

Moving towards Open Access to research data in health sciences: *a changing landscape* for **grey literature**



Paola De Castro

Istituto Superiore di Sanità, Rome (Italy)

16. International Conference

Grey Literature • Washington DC, 8-9 December 2014

Share considerations on

- Traditional role of GL and new perspectives
- LG in OA policies and Open science
- Examples of GL in health sciences
- Profile of stakeholders
- Why taking action



A helicopter view

GL in health sciences (stage 1)

20th century

Traditional role

before Internet and soon after

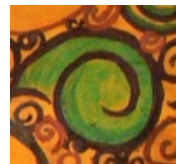
- Technical reports
- Preprints
- Proceedings
- Working papers
- Patents
- Posters
- Case studies

□ Seminar of York (definition)

→ SIGLE database □

MAIN CONCERNS FOR PRINTED GL


- Bibliographic control
- Retrievability
- Accessibility
- Poor layout
- Unique and precious content (information & data)
- Peer review



Useful, BUT GREY
“Cinderella”

How to write and distribute GL

in between the traditional role and new perspectives

 GREY LITERATURE INTERNATIONAL STEERING COMMITTEE	GREY LITERATURE INTERNATIONAL STEERING COMMITTEE
About GLISC Contact us Credits	Translations The Documents citing the Guidelines NEWS
Contents of the Guidelines	
1. Statement of purpose 1.1. About the Guidelines 1.2. Potential users of the Guidelines 1.3. How to use the Guidelines	Guidelines for the production of scientific and technical reports: how to write and distribute grey literature Version 1.1
2. Ethical considerations 2.1. Authorship and contributorship 2.2. Issuing organization 2.3. Peer review 2.4. Conflicts of interest 2.5. Privacy and confidentiality	
3. Publishing and editorial issues 3.1. Copyright 3.2. Correspondence 3.3. Electronic publishing and institutional repositories 3.4. Advertising	
4. Report preparation 4.1. Instructions to authors 4.2. Report structure 4.3. Revision editing 4.4. Sending the report	
5. General information on the Guidelines 5.1. Steering committee 5.2. Use, distribution, translation and inquiries	
References	
Annex. List of institutions adopting the Guidelines	
Annex 2. Report check list	
© GLISC 2006 Webmaster	July 2007

Updating
ISO Standard 5966/1982

ANSI Standard
Z39.18/2005
Preparation, Presentation,
and Preservation

A good structure
as a pre-requisite for online
dissemination
□ **next step**
to be translated into a
metadata scheme

GL in health sciences (stage 2)

21th century

New perspectives

Open access movement (2000)

Open access to publications

OA Declarations

OA Policies

Open data

Open science

Altmetrics

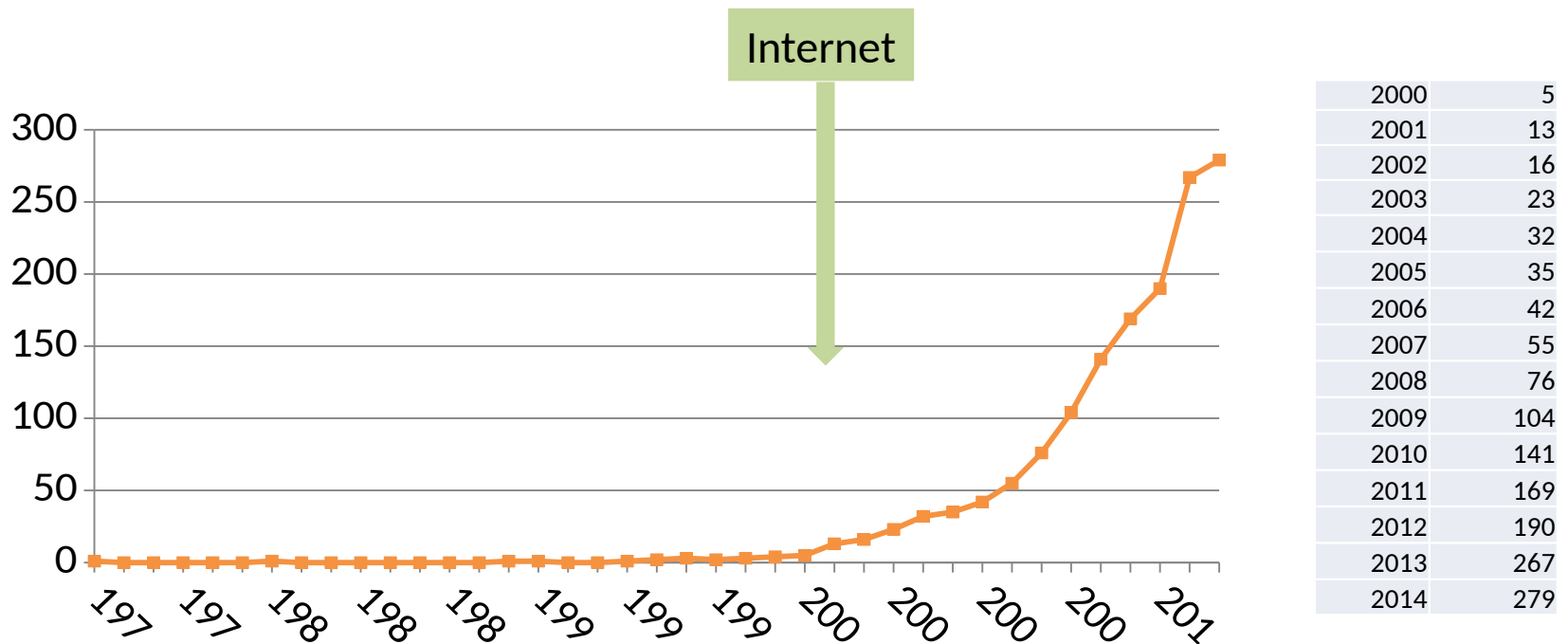
MAIN CONCERNS FOR DIGITAL GL

- Quality & online dissemination
- Inclusion in Repositories
- Interoperability & Metadata
- Recognition & Opportunities
- Curation & Preservation

Useful, AND GREY
“Princess”



Occurrences of the term “grey literature” in titles or abstracts of articles indexed in **PubMed** (1976-2014) Tot. 1466



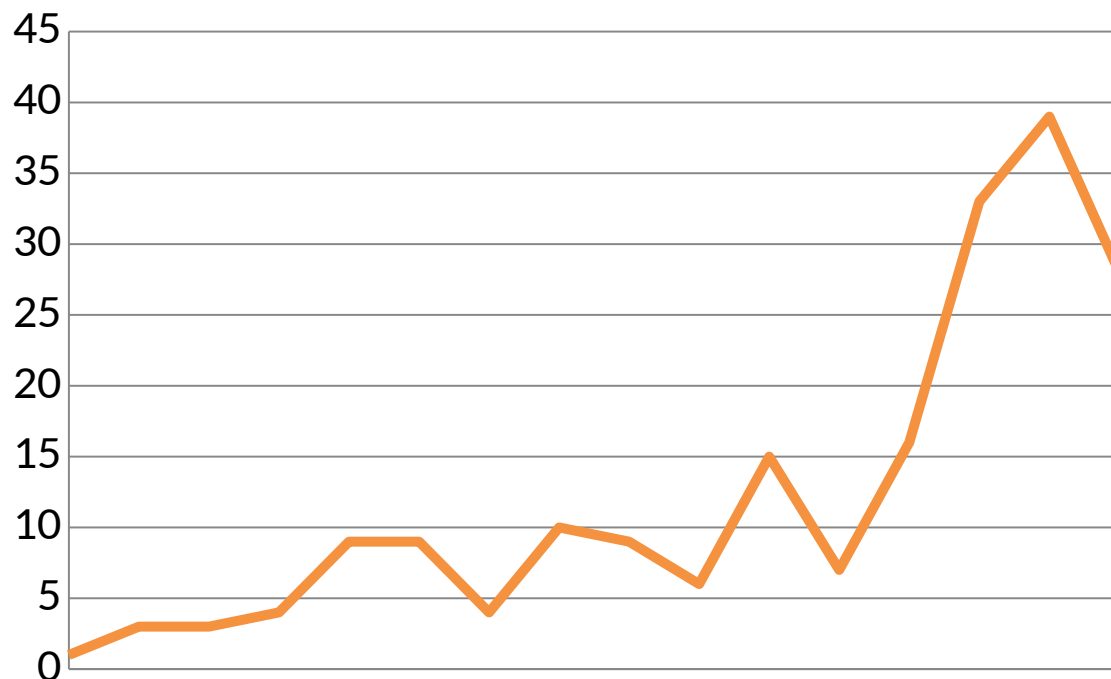


THE COCHRANE LIBRARY

Independent high-quality evidence for health care decision making

from [The Cochrane Collaboration](#)

**Occurrences of the term “grey literature”
in titles or abstracts of articles (1999-2014) Tot. 196/8741**

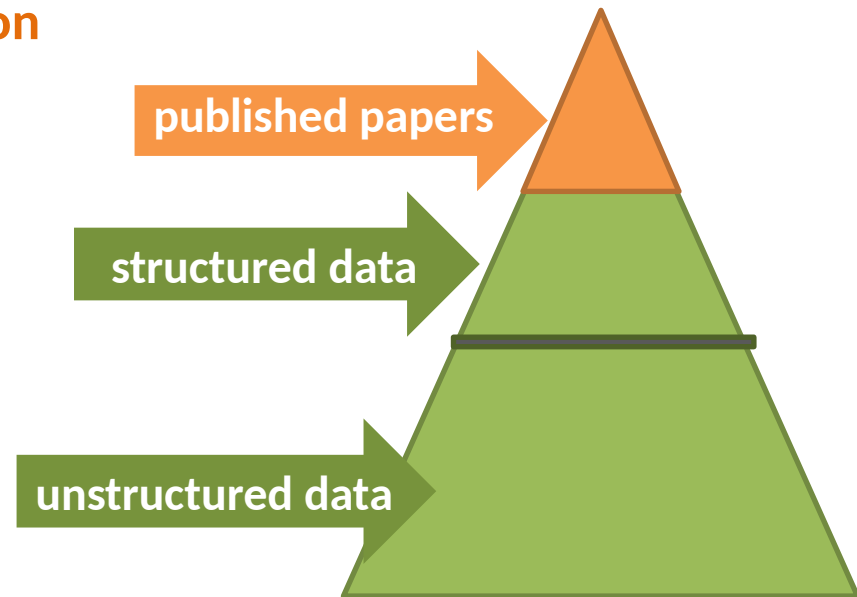


General trend from the year 2000

Discover the value of hidden research

EBM. «Published» information is not representative of the big sample.

Results of studies published in peer-reviewed journals are not representative of the entire samplings of research results on a topic



Grey Literature in Meta-Analyses

V S. Conn ,JC. Valentine,
HM Cooper, MJ. Rantz
Nursing Research July/August 2003

Alan Winfield. (personal communication)
Science, Innovation and Society
Conference . Rome 19-21 November 2014

New borderlines, new challenges for digital grey literature

The **traditional borderline** between commercial and non-commercial or grey literature, (very clear in the past century), is now becoming **obsolete**, in fact, the OA movement to scientific publications implies the **free** (libre) use of **different types research output**, and therefore it also includes **GL**, as stated in the numerous declarations in support of OA, stating from the year 2003.



OA MOVEMENT, 3 B_s



Budapest OA initiative, 2002
...unrestricted, free access to **scholarly research...**

Bethesda, 2003
...goal of providing open access to the **primary scientific literature**


Berlin Declaration, 2003
... Internet has fundamentally changed the practical and economic realities of distributing **scientific knowledge and cultural heritage...**

Why sharing knowledge is so critically important?

MAIN CONCEPTS



Human rights
Equity
Progress
Etc.



ITC and
straightforward
information policies
can makes it possible

LG is just one of the different ways currently
utilized for knowledge sharing

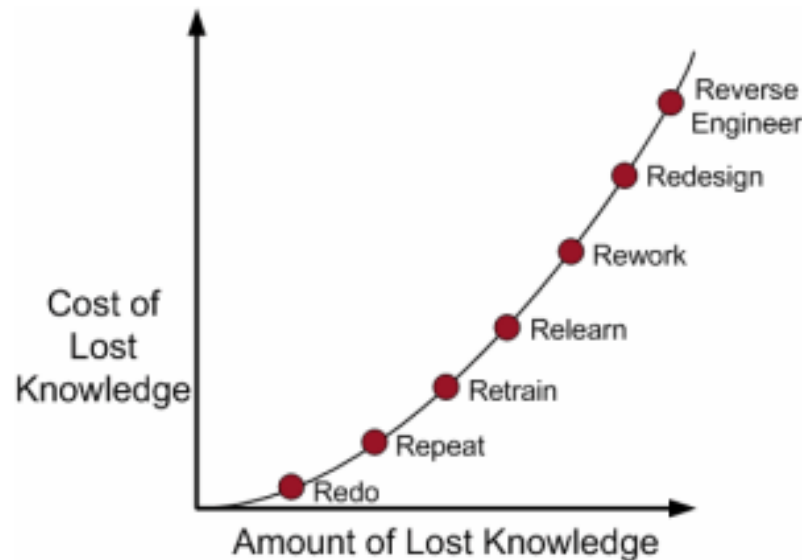


Advantages of sharing information and data in health sciences

- Maximize **return on investment**
- Create **awareness** among different stakeholders (patients, policy makers, general public, etc.)
- Improve a general **understanding** of research efforts
- **Contribute to improving global health** (prevention and treatment)
- Create **trust** in research activity
- Accelerate **progress** in science
- Develop of **cross-disciplinary** research
- Increase **collaboration**

Economic impact of lost knowledge

also GL plays a relevant role



Correct knowledge management

- Reduces duplication of activities
- Limits the amount of lost knowledge
- Saves time and money
- Allows participation of all stakeholders

GL is part of lost knowledge, therefore we urge

□ **grey community to take action** and make all **quality GL** publicly available

What do we need to make the best use of existing knowledge?



Mind the quality (shared responsibility between producers and users)



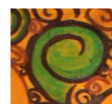
Provide access knowledge, no matter if grey or white (□all)



Develop and implement **large scale collection** services able to host and make available relevant, high quality **resources** quickly, efficiently, persistently and **free** online to everybody (repositories)



Acquire the **ability to use** such resources at different levels and distinguish good from misleading information (**information literacy**)



Actors of change

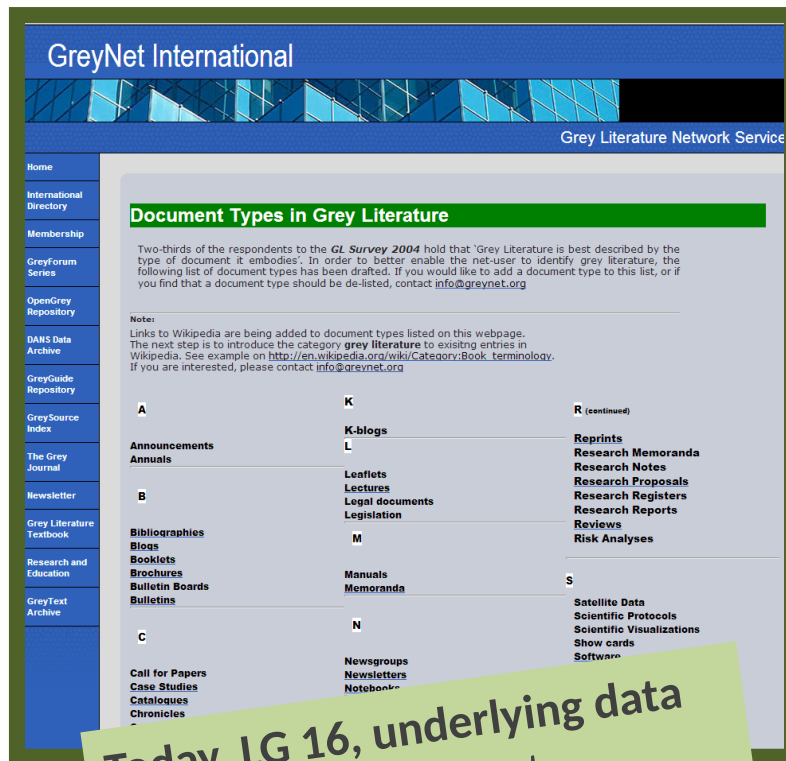
Authors
Information specialists
Librarians
Editors
Publishers

Policy makers
General public
Citizens
Patients

Open science
Benefits from
multi-
dimensional
approach

New types of digital grey literature

2004, GreyNet Survey 2004 to enable net users to identify GL



Today, LG 16, underlying data represent the greatest **added value** behind all these new types of documents

- Blogs
- Databases
- Datasets
- Datasheets
- Digital Documents
- Digital Educational Material
- E-prints
- E-texts
- Enhanced Publication
- ETD (Electronic Theses and Dissertations)
- Website Reviews
- Websites

GL and data rich publications

past and present

In the past century, technical reports represented the only type of documents able to host large amounts of data which journal articles, mainly for economic reasons, were not able to print.

Today, journals are even encouraged to publish enriched publications including datasets and they are looking for the best way to share data so as to be re-used for different purposes.

Also MEDLINE allows a conversation



The grey community starts a **NEW DIALOGUE** with different stakeholders

New alliances are developed with OA supporters and the promoters of Open Data

□ The PISA DECLARATION

Open Access □ Open research data □ Open Science

PISA Declaration

on Policy Development for GL Resources, May 16, 2014

Commitment on different grounds... ..

Organizational: commitment to OA, cooperation among organizations engaged in the GL production, use, collection, management, persistent identifiers and open metadata standards

Research/Educational: recognition and reward for quality GL, standards in production and bibliographic control, development of good practice guides

Legal: Addressing legal obstacles to GL dissemination (copyright)

Financial/Sustainable: funding for research involving GL, investment in infrastructure and new technologies

Technical: accessibility of online content, linking data, interoperability standards for sharing GL

TWO SEMINARS



1. GL, 7 April 2014
2. Open data, 8 April 2014

debating
similar issues

GL and OA policies

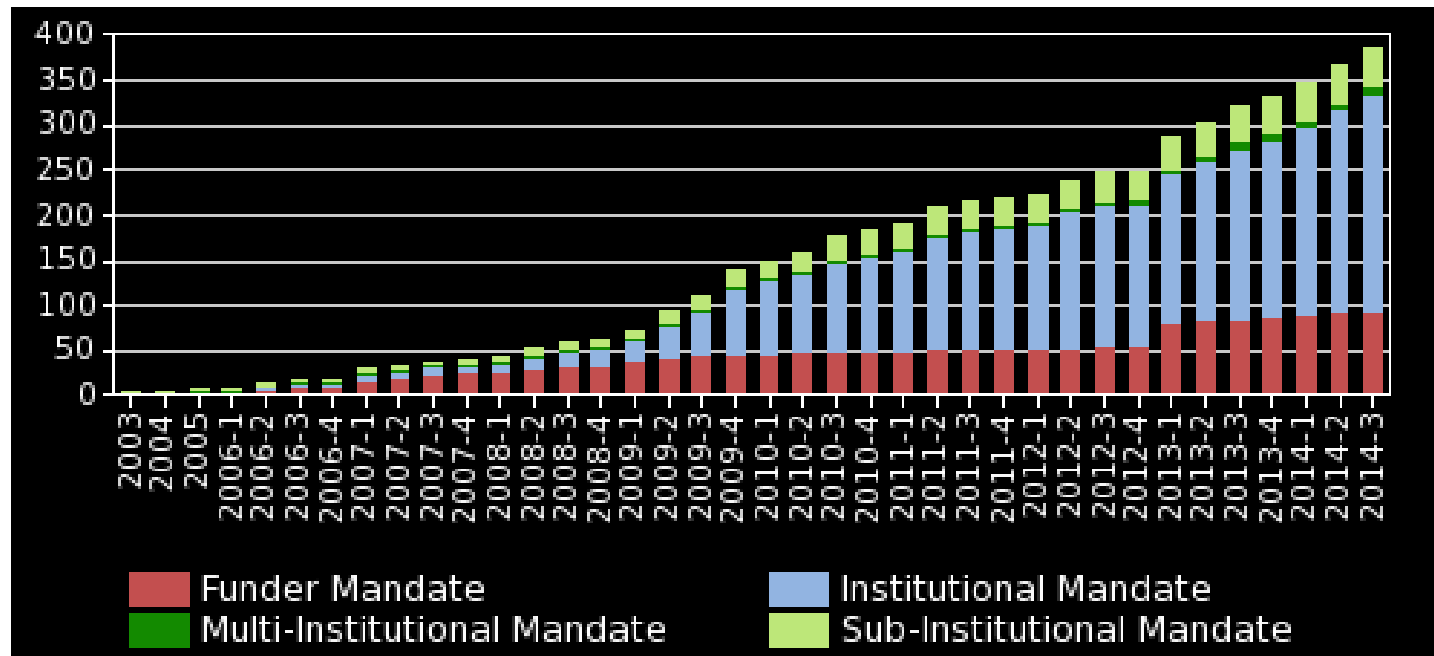
- **Most current open access policies,** recommendations and regulations issued by research institutes, funding agencies and governments **include GL as a relevant primary source** of research information.

- New trend in OA in **health sciences** is moving towards **open data sharing**



- **Pilot Open Access in FP7**
 - **Pilot Open Data in Horizon 2020**
 - recognises that research data is as important as publications
 - requires data management plans
- GL is directly involved**

ROARMAP: Registry of Open Access Repositories Mandatory Archiving Policies



Apply to:

the results of all scientific research, all research output, **thesis**, preprints

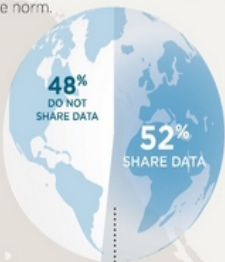
RESEARCHER DATA SHARING INSIGHTS

WILEY

- Wiley's Researcher Data Insights Survey was launched earlier this year to understand how and why researchers make their research data publicly available. The study's results, highlighted below, are intended to advance the global conversation about data sharing and help Wiley better meet the needs of our researchers, authors, and partners in the rapidly evolving landscape of scientific research and communications.
- The survey was deployed in March 2014 and received more than 2,250 responses from researchers around the world.

GLOBAL DATA SHARING TRENDS

Data sharing practices vary widely across research fields and geographic areas. Just over half of researchers report making their data publicly available, though archiving results in repositories is not yet the norm.

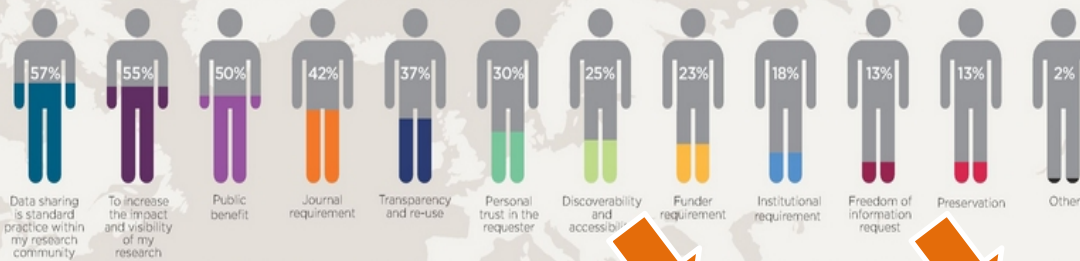


WAYS DATA IS SHARED

- 67%** As supplementary material in a journal
- 37%** Personal, institutional or project webpage
- 26%** Institutional data repository (i.e. university or institute-sponsored)
- 19%** Discipline-specific data repository
- 6%** General-purpose data repository (e.g. Dryad, figshare)
- 5%** Other

Globally, researchers also report sharing their data in limited and non-permanent ways: 57% are sharing data at a conference while 42% of researchers share their data upon informal request (e.g. email, direct contact, etc.).

RESEARCHER MOTIVATIONS FOR SHARING DATA



DATA SHARING TRENDS BY COUNTRY



REASONS WHY RESEARCHERS ARE HESITANT TO SHARE THEIR DATA

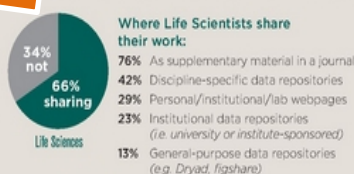
- 42%** Intellectual property or confidentiality issues
- 36%** My funder/institution does not require data sharing
- 26%** I am concerned that my research will be scooped
- 26%** I am concerned about misinterpretation or misuse
- 23%** Ethical concerns
- 22%** I am concerned about being given proper citation credit or attribution
- 21%** I did not know where to share my data
- 20%** Insufficient time and/or resources
- 16%** I did not know how to share my data
- 12%** I don't think it is my responsibility
- 12%** I did not consider the data to be relevant
- 11%** Lack of funding
- 7%** Other

DATA SHARING BY DISCIPLINE

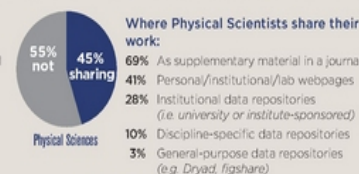
Data sharing, specifically by way of data repositories, is most prevalent amongst life scientists, particularly those in the earth and environmental and agriculture and food sciences.



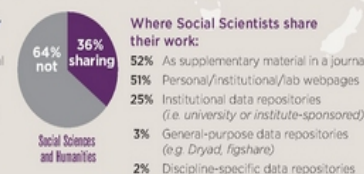
A typical **Health Science researcher** says she would be motivated to share her data in the future in order to benefit the public, so long as privacy and ethical concerns are assuaged.



A typical **Life Science researcher** says she would be motivated to share more of her data in the future if she was guaranteed proper credit.



A typical **Physical Science researcher** says she would be motivated to share her data in the future because it is standard practice within her research community and because it increases the impact and visibility of her work.



A typical **Social Science and Humanities researcher** says she would be motivated to share her data in the future if it increased the impact and visibility of her work or if she was required to by her funder.



Key words of the changing landscape

open access, **open** data, **open** standards,
open science, innovation, **evidence-based** policy,
knowledge transfer, rights, licences, equality,
interoperability, **responsible research and innovation**

*What is the impact
of such concepts on GL
producers and users?*

- Authors, readers
 - issuing organizations
 - librarians, archivists, data curators
 - other information professionals
- in government, academy, business and industry

similar challenges as
those affecting
producers and users of
“traditional” literature

- We all participate in the same debate
- Need to join efforts
- Find common solutions

Why direct involvement of the grey community is necessary?

The involvement in the international debate on knowledge sharing and **open data** becomes fundamental to

- **learn** from the others
- **share** positive and negative experiences
- **contribute** to speed up progress towards new forms of information dissemination.



definitively **cleaning up the grey dust and poor quality** often associated with GL.

GL moving towards OPEN SCIENCE

OPEN SCIENCE is the movement to make scientific research, data and dissemination **accessible to all levels** of an inquiring society, amateur or professional

an example,
among many



**Opens VIVO Research Networking Tool
to Public participation, November 2014**

1. **Improve** production standards and transparency
2. Ensure greater **discoverability and accessibility**
3. **Recognise the value of grey literature** for scholarly communication
4. Improve collection and curation of policy resources
5. Reform copyright and legal deposit legislation.

Altmetrics and GL, virtue rewarded

Find alternative metrics to the use of Impact Factor

Examples

DORA declaration (Declaration On Research Assessment)
pledging to find alternative and better modes of assessing academics
(American Society of Cell Biology), Dec 2012

UK research excellence framework REF, 2014

“No sub-panel will make any use of journal impact factors, rankings, lists or the perceived standing of publishers in assessing the quality of research outputs. An underpinning principle of the REF is that **all types of research and all forms of research outputs across all disciplines shall be assessed** on a fair and equal basis.”

NISO standard on Altmetrics in preparation
public comments in July 2014 on Draft including GL

WHAT COLOUR FOR REASEARCH DATA ?

Examples of «underlying data» in health sciences

Experimental data

Results of the action of a drug on experimental model of disease, *in vivo* o *in vitro*

Clinical or biomonitoring data

es. TAC images , biological samples, burden of chemical pollutants in the body

Surveillance data

Epidemiological data: incidence, mortality, demographic data, data of population affected by a disease

Simulation data

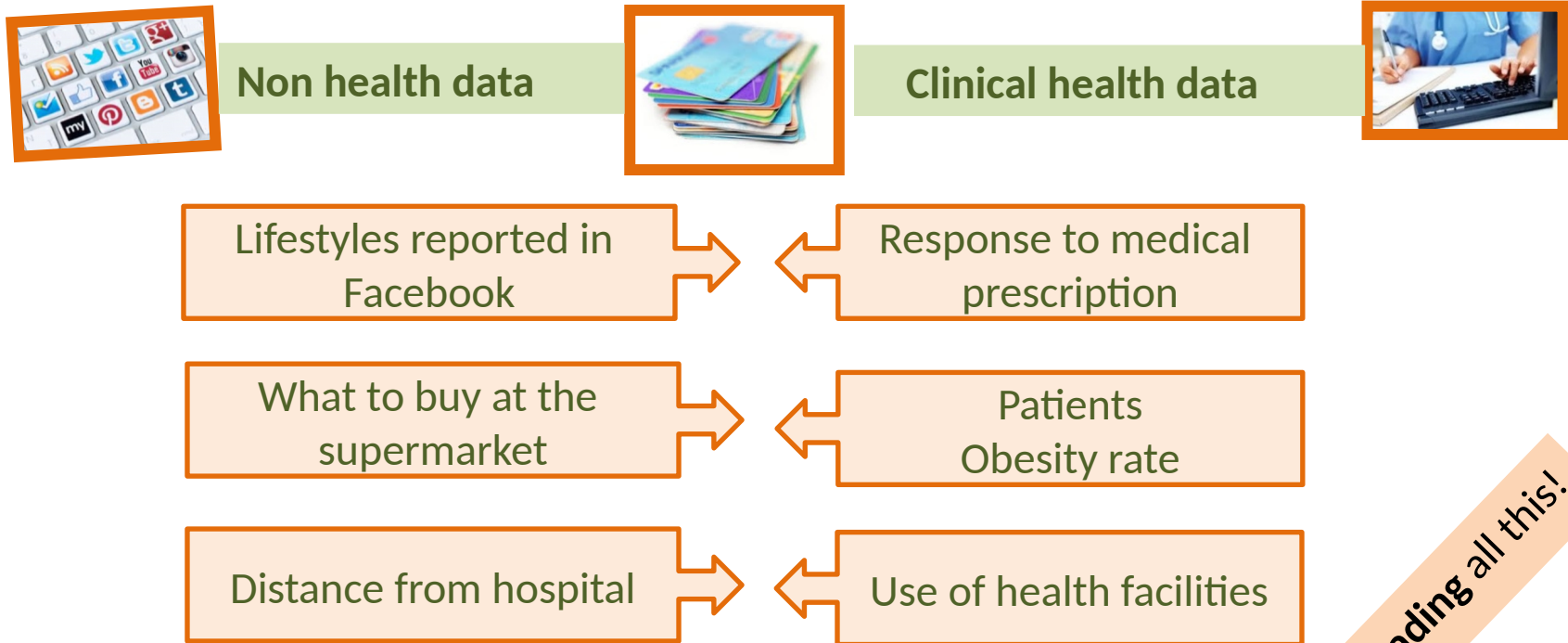
in silico: data generated by a program reproducing biological processes

Privacy
of personal data

a new challenge for GL

Big data in biomedicine, finding the missing link

Integration of etherogeneous data



Weber GM, Mandl KD, Kohane IS. Finding the Missing Link for Big Biomedical Data. *JAMA* 2014;311:2479-2480.

GL pervading all this!

The case of clinical trials

All trials should be registred and the results published.

Indeed, for about half of the trials registered in ClinicalTrials.gov,

published data are only partial

([Chalmers I, et al. All trials must be registered and the results published. BMJ 2013](#))

Therefore, once again, GL can help to...

- Meet patients expectations (contribute to scientific progress / test new treatments)
- Support clinical decisions based on new trials
- Meet the requirements of Helsinki declaration (1964 and updatings)
- Make the results of human experimentation publicly available
- Publish negative results in medicine
- Avoid duplication and waste (human and economic resources)

A red rectangular logo with a white plus sign and the text "AllTrials" in white.

A petition was recently launched in support of the above: *alltrials.net*

Currently signed by 81,732 people, 531 organizations. **Why not Textrealease?**

(GL) Data from regulatory agencies

European Medicines Agency (EMA) -

October 2014: [policy](#) on publication of clinical data



Access and use of

clinical reports

by the scientific community
(download, save and print data
except for commercial purpose)

Patient Privacy

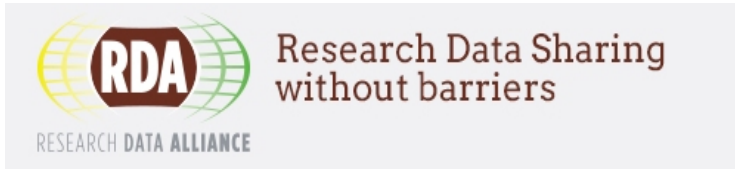


Commercial issues



Research Data Alliance (RDA)

Research data sharing without barriers



Started in 2013, financed by Australian Government, the *National Science Foundation* and the *Commission of the EU*

Vision □ *researchers and innovators openly **sharing data** across technologies, disciplines, and countries to address the grand challenges of society.*

Strives for: openness, consensus, balance, harmonization, with a community driven and non-profit approach.

Advocates: incentives for open data attitudes; recognition of OD- related work, development of infrastructures; services; policies; training for data re-use *across place and time*, data traceability, etc.

once again, GL is there!

Final considerations

1. Recognize the value of digital & open GL

as a primary source of quality information **affecting different stakeholders** in its major potential to contribute to

- advancement of research
- economic growth
- citizens' empowerment

2. Include GL in evaluation

- career advancement
- ranking of institutions
- funding of research

3. Consider GL within the Open science movement

sharing technical and ethical considerations in a new wide vision of knowledge transfer



THANK YOU
paola.decastro@iss.it

Special thanks
Gustav Klimt
for inspiring this
presentation.

A final fairy tale

Like **Cinderella**, Klimt lived in poverty for most of his childhood; then he made a **Team**, «The company of artists». His life started to change; yet, for a while, his art was still considered **rather disturbing**, and some of his precious works were even destroyed...

Next chapter,
next Conference...