



BioSharing workshop

Bringing catalogues of bio-resource and standards together

June 8th and 9th 2011

Location

University of Oxford, Oxford e-Research Centre,
7 Keble Rd, Oxford, OX1 3QG, UK
(*Conference Room*)



*Final agenda, presentations and final summary are in the
Nature Precedings' BioSharing Collection:*

<http://precedings.nature.com/collections/biosharing-meetings>

Lodging and meals sponsored by:



INTRODUCTION

This workshop brings together representatives of groups developing catalogues of bio-resources and its aims are twofold:

1. Ensure that **BioSharing**¹, a catalogue of minimal information checklists, terminologies and exchange formats (hereafter called standards) and policies, complements and links to existing catalogues of tools and databases, and also to publications and related material;
2. Outline the technical implementations of **bioDBcore** checklist², the proposed uniform system for describing the catalogues of tools and databases, in particular, how to ensure (bi-directional) link to the BioSharing catalogue.

Which are the challenges we need to address?

Funding agencies, journals and the research community continue to participate in the development of policies, standards, tools and databases to support data sharing, as they must if it is to become practicable and commonplace³. Proliferation of resources is a positive sign of stakeholders' engagement, but how much do we know about a standard, for example? Or which one is mature and stable enough to be used or recommend by a policy? Which tools and databases implement which standard? Which tools, databases and standards are suited for a certain data type or field of study? What functionalities each tool, database provides users with?

Several efforts have emerged to make these resources more discoverable, by registering and cataloguing tools, databases, often also publications and other educational material^{4,5,6,7,8,9,10}. However, due to the diverse motivation behind their development, their coverage, target users and functionalities, these bio-resources often duplicate or lack of core information, for example which standard and policy the tools, databases and publications comply to.

Which steps should we take at the workshop?

The bioDBcore checklist already outlines the core set of information these resources should represent; it is now time to move to the implementation phase (aim 2, above), ensuring that these resources also 'cross reference' to the BioSharing catalogue when describing which standard and policy the tools, databases and publications comply to (aim 1, above).

This fits perfectly with the objectives of the BioSharing catalogue, that in partnership with BioMedCentral¹⁰ and jointly with Nature Precedings¹¹ is set to:

1. centralize bioscience standards and policies;
2. link to other portals and open access resources and lists of tools and databases implementing the standards;
3. develop a set of criteria for assessing the quality of the standards, including their uptake by tools, databases, journals and funders;
4. foster interoperability, addressing overlaps and duplication of efforts that hamper their wider uptake and interfere with the implementations of policies and the creation of standards-compliant systems.

To enable such 'interoperability', we need a mechanism for representing and exchanging the (descriptive) elements of the bioDBcore checklist. An initial RDF-based representation format has been drafted and opened for comments to the bioDBcore list¹² used to engage with any interested parties. At the workshop, we will use this draft, evaluate other formats and outline the final proposal. We will also address use of terminologies (to describe the elements of the bioDBcore checklist) and identify the mechanism to reference the BioSharing catalogue's entry. Lastly, we will outline the implementations steps each participant needs to take to 'retrofit' the bio-resources to become bioDBcore-compliant and achieve bidirectional links with the BioSharing catalogue.

What are the goals and who will benefit from our work?

Ultimately, our work will i) assist a variety of users to locate and access the information distributed in these bio-resources; ii) inform journal editors and funders, implementing data preservation, management and sharing policies, when recommending or requiring that certain standards are met and that an existing tool or database is used, and ii) guide software developers to adopt, get involved and extend a given standard or (open source) tool. The grand vision is to reduce unnecessary duplications, ease data management and enable data exchange, shielding users from unnecessary complexity.

¹ BioSharing: <http://biosharing.org>

² bioDBcore initiative, joint initiative between the International Society for Biocuration and BioSharing, : <http://biocurator.org/biodbcore.shtml>

³ Field, Sansone *et al.* 'Omics Data Sharing, *Science* 9, 234 (2009).

⁴ Casimir DDF: http://casimir1.pdn.cam.ac.uk/casimir_ddf/

⁵ NIF registry: <http://neuinfo.org/nif/nifgw.html?tab=registry>

⁶ MIRIAM: <http://www.ebi.ac.uk/miriam/main>

⁷ Biositemaps Resource Discovery System: <http://biositemaps.ncbcs.org/rds/index.html>

⁸ Bioinformatics Links Directory: http://www.bioinformatics.ca/links_directory

⁹ BioCatalogue: <http://www.biocatalogue.org/>

¹⁰ BioMedCentral: http://blogs.openaccesscentral.com/blogs/bmcblog/entry/bmc_research_notes_and_biosharing

¹¹ Nature Precedings: <http://precedings.nature.com/>

¹² biodbcore@googlegroups.com

OUTCOMES

General:

- Sharing experiences on collecting and cataloguing resources for the bioscience community
- Strengthen the relations between the resources represented at the workshop

Specific deliverables:

- Finalize the RDF-based representation formats, reviewing and identifying suitable:
 - data schema¹³, unique identifiers¹³, terminologies¹³
 - mechanism for linking to BioSharing catalogue's entries
- Define a plan to implement the bioDBcore representation
- Produce examples of the bioDBcore representation with links to the BioSharing catalogue's entries
- Outline agenda for the ISMB workshop¹⁴
- Agree to publication plans
- Set actions, identify lead persons/groups, and communication strategy to ensure the work progresses

AGENDA

Tuesday 7th, June

18.30 - 20.30	WELCOME EVENT - <i>The Old Parsonage</i>
---------------	--

Wednesday 8th, June (Conference Room)

08.30	Arrival and coffee
08.45 - 09.00	Welcome - <i>Susanna-Assunta Sansone, Pascale Gaudet, Dawn Field</i>
	INTRODUCTIONS
09.00 - 09.30	Scoping the workshop
09.00 - 09.15	• Roundtable introductions - <i>All</i>
09.15 - 09.30	• Purpose of the meeting - <i>Susanna Sansone</i>
09.30 - 12.30	Participating resources Expectations and contributions towards the overall goal of the meeting (10 min + 5 min Q&A); chair, <i>Dawn Field</i>
9.30 - 9.45	• BioSharing - <i>Susanna Sansone</i>
9.45 - 10.00	• bioDBcore - <i>Pascale Gaudet</i>
10.00 - 10.15	• NIF and NeuroLex - <i>Anita Bandrowski</i>
10.15 - 10.30	• CASIMIR DDF and ENFIN - <i>John Hancock</i>
10.30 - 10.45	• MIRIAM - <i>Nick Juty</i>
10.45 - 11.00	• Biositemaps and NCBO - <i>Trish Whetzel</i>
11.00 - 11.15	BREAK
11.15 - 11.30	• LinksDirectory and DATABASE journal - <i>Francis Ouellette</i>
11.30 - 11.45	• BMC Genome Biology journal - <i>Naomi Attar</i>
11.45 - 12.00	• Elsevier journals - <i>Adriaan Klinkenberg</i>
12.00 - 12.15	• Nature PG and Nature Precedings - <i>Ruth Wilson</i>
12.15 - 12.30	• F1000 - <i>Rebecca Lawrence</i>
	• NAR - <i>Michael Galperin</i> (via skype)
12.30 - 13.30	LUNCH
	REVIEW OF EXISTING REPRESENTATIONS AND ANALYSIS OF THE REQUIREMENTS
13.30 - 14.30	Data schema Presentations and discussions; chair <i>Philippe Rocca-Serra</i>
13.30 - 14.00	• Draft work to date: DOAP, FOAF, DC, DWC based
14.00 - 14.30	• Existing data schema: NIF and Biositemaps RDFs
14.30 - 15.30	Unique identifiers Presentations and discussions; chair <i>Pascale Gaudet</i>
14.30 - 14:50	• Existing options for resources: UUID, LSRN, LSID, DOI, MIRIAMID, PURL - <i>Nick Juty</i>

¹³ Many options exist for schemas, identifiers and terminologies; often there is an unnecessary overlap and duplication of effort. Trying to harmonize them would be ideal but *outside* the scope of this workshop. Efforts should be made to identify ways for dealing with the current scenario and, where possible, select the most suitable option at the present.

¹⁴ [Submitted proposal and current agenda.](#)

14.50 - 15.10	<ul style="list-style-type: none"> IDs for persons and organizations: ORCID
15.10 - 15.30	<ul style="list-style-type: none"> Elsevier journals's perspective - <i>Adriaan Klinkenberg</i>
15.30 - 15.45	BREAK
15.45 - 16.15	Links to entry in BioSharing Reference system to standards in the BioSharing catalogue; presentation by <i>Philippe Rocca-Serra</i>
16.15 - 16.30	Terminologies Presentations and discussions; chair <i>Pascale Gaudet</i>
15.45 - 16.30	<ul style="list-style-type: none"> Existing options: BRO - <i>Trish Whetzel - 10 min</i> NIFSTD/EagleI - <i>Anita Bandrowski - 10 min</i> MIRIAM : <i>Nick Juty - 10 min</i>
SUMMARY OF THE DECISIONS	
16.30 - 17.30	Agreed points Summary of the desirable features and agreeremnts (from notes taken during the day) - <i>Susanna Sansone</i>
19.30 - 21.00	DRINKS - <i>The Old Parsonage</i>

Thursday 9th, June (*Conference Room*)

08.30 - 09.00	Arrival and coffee
MOVE TOWARDS IMPLEMENTATIONS	
09.00 - 09.30	Summary of day 1 Recap key points and tie up any loose ends; chair <i>Susanna Sansone</i>
09.30 - 11.00	Steps required for implementing the bioDBcore representation Discuss what each resources represented at the workshop is willing to do; chair <i>Pascale Gaudet</i> <ul style="list-style-type: none"> agree to one or more mechanisms for collecting the missing information, also to e.g. ensure updates evaluate the benefit and feasibility of a coordinated operation vs individual efforts
11.00 - 11.15	BREAK
COMMUNICATION AND DISSEMINATION	
11.15 - 11.45	Publication plans Discuss different options; chair <i>Dawn Field</i> <ul style="list-style-type: none"> e.g. meeting summary, open letter, specification/technical document, roadmap for continuing the engagement and involving other stakeholders
11.45 - 12.15	ISMB workshop Discuss and review and finalize the agenda; chair <i>Susanna Sansone</i>
12.15 - 12.30	Action plan to ensure the work progresses Review actions, lead persons and communication strategies (from notes taken during the day) - <i>Susanna Sansone</i>
12.30 - 13.30	LUNCH
13.30 - 17.00	Hands-on session Produce examples of the bioDBcore representation with links to BioSharing; chair <i>Philippe Rocca-Serra</i> <ul style="list-style-type: none"> test the decisions for the technical implementations by creating real examples
17.30	CLOSE OF MEETING

Organizers

Susanna-Assunta Sansone; Pascale Gaudet; Dawn Field; Philippe Rocca-Serra. *Logistics: Clementine Harris*

Meeting location and travel directions

University of Oxford, Oxford e-Research Centre, Oxford, OX1 3QG, UK. June 8th: Access Grid Room, and June 9th: Conference Room. Directions and transportation at: <http://www.oerc.ox.ac.uk/about/how-to-find-us>

Sponsored lodging

Rooms have been reserved at the *Cotswold Lodge* (<http://www.cotswoldlodgehotel.co.uk>) located within 5-10 mins walking distance from the meeting location.

Sponsored events

Coffee breaks and lunches will be provided during the meeting. On Tuesday, a welcome event will be held at 7pm, and a dinner organized on Wednesday at *The Old Parsonage*: <http://www.oldparsonage-hotel.co.uk/>

CONFIRMED PARTICIPANTS

Name	Affiliation	Group/resource represented
1. Pascale Gaudet	Swiss Institute of Bioinformatics	International Society for Biocuration
2. Susanna-Assunta Sansone	University of Oxford	BioSharing
3. Philippe Rocca-Serra	University of Oxford	BioSharing
4. Dawn Field	NERC-NEBC	BioSharing
5. Francis Ouellette	Ontario Institute for Cancer Research	a. LinksDirectory b. Associate Editor for DATABASE , the journal of Biological Databases and Curation
6. Anita Bandrowski	University of California at San Diego	a. Neuroscience information framework b. NeuroLex
7. Adriaan Klinkenberg	Elsevier Life Sciences	Publisher of biochem journals eg. BBA, J Proteomics, BBRC, FEBS Letters (SDA experiment)
8. John Hancock	TBC	a. CASIMIR , b. ENFIN
9. Nick Juty	EMBL-EBI	MIRIAM (identifiers and Resources)
10. Trish Whetzel	Stanford University	a. Biositemaps b. NCBO
11. Naomi Attar	BioMed Central	Genome Biology journal, Assistant Editor
12. Rebecca Lawrence	F1000	F1000
13. Annapaola Santarsiero	University of Oxford	BioSharing, developer
14. Eamonn Maguire	University of Oxford	BioSharing, developer
(<i>apologies</i>) Ruth Wilson	Nature Publishing Group	Nature Publishing Group
(<i>via skype</i>) Michael Galperin	NCBI, National Library of Medicine	NAR Database Issue