European Parliament,, October 10, 2012

Open Data: what and why

Elena Giglia
University of Turin
elena.giglia@unito.it

Opening story: Fukushima

After Fukushima publishers unlocked their data on nuclear radiation and nuclear pollution [see http://www.earlham.edu/~peters/fos/newsletter/04-02-11.htm]

Wordlwide applause.

But that means that those **data** where, until then, **locked**…

What data?

Data from scientific/academic researches funded by public funds

Data from PRIMARY RESEARCH not applied research

Journal articles = **a summary** of the work, all that could be published in a **print-on paper age**

In the **Web age,** we can make a **better, full record** of the work available

Data are **digitally created,** they **can be published** alongside the article

Aim: to publish the raw data as well as the summary that is in the article

… these are the data that the article is ABOUT anyway

DEFINITELY NOT
- COMMERCIAL SECRETS
- CONFIDENTIAL DATA
- PERSONAL DATA

ONLY the data already represented and reported on in the published article

Is it a new idea? NO

Researchers
[who are **primary** actors in research]
do share data e do want to do it

It´s a matter of
supporting them
not to forcing them…

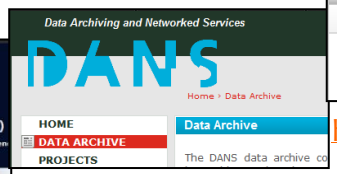
Some journals already requires data to be published alongside the article

There are many public databases

Humanities

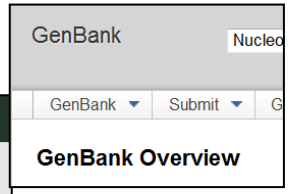
http://echo.mpiwg-berlin.mpg.de/home

Social sciences



http://www.dans.knaw.nl/en

Genomics



http://www.ncbi.nlm.nih.gov/genbank

…and many others

Also an EU funded one…

All the large life science research organisations are part of this effort to **collect, share and preserve** experimental data from life science research

Why? /1

Access for other scientists
(to compare, contrast, with their own data)

… better to **rely on data than on data interpretation**
(as scientific articles are…)

Reuse by other scientists
(to add to their own data and **create new knowledge**)

Transparency
(checking, detection of fabrication)


Retraction Watch

Salzburg University fires crystallographer Robert Schwarzenbacher for faking data in Journal of Immunology paper

with 22 comments

The crystallographer who confessed to data fabrication that has forced the retraction of a structure in a *Journal of Immunology* paper on birch pollen allergen — but later recanted — has been fired by the University of Salzburg.

Robert Schwarzenbacher, 39, was awarded a 1.7 million-euro Marie Curie fellowship, the highest individual European research award, six years ago. According to Salzburg's ORF.at:

Retraction watch blog on scientific retracted papers never runs out of posts…

http://retractionwatch.wordpress.com/

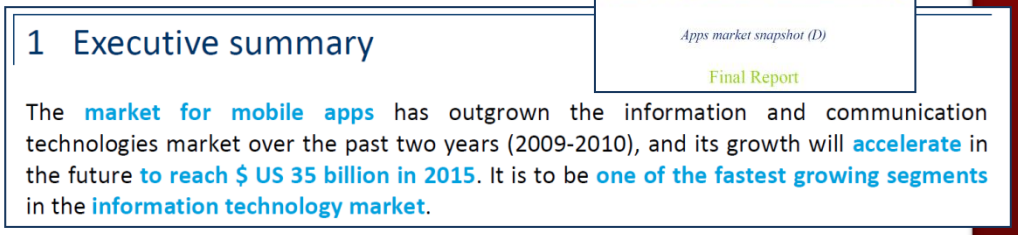
Why ? / 2

New technologies can work on data to create new data (data-mining)

http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation

MGI predict that effective and creative use of these large data sets in the US health care sector could generate more than $300bn in value per annum and reduce national health care expenditures by around 8%.

http://www.jisc.ac.uk/publications/reports/2012/value-and-benefits-of-text-mining.aspx



http://ec.europa.eu/information_society/policy/psi/docs/pdfs/report/11_2012/apps_market.pdf

Reuse for remix (Apps / APIs market)

The benefits

Better science

More reliable (data integrity is checkable)

Some science can ONLY work this way (e.g. genomics)

New science
(new technologies, new findings)

Innovation in the SME community

Economic growth

Better-informed citizenry

Knowledge society

Final story

Ilaria Capua – virologist, Rome http://ec.europa.eu/research/profiles/index_en.cfm?p=1_capua

- In 2006 she sequenced the genome of the first strain of the H5N1 flu virus (avian flu).
- Against the WHO suggestion to deposit the virus genetic sequence in its closed database, **she decided to challenge the system and shared her data, depositing the genetic sequence** in GenBank, **an open access database**.
- In 2006, she launched in a letter to Nature the **GISAID** (Global Initiative on Sharing All Influenza Data), an international network aimed at sharing online genetic data on avian flu. 70 researchers supported the initiative. Among them, 7 Nobel Prize winners.
- Since then, WHO, FAO and OIE **adopted her open and transdisciplinary approach** which now is a core part of the **Global influenza preparedness strategy**

…a MAJOR component of public health policy worldwide (from which all of us will benefit), all from one scientist's vision about Open Data

More on Open Access / Open Data

http://oad.simmons.edu/oadwiki/Main_Page

http://www.openoasis.org/

European Report on Open Data:

Riding the wave. How Europe can gain from the rising tide of scientific data

http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf