

Opening by **Minister for Small Business John Perry**

- video recording <http://media.heanet.ie/page/d449b30f1f5d440d8700729eb7fde1d0>
- fragmented and diverse data all over world
- small businesses should develop thanks to sharing data
- mentioning Horizon 2020 (biggest European research funding program ever), wants Ireland to become a leading partner on big data in Europe
- Ireland is the best country for small business in international ranking within Europe (info from Dec 2013)

Professor Mark Ferguson

- video recording <http://media.heanet.ie/page/32353f411e5f4c18ac247455c6f9cb4f>
- Director General of Science Foundation Ireland & Chief Scientific Adviser to the Government of Ireland
- all countries are sending scientists to other countries (abroad) to increase the quality of the science – (optimal time for staying abroad is for about 2 years – than the performance is the maximum)
- why should we fund internships for months, if it makes sense only for 2 years?
- the mostly cited papers (most powerful papers) are based on this points:
 1. - collaboration between academia and industry
 2. - international collaboration
 3. - collaboration within own country
 4. - activity in your own lab

The papers, which are leading to patents, are mostly **funded via jointly funded academic and commercial bodies.**

It's about getting joint language, the architecture, the infrastructure, to get everything right and to achieve common metrics for data.

Professor Ian William Chubb

- video recording <http://media.heanet.ie/page/829f40520371455bb81e96de81c4f1bd>
- Australia's Chief Scientist
- Why should the Australian citizens invest into science?
- the purpose isn't just to do the research
- research is requiring increased cooperation – this is increasing last 5 years
- global challenges can be solved only by global cooperation
- new collaborative technologies (e.g. shared telescopes in Chile)
- international and interdisciplinary collaboration needed
- some countries are expecting to get access to the results
- community driven initiatives, memorandums of understanding
- researchers needs new skills to be able to play in the new landscape (know-how to collaborate effectively, maximum the benefits, share the data!)
- developed support for researchers – national collaborative approach – easily accessible for every Australian scientist
- world leading climate science in Australia – developing infrastructure, which enables this – Australia is highly influenced by changing climate– it's extremely important to have good storage capacity (computer capacity) to analyze the data

- **in Australia the data must be published in an online available repository within 12 month period after related publication is published**
- preparing guidelines and guides for researchers and institution within whole Australia

Panel on data policy challenges

- video recording <http://media.heanet.ie/page/99bd9788c410472bb67d40a1bdba3b3c>

Patricia Clarke (Senior Policy Analyst, Irish Health Research Board)

- Health Research Board's policy requiring data deposit & linkage to publications adopted phased implementation
- minimal metadata about the research data are requested
- **It's easy to write the policies and procedures. It's hard to change the culture.**

Merce Crocas (Director of Data Science, IQSS, Harvard University)

- Dataverse repository hosted at Harvard is open to all researchers and all disciplines
- **guidelines for Data Publishing:**
 - A published dataset must contain at a minimum a required subset of citation metadata and one or more data files
 - A dataset is assigned a data citation compliant with the Joint Declaration of Data Citation Principles
 - We encourage and facilitate sharing sufficient information (domain-specific metadata and complementary files) to ensure reusability
 - We encourage and facilitate free, accessible data by assigning default open data licenses to both metadata and files, but allowing terms of use and restrictions to files when needed (metadata must remain open)
 - Once a dataset is published, it cannot be unpublished. Versioning is supported for changes in files and metadata. Deaccession is supported when a dataset must be retracted, but citation information remains.
 - A dataset might be associated to zero, one or more publications.

- guidelines for connecting journals to data

- We encourage and facilitate integration between journals and Data through Dataverse by:
 - supporting multiple workflows
 - supporting multiple journal data policies
 - dataset may be either required or optional upon article submission
 - dataset may be published before, at the time or after article is published
 - dataset is peer-reviewed during article review (less common)
 - dataset is published even if article is not published (not common)

Simon Hodson (Executive Director CODATA)

Types of Data Policies:

- International Quasi-Governmental Principles
- Research Funder Policies
- Research institution Policies or Codes
- International Data Sharing or Research Initiative Agreements or Principles
- Community Norms in Journal Data Availability Policies

- it's important to engage with the policies
- definition of data is needed: **We just want the data, which supports the argument of the article.**

Carlos Morais-Pires (DG Connect, European Commission)

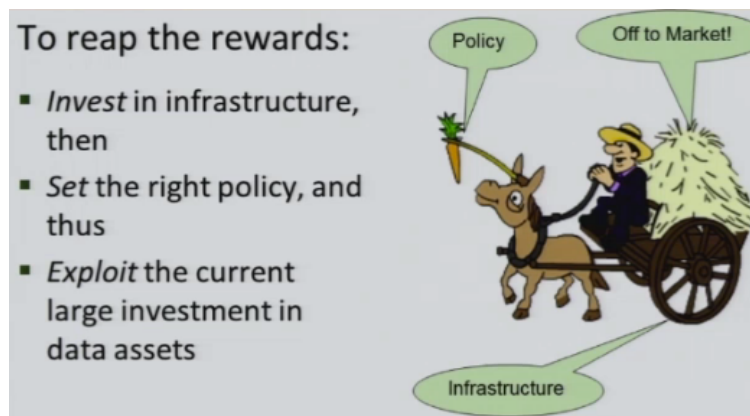
- based on G8+O6 working group on data infrastructure "White paper" was created
- includes 5 principles describing the benefits of global data infrastructure
- Data is:
 - Discoverable –ID's, descriptive metadata
 - Accessible –acknowledgement, licenses, terms of use
 - Understandable – semantics, analysis, quality-
 - Manageable – responsibility, costs, preservation

Data and Computing are two sides of the same coin, is RDA able to tackle this challenge?

"The infrastructure we need, cannot be designed from a bureaucrat from a room in Brussels."

Ross Wilkinson (Executive Director, Australian National Data Service)

- return of investment is the big question
- invest in infrastructure, set the right policy and thus exploit the current large investment in data assets
- stresses the importance of not attaching the carrot before the cart



QA session

Who should pay for the long-term preservation? Long living institutions.

"Only a long lived institution can provide long term preservation."

Research data should be considered as an asset, not a liability, (because it could deteriorate, get lost...) might help preserving

[@simonhodson99](#) Not just gold [#OA](#), whole [#OpenScience](#) agenda relevant to developing sustainable RD infrastructure

Publish Data Interest Group: introductory session

Most researchers are not aware of what to do to publish their data.

- presentation of the subgroups on Publishing data

2. WG Bibliometrics

- why Bibliometrics for Data?
- conceptualize data metrics and corresponding services to overcome barriers
- compilation of existing work
- summarize current practice and policies
- evaluation of possible approaches (potentially as a survey)
- develop recommendations

3. WG Publishing Data services

Data Cite – Jan Brase

Re3Data and DataBib to include their service in 2016

DataCite Annual Conference 2014

25-26 August 2014, INIST-CNRS, Nancy, France

The DataCite 2014 Annual Conference (formerly Summer Meeting) will be held at [INIST-CNRS](#) in Nancy, France, end of August, just after the [IFLA World Library and Information Congress](#) in Lyon.

4. WG Publishing Data Workflow

Having information about workflows is crucial for researchers to know what to do.

Define workflows and look at them more in details, how they work.

- including deposit and citation
- provide reference models, a “classifications”
- implementations of key components for application in new workflows
- existing initiatives: SCOR/IODE/MBLWHOI, PREPARDE, KomFor, ODIN and many others
- current timeline is presented, and at the end of the year there should be some best practices known, start of 2015 – implementation

Categories we are looking at?

Based on standards such as OASIS

Includes PID assignment to data set

Includes/links to author PIDs

Peer review of data

Peer review of metadata

Refers to a data life cycle

Number of people involved (incl. author, what are the roles and the responsibilities)

Links to grants

Link to paper

Standalone data

Applicable for both cases (standalone and linked)

Comments

- study of specific disciplines
- identify the general “components” the building blocks

- **practical Cookbook on Data Publication** by Woods Hole Oceanographic Institution <http://www.nodc.noaa.gov/media/pdf/oceanacidification/mg64.pdf>