Pre-competitive Research Collaboration

National Academies & IOM
Washington, D.C.
February 9, 2010

Thinh Nguyen, Science Commons

What is Pre-Competitive?

- Private Property or the Commons?
  - Industry model: internalized R&D, data generation, curation, tools and analytics
  - Academic model: external funding, journal publication, public curation, open source tools, and published analytics

Changes for Industry

- Shrinking drug pipelines & new drug approvals
- Increased costs for R&D and regulatory approvals
- Reductions in budgets for internal R&D
- Lack of revenue source diversity (reliance on a few blockbuster drugs)
- Value-added shifts from data generation to data exploitation

Changes for the Commons

- Quantity of public domain data is growing rapidly
- Quality of the public domain data has improved vastly
- Public tools for data analysis and exploitation have grown more sophisticated

nature International weekly journal of science

Opinion

Nature **461**, 171-173 (10 September 2009) | doi:10.1038/461171a; Published online 9 September 2009

Post-publication sharing of data and tools

Paul N. Schofield[1], Tania Bubela[2], Thomas Weaver[3], Lili Portilla[4], Stephen D. Brown[5], John M. Hancock[5], David Einhorn[6], Glauco Tocchini-Valentini[7], Martin Hrabe de Angelis[8], Nadia Rosenthal[9] & CASIMIR Rome Meeting participants[10]

▲ Top

Despite existing guidelines on access to data and bioresources, good practice is not widespread. A meeting of mouse researchers in Rome proposes ways to promote a culture of sharing.

Sharing scientific data through publication has long underpinned the cycle of discovery and is the dominant means by which scientists earn credit for their work. More recently, technologies generating very large data sets and novel biological materials have given rise to principles under which communities share data and materials (pre-and post-publication), and to a new sharing infrastructure — large public databases and repositories. Although much attention has been given to practical and ethical guidelines for prepublication data release from large-scale 'community resource projects', summarized in the Bermuda Principles[1] and the Fort Lauderdale report[2], sharing of data and resources from hypothesis-driven research has largely been addressed piecemeal by individual

Obstacles

- Technical interoperability challenge: how do we use and make sense of all this content (i.e., varying formats, ontologies, and analytic platforms)?
- Social / Cultural Challenge: how should we collaborate in public? Do we contribute to the public domain and how? How do we define (internally) what is pre-competitive?
- Legal barriers to sharing: intellectual property, contract, corporate law, and liability

What is Pre-Competitive?

- Varies by company: culture, products, business model, and comparative advantage
- Varies by field (e.g., physics vs. chemistry vs. biology)
- Can be subjective
- Can change over time
- No universal answer

Likely Candidates for Pre-Competitive Collaboration

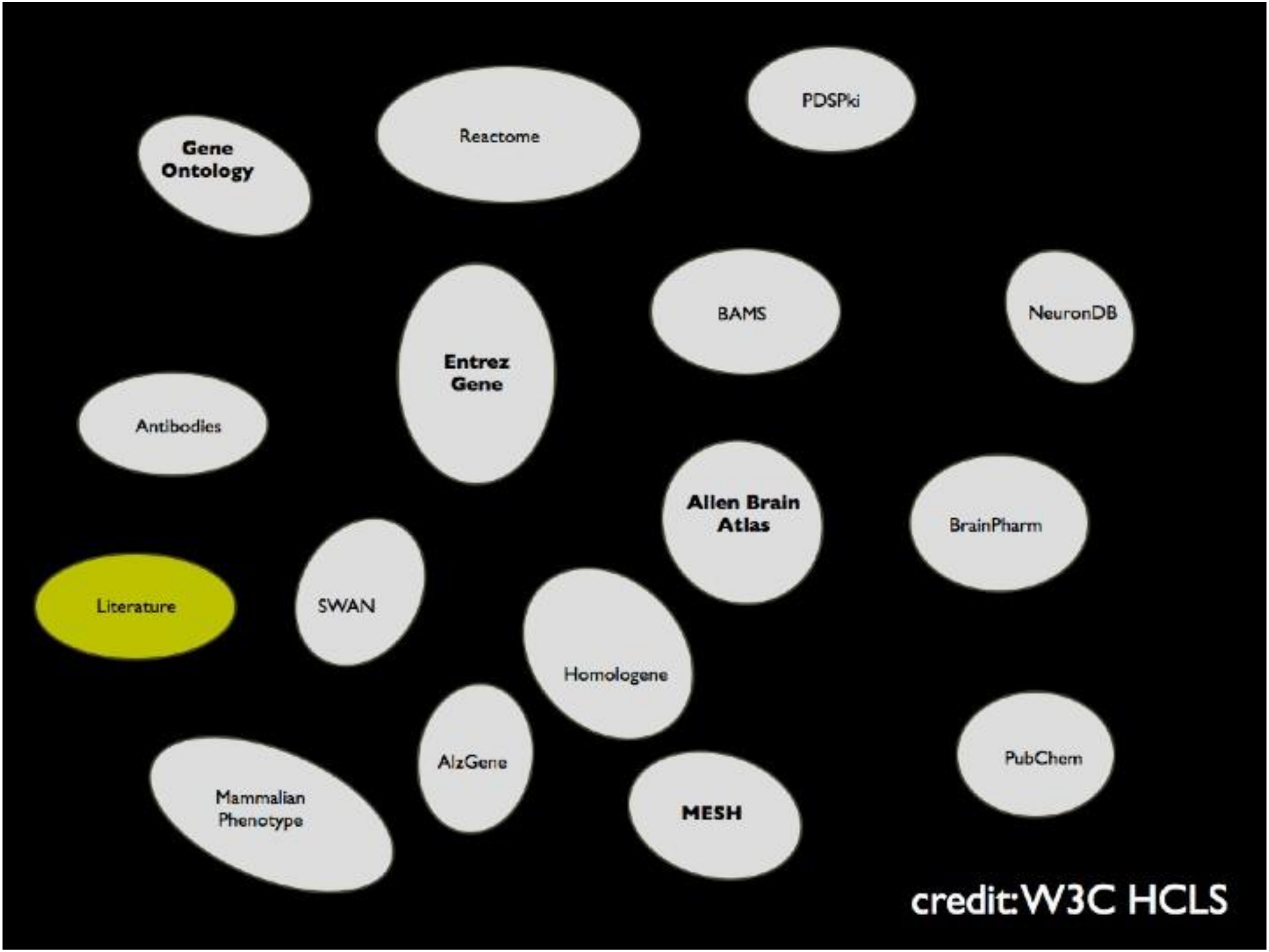
- Informatics resources and tools
- Ontologies and metadata for data integration (e.g., Semantic Web)
- Common data standards and formats
- Biomarker data (e.g., toxicity biomarkers)
- Biological pathway and target information

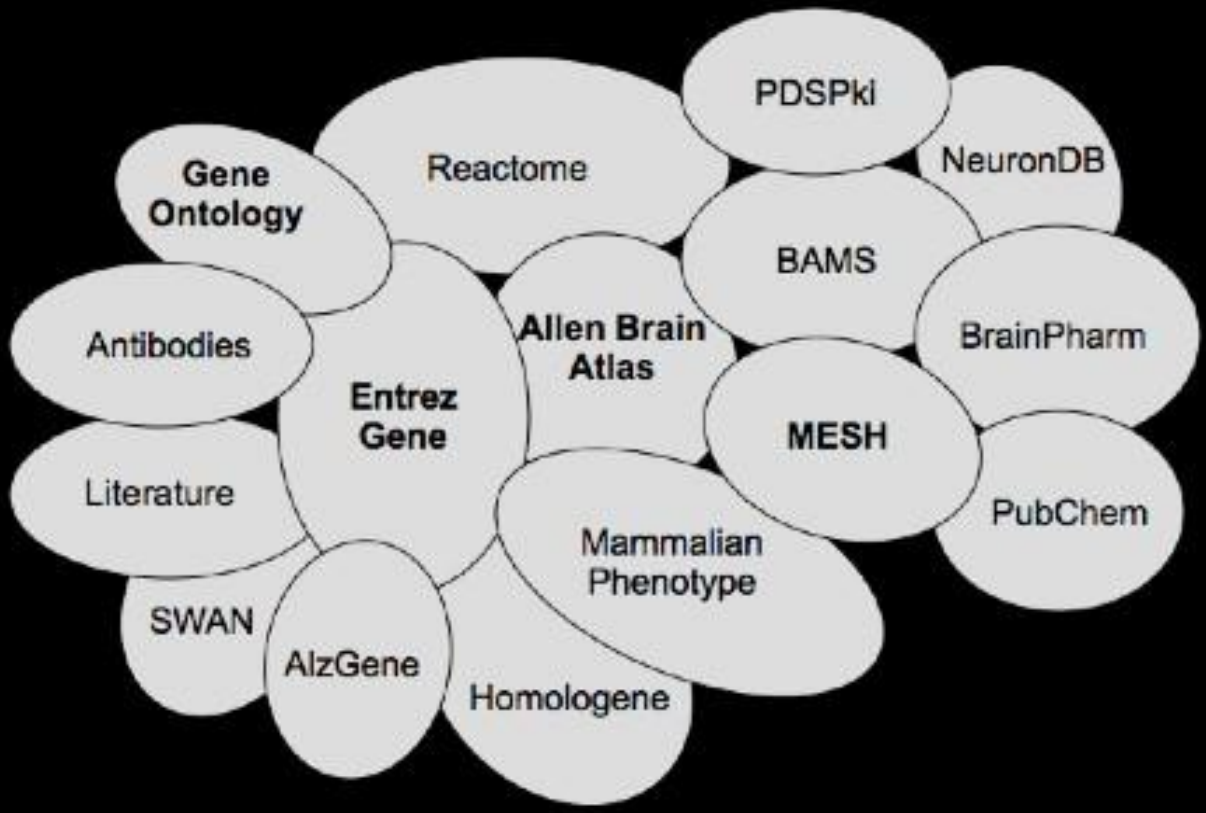
Collaborative Data Integration

- Pharmas spend large amounts of time and $$ on internally developing or licensing data integration and analysis tools that analyze public domain data
- “Re-inventing the wheel” in bio-informatics is costly and time-intensive
- Shifting internal funding priorities can result in abandonment of tools and data resources

Science Commons Use Case

- Science Commons NeuroCommons Project
- Uses Semantic Web technology as a data integration tool:
- http://neurocommons.org


Gene Ontology
Reactome
PDSPki
BAMS
NeuronDB
Entrez Gene
Antibodies
Allen Brain Atlas
BrainPharm
Literature
SWAN
Homologene
AlzGene
PubChem
Mammalian Phenotype
MESH
credit: W3C HCLS


PDSPki
Reactome
NeuronDB
Gene Ontology
BAMS
Antibodies
Allen Brain Atlas
BrainPharm
Entrez Gene
MESH
Literature
PubChem
Mammalian Phenotype
SWAN
AlzGene
Homologene

Some Examples

- Sage Bio-networks
- Life Science Grid - Eli Lily
- Health Commons
- Neurocommons

Important Future Questions

- Funding: who pays to support and maintain public domain repositories and infrastructure? What are long-term funding models?
- Ownership & IP: are there proprietary rights? Who owns any IP? Have IP rights been waived (e.g., CC0)? What are barriers to entry?
- Structure: how will the efforts be organized? Will it be hosted by a non-profit organization? How is government involved?
- Who sets standards and how is consensus reached? W3C-like process? What is the membership structure?
- Will companies contribute back to the public domain? Will collaboration really two-way?

creative commons

CC0 1.0 Universal

No Copyright

The person who associated a work with this document has dedicated this work to the Commons by waiving all of his or her rights to the work under copyright law and all related or neighboring legal rights he or she had in the work, to the extent allowable by law.

Other Rights — In no way are any of the following rights affected by CC0:

- Patent or trademark rights held by the person who associated this document with a work.
- Rights other persons may have either in the work itself or in how the work is used, such as **publicity** or privacy rights.

Questions and Comments

- Thinh Nguyen
- Email: thinh@creativecommons.org
- Website: www.sciencecommons.org