribution is one of many steps toward sharing the vital cultural knowledge held by libraries with all," said Mary Lee Kennedy, Senior Associate Provost for the Harvard Library.

The catalog records are available for bulk download from Harvard, and are available for programmatic access by software applications via API's at the Digital Public Library of America (DPLA). The records are in the standard MARC21 format.

"By instituting a policy of open metadata, the Harvard Library has expressed its appreciation for the great potential that library metadata has for innovative uses. The two metadata releases today are prime examples," said Stuart Shieber, Library Board Member, Director of the Office for Scholarly Communication and Professor of Computer Science at Harvard.

John Palfrey, Chair of the DPLA, said, "With this major contribution, developers will be able to start experimenting with building innovative applications that put to use the vital national resource that consists of our local public and research libraries, museums, archives and cultural collections." He added that he hoped that this would encourage other institutions to make their own collection metadata publicly available.

http://goo.gl/QXAPk

Dati collegati agli articoli

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.

BioMedCentral, Press release, 4 Aprile 2012

Come citare i dati e collegarli all'articolo

A Digital Curation Centre 'working level' guide

DCC | JISC

How to Cite Datasets and Link to Publications

Alex Ball (DCC) and Monica Duke (DCC)

Requirements for data citations

The SageCite Project has identified a set of requirements for dataset citations and any services set up to support them.[6]

The citation itself must be able to identify uniquely the object cited, though different citations might use different methods or schemes to do so.

It must be able to identify subsets of the data as well as the whole dataset.

It must provide the reader with enough information to access the dataset; indeed, when expressed digitally it should provide a mechanism for accessing the dataset through the Web infrastructure.

It must be usable not only by humans but also by software tools, so that additional services may be built using these citations. In particular, there need to be services that use the citations in metrics to support the academic reward system, and services that can generate complete citations.

Elements of a data citation

The elements that would make up a complete citation are a matter of some debate. The following list is a superset taken from four different papers on the subject.[7,8,9,10]

Author. The creator of the dataset.[7,8,9,10]

Publication date. Whichever is the later of: the date the dataset was made available,[7] the date all quality assurance procedures were completed,[8,9] and the date the embargo period (if applicable) expired.[10]

Title. As well as the name of the cited resource itself,[7,10] this may also include the name of a facility[8] and the titles of the top collection and main parent sub-collection (if any) of which the dataset is a part.[9]

Edition. The level or stage of processing of the data, indicating how raw or refined the dataset is.[8]

Version. A number increased when the data changes, as the result of adding more data points or re-running a derivation process, for example.[10]

Feature name and URI. The name of an ISO 19101:2002[11] 'feature' (e.g. GridSeries, ProfileSeries) and the URI identifying its standard definition, used to pick out a subset of the data.[8]

Resource type. Examples: 'database',[9] 'dataset'.[10]

Publisher. The organisation either hosting the data[10] or performing quality assurance.[8]

Unique numeric fingerprint (UNF). A cryptographic hash of the data, used to ensure no changes have occurred since the citation.[7]

Identifier. An identifier for the data, according to a persistent scheme.[7,8,9,10]

Location. A persistent URL from which the dataset is available. Some identifier schemes provide these via an identifier resolver service.[7,8,9,10]

The most important of these elements – the ones that should be present in any citation – are the author, the title and date, and the location. These give due credit, allow the reader to judge the relevance of the data, and permit access the data, respectively. In theory, they should between them uniquely identify the dataset; in practice, a formal identifier is often needed. The most efficient solution is to give a location that consists of a resolver service and an identifier (for an example, see Figure 3 on page 4).

Note that the way in which these elements would

http://goo.gl/ZY55O

Data management

Luglio 2012, http://goo.gl/AikAN

Ten recommendations for libraries to get started with research data management

1. Offer research data management support, including data management plans for grant applications, intellectual property rights advice and information materials. Assist faculty with data management plans and the integration of data management into the curriculum.
2. Engage in the development of metadata and data standards and provide metadata services for research data.
3. Create Data Librarian posts and develop professional staff skills for data librarianship.

Data management / 2

4. Actively participate in institutional research data policy development, including resource plans. Encourage and adopt open data policies where appropriate in the research data life cycle.
5. Liaise and partner with researchers, research groups, data archives and data centers to foster an interoperable infrastructure for data access, discovery and data sharing.
6. Support the lifecycle for research data by providing services for storage, discovery and permanent access.

Data management / 3

7. Promote research data citation by applying persistent identifiers to research data.
8. Provide an institutional Data Catalogue or Data Repository, depending on available infrastructure.
9. Get involved in subject specific data management practice.
10. Offer or mediate secure storage for dynamic and static research data in co-operation with institutional IT units and/or seek exploitation of appropriate cloud services.

DataCite

Helping you to find, access, and reuse data

Research data management

ARIADNE

Web Magazine for Information Professionals

HOME | CURRENT ISSUE | TODAY'S CHOICE | ARCHIVE | AUTHORS | ARTICLI

necessarie nuove competenze…
in UK formazione in Research Data Management

Upskilling Liaison Librarians for Research Data Management

6 December 2012 - 7:27pm

Printer-friendly version

For many UK HEIs, especially research-intensive institutions, Research Data Management (RDM) is rising rapidly up the agenda. Working closely with other professional services, and with researchers themselves, libraries will probably have a key role to play in supporting RDM. This role might include signposting institutional expertise in RDM; inclusion of the topic in information literacy sessions for PhD students and other researchers; advocacy for open data sharing; or contributing to the management of an institutional data repository. It seems that there are choices for each librarian to make, largely shaped by their existing role. For some, RDM may rapidly become a core part of their job. For others it may be something of which they simply need a greater awareness. New graduates entering the profession require a grounding in RDM-related knowledge and skills, but there is also a need for established professionals to update their competencies too.

Andrew Cox, Eddy Verbaan and Barbara Sen explore the design of a curriculum to train academic librarians in the competencies to support Research Data Management.

Table of Contents [hide]

1. Existing Literature
2. Existing Curricula
3. The RDMRose Project Vision
4. Conclusion
5. References
6. Author Details
7. Appendix 1

In this context, JISC have funded the White Rose consortium of academic libraries at Leeds, Sheffield and York, working closely with the Sheffield Information School, in the RDMRose Project, to develop learning materials that will help librarians grasp the opportunity that RDM offers. The learning materials will be used in the Information School's Masters courses, and are also to be made available to other information sector training providers on a share-alike licence. A version will also be made available (from January 2013) as an Open Educational Resource for use by information professionals who want to update their competencies as part of their continuing professional development (CPD). The learning materials are being developed specifically for liaison librarians, to upskill existing professionals and to expand the knowledge base for new entrants to librarianship. It is hoped to accommodate the perspectives of any information professional, but the scope is not intended to encompass a syllabus for a data management specialist role (following the distinction made by Corrall [1]).

http://goo.gl/UkAVx

Digital curation bibliography

Digital Curation Bibliography: Preservation and Stewardship of Scholarly Works, 2012 Supplement

Charles W. Bailey, Jr.

HOUSTON: DIGITAL SCHOLARSHIP

3/18/2013

Introduction

In a rapidly changing technological environment, the difficult task of ensuring long-term access to digital information is increasingly important. The *Digital Curation Bibliography: Preservation and Stewardship of Scholarly Works, 2012 Supplement* presents over 130 English-language articles, books, and technical reports published in 2012 that are useful in understanding digital curation and preservation. This selective bibliography covers digital curation and preservation copyright issues, digital formats (e.g., media, e-journals, research data), metadata, models and policies, national and international efforts, projects and institutional implementations, research studies, services, strategies, and digital repository concerns.

It is a supplement to the *Digital Curation Bibliography: Preservation and Stewardship of Scholarly Works*, which covers over 650 works published from 2000 through 2011. All included works are in English. The bibliography does not cover conference papers, digital media works (such as MP3 files), editorials, e-mail messages, letters to the editor, presentation slides or transcripts, unpublished e-prints, or weblog postings.

The bibliography includes links to freely available versions of included works. If such versions are unavailable, italicized links to the publishers' descriptions are provided.

Links, even to publisher versions and versions in disciplinary archives and institutional repositories, are subject to change. URLs may alter without warning (or automatic forwarding) or they may disappear altogether. Inclusion of links to works on authors' personal websites is highly selective. Note that e-prints and published articles may not be identical.

Bibliography

Adams, Sam, and Peter Murray-Rust. "Chempound—A Web 2.0-Inspired Repository for Physical Science Data." *Journal of Digital Information* 13, no. 1 (2012). http://journals.tdl.org/jodi/index.php/jodi/article/view/5873

Aitken, Brian, Patrick McCan, Andrew McHugh, and Kerry Miller. "Digital Curation and the Cloud: Final Report." London: JISC, 2012. http://www.jisc.ac.uk/media/D/0/1/{BD01C1CDB-AF99-4A20-A9BC-12E73DB224DD}Curation-in-the-Cloud.pdf

Allard, Suzie. "DataONE: Facilitating eScience through Collaboration." *Journal of eScience Librarianship* 1, no. 1 (2012). http://dx.doi.org/10.7191/jeslib.2012.1004

Altenhöner, Reinhard. "Tailoring Digital Preservation Services for Practice: Workflow Development and Cost Modeling." *DESIDOC Journal of Library & Information Technology* 32, no. 4 (2012): 302-305. http://publications.drdo.gov.in/ojs/index.php/djlit/article/view/2523

Ball, Alexander, Mansur Darlington, Thomas Howard, Chris McMahon, and Steve Culley. "Visualizing Research Data Records for Their Better Management." *Journal of Digital Information* 13, no. 1 (2012). http://journals.tdl.org/jodi/index.php/jodi/article/view/5917

Bardyn, Tania P., Taryn Resnick, and Susan K. Camina. "Translational Researchers' Perceptions of Data Management Practices and Data Curation Needs: Findings from a Focus Group in an Academic Health Sciences Library." *Journal of Web Librarianship* 6, no. 4 (2012): 274-287. http://www.tandfonline.com/doi/abs/10.1080/19322909.2012.730375

Beagrie, Neil, Monica Duke, Catherine Hardman, Dipak Kalra, Brian Lavoie, Manjula Patel, Liz Lyon, and Matthew Woollard. "The KRDS Benefit Analysis Toolkit: Development and Application." *International Journal of Digital Curation* 7, no. 2 (2012):

http://goo.gl/XkUl2

rilasciata il 18 febbraio 2013, riguarda anche i research data

Open science now!

amplificare l'intelligenza collettiva per risolvere le sfide globali

Michael Nielsen calls for scientists to embrace new tools for collaboration that will enable discoveries to happen at the speed of Twitter.

http://www.ted.com/talks/michael_nielsen_open_science_now.htm

Grazie, buon lavoro!