

The emerging complexity of open data dynamics

lessons from human evolutionary Genetics & Genomics

Giovanni Destro Bisol

Sapienza Università di Roma

Istituto Italiano di Antropologia

1. Just to remind you that...

a third view of data sharing... is possible

...two opposite ideas about
data sharing



Intelligent Openness*

transparency

assessability

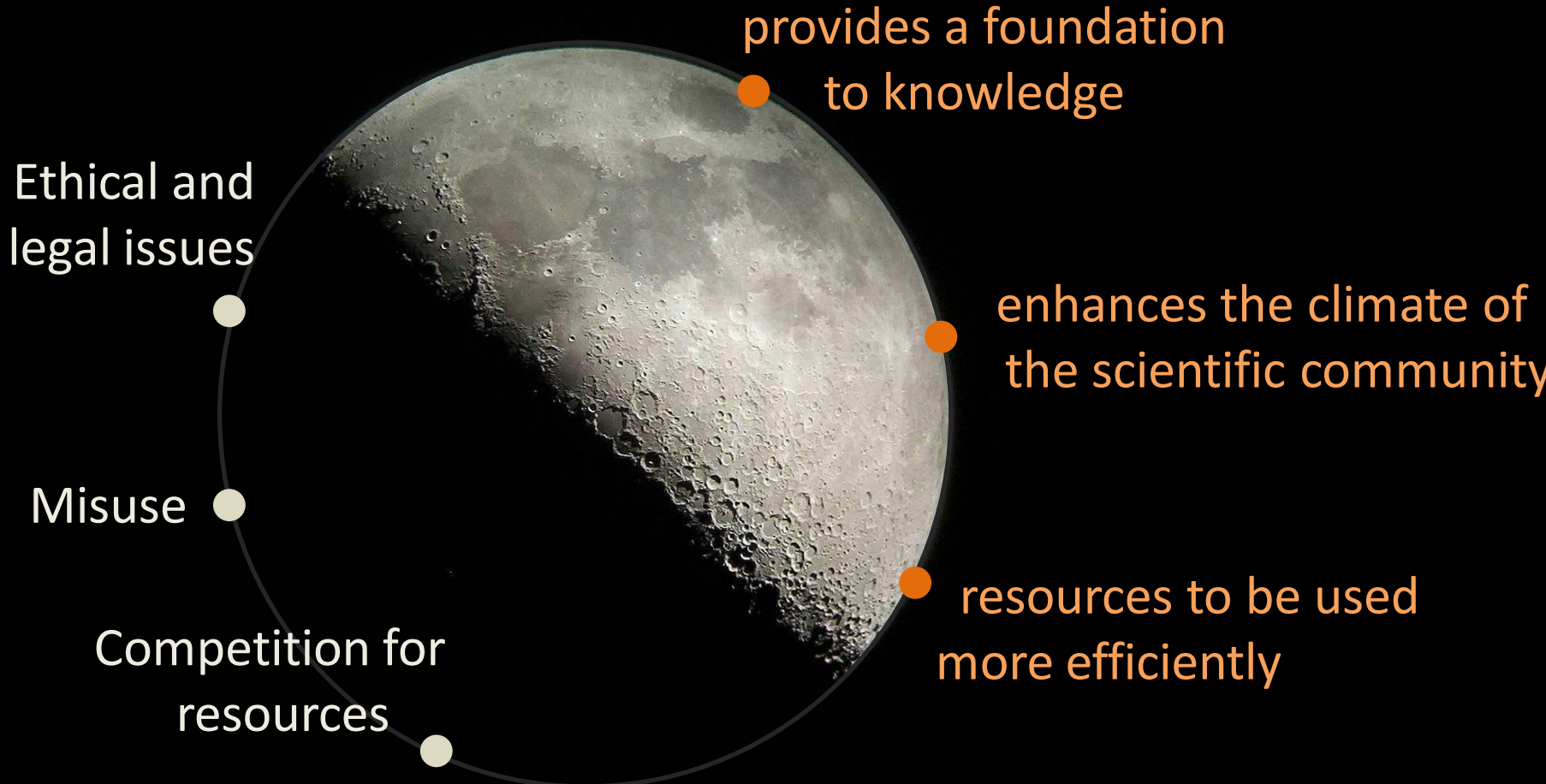
information flow

**accessibility
useability**

inclusiveness

intelligibility

* <https://royalsociety.org/~media/policy/projects/sape/2012-06-20-saoe.pdf>



provides a foundation
to knowledge

enhances the climate of
the scientific community

resources to be used
more efficiently

Ethical and
legal issues

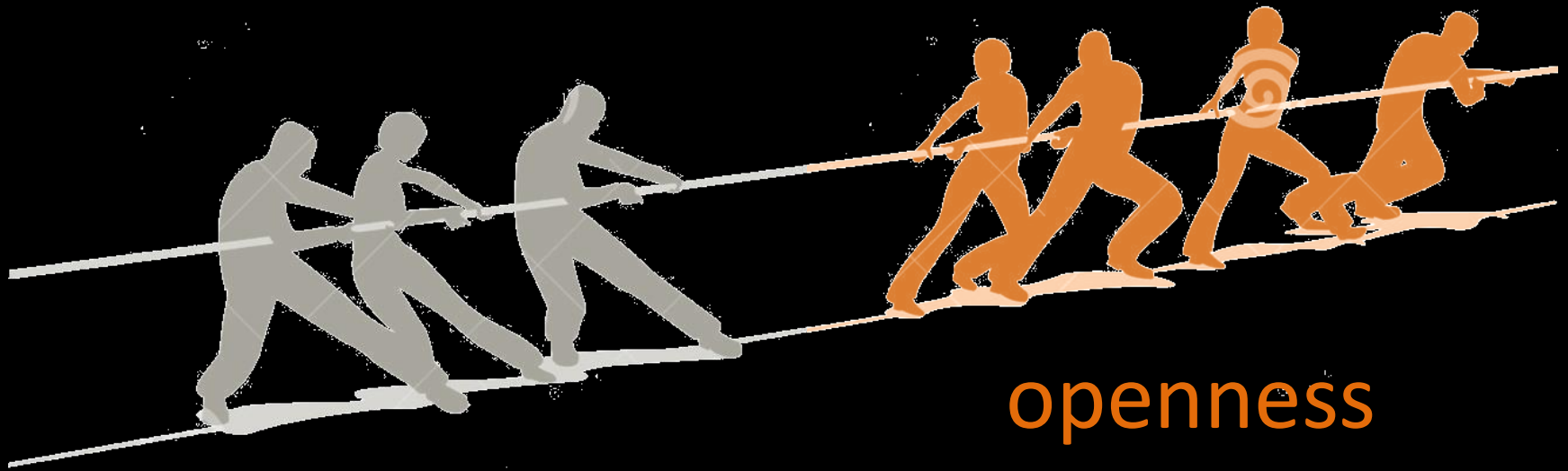
Misuse

Competition for
resources

....the third view

tension

secrecy



openness

BUT



The round table model.

Soranno et al., *Bioscience*, 2015.

doi: 10.1093/biosci/biu169

2. Stepping ahead

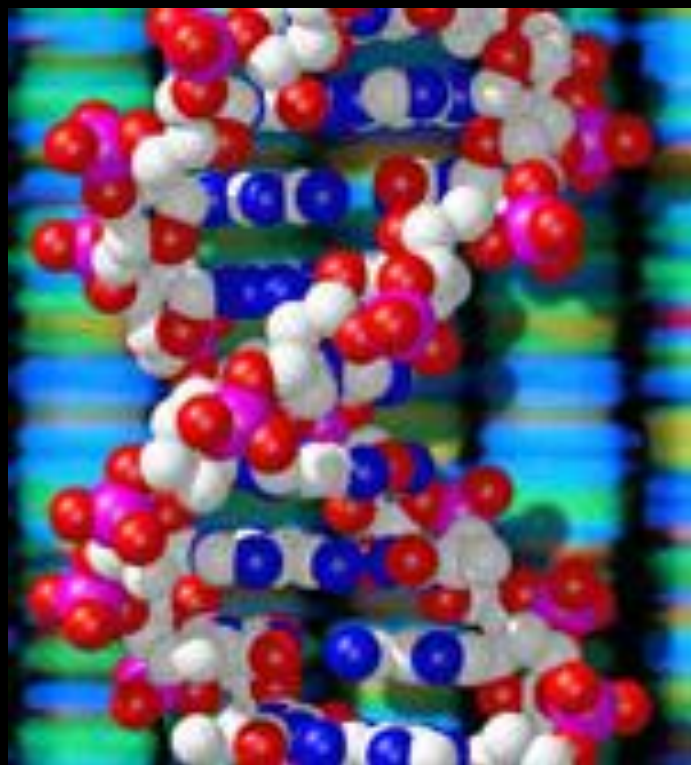
from the big picture to a single research field

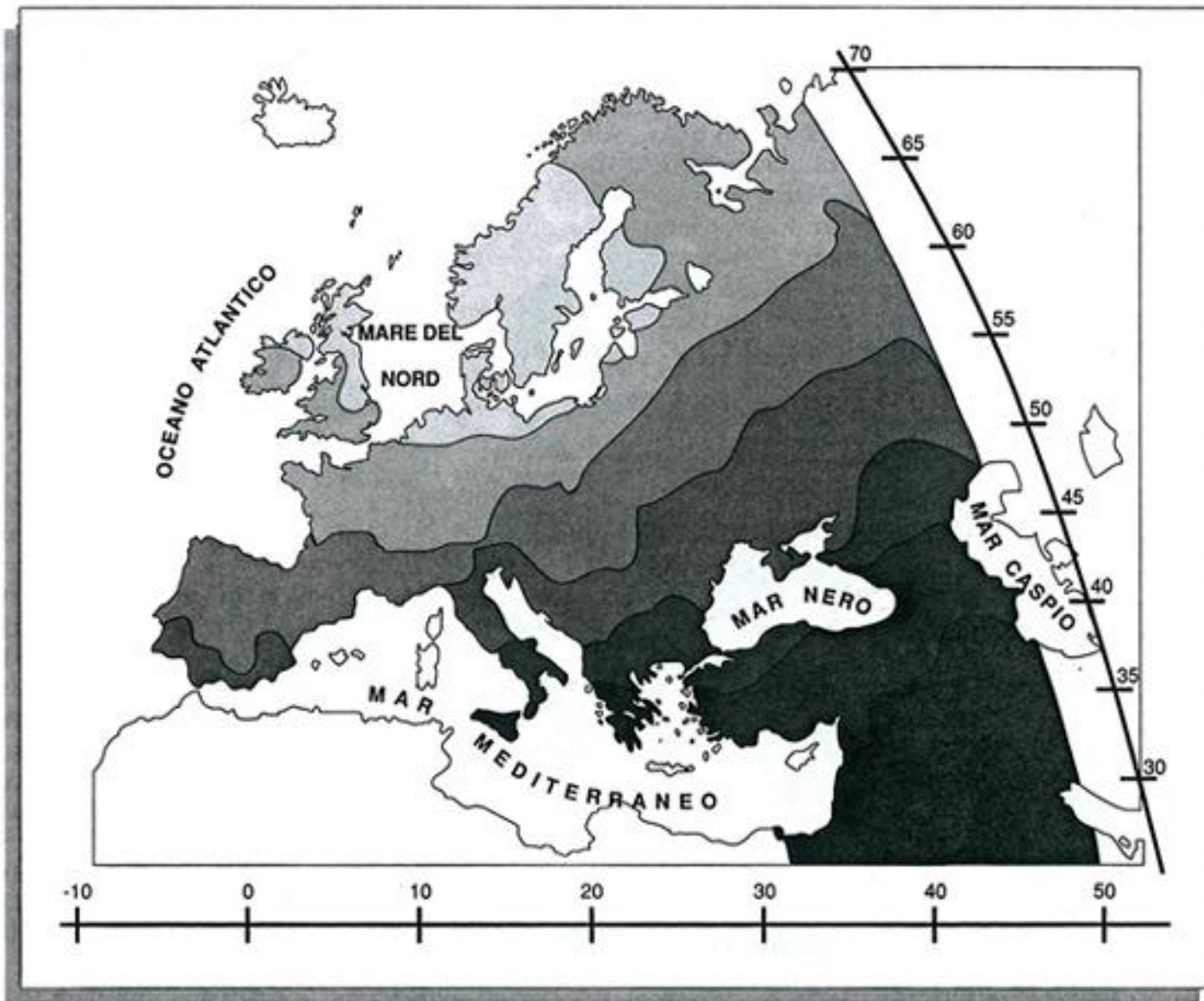
...Human evolutionary

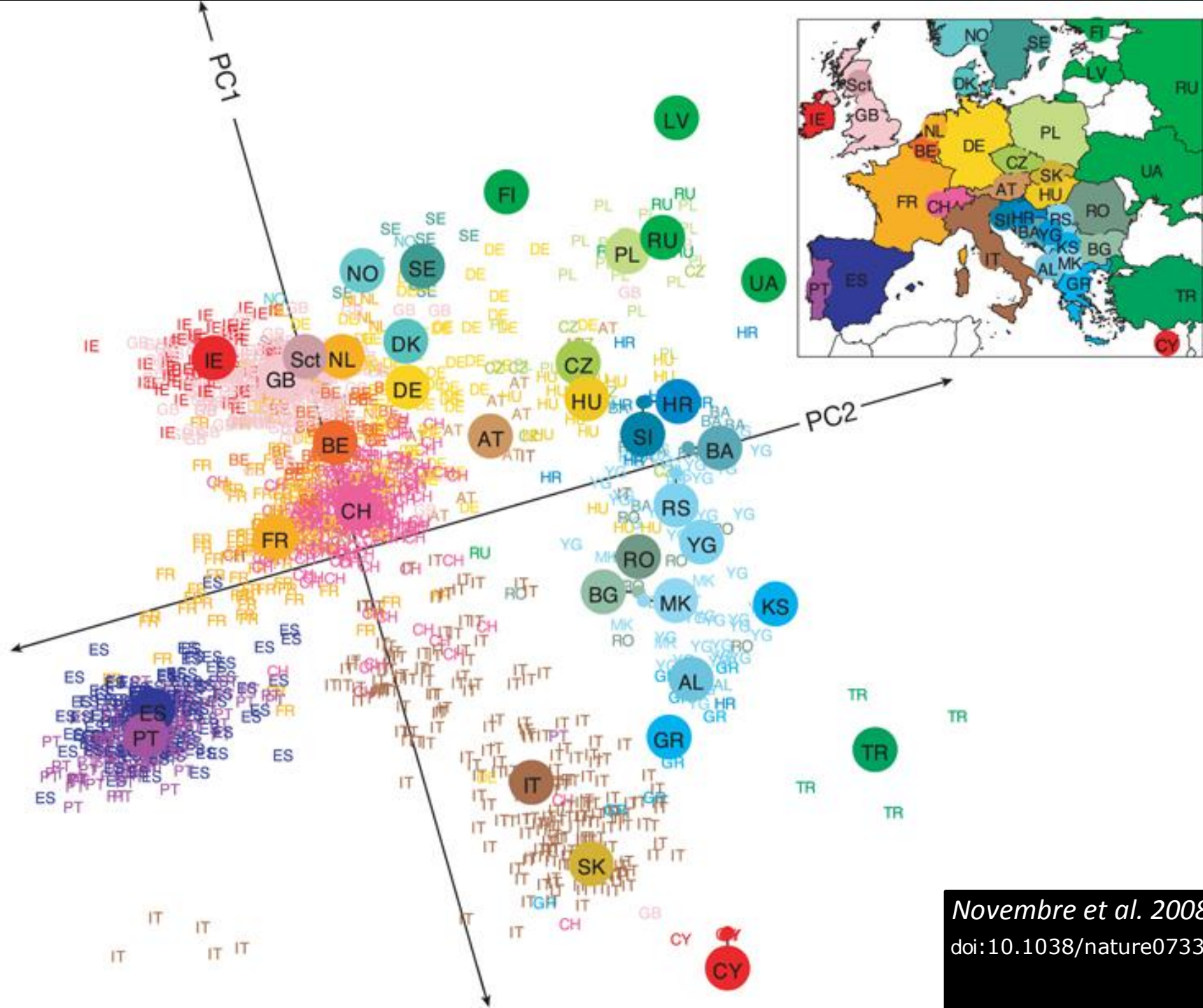
Genetics & Genomics

HUMAN EVOLUTION

Edited by Steve Jones, Robert Martin and David Pilbeam • Co-volume editor Sarah Hedges
Foreword by Richard Dawkins

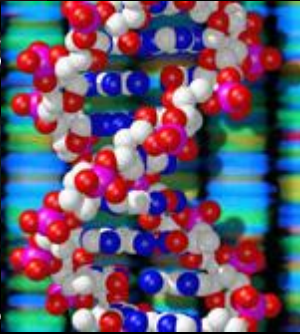




a

Novembre et al. 2008.
doi:10.1038/nature07331

HUMAN
EVOLUTION



Human evolutionary Genetics & Genomics

- ✓ codified nature of genetic information
- ✓ on the wave of methodological innovations
- ✓ encompasses small and big sciences (data)
- ✓ interdisciplinary nature

3. ... complexity?

JOURNAL OF THE AMERICAN SOCIETY FOR
INFORMATION SCIENCE AND TECHNOLOGY

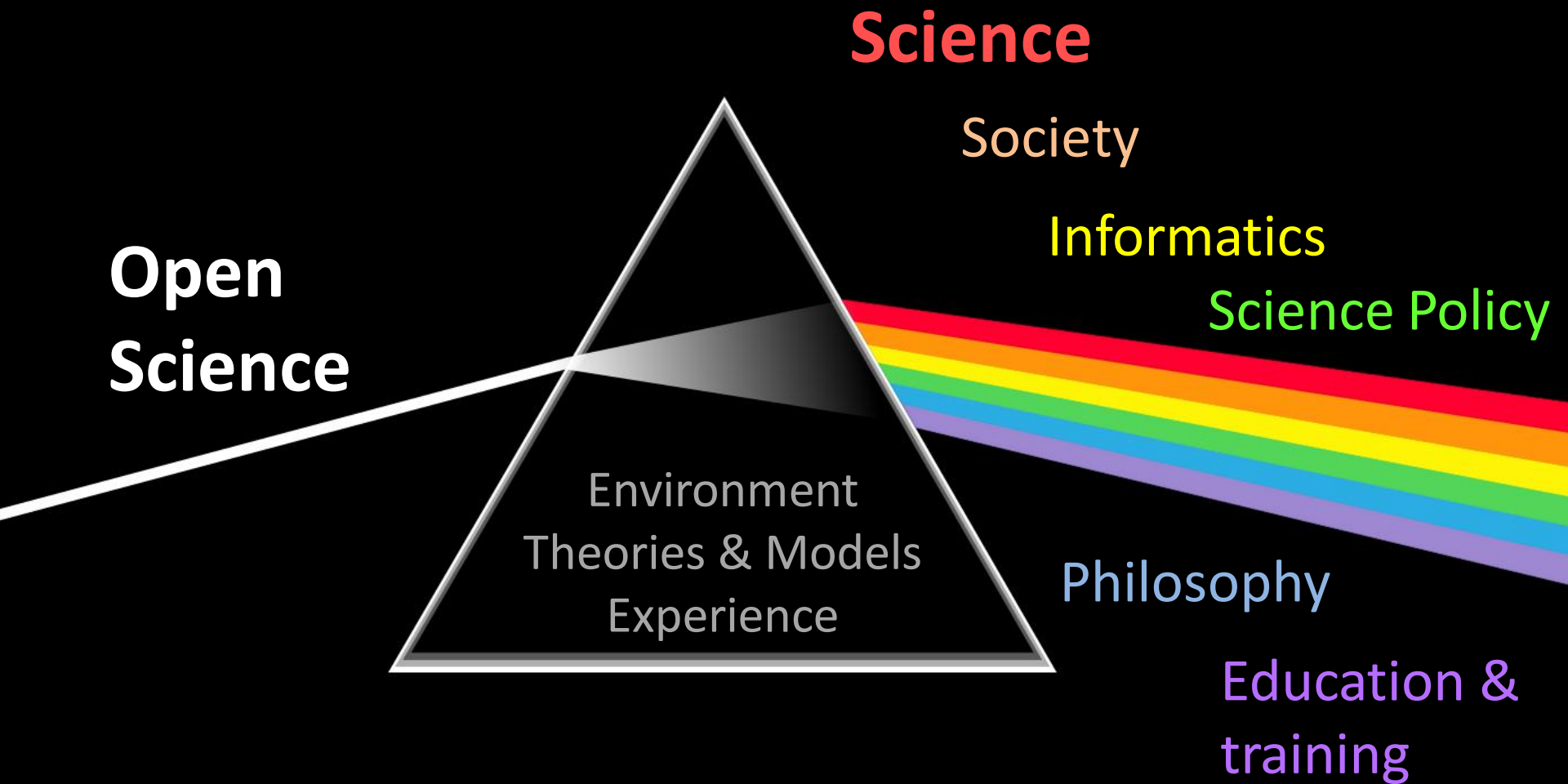
2012

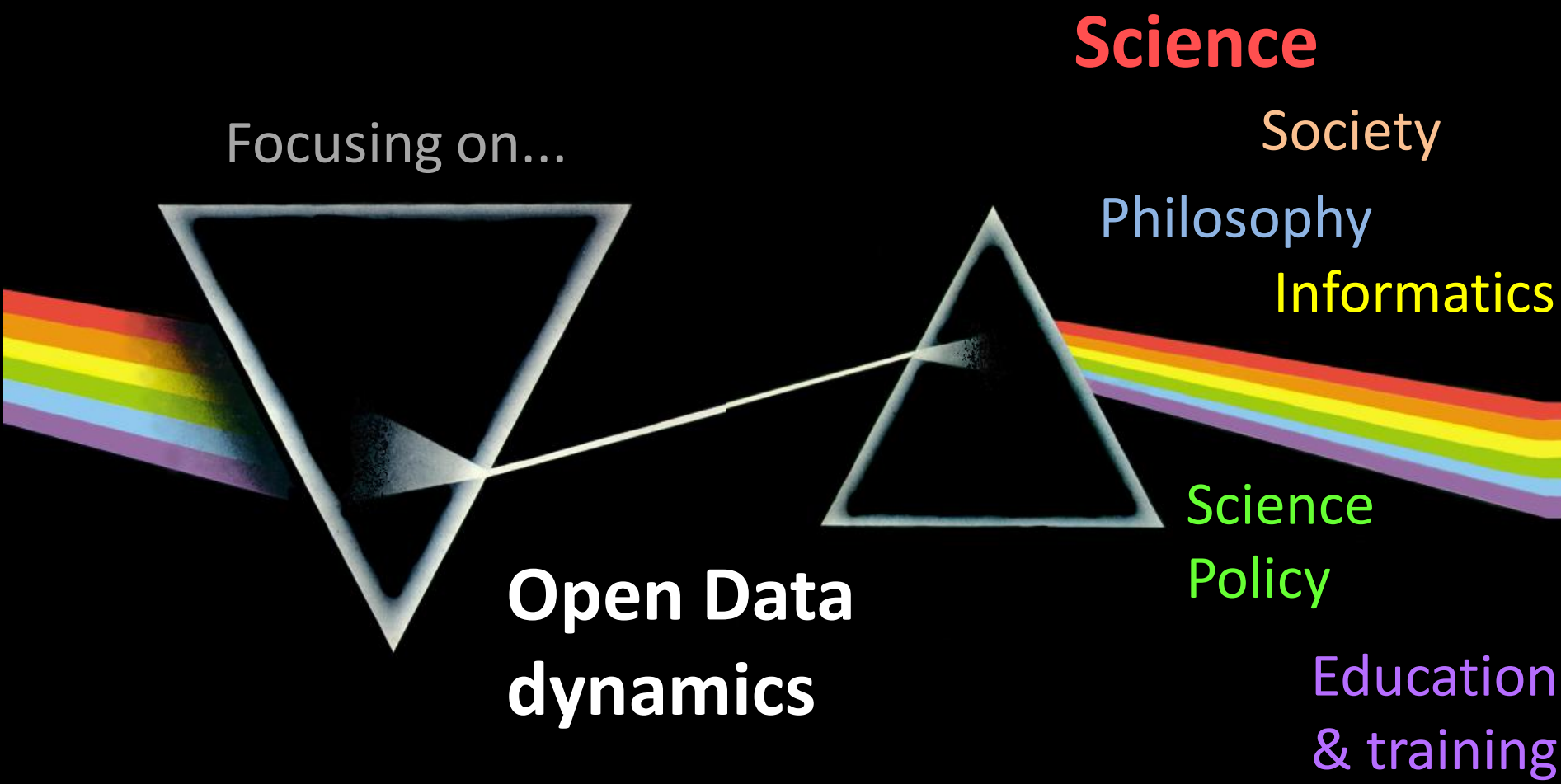
ADVANCES IN INFORMATION SCIENCE

The **Conundrum** of Sharing Research Data

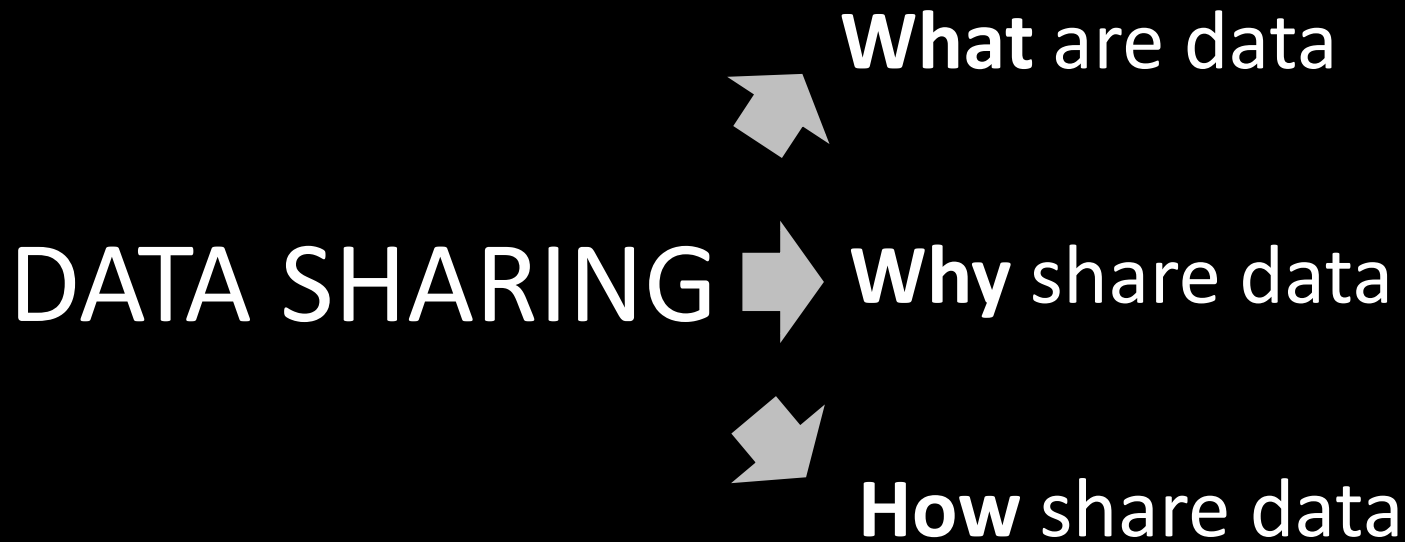
Christine L. Borgman
UCLA Department of Information Studies, GSEIS Bldg, Rm. 235, Box 951520, Los Angeles, California
90095-1520. E-mail: borgman@gseis.ucla.edu

doi 10.1002/asi.22634

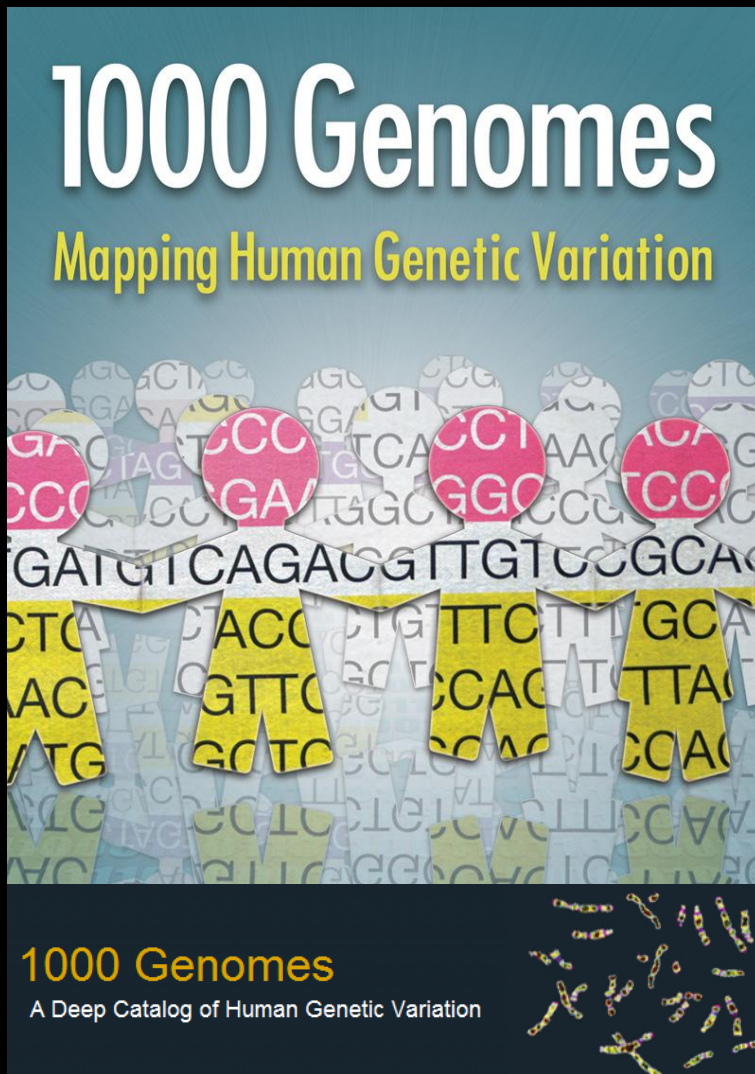




3. ... complexity?



Big Science vs Small Science



<http://www.1000genomes.org/>



What are data?

	BIG SCIENCE Genomics	SMALL SCIENCE Genetics
Data type	derivative of experimental data	
Data replicates	yes (~ 4 – 50)	not necessary
Bioinformatic processing	Intensive	limited
Data amount	large – huge (~GB)	small-large (~MB)

Why share data (...or withhold)

	BIG SCIENCE Genomics	SMALL SCIENCE Genetics	
Human and economic resources available for DS	large	small	SP
Expectations from funders and Journals for DS	high	low	SP
Research environment (competitiveness)	high	low	SO
Exploitation of data produced	rapid, may be not exhaustive	slow, usually more complete	SP SO
Research environment (education & training for DS)	high	low	ET
Potential for <u>data reuse</u>	high	low	IN ET

How share data

	BIG SCIENCE Genomics	SMALL SCIENCE Genetics	
Accessibility			
Online primary databases	available	available*	
Other sharing modes (text, downloadable files, supplementary material)	Not useable	Not infrequently used	ET
Assessability			
Online databases with quality control	not available	available	SP
Useability			
Metadata standards	not available	not available	SP
Intelligibility			
Societal engagement	not often (freezer anthropology)	sometimes	PH

4. Three simple questions

Genetics vs Genomics: Who share more?

How DS rates compare with other research fields?

Are there examples of good practice?

Empirical approaches



Paper scrutiny

[Public Availability of Published Research Data in High-Impact Journals](#)
Abdul K. Mulla^{1,2,3*}, Roger Gandy⁴, Mousa K. Alkhatib⁵, John B. A. Iqbal^{6,7,8,9}

[Empirical Study of Data Sharing by Authors Publishing in PLoS Journals](#)
Caroline J. Savage, Andrew J. Valleron

[Who Shares? Who Doesn't? Factors Associated with Openly Archiving Raw Research Data](#)
Markus A. Blomker

[CONCEPTS AND METHODS](#)
Big data and the future of ecology
Dimitris S. Hountalas¹, Gabe A. Street², Joshua J. Tonks³, Wendy S. Gilmer⁴, Ashli E. Fielden⁵, Walter L. Branchini⁶, Michael S. Drake⁷, and John H. Brown



Questionnaire based surveys

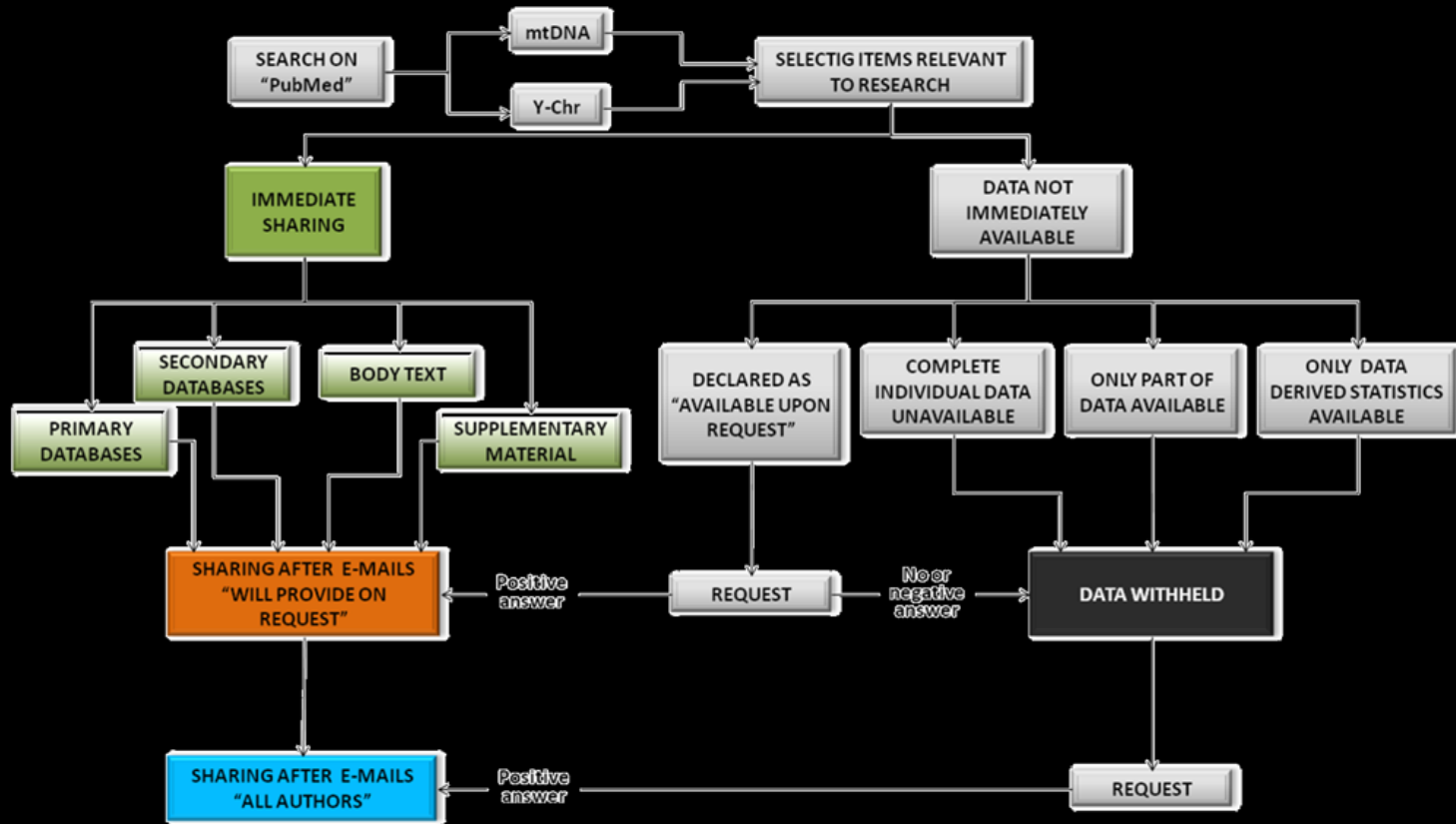
[Empirical Study of Data Sharing by Authors Publishing in PLoS Journals](#)
Caroline J. Savage, Andrew J. Valleron

[Data Withholding in Academic Genetics: Evidence From a National Survey](#)

[The world's most ecobiodiversity data sharing — Identifying best practices and requirements to enable a sustainable use of research data](#)

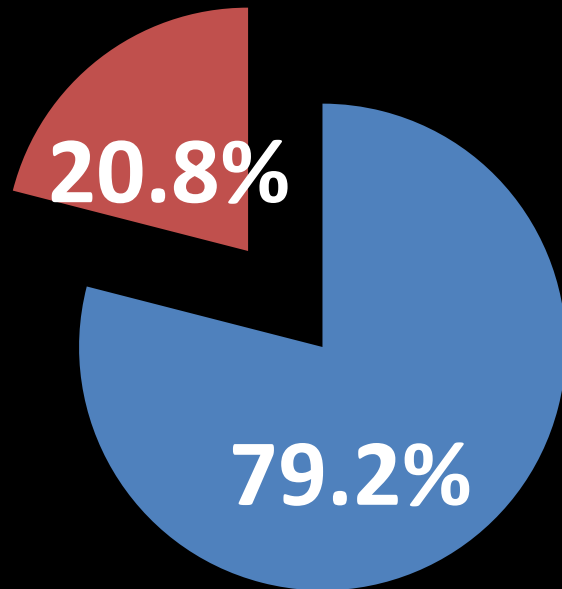
[LETTER](#)
Willing or unwilling to share primary biodiversity data: results and implications of an international survey
Xiaohu Zhang^{1,2}, Bradford A. Squires³, Jason LaF⁴, Guy L. Miller⁵, Colin Garver⁶, Xinyi Zhang⁷, & Shih-Ding

Paper scrutiny



Milia et al. 2012. doi 0.1371/journal.pone.0037552

Q1 - Who share more?

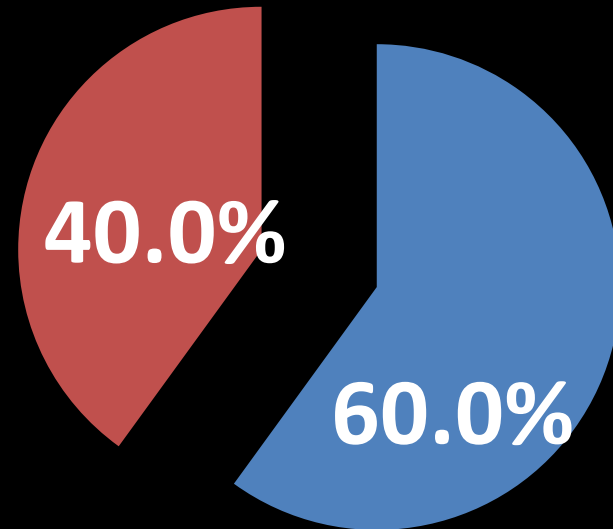


Genetics*

*Milia et al. 2012.

doi 0.1371/journal.pone.0037552

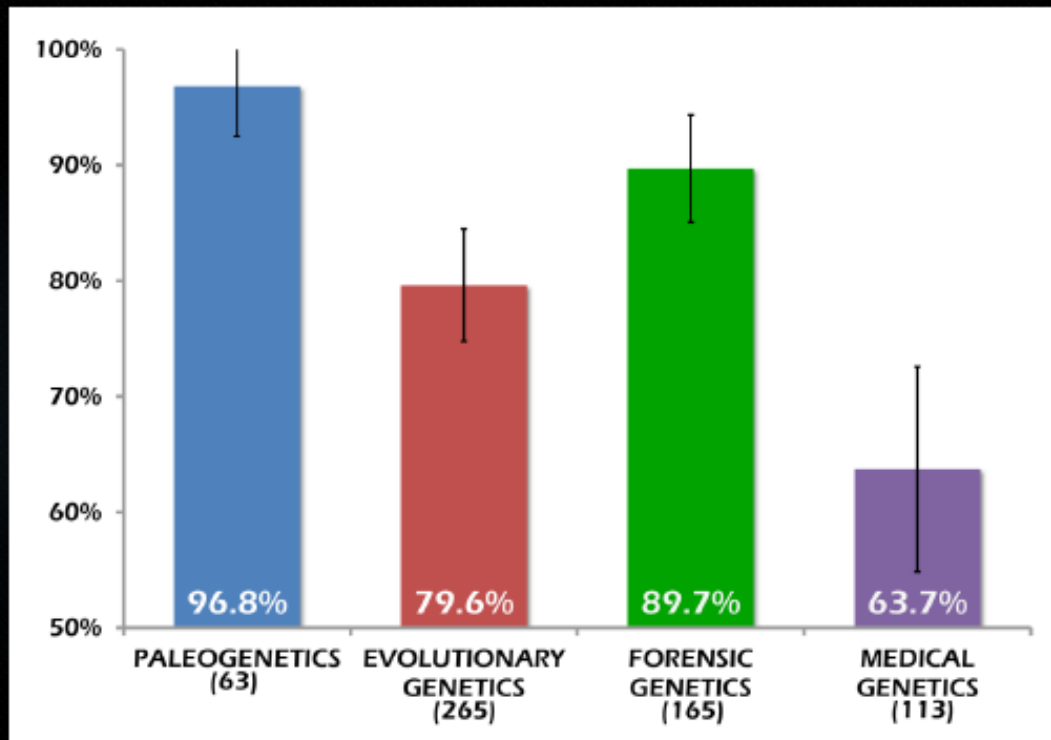
vs



Genomics**

** preliminary data

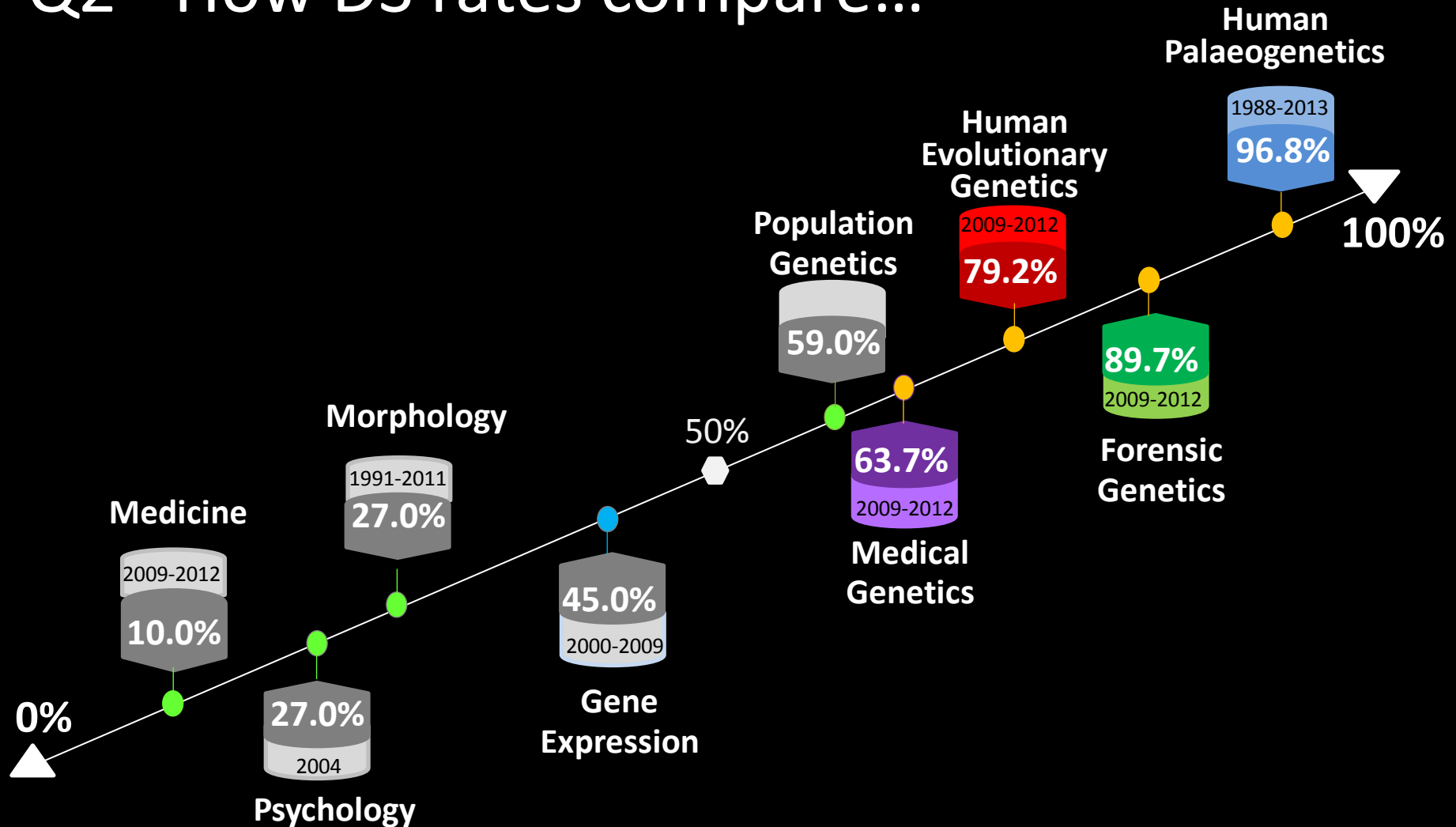
Q2 - How DS rates compare...



Sharing rates in papers concerning mitochondrial and Y chromosomal polymorphisms in humans. All papers were indexed in Medline from 1/1/2008 to 31/12/2011.

Anagnostou et al. 2015. doi:10.1371/journal.pone.0121409

Q2 - How DS rates compare...



Data sharing rates
inferred from

● e-mail requests to authors

● Availability through online
databases

● Scrutiny of scientific
publications

Q3 - good practice: Forensic Genetics

Forensic Science International: Genetics 19 (2015) 56–67



Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig



Forensic ancestry analysis with two capillary electrophoresis ancestry informative marker (AIM) panels: Results of a collaborative EDNAP exercise



C. Santos^{a,1}, M. Fondevila^{a,1},
W. Branicki^{f,2}, F. Brisighelli^g,
V. Decroyer^l, R. England^m, K.
P. Hoff-Olsen^r, A. Hoffmann^f,
C. McGovern^m, N. Morling^{e,1},
A. Roseth^r, P.M. Schneider^{a,1},
J.E. Templeton^s, M. Turanská
EUROFORGEN-NoE Consorti



Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig



Letter to the Editor

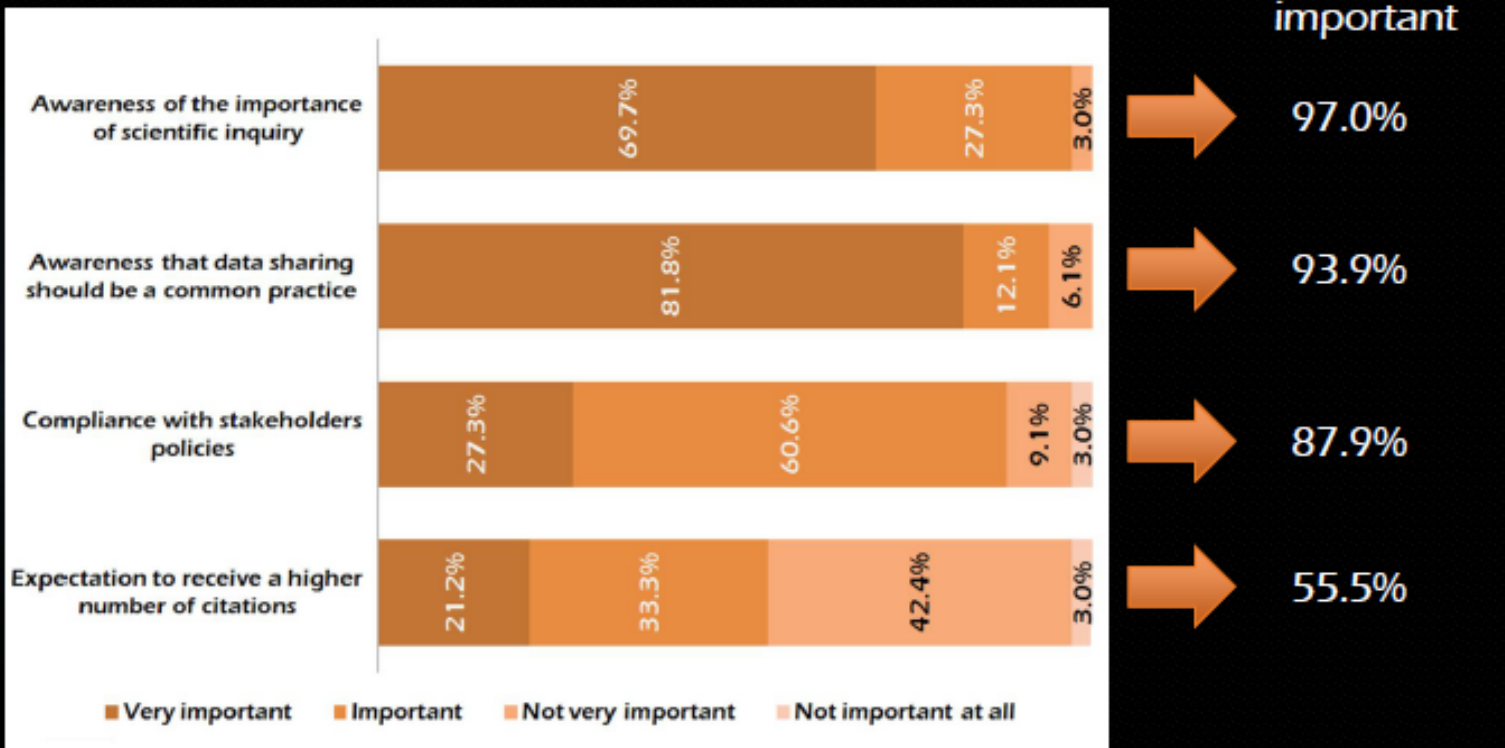
Research data sharing: Lessons from forensic genetics

Anagnostou et al. 2013. doi:10.1016/j.fsigen.2013.07.012

Q3 - good practice: Paleogenetics

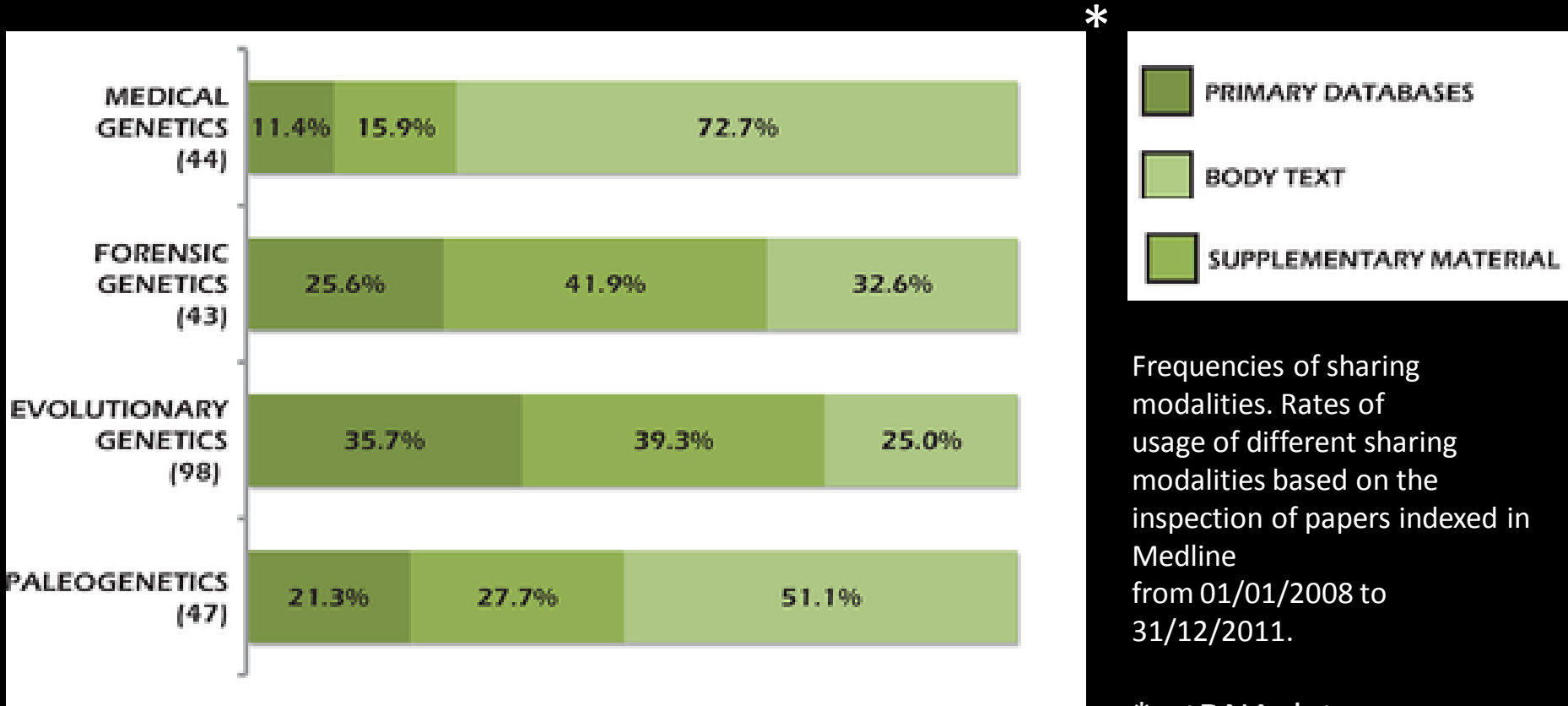
“Focusing on your overall publication experience, what is the contribution of the following factors to your choice of sharing ancient human DNA data?”

33 respondents, 24.0% of the total sample



Anagnostou et al. 2015. doi:10.1371/journal.pone.0121409

Q3 - good practice: Paleogenetics



Anagnostou et al. 2015. doi:10.1371/journal.pone.0121409

Frequencies of sharing modalities. Rates of usage of different sharing modalities based on the inspection of papers indexed in Medline from 01/01/2008 to 31/12/2011.

...to conclude

1.

Understanding data sharing dynamics (DSDs) in
Human evolutionary Genetics /Genomics and
closely related fields

- working on DSDs is not simply a
methodological/informatic issue
- DSDs are driven by social, cultural and
political factors

...to conclude

2.

Examples of **good practice**

- Forensic Genetics (collaborative studies)
- Paleogenetics (awareness)

...to conclude

3.

What **AISA** can do

→ foster empirical studies

→ foster educational activities

Acknowledgements

Paolo Anagnostou (Sapienza Università di Roma)

Marco Capocasa (Istituto Italiano di Antropologia)

Nicola Milia (Università di Cagliari)