Research Data Sharing and Frameworks

Yasuhiro Murayama
(National Institute of Information and Communications Technology,
ICSU-World Data System ex officio,
Kyoto University)

International Programme Office Hosted by NICT
Based in Tokyo, Japan

Recent Situation of Scientific Research Data

G8 Science Ministers Statement London UK, 12

Introduction

We, the G8 Science Ministers met in London on Wednesd
of our respective national science academies, as part of th
this unique meeting we discussed how our nations could le
transparency, coherence and coordination of the global sc
in order to address global challenges and maximise the so
of research.

3. Open Scientific Research Data

Open enquiry is at the heart of scientific endeavour, and rapid technological change has profound implications for the way that science is both conducted and its results communicated. It can provide society with the necessary information to solve global challenges. We are committed to openness in scientific research data to speed up the progress of scientific discovery, create innovation, ensure that the results of

4. Expanding Access to Scientific Research Results

G8 Open Data Charter will 'increase transparency' and 'fuel innovation'

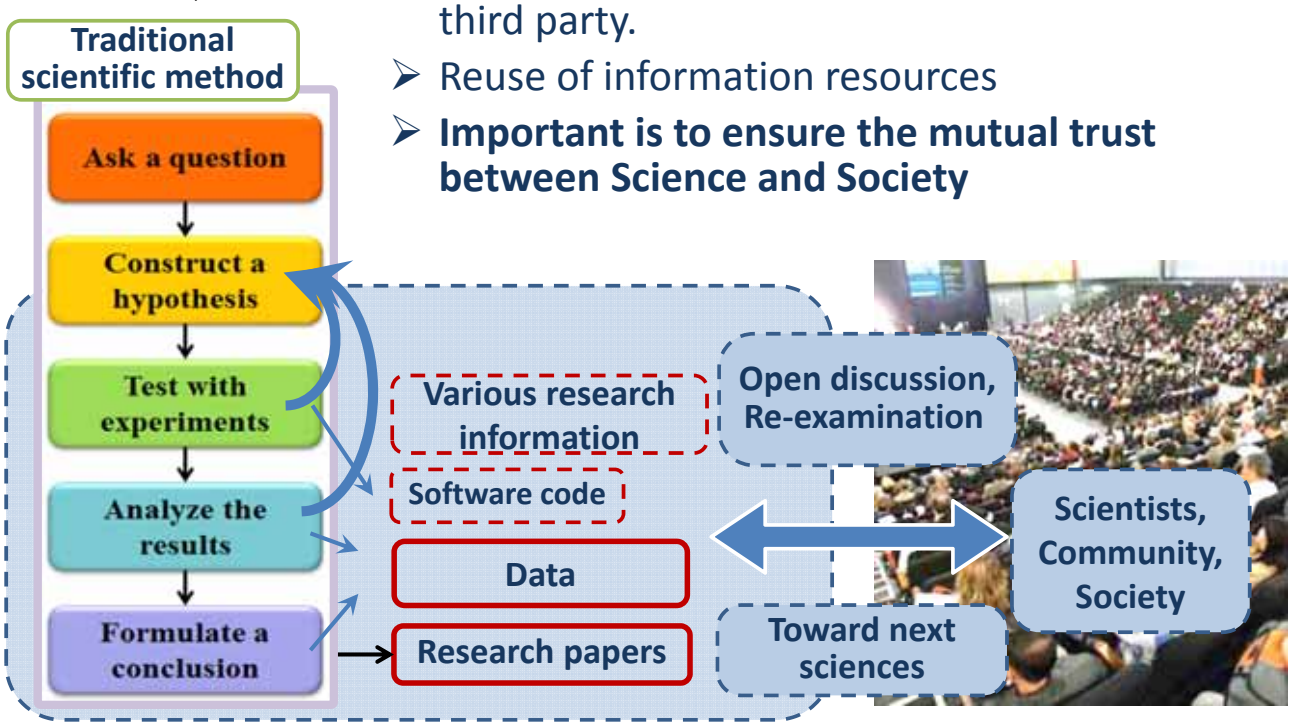
Five key principles outlines how governments should release datasets for economic and social benefits

Why “DATA” now?
改めて、「いま、なぜ、データか？」

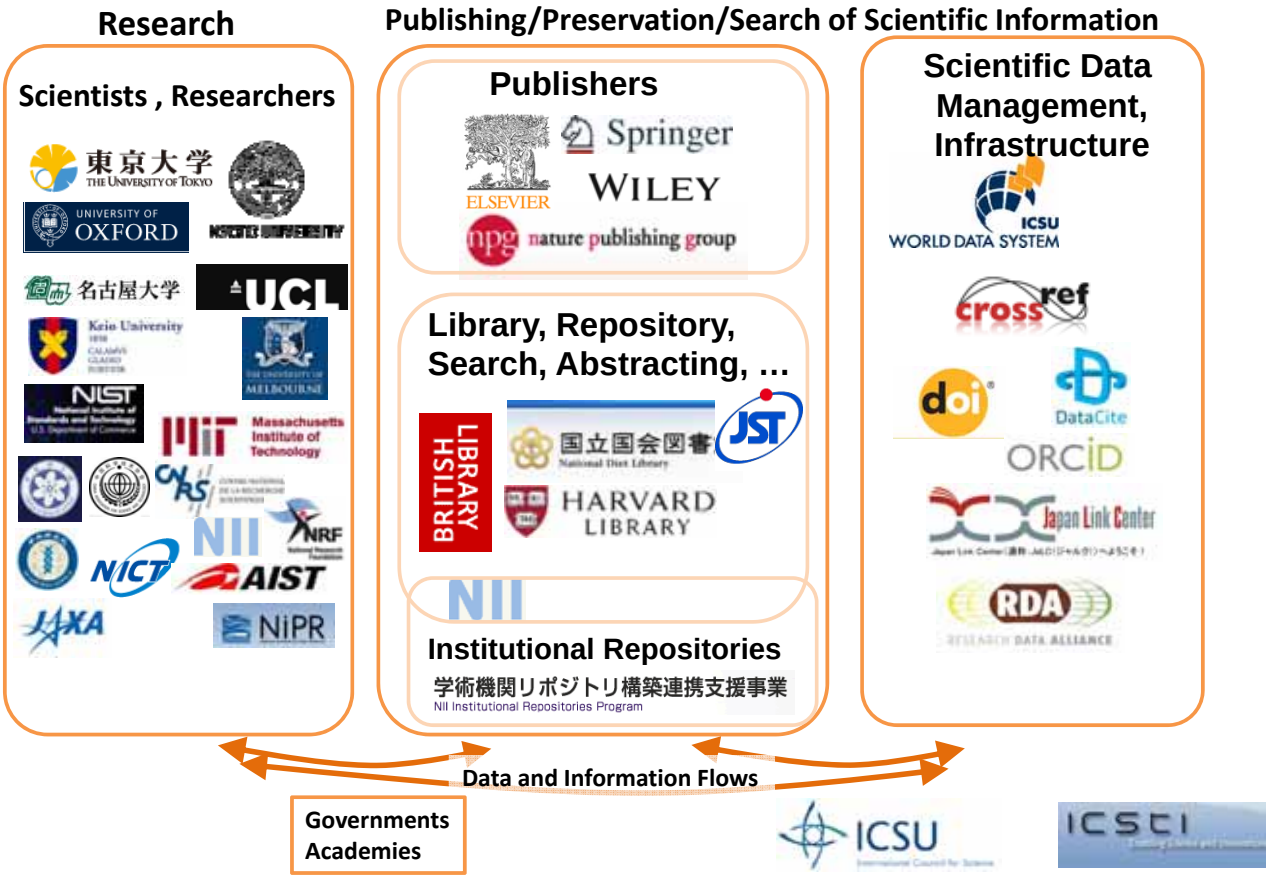
- Science and Society
  - Role of Science and Scientists in Society
    近年、社会と科学者の関わりが問われている
- Sharing data and information as part of "Science”
  科学技術活動の一部としての「データ（または情報）の共有」

Why Open Data, Open Access?

- Sharing research information and evidence
- For open discussion and re-examination by third party.
- Reuse of information resources
- **Important is to ensure the mutual trust between Science and Society**



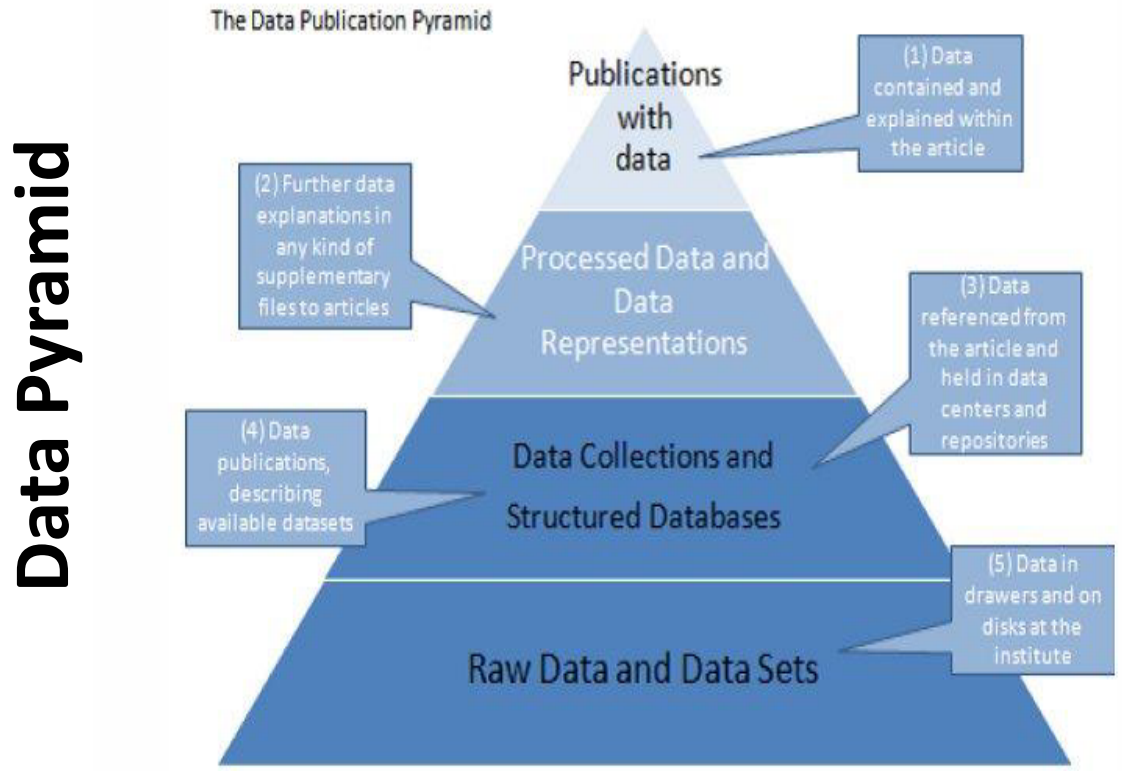
Science as a Social System with “Print” Publication



Value of Data

- Proof/evidence of scientific finding and understanding (as of original scholarly paper)
  - –➔ Data should be shared with everyone for proof and discussion!
- Resource for research and innovation
  - –➔ I don’t want to share my data with other scientists since my research is based on it!

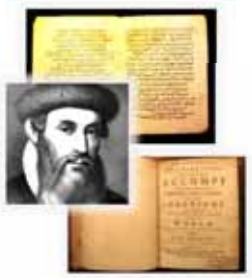
[H. Frederick Dylla, 2012]

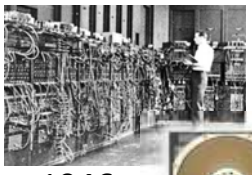


History: scientific record & communication

- Public library (paper media)：8c
- Printing press/Gutenberg: 1445
- First scientific journal: 1665
- Intl. Assoc. Academies: 1899
- ICSU established: 1931
- World Data Center system : 1957

348 years





- ENIAC, von Neumann: 1946
- Hard Disk Drive: 1956
- TCP/IP, dial-up (64kbps): 1982
- WWW (CERN): 1991
- Broadband internet (>1Mbps)：～2000

67 years

- New global data initiatives: ICSU-WDS、RDA etc.：2008～2013

Changing standards and culture takes long time.

- Metric system
  - Introduced 1795
  - Convention du Mètre, 1875
  - International System of Units (SI) 1960
  - One big holdout
- Time zones
  - 1st use of standard ("railway") time 1847
  - International Meridian Conference 1884 established GMT but did not alter local times
  - Final adoption of "standard offset" from GMT/UTC 1986
  - Current number of time zones in China and India: 1

[Mark Parsons, 2013]

Creation of ICSU-World Data System

ICSU 29th General Assembly decision (October 28, 2008):

PAST (since 1950's)

WDC (World Data Center) : 50 WDSs at max.
FAGS (Federation of Astronomical and Geophysical Data Analysis Services)

PRESENT (2008~)

ICSU WORLD DATA SYSTEM

ICSU

ICSU International Scientific Unions data bodies
ICSU National Members data bodies
ICSU Interdisciplinary Bodies data activities

82 Members (April 2014)

54 Regular	Data curation & data analysis services
9 Network	Networks of Regular Members & umbrella organizations
3 Partner	Do not deal directly with data stewardship, but support to ICSU-WDS
16 Associate	Organizations interested in the WDS endeavour

Research Data Alliance Created to Accelerate Development of Research Data Sharing Infrastructure Worldwide

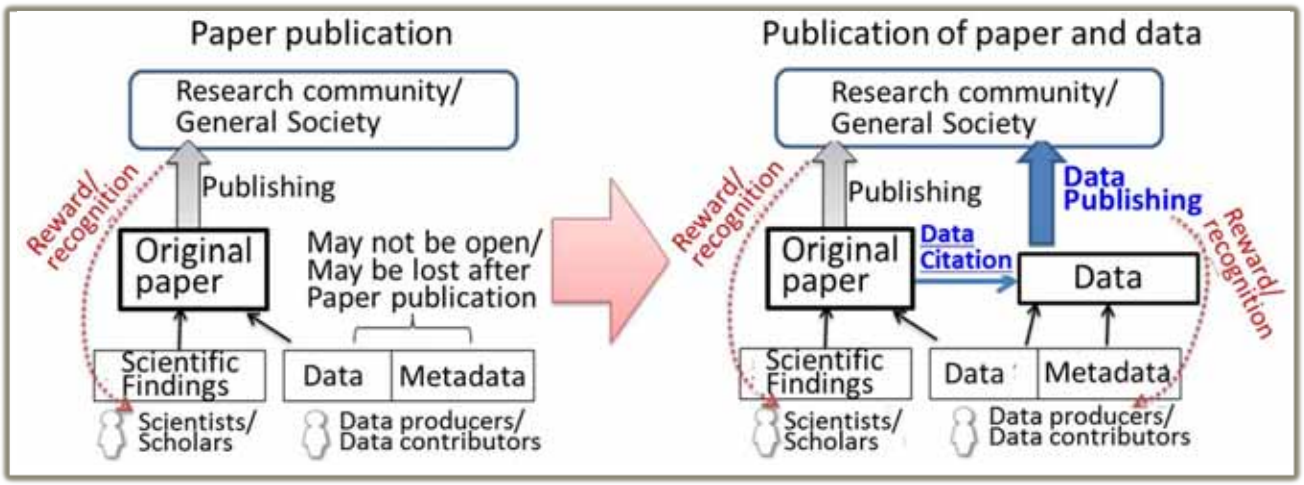
- RDA community efforts focus on building **social, organizational and technical infrastructure** to
  - *reduce barriers to data sharing and exchange*
  - *accelerate the development of coordinated global data infrastructure*

RDA and RDA/US are supported in part by the National Science Foundation.

Fran Berman

[Fabrizio Gagliardi , 2014]

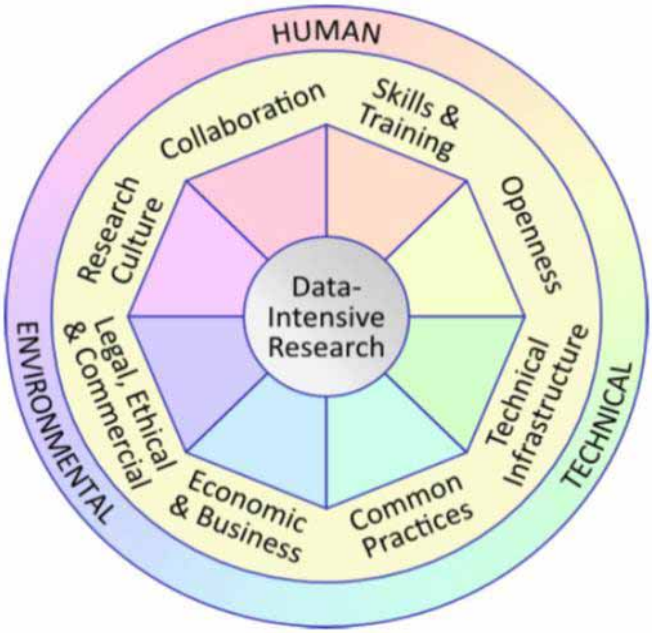
"Data Publication" and "Data Citation"



[Society of Geomagnetism, Earth, Planetary and Space Sciences, 2013]

- Data Publications
  cf. journal publication: review, fix (print), publish with DOI…, metrics (citation index etc.)
- Data Citation
  – ID of dataset ("DOI" is OK?), citation standards? metrics?…
- More outputs from scientists to Society

Toward Data Intensive Science



https://www.rd-alliance.org/filedepot_download/383/230

- RDA Community Capability Model Interest Group
  - Secretary: Univ. of Bath & Microsoft Research Connections
- Big data science/data intensive science become reality when the human, environmental, and technical difficulties are overcome.

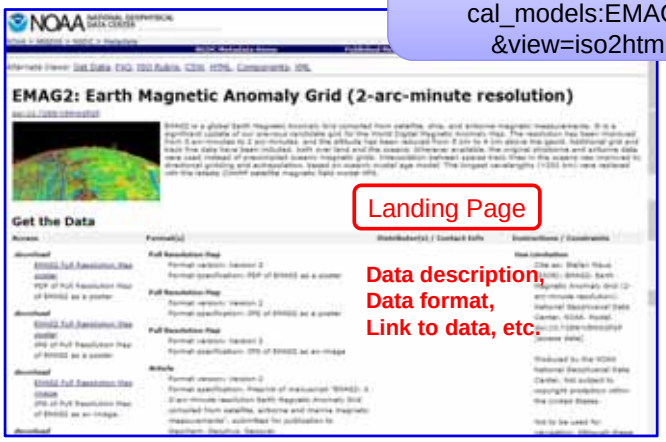
[Nose et al., 2013]

Example of DOI-minting to Earth Science database in NOAA/NGDC

✓ EMAG2: Earth Magnetic Anomaly Grid (2-arc-minute resolution)

doi:10.7289/V5MW2F2P

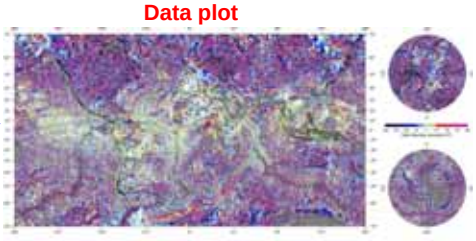
http://www.ngdc.noaa.gov/nmmrview/metadata.jsp?id=gov.noaa.ngdc.mgg.geophysical_models:EMAG2&view=iso2html



Digital data

-55.033333 -89.900000 -56.134989
-55.000000 -89.900000 -56.127400
-54.966667 -89.900000 -56.119806
-54.933333 -89.900000 -56.112213
-54.900000 -89.900000 -56.104616
-54.866667 -89.900000 -56.097016
-54.833333 -89.900000 -56.089413
-54.800000 -89.900000 -56.081806
-54.766667 -89.900000 -56.074197
-54.733333 -89.900000 -56.066584
-54.700000 -89.900000 -56.058968
-54.666667 -89.900000 -56.051348
-54.633333 -89.900000 -56.043725
-54.600000 -89.900000 -56.036097
-54.566667 -89.900000 -56.028466
-54.533333 -89.900000 -56.020832
-54.500000 -89.900000 -56.013193
-54.466667 -89.900000 -56.004623

Data plot



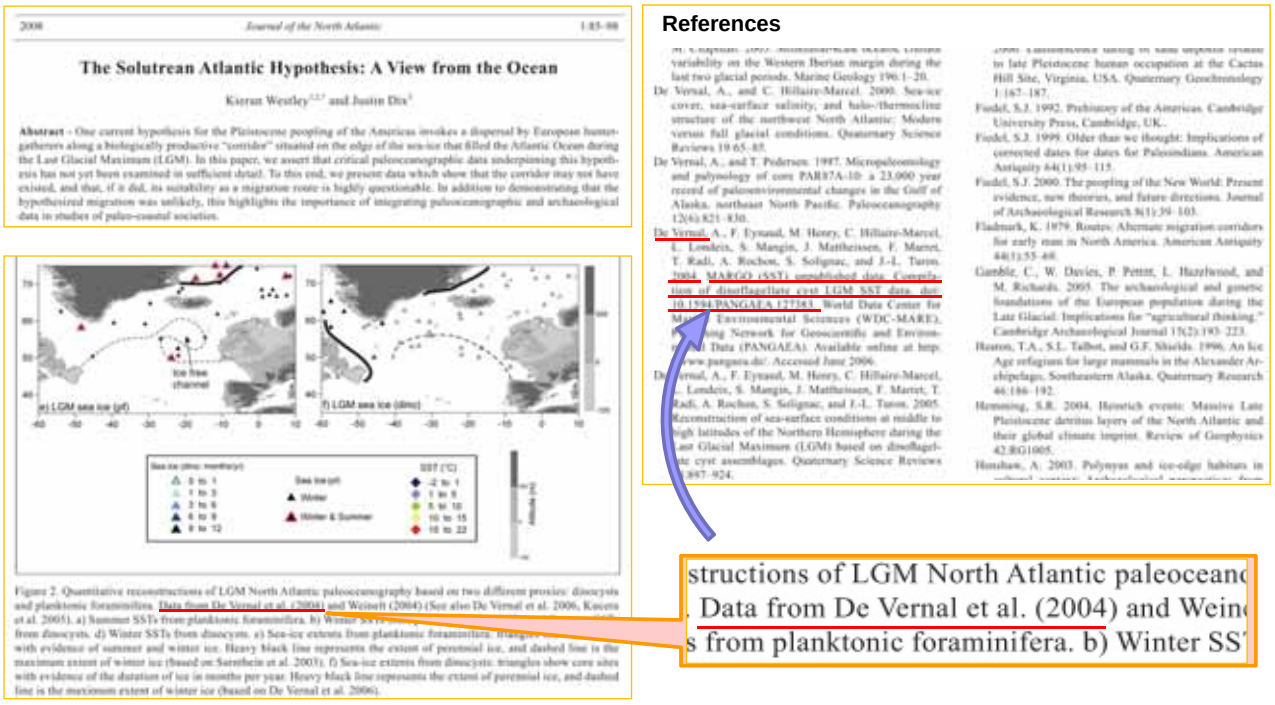
Instruction of data citation

Maus (2009): EMAG2: Earth Magnetic Anomaly Grid (2-arc-minute resolution). National Geophysical Data Center, NOAA. Model, doi:10.7289/V5MW2F2P [access date]

[Nose et al., 2013]

Example of data citation

Evaluation of the Solutrean hypothesis



2008 *Journal of the North Atlantic* 1:85–98

The Solutrean Atlantic Hypothesis: A View from the Ocean

Kieran Westley[1,2,*] and Justin Dix[1]

Abstract – One current hypothesis for the Pleistocene peopling of the Americas invokes a dispersal by European hunter-gatherers along a biologically productive "corridor" situated on the edge of the sea-ice that filled the Atlantic Ocean during the Last Glacial Maximum (LGM). In this paper, we assert that critical paleoceanographic data underpinning this hypothesis has not yet been examined in sufficient detail. To this end, we present data which show that the corridor may not have existed, and that, if it did, its suitability as a migration route is highly questionable. In addition to demonstrating that the hypothesized migration was unlikely, this highlights the importance of integrating paleoceanographic and archaeological data in studies of paleo-coastal societies.

Figure 2. Quantitative reconstructions of LGM North Atlantic paleoceanography based on two different proxies: dinocysts and planktonic foraminifera. Data from De Vernal et al. (2004) and Weinelt (2004) (See also De Vernal et al. 2006, Kucera et al. 2005). a) Summer SSTs from planktonic foraminifera. b) Winter SSTs from planktonic foraminifera. c) Winter SSTs from dinocysts. d) Winter SSTs from dinocysts. e) Sea-ice extents from planktonic foraminifera. Triangles show core sites with evidence of summer and winter ice. Heavy black line represents the extent of perennial ice, and dashed line is the maximum extent of winter ice (based on Sarnthein et al. 2003). f) Sea-ice extents from dinocysts. triangles show core sites with evidence of the duration of ice in months per year. Heavy black line represents the extent of perennial ice, and dashed line is the maximum extent of winter ice (based on De Vernal et al. 2006).

References

variability on the Western Iberian margin during the last two glacial periods. Marine Geology 196:1–20.
De Vernal, A., and C. Hillaire-Marcel. 2000. Sea-ice cover, sea-surface salinity, and halo-/thermocline structure of the northwest North Atlantic: Modern versus full glacial conditions. Quaternary Science Reviews 19:65–85.
De Vernal, A., and T. Pedersen. 1997. Micropaleontology and palynology of core PAR87A-10: a 23,000 year record of paleoenvironmental changes in the Gulf of Alaska, northeast North Pacific. Paleoceanography 12(6):821–830.
De Vernal, A., F. Eynaud, M. Henry, C. Hillaire-Marcel, L. Londeix, S. Mangin, J. Matthiessen, F. Marret, T. Radi, A. Rochon, S. Solignac, and J.-L. Turon. 2004. MARGO (SST) unpublished data. Compilation of dinoflagellate cyst LGM SST data. doi: 10.1594/PANGAEA.177063. World Data Center for Marine Environmental Sciences (WDC-MARE), Marine Network for Geoscientific and Environmental Data (PANGAEA). Available online at http://www.pangaea.de/. Accessed June 2006.
De Vernal, A., F. Eynaud, M. Henry, C. Hillaire-Marcel, L. Londeix, S. Mangin, J. Matthiessen, F. Marret, T. Radi, A. Rochon, S. Solignac, and J.-L. Turon. 2005. Reconstruction of sea-surface conditions at middle to high latitudes of the Northern Hemisphere during the last Glacial Maximum (LGM) based on dinoflagellate cyst assemblages. Quaternary Science Reviews 24:897–924.
to late Pleistocene human occupation at the Cactus Hill Site, Virginia, USA. Quaternary Geochronology 1:167–187.
Fiedel, S.J. 1992. Prehistory of the Americas. Cambridge University Press, Cambridge, UK.
Fiedel, S.J. 1999. Older than we thought: Implications of corrected dates for dates for Paleoindians. American Antiquity 64(1):95–115.
Fiedel, S.J. 2000. The peopling of the New World: Present evidence, new theories, and future directions. Journal of Archaeological Research 8(1):39–103.
Fladmark, K. 1979. Routes: Alternate migration corridors for early man in North America. American Antiquity 44(1):55–69.
Gamble, C., W. Davies, P. Pettitt, L. Hazelwood, and M. Richards. 2005. The archaeological and genetic foundations of the European population during the Late Glacial: Implications for "agricultural thinking." Cambridge Archaeological Journal 15(2):193–223.
Heaton, T.A., S.L. Talbot, and G.F. Shields. 1996. An Ice Age refugium for large mammals in the Alexander Archipelago, Southeastern Alaska. Quaternary Research 46:186–192.
Hemming, S.R. 2004. Heinrich events: Massive Late Pleistocene detritus layers of the North Atlantic and their global climate imprint. Review of Geophysics 42:RG1005.
Henshaw, A. 2003. Polynyas and ice-edge habitats in

Westley and Dix [2008]

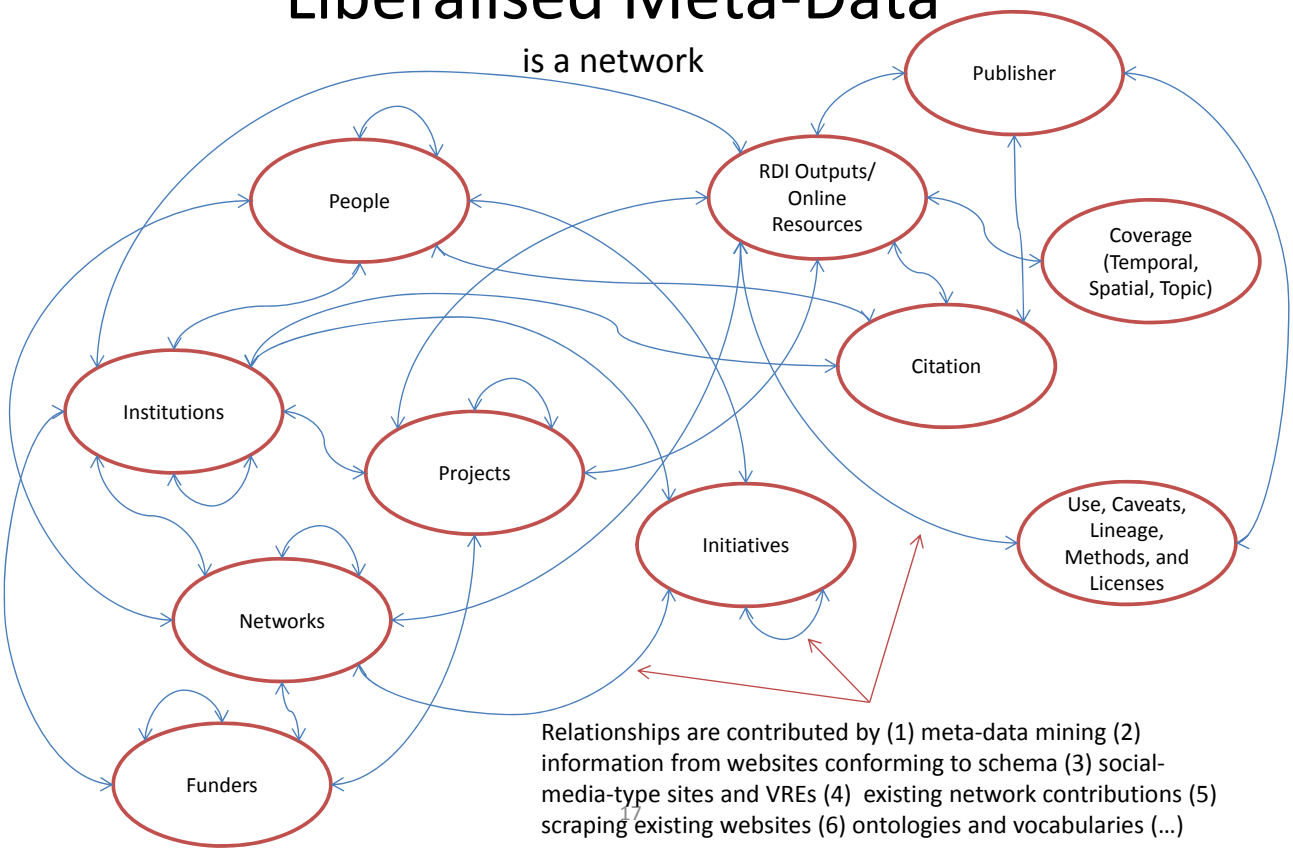
Steps by Major scientific publishers
encouraging data deposition

- **Willey/AGU** publication policy:
  "…in AGU's journals, **all data** necessary to understand, evaluate, replicate, and build upon the reported research **must be made available and accessible whenever possible**…"
- **SpringerOpen**/"Earth, Planets and Space", "Geoscience Letters"…
  "…Electronic archiving of data enables readers to replicate, verify and build upon the conclusions published in papers in the journal. **It is recommended that all data** which are not directly attached to a publication as electronic supplementary files **be deposited**…"
- **Elsevier**/JASTP:
  "…Elsevier encourages **authors to deposit raw experimental data sets** underpinning their research publication in data repositories, and to enable interlinking of articles and data…"

[Win Hugo, JpGU, May 2013]

Liberalised Meta-Data

is a network



Relationships are contributed by (1) meta-data mining (2) information from websites conforming to schema (3) social-media-type sites and VREs (4) existing network contributions (5) scraping existing websites (6) ontologies and vocabularies (…)

17

ご清聴ありがとうございました。