The Path to Building Consensus: Can good minimum information checklists be created by the community?

Dawn Field

NERC Centre for Ecology and Hydrology &
the University of Oxford e-Research Centre

Steps toward a standard of Minimum Information About a Phylogenetic Analysis (MIAPA)

Hilmar Lapp, Nico Cellinese, Jim Leebens-Mack, Enrico Pontelli, Arlin Stoltzfus

ABSTRACT Many **phylogenetic analysis results** are published in ways that present serious barriers to their reuse in numerous research applications that would stand to benefit from them. While some of these barriers are well understood, such as issues with adherence to standard exchange formats, those centering on the associated **metadata** necessary for researchers to evaluate or reuse a published phylogeny have only recently begun to be articulated. One of the critical next steps towards formalizing these metadata requirements as a **minimum reporting standard** is to convene meetings of key stakeholder communities with the goal to identify information attributes necessary and desirable for facilitating reuse, and to **build consensus** on their priority. To this end, we are holding a workshop at the 2011 Biodiversity Information Standards (TDWG) Conference to determine how a future reporting standard for phylogenetic analyses can best serve biodiversity science and related research applications. We invite **all interested colleagues to participate**.

What are Standards?

- A standard is a convention that gives uniformity to an area of research or innovation.
- Standards unite groups and enable collective change.
- Standards provide the language in which innovation is written.

standards

Principles:

Not everything should be ‘standardized’

Aggregation of data, information, and knowledge requires standard ways of doing things

Standards provide foundations; Standards should drive innovation (think of electrical plugs or the internet)

Pick the right concepts to standardize – at the right time, with the right people

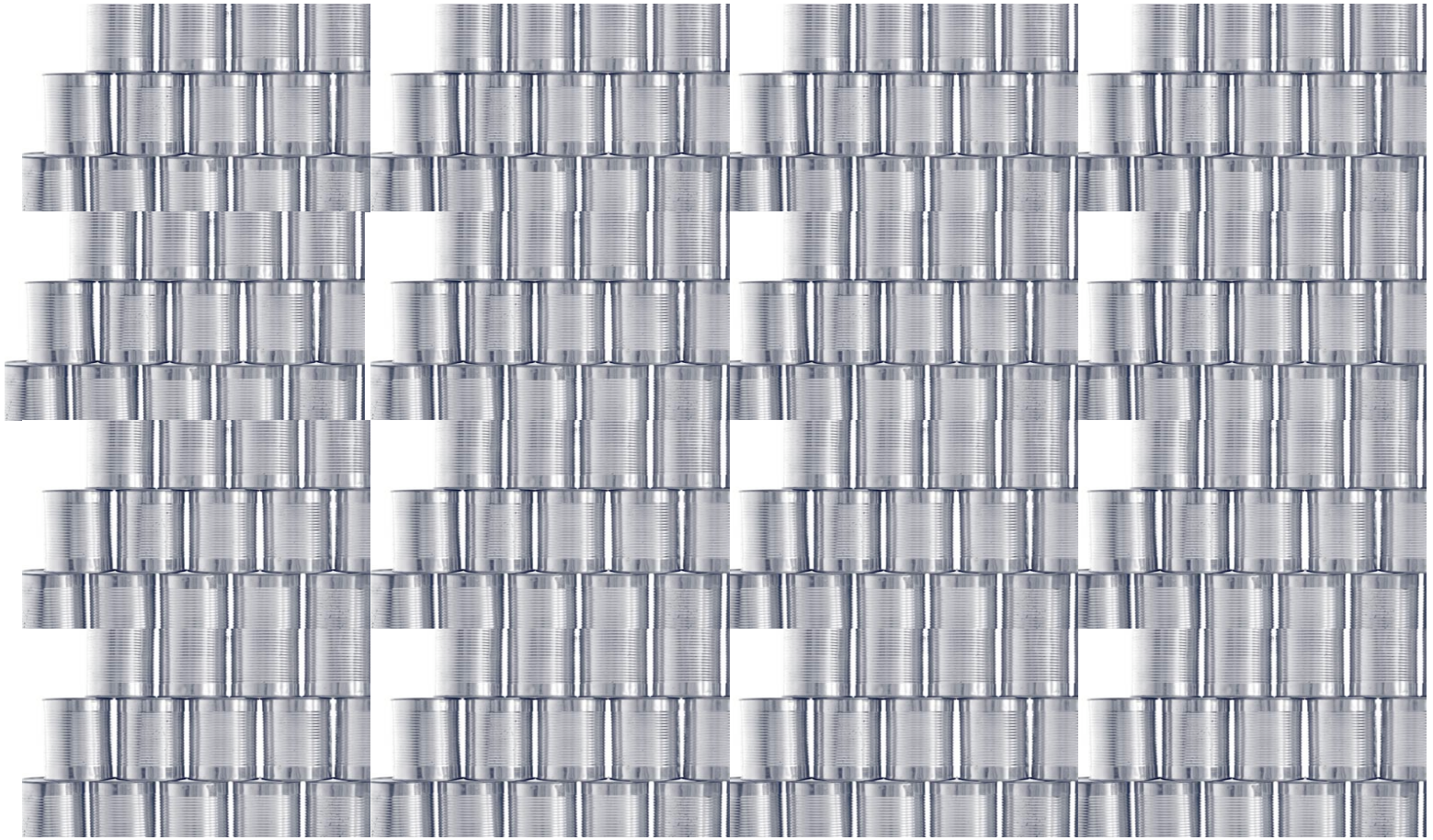
Requires good ‘group think’ – or ‘systems thinking’

SuperMarket

NEBC Envo genomic gsc STANDARDS consortium genomic observatories isainfrastructure MIBBI biosharing NETWORK

DataMarket

Norman Morrison



NEBC Envo gsc genomic STANDARDS consortium genomic observatories isainfrastructure MIBBI biosharing NETWORK

Packaging data

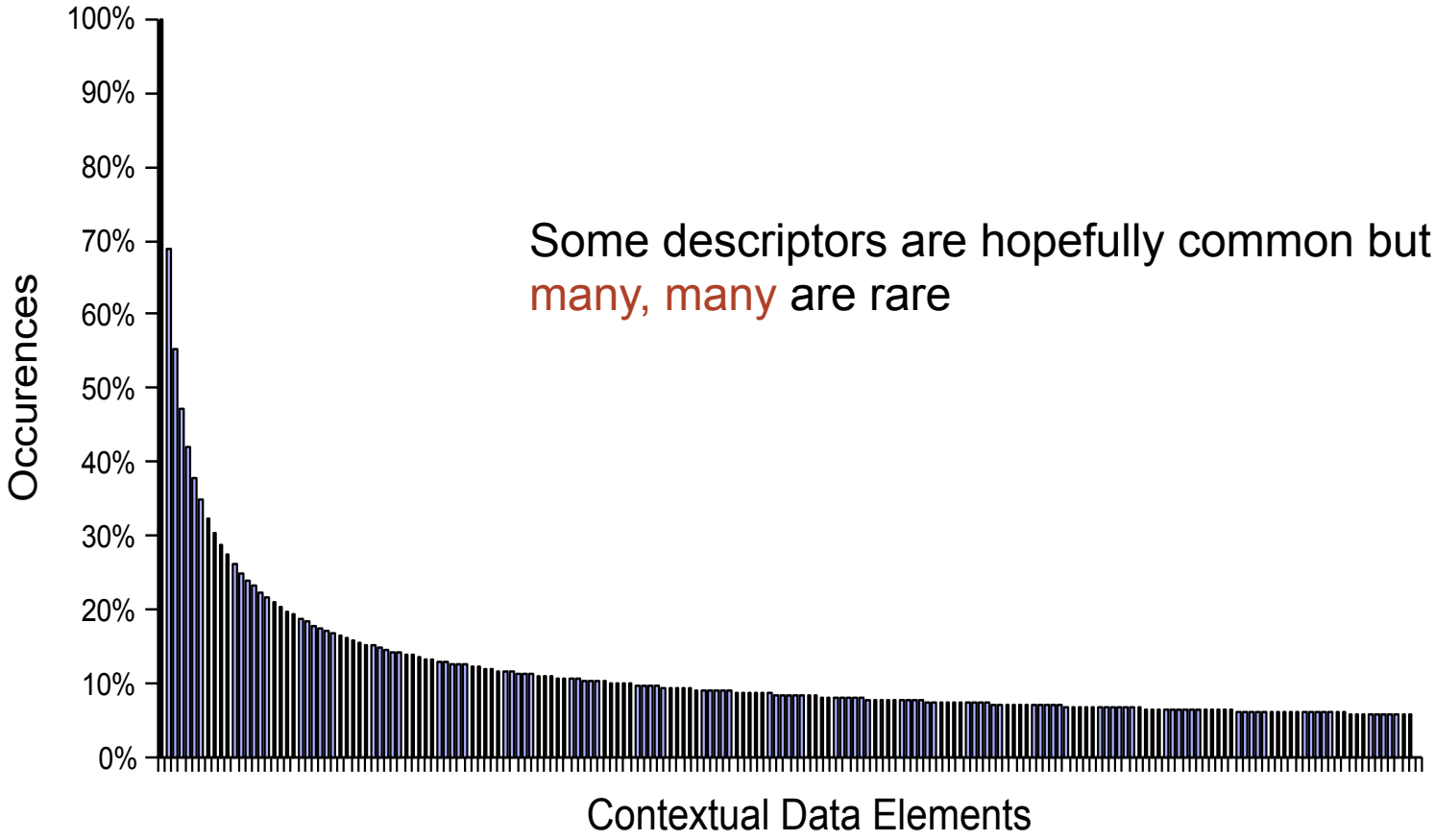
NEBC Envo gsc genomic STANDARDS consortium genomic observatories isainfrastructure MIBBI biosharing NETWORK

Labels for data

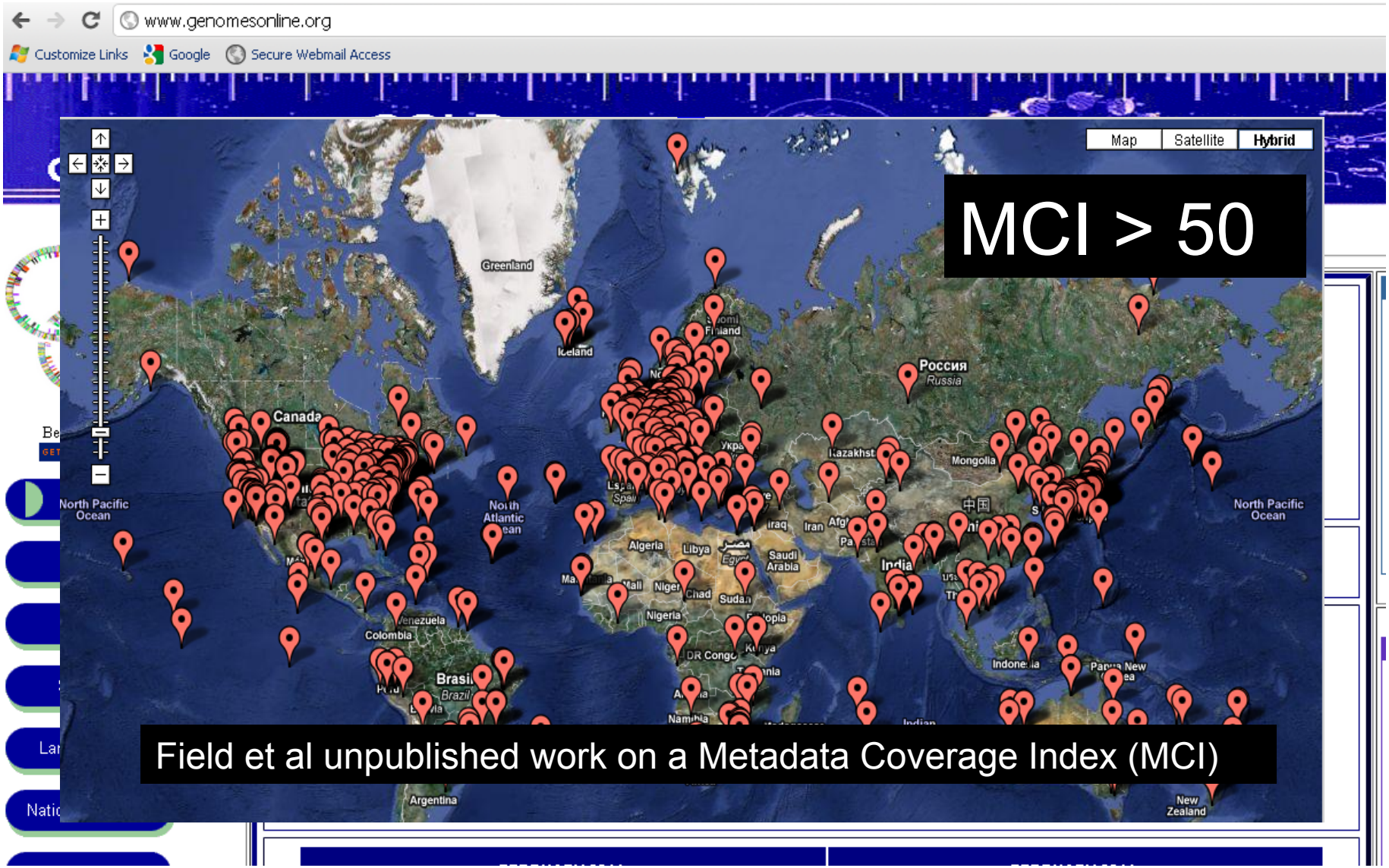
What does your metadata look like?

- How easily will you be able to build consensus?
- If you polled everyone in the room for the top 5-10 descriptors how many would be the same?
- Surveys are helpful
- When you start to reach consensus, polling new people largely yeilds the same results
- Can we make this an objective process?

Truth: Prevalence of descriptors



Some descriptors are hopefully common but many, many are rare



NEBC Envo genomic gsc STANDARDS consortium genomic observatories isainfrastructure MIBBI biosharing NETWORK

Community-driven solutions:

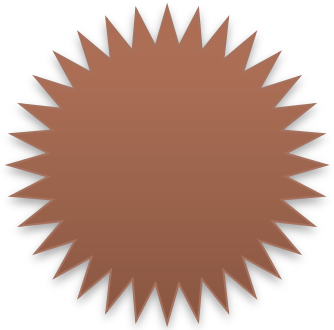
The Common Path:

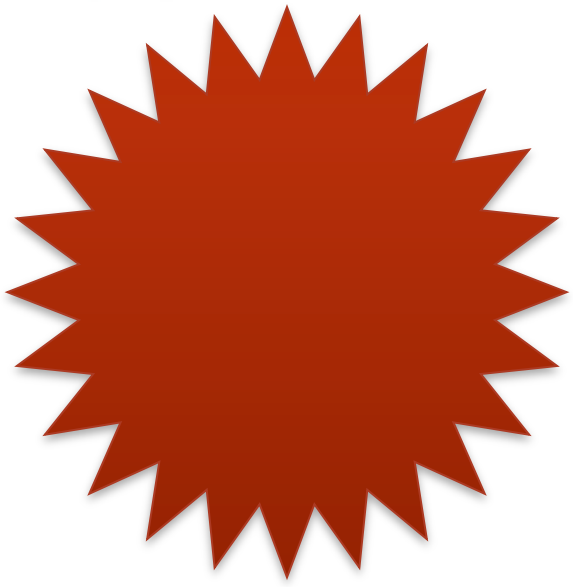
- Identify the problem
- Define a community to address it
- Define scope of the solution
- Implement solution
- Gain adoption of solution

Progression of ideas

- Pre-meeting thoughts – wish lists
- Beer discussions
- Good suggested concepts
- Working groups – with at least one champion
- Formal launch at a meeting – ideally, recorded in meeting report publication
- Actively producing outputs (papers, standards, tools)
- Funded (workshops, development)
- Changing the landscape of your field
- Changing the landscape of **science**

Joining up Communities





The Genomic Standards Consortium

Innovation through Collaboration

- Established in September 2005
- Governed by a Board
- 100+ members
- Representatives from INSDC (EMBL/DDBJ/GenBank), EBI, Sanger, JGI, JCVI, GOLD, CAMERA, MG-Rast, RDP, Silva, Greengenes, VAMPS, and many more
- Open membership defined by participation

GSC 10
Argonne, 2010

GSC 11,
Hinxton, 2011

GSC 12
Bremen, 2011

GSC 13
BGI 2012

nature biotechnology

PERSPECTIVE

PERSPECTIVE

nature biotechnology

The minimu (MIGS) spec

Dawn Field*[1], George Gar
Nicholas Thomson[8], Mich
Sandra Baldauf[12], Stuart B
Claude dePamphilis[18], Rol
Frank Oliver Glöckner[23], F
Henning Hermjakob[6], Chr
Jessie Kennedy[27], George F
Jim Leebens-Mack[33], Suza
Victor Markowitz[37], Jenni
Julian Parkhill[8], Lita Proct
Paul Swift[1], Chris Taylor[6],
Naomi Ward[45], Trish Whe

Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications

Pelin Yilmaz[1,2*], Renzo Kottmann[1], Dawn Field[3], Rob Knight[4,5], James R Cole[6,7], Linda Amaral-Zettler[8], Jack A Gilbert[9–11], Ilene Karsch-Mizrachi[12], Anjanette Johnston[12], Guy Cochrane[13], Robert Vaughan[13], Christopher Hunter[13], Joonhong Park[14], Norman Morrison[3,15], Philippe Rocca-Serra[16], Peter Sterk[3], Manimozhiyan Arumugam[17], Mark Bailey[3], Laura Baumgartner[18], Bruce W Birren[19], Martin J Blaser[20], Vivien Bonazzi[21], Tim Booth[3], Peer Bork[17], Frederic D Bushman[22], Pier Luigi Buttigieg[1,2], Patrick S G Chain[7,23,24], Emily Charlson[22], Elizabeth K Costello[4], Heather Huot-Creasy[25], Peter Dawyndt[26], Todd DeSantis[27], Noah Fierer[28], Jed A Fuhrman[29], Rachel E Gallery[30], Dirk Gevers[19], Richard A Gibbs[31,32], Inigo San Gil[33], Antonio Gonzalez[34], Jeffrey I Gordon[35], Robert Guralnick[28,36], Wolfgang Hankeln[1,2], Sarah Highlander[31,37], Philip Hugenholtz[38], Janet Jansson[23,39], Andrew L Kau[35], Scott T Kelley[40], Jerry Kennedy[4], Dan Knights[34], Omry Koren[41], Justin Kuczynski[18], Nikos Kyrpides[23], Robert Larsen[4], Christian L Lauber[42], Teresa Legg[28], Ruth E Ley[41], Catherine A Lozupone[4], Wolfgang Ludwig[43], Donna Lyons[42], Eamonn Maguire[16], Barbara A Methé[44], Folker Meyer[10], Brian Muegge[35], Sara Nakielny[4], Karen E Nelson[44], Diana Nemergut[45], Josh D Neufeld[46], Lindsay K Newbold[3], Anna E Oliver[3], Norman R Pace[18], Giriprakash Palanisamy[47], Jörg Peplies[48], Joseph Petrosino[31,37], Lita Proctor[21], Elmar Pruesse[1,2], Christian Quast[1], Jeroen Raes[49], Sujeevan Ratnasingham[50], Jacques Ravel[25], David A Relman[51,52], Susanna Assunta-Sansone[16], Patrick D Schloss[53], Lynn Schriml[25], Rohini Sinha[22], Michelle I Smith[35], Erica Sodergren[54], Aymé Spor[41], Jesse Stombaugh[4], James M Tiedje[7], Doyle V Ward[19], George M Weinstock[54], Doug Wendel[4], Owen White[25], Andrew Whiteley[3], Andreas Wilke[10]

_computational BIOLOGY

nature biotechnology (2008) Vol 26 No 8

COMMENTARY

Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project

Chris F Taylor[*1,2], Dawn Field[2,3], Susanna-Assunta Sansone[1,2], Jan Aerts[4], Rolf Apweiler[1], Michael Ashburner[5], Catherine A Ball[6], Pierre-Alain Binz[7,8], Molly Bogue[9], Tim Booth[2], Alvis Brazma[1], Ryan R Brinkman[10], Adam Michael Clark[11], Eric W Deutsch[12], Oliver Fiehn[13], Jennifer Fostel[14], Peter Ghazal[15], Frank Gibson[16], Tanya Gray[2,3], Graeme Grimes[15], John M Hancock[17], Nigel W Hardy[18], Henning Hermjakob[1], Randall K Julian Jr[19], Matthew Kane[20], Carsten Kettner[21], Christopher Kinsinger[22], Eugene Kolker[23,24], Martin Kuiper[25], Nicolas Le Novère[1], Jim Leebens-Mack[26], Suzanna E Lewis[27], Phillip Lord[16], Ann-Marie Mallon[17], Nishanth Marthandan[28], Hiroshi Masuya[29], Ruth McNally[30], Alexander Mehrle[31], Norman Morrison[2,32], Sandra Orchard[1], John Quackenbush[33], James M Reecy[34], Donald G Robertson[35], Philippe Rocca-Serra[1,36], Henry Rodriguez[22], Heiko Rosenfelder[31], Javier Santoyo-Lopez[15], Richard H Scheuermann[28], Daniel Schober[1], Barry Smith[37], Jason Snape[38], Christian J Stoeckert Jr[39], Keith Tipton[40], Peter Sterk[1], Andreas Untergasser[41], Jo Vandesompele[42] & Stefan Wiemann[31]

mibbi
http://mibbi.org

BBSRC bioscience for the future

NATURAL ENVIRONMENT RESEARCH COUNCIL

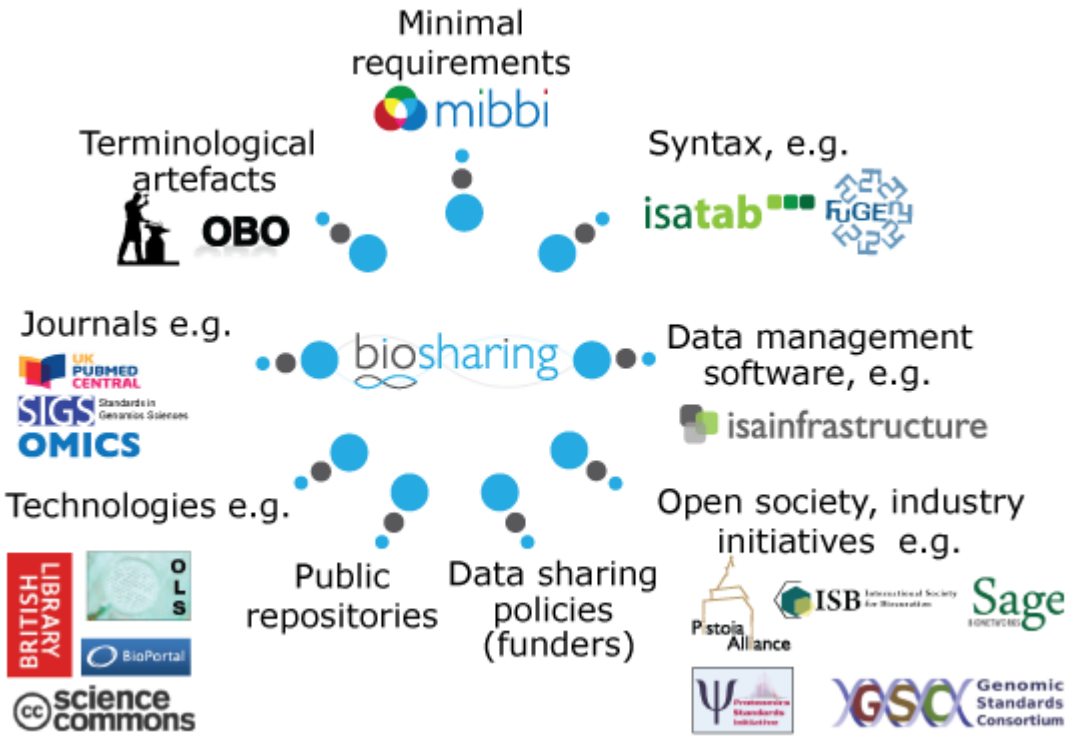
POLICYFORUM

http://biosharing.org

MEGASCIENCE

'Omics Data Sharing

Dawn Field,[1*†‡] Susanna-Assunta Sansone,[1,2†] Amanda Collis,[3†] Tim Booth,[1] Peter Dukes,[4] Susan K. Gregurick,[5] Karen Kennedy,[6] Patrik Kolar,[7] Eugene Kolker,[8] Mary Maxon,[9] Siân Millard,[10] Alexis-Michel Mugabushaka,[11] Nicola Perrin,[12] Jacques E. Remacle,[7] Karin Remington,[13] Philippe Rocca-Serra,[12] Chris F. Taylor,[12] Mark Thorley,[14] Bela Tiwari,[1] John Wilbanks[15]



Data sharing, and the good annotation practices it depends on, must become part of the fabric of daily research for researchers and funders.

[1]U.K. Natural Environment Research Council (NERC), Environmental Bioinformatics Centre. [2]European Molecular Biology Laboratory (EMBL) Outstation, The European Bioinformatics Institute (EBI). [3]U.K. Biotechnology and Biological Sciences Research Council. [4]U.K. Medical Research Council. [5]U.S. Department of Energy. [6]Genome Canada. [7]Unit for Genomics and Systems Biology, European Commission. [8]Seattle Childrens Hospital. [9]Marine Microbiology Initiative, Gordon and Betty Moore Foundation. [10]U.K. Economic and Social Research Council. [11]European Science Foundation. [12]The Wellcome Trust. [13]U.S. National Institute of General Medical Sciences, NIH. [14]NERC. [15]Science Commons.
* The first three authors contributed equally to this article.

NEBC Envo genomic gsc STANDARDS consortium genomic observatories isainfrastructure MIBBI biosharing NETWORK

biosharing.org

Standards, Policies and Communication

www.biosharing.org

combine computational modelling in biology network

BioMed Central The Open Access Publisher

F1000 FACULTY OF 1000

equator network

(GIGA)n SCIENCE

nature precedings Pre-publication research and preliminary findings

BioSharing work
data sharing poli
the communicat
sharing of high-t

We work with c
1. develop ca
- enrich
seeking
2. moderate a
- promo
policies

BRITISH LIBRARY

UK PUBMED CENTRAL

mibbi

OBO

isatools

genomic standards consortium

Proteomics Standards Initiative

ISB International Society for Biocuration

BioPortal

POLICIES

A catalogue of data preservation, management and sharing policies from international funding agencies and regulators.

OLS

DRYAD

Sage BIONETWORKS

STANDA

formats te

A catalogue of
(minimum repo
formats and te
organizations t

Pistoia Alliance

science commons

FuGE

NIF

BioCatalogue

NCRI INFORMATICS INITIATIVE

THE NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY

in partnership with

BioMed Central The Open Access Publisher

BioMed Central and the journal BMC Research Notes proudly support BioSharing and its catalogue of domain-specific data standards.

developed by

the same people who brought you

isatools mibbi

follow us

twitter facebook rss feed

NEBC

biosharing NETWORK

A CATALOGUE OF STANDARDS

You can **sort** columns and **browse** the reporting guidelines content, or you can view ***all the standards***, or terminological artifacts or exchange formats only; or go back to the **catalogue main page**.

ACRONYM	FULL NAME	TYPE	DOMAIN	VERSION	PUBLICATION	CONTACT
BioPAX	Biological Pathway Exchange	exchange format	biological pathway	Level 3	Demir et al; Nat Biotech; 2010	BioPAX community
CellML	Cell Markup Language	exchange format	cell modelling	v 1.1	Cu… Sim…	…y
SBML	System Biology Markup Language	exchange format	computational modelling (biochemical reaction networks)	level 3, v 1 core	Huc… Bio… 200…	
FuGE	Functional Genomics Experiment Markup Language	exchange format	experimental description	v 1.0	Jon… Bio…	
ISA-Tab	Investigation/Study/Assay Tabular	exchange format	experimental description	v 1.0	Roc… Bio… 201…	
MINiML	MIAME Notation in Markup Language	exchange format	experimental description (functional genomics)	v 1.0		
SOFT						
GCDML						
MAGE-Tab						
GelML						
mzML						
MIABE						
MIPFE						
BioCoreDB						
GIATE	About Therapy Experiments	guideline	experiments	not specified	Eng Des Sel; 2009	Antibody Society
MIRIAM	Minimal Information Required In the Annotation of biochemical Models	reporting guideline	computational modeling	not specified	Le Novère et al; Nature Biotech; 2005	BioModels.net

SBML

System Biology Markup Language

ID
bsg-000052

TYPE
exchange format

DOMAIN(S) COVERED
computational modelling (biochemical reaction networks)

PUBLICATION(S)
Hucka et al; Bioinformatics; 2003

ORGANIZATION
SMBL community

MAIN CONTACT(S)

Like

Tweet 0

» Login or register to post comments

Escalating number of standardization efforts in bioscience, *e.g.:*

formats

terminologies

guidelines

To exploit fully the promise of these data we need both scientific innovation and community agreement on how to provide appropriate stewardship of these resources for the benefit of all.

Requires the evolution of our scientific, technological and sociological thinking….

Conclusions

- The era of real data sharing is just beginning…
- Self-organization by the scientific community can pay dividends (i.e. consensus building, large-scale co-ordination)
  - Standards are keys to unlocking data
  - Group thinking overcomes the tragedy of the commons
- Many communities and 'solutions' to learn from and work with