



Research Data Alliance Update

**John Wood/
RDA-Europe Chair**

research data sharing without barriers
rd-alliance.org



An ERA driven by societal needs to address the 'Grand Challenges'

3

The ERA Milestones

We will know the ERA is driven by societal problems in 2030 when we see:

- A third of public, non-military research is geared to grand societal challenges, with a multi-disciplinary approach.
- 30% of all scientists, including humanities and social sciences, are trained in research fields relevant to the Grand Challenges.
- Multi-disciplinary academic training is generalized to educate our research community into the complexity of the Grand Challenges, without diminishing the importance of discipline-based expertise.
- The tools of 'e-science' are deployed throughout the ERA, permitting international collaboration so that all researchers will see themselves as part of the global research system.

From “Riding the Wave”

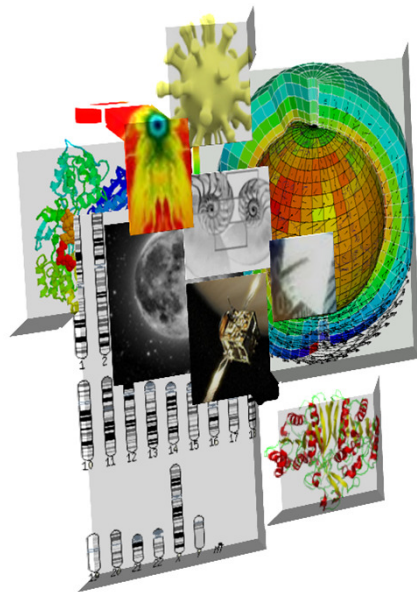
4

“Our vision is a scientific e-Infrastructure that supports seamless access, use, re-use and trust of data. In a sense, the physical and technical infrastructure becomes **invisible and the data themselves become the infrastructure** – a valuable asset, on which science, technology, the economy and society can advance.”



High-Level Group on Scientific Data

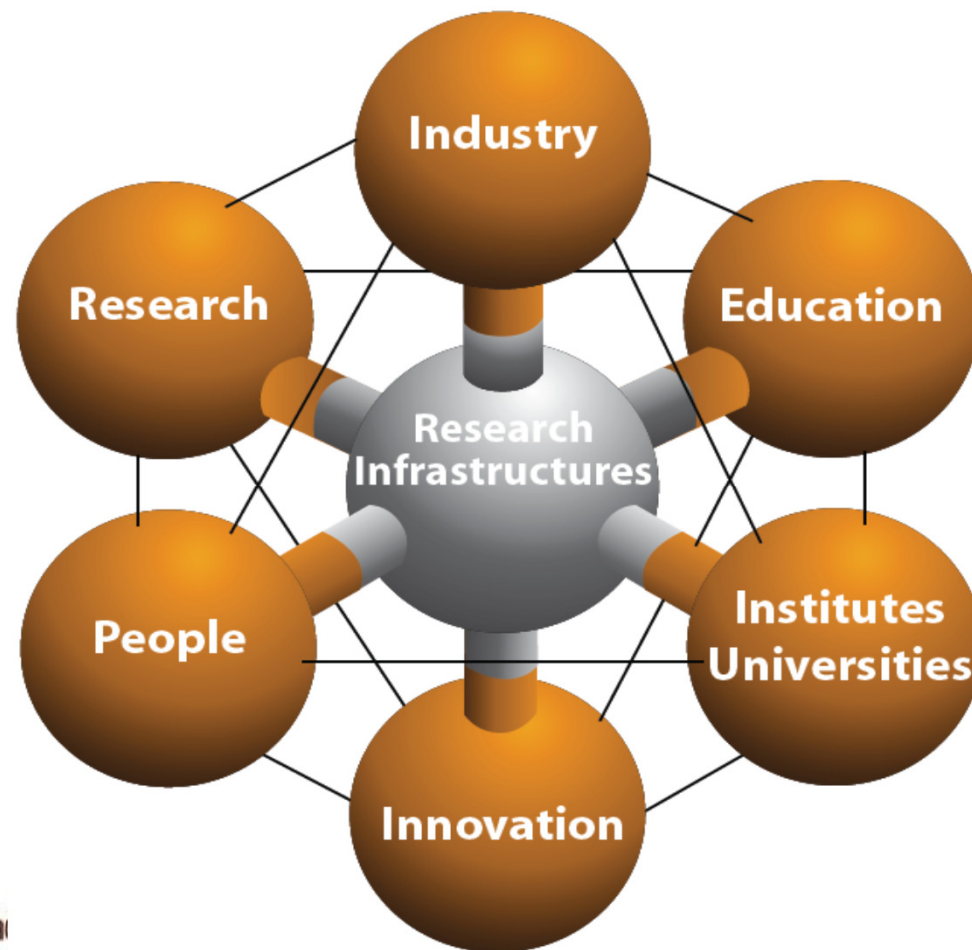
“Riding the Wave: how Europe can gain from the raising tide of scientific data”



- They can engage in whole new forms of scientific inquiry and treat information at a scale we are only beginning to see.
- ... and help us solving today's Grand Challenges such as climate change and energy supply.

The Centrality of Research Infrastructures for Innovation

6



oldWeather

HOME VESSELS TUTORIAL TRANSCRIBE ABOUT BLOG FORUM GET STARTED

old



HMS Empress of Britain

Active: Cape Verde Islands

85% COMPLETE

ElisabethB is the Captain and 136 volunteers are following this vessel

Jump Aboard

Google

Datos de mapa ©2011 Google - Términos de uso

Old Weather: Our Weather's Past, the Climate's Future

Introduction

Help scientists recover worldwide weather observations made by Royal Navy ships around the time of World War I. These transcriptions will contribute to climate model projections and improve a database of weather extremes. Historians will use your work to track past ship movements and the stories of the people on board.



1. Follow vessels



2. Digitise pages



3. Get promoted

Project Statistics

Old Weather transcriptions so far

48 % OF THE LOGS COMPLETED

371713 PAGES DONE
71 SHIPS COMPLETE

Why Scientists Need You

Scientists explain why your contributions are vital, and what they're doing with your results behind the scenes. [More information](#)



Using old weather observations to predict the climate's future.

TRANSCRIBE

Help us find the weather measurements to build a climate model.



Learn about our log books and the people that kept them.

VESSELS

Discover the ships these logbooks came from and how the data is contributing to our knowledge.



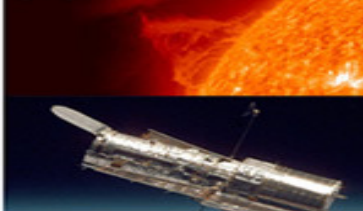
How the data from logs is used to build climate models.

WHO WE ARE

Find out about some of the people behind the project and how they will be using the data you are finding.

More Zooniverse

SOLAR STORMWATCH



GALAXY ZOO

THE PROJECT

Old Weather aims to recover worldwide weather observations to help improve climate model predictions. [Learn more](#)

ABOUT

Meet the team members and partner institutions behind Old Weather. [Read more](#)

FAQ

Questions about Old Weather? Visit the [FAQ](#) page or ask a question on the [Old Weather forum](#).

SKA

8



31/10/

research data sharing without barriers
rd-alliance.org



Scary Facts

9



> Did you know?

- + The data collected by the SKA in a single day would take nearly two million years to playback on an ipod.

Data is the new oil: Expert

New media will make consumers more powerful, says media futurist

BY CHRIS V
PANGANIBAN

DOHA: The world is becoming more interconnected thanks to broadband Internet and social media are playing a big role in bringing about a major generation shift, according to media futurist Gerd Leonhard (pictured).

Leonard, who was here to speak at ictQatar's Connected Speakers Series, said every content and media industry, including film, television, radio, gaming, software, music, publishing and news, had been severely impacted by the explosive growth of the Internet and the empowerment of the hyper-connected consumer.

Speaking before some 500 people at the Ritz-Carlton, Doha, Leonard said the move from the pre-Internet broadcast culture to the broadband culture had changed pretty much everything, from



economic models and their underlying business logic to the very definition of copyright and fair use.

He said such a shift was already in the midst of people around the world, noting that Internet content earned billions of dollars more in the US than content on television and newspapers.

He said the broadcast and print media were threatened

by this trend, citing the example of Facebook, which reaches more people in the Middle East than newspapers.

"In the content industries, we are going from selling copies, whether physical or digital such as books and CDs, to selling access, such as bundled music offerings where music is included in Internet access," said Leonhard.

"Without a doubt, data is the new oil. Over four billion connected users will generate zettabytes of data, every single day, by commenting, rating, tagging, forwarding, uploading and sharing content. Every marketer, every brand, every telco and every mobile operator will want to get to this data, and be allowed to use it," he said.

He said the consumer, better known in digital media as "user", will be more powerful than ever before, and — just like oil — many difficult

situations will arise from the use, discovery, mining and refining of data.

"Data is the most powerful thing in media", he said, predicting that by 2020 there will be an explosive amount 35 zettabytes (1 zettabyte = approximately 1 trillion gigabytes) of data on the Internet.

Leonhard is the co-author of the influential book, *The Future of Music*, as well as the author of *Music2.0* and the blog-book, *The End of Control*.

In concluding his presentation, Leonhard said: "The era of 'egosystems' such as walled, controlled, centralised and RoI-obsessed businesses, is rapidly closing. Now it's all about creating powerful ecosystems that are built on openness, standards, transparency, trust and decentralisation. Think Google versus AT&T, or Twitter versus FoxNews."

THE PENINSULA

Data Driving Solutions to Complex Science and Societal Challenges

11

What is the likely impact of a large-scale earthquake?



Earthquake simulations integrate
Fault geometry data
+ Subsurface data
+ Structure data
+ Population data
+ Physical infrastructure data

research data sharing without barriers
rd-alliance.org

Disease spread models couple
Medical data
+ Population data
+ Env't. data

How does disease spread in large urban populations?



How can we increase wheat yields?



Productivity analysis leverages interoperability of
Germplasm data
+ Genetic and phenotypic data
+ Statistical data
+ Bibliographic data



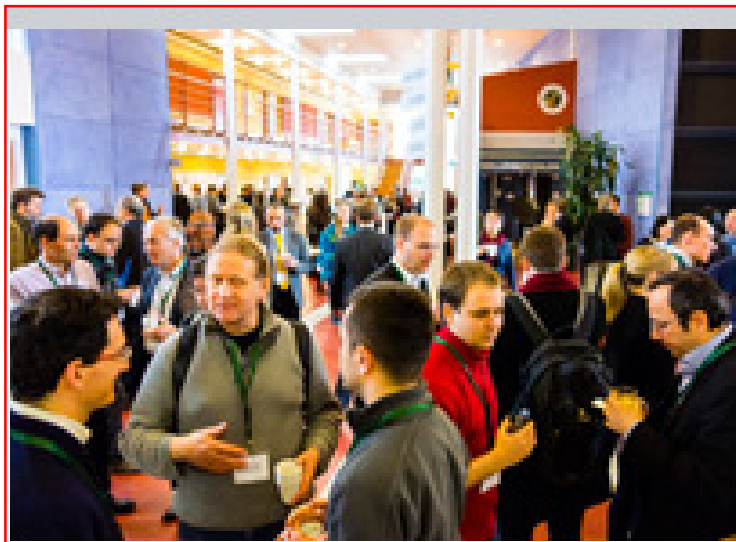
RDA Launch March 18-20, 2013

Gothenburg, Sweden


- Over 200 participants
- 31 countries
- 5 continents
- > 6,400 tweets
- Public, private, academic sectors
- High-profile Govt. and Science speakers



12



 **najla**
@najlaao
Jahanian #NSF: the #RDALaunch promises to accelerate pace of scientific discoveries. Data is new currency for science
4 DAYS AGO · REPLY · RETWEET · FAVORITE

 **ResearchDataAlliance**
@resdatall
Another theme emerging: RDA as a "neutral" space across domains, organizations, nations, etc. #RDALaunch
2 DAYS AGO · REPLY · RETWEET · FAVORITE

 **ResearchDataAlliance**
@resdatall
South Africa can't afford not to participate in RDA #RDALaunch
4 DAYS AGO · REPLY · RETWEET · FAVORITE

RDA Plenary 2 -- September 16-18, Washington D.C.

13

■ RDA Plenary 2

- 368 participants from 22 countries and all sectors
- All-hands stakeholder talks and RDA working meeting
- Data Citation Summit convened by DataCite, force11, CODATA/ICST, ESIP, DCC, etc. to create a common agenda
- ~5000 tweets over 3 days



RDA Plenary 2 -- Stakeholders and Invited Speakers

14

■ Keynotes:

- Tom Kalil, OSTP (introduced by Farnam Jahanian, NSF)
- John Wilbanks, Chief Commons Officer, Sage Bionetworks
- Carole Palmer, Professor, UIUC School of Library and Information Science

■ Global Partnerships Panel:

- Chris Greer, NIST
- Mark Suskin, NSF CISE/ACI
- Kostas Glinos, European Commission
- Clare McLaughlin, Australian Embassy
- Mike Stebbins, OSTP

■ Affiliate Organization Panel:

- Sara Graves, CODATA
- Curt Tilmes, ESIP
- Phil Archer, W3C
- Jan Braase, DataCite



Sara!

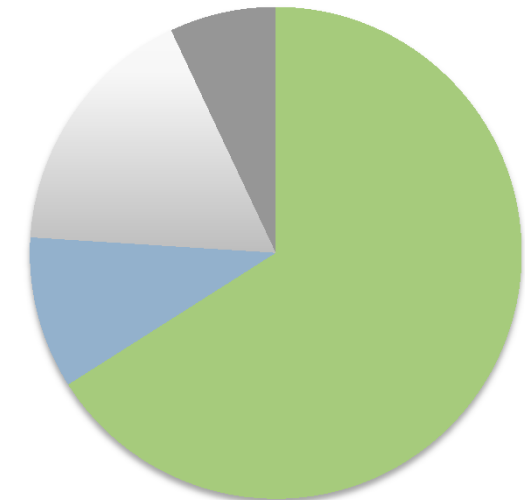


RDA Community Current Status: ~1300 participants from 50+ countries

15

- | | | |
|----------------------------|-------------------|--------------------------|
| 1. Albania | 18. France | 37. Russian Federation |
| 2. Australia | 19. Germany | 38. Rwanda |
| 3. Austria | 20. Greece | 39. Serbia |
| 4. Bangladesh | 21. Iceland | 40. Singapore |
| 5. Belgium | 22. India | 41. Slovenia |
| 6. Bolivia | 23. Iran | 42. South Africa |
| 7. Botswana | 24. Ireland | 43. South Korea |
| 8. Brazil | 25. Ireland {Rep} | 44. Spain |
| 9. Bulgaria | 26. Italy | 45. Sweden |
| 10. Canada | 27. Japan | 46. Switzerland |
| 11. China | 28. Krygrystan | 47. Taiwan |
| 12. Congo {Democratic Rep} | 29. Kuwait | 48. Turkey |
| 13. Costa Rica | 30. Mexico | 49. United Arab Emirates |
| 14. Czech Republic | 31. Netherlands | 50. United Kingdom |
| 15. Denmark | 32. New Zealand | 51. United States |
| 16. Estonia | 33. Norway | 52. Vatican City |
| 17. Finland | 34. Palestine | 53. Venezuela |
| | 35. Poland | |
| | 36. Portugal | |

RDA by Sector



- Academics (66%)
- Private Sector (10%)
- Public Sector (17%)
- Unknown (7%)

RDA Community Building Momentum

16

- Growth in number and scope of Interest Groups and Working Groups
 - New: BOFs for groups as precursor to Interest Groups
 - Groups beginning to “self-monitor” to promote concrete deliverables to be used and adopted
 - Increasing interest in more interaction and “connective tissue” between groups
- Pressing To-Dos before Plenary 3:
 - Develop an RDA policy for IP that comes up in Interest and Working Groups
 - Determine the form of RDA deliverables and what’s needed in terms of an “RDA archive”



Groups that Met at the RDA Plenary

**BOLD =
new since
last Plenary**

■ **Birds-of-a-Feather**

- **Linked Data**
- **Chemical Safety Data**
- **Education and Skills Development in Data Intensive Science**
- **Libraries and Research Data**
- **Cloud Computing and Data Analysis Training for the Developing World**

■ **Working Groups**

- **Data Type Registries**
- **Metadata Standards**
- **Practical Policy**
- **Persistent Identifier Types**
- **Data Foundations and Terminology**
- **Data Categories and Codes**

■ **Interest Groups**

- **Agricultural Data**
- **Big Data Analytics**
- **Data Brokering**
- **Certification of Trusted Repositories (joint with ICSU-WDS)**
- **Long tail of Research Data**
- **Marine Data Harmonization**
- **Community Capability Model**
- **Data Publishing (joint with WDS)**
- **Toxicogenomics Interoperability**
- **Research Data Provenance**
- **Data Citation**
- **Metadata**

■ **Economic Models and Infrastructure for Federated Materials Data Management**

- **Engagement**
- **Preservation e-Infrastructure**
- **Legal Interoperability (joint with CODATA)**
- **Global Registry of Trusted Data Repositories and Services**
- **Digital Practices in History and Ethnography**

■ **Data Citation Harmonization Summit**

- **DataCite, force11, CODATA/ICST, ESIP, DCC, etc.**

RDA Organizational Partners

New RDA
constituencies
/ stakeholders

- **Organizational Assembly** = Organizational Members (subscription) + Organizational Affiliates (MOUs).
- **Organizational Advisory Board** will govern Organizational Assembly.
- *Current Status:*
 - **Organizational Membership** under discussion with Microsoft, IBM, ANDS, Australian Antarctic Data Center, Intersect, Terrestrial Ecosystems Research Network, CSC – IT, Center for Science Ltd., Oracle, STFC, CNRI, STM, EUDAT, Barcelona Supercomputer Center, Columbia University Libraries / Information Services, and many more after the Plenary
 - **Organizational Affiliation** under discussion with CODATA, WDS and others
- *Next 6 months (before Plenary 3)*
 - Firm up model for Affiliates (how many, how substantive should the interaction be?)
 - Complete creation of legal entity to host subscriptions for Organizational Members

RAL to be the Registered Office

CLG UK Charity

- Elect Organizational Advisory Board at Plenary 3

RDA Constituent Groups Coming Together

New Position: RDA
recruiting for full-
time Secretary-
General

RDA Colloquium

(National Research Agencies, EC and Funders)

RDA Council

(overarching leadership)

Technical Advisory Board

(Technical oversight)

Secretary-General and Secretariat

(Administration and
Operations)

Organizational Advisory Boards and Organizational Assembly

(Organizational
partnerships and
guidance)

Working Groups and Interest Groups

(impact - focused infrastructure)

RDA Membership

RDA Leadership since Plenary 1

20

■ **Council:**

- **Patrick Cocquet** (France)
- **Doris Wedlich** (Germany)
- **Kaye Raseroka** (Botswana)
- **Tony Hey** (US)
- **Ross Wilkinson** (AU)
- **John Wood** (UK), co-Chair
- **Fran Berman** (US), co-Chair

● **Organizational Assembly**

- **Juan Bicarregui**, co-Chair
- **Leif Laaksonen**, co-Chair

■ **Technical Advisory Board**

- **Peter Wittenburg**
- **Francoise Genova**
- **Andrew Treloar**
- **Bill Michenor**
- **Beth Plale, Chair**
- 6 new TAB members to be elected this month (14 candidates)
- 12th TAB member to be appointed by Council (for balance)

Hot News!

21

- Advertisement for the appointment of a Secretary-General announced last week
 - Senior post to run the RDA activities.

See RDA website for details – closing date 25th
November 2013

- Other countries have indicated they wish to join RDAC,

Next Plenaries (2X a year)

22

- Plenary 3 will be in Dublin March 26-28 in 2014, hosted by Australia and Ireland
- Plenary 4 will be in the Netherlands – late September in 2014
- Plenary 5 or 6 likely back in the U.S. (west coast?)



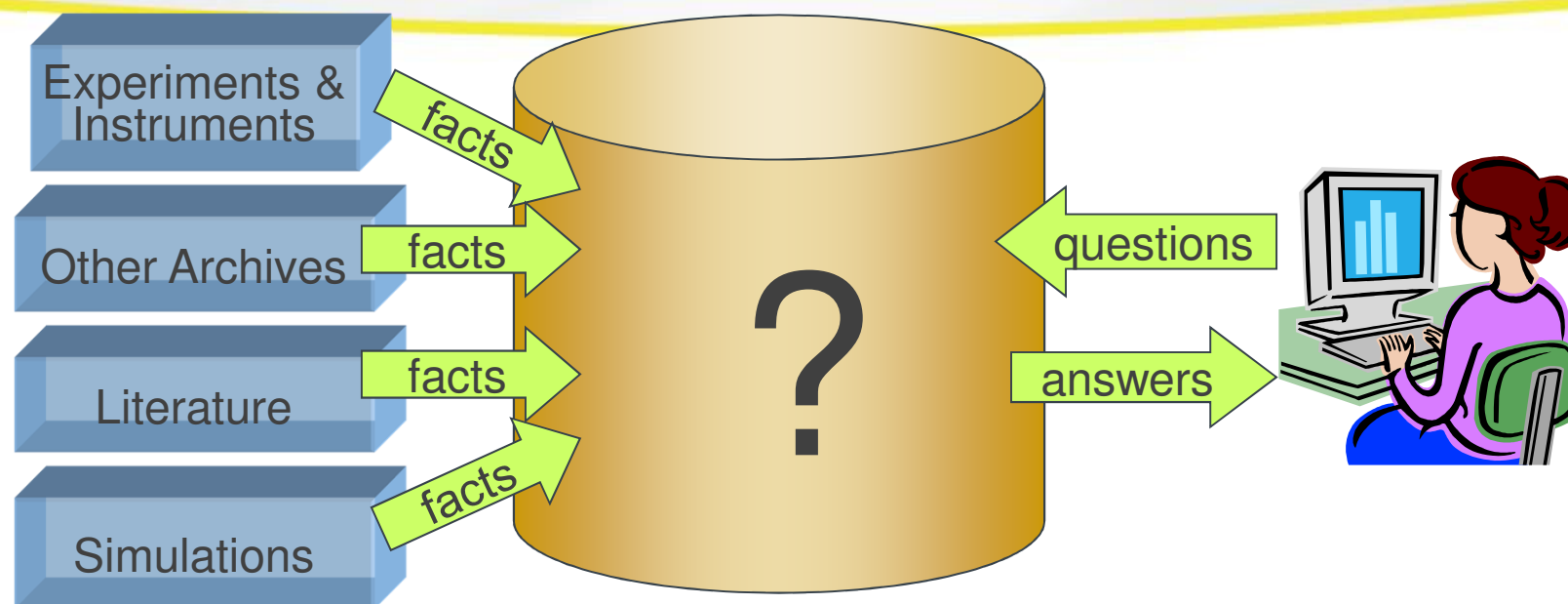
Council Meetings – Every 3 months

23

- Co-incide with plenaries
Palo Alto – February
Vancouver – June
- Next – London to coincide with G8+6 group in December – expected to endorse
- **White Paper: 5 Principles for an Open Data Infrastructure**
Draft v2.0, 21 May 2013
(editors: Alan Blatecky, Juan Bicarregui, Carlos Morais Pires)
- Discoverable, Accesible, Understandable, Manageable, People, Metrics

The Problem for the e-Scientist Student

24



Data ingest
Managing petabytes+
Common schema(s)
How to organize?
How to re-organize?

How to coexist & cooperate with
other scientists and researchers?
Data query and visualization tools
Support/training
Performance

Execute queries in a minute
Batch (big) query scheduling



Info:

enquiries@rd-alliance.org

research data sharing without barriers
rd-alliance.org

Fran Berman