International Society for Biocuration

In This Issue

Biocuration 2013 Report

After Cambridge

Thank you for attending the 6th International Conference in Biocuration! We were delighted to see known faces and meet new members, and we hope you had a very productive time networking with fellow biocurators.

Cambridge University Press sponsored Best Talk prizes for the top-three best talks of the conference: Daniel Jamieson received first place with a discussion on Cataloguing the biomedical world of pain through semi-automated curation of molecular interactions. Prudence Mutowo-Meullenet earned second place speaking about the Use of Gene Ontology Annotation to understand the peroxisome proteome in humans. Last, but not least, Heiko Dietze earned third

prize sharing TermGenie with us, a story about Granting Biocurator's Wishes for the Gene Ontology.

The Heidelberg Institute for Theoretical Studies (HITS) sponsored a 'Best Poster' award. Attendants to the conference voted for their favorite posters over the course of the conference and the Top Three Poster Awards went to 1) Scott Cain for his GMOD poster, 2) Jerven Bolleman with Catching inconsistencies in UniProtKB/Swiss-Prot with the semantic web and 3) Siew-Yit Yong for her poster about The InterPro web site - refreshed, revamped, refined.

ISB awarded travel fellowships to Helen Benson, Tomas Di Domenico, Andrew Oberlin, Daniela Ledezma-Tejeida, Marie-Jean Meurs and Francesco Russo.

Daniel Jamieson, Prudence Mutowo-Meullenet and Heiko Dietze (right to left) received awards for the top-three Best Talks of the Conference (in that order).

"Compared to other data-rich sciences, biology has a major process of consensus formation through curation."

Ewan Birney, during his keynote presentation at Biocuration 2013.

Twitted by @tharris to #ISB2013

Clockwise from the top: Poster Awards winners (right to left) Scott Cain, Jarven Bolleman and Siew-Yit Yong. ISB2013 Chair Alex Bateman welcomes attendees to the 6th International Biocuration Conference. ISB2013 Participants take a walking tour of Cambridge, UK. Travel Fellowship awardees, (right to left) Marie-Jean Meurs, Andrew Oberlin and Helen Benson.

Next up: ISB 2014 in Toronto!

The Seventh International Biocuration Conference will be held on April 6-9, 2014 at the University of Toronto in Toronto, Canada. Hosted by the Ontario Institute for Cancer Research, Biocuration 2014 will focus our attention on **bridging the gap between genomes and phenomes**. More information is available at http://biocuration2014.events.oicr.on.ca/

Join us in Toronto!

Training Opportunities

Bioinformatic tools for NRPS discovery: from genomic data to the products. International workshop for young researchers.

The workshop will be held on July 10-12, 2013 at University of Lille1-Villeneuve d'Ascq, France.

Aim and Scope

Being able to rapidly identify, annotate and analyze polyketide and non-ribosomal peptide biosynthesis gene clusters in bacterial and fungal genomes, managing the prediction of the most probable structure and activity of the peptides, getting familiar with bioinformatic tools such as Norine (http://bioinfo.lifl.fr/norine/) and antiSMASH (http://antismash.secondarymetabolites.org/).
Registration: http://goo.gl/YohBx

News & Views from around the ISB Community

Check if your database or repository is missing

Elsevier is keen to work with researchers and data repositories to ensure that data that is relevant for scientific, technical and medical research can be easily discovered and accessed. One of the ways in which we are doing this is by creating bidirectional links between data repositories and online articles on ScienceDirect. This provides ScienceDirect's readers with one-click access to relevant, trusted data that may help to validate research or drive further investigations. Linking helps to make articles and data better discoverable, attracting more usage. Sharing the data that underpins conclusions is not only good scientific practice, but also increasingly required by funding bodies.

Specific journal instructions for authors depend on the data repository: in some cases data is extracted from the article by curators, while in other cases authors need to upload their data manually.

We need to hear from you with your detailed instructions. Detailed information for already established partnerships is available in the (recently revised and updated) listing of supported databases. Alternatively, copy & paste this link: http://goo.gl/8Jh8k

If you are a data repository manager interested in setting up bidirectional linking with Elsevier publications, please contact us at articleofthefuture@elsevier.com

Submitted by Adriaan Klinkenberg.

## The M:N Project at MGD: Beyond 1:1 Orthology Assertions

The Mouse Genome Database (MGD) curates, integrates, and provides comprehensive genetics, genomic, and phenotypic information for the laboratory mouse, a primary model organism for experimental investigation of human biology and disease. MGD is found at http://www.informatics.jax.org.

A core component of MGD data for over 20 years has been the curated assertion of 1:1 orthology between mouse, human, and rat protein-coding genes. Now, with completely sequenced genomes available for comparative analysis, phylogenetic analysis clearly identifies cases where descent from common ancestor does not always define a 1:1 relationship, but rather that gene duplication following an ancestral speciation event more correctly results in M:N relationship between genes in different species.

This has implications for the study of human biology in the mouse system and for the presentation of inferential functional and disease associated assertions based on comparative analysis. MGD has recently restructured its database to accommodate such homology classes with concurrent changes in presentation of data related to homology classes and in the representation of human diseases associated with mouse genes by curation of comparative or experimental data. We load data from all mammalian species with completed genome sequences, and will next extend our

data to include chicken and Zebrafish protein-coding gene classes. While 1:1 assertions predominate (~80%), we now more clearly represent cases such as the Serpina1 gene class (1 human, 5 mouse, 1 rat), and provide better cross-referencing among related genes, the diseases that have been studied in respect to those genes, and the relationship between genomic features in related genomes. This work is supported by NIH NHGRI grant HG-000330.

Submitted by Judith A Blake, Richard Baldarelli, Mary Dolan, Mark Airey, Jon Beal, Sharon Giannatto, David Miers, Jill Lewis, Carol J. Bult, Janan T. Eppig and James Kadin.
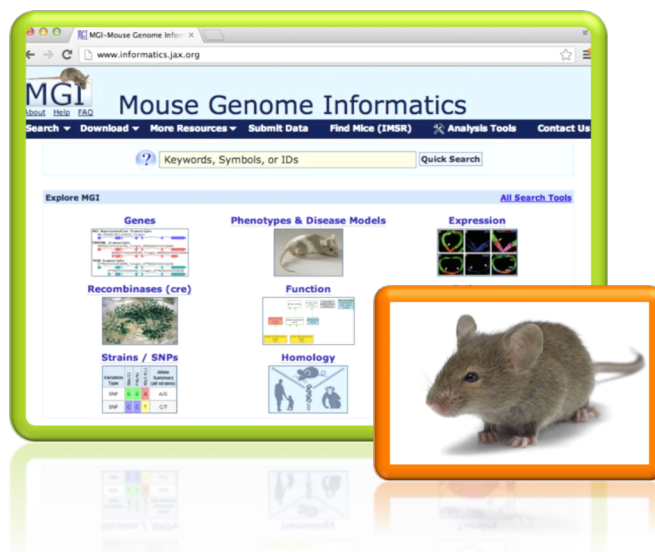
The Comparative Toxicogenomics Database Text-Mining Pipeline

The Comparative Toxicogenomics Database (CTD; http://ctdbase.org/) confirmed the effectiveness of its text-mining pipeline in evaluating and prioritizing scientific literature for the manual curation of chemical-gene-disease information. The results were selected for a special issue of PLoS Text Mining Collection in April (http://goo.gl/f1ubr).
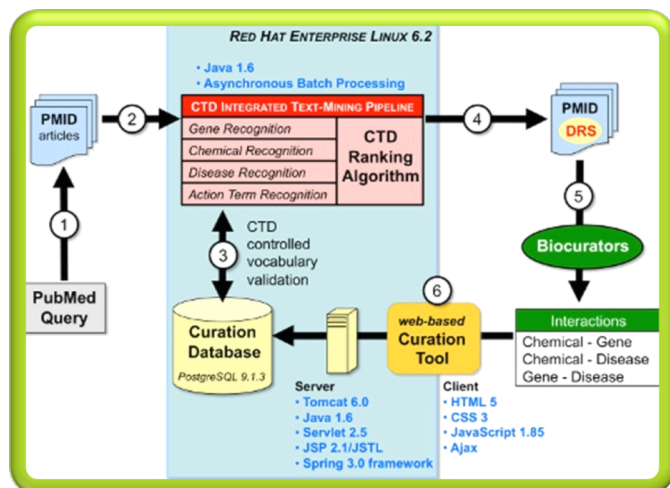
For the study, CTD tested their sophisticated text-mining algorithm by using it to evaluate the text from 15,000 articles and assign a relevancy score to each document. A representative sample of the corpus was sent to their team of biocurators to manually read and evaluate on their own, blind to the computer's score. The biocurators concurred with the algorithm 85% of the time with respect to the highest-scored papers, and there was a clear step-wise progression, wherein the likelihood of an article's true relevancy decreased linearly as the text-mining scores declined.

Ranking papers by text mining allowed biocurators to focus on the most relevant papers and avoid the extraneous ones, increasing productivity by 27% and novel data content by 2-fold. The curated articles were also broad and encompassing with respect to data coverage, finding both shared as well as unique biological processes, pathways, and toxicological end-points, confirming that the ranking system could help identify articles that contribute to a mechanistic understanding of toxicity.

By incorporating similar text mining-based scoring, other databases may also be able to enhance their manual curation by prioritizing more relevant articles, thereby increasing data content, productivity, and efficiency.

Submitted by Allan Peter Davis.

GeneView

GeneView is a web-based interface for searching and visualizing biomedical articles. The underlying text repository consists of more than 21 million PubMed abstracts and approximately 300,000 full text bodies from PubMed Central. All articles are automatically searched for biological entities using 10 different state-of-the-art named entity recognition tools. Recognized entities are associated with unique database identifiers if possible. For instance, the haemochromatosis causing polymorphism Cys282Tyr is associated with the dbSNP identifier rs1800562. Similarly, other mentions referring to the same polymorphism (e.g., p.Cys102Tyr, c.569G>A, c.77-206G>A, etc.) are mapped to the same dbSNP identifier. Using named entity recognition and normalization to database identifiers, GeneView facilitates retrieval of relevant articles (e.g., http://goo.gl/HLbEK). In addition to its broad coverage of entities, GeneView offers a number of unique features for searching articles. For instance, users may rank query results based on the content of articles with respect to a personalized gene list or based on the number of genes or SNPs they contain.

Selected articles are visualized in GeneView. Entity markup is

directly provided in the text and additional entity-specific information is provided on mouse click. This encompasses link-outs to entity-specific databases (e.g. EntrezGene, DrugBank, IntAct, etc.) or additional information such as the number of protein-protein-interactions according to IntAct or the official gene name.

The text repository is regularly updated and the underlying text-mining tools are constantly improved. GeneView can be freely accessed at http://bc3.informatik.hu-berlin.de/. More information can be found at Thomas P., Starlinger J., Vowinkel A., Arzt S., and Leser U. (2012) "GeneView: A comprehensive semantic search engine for PubMed". Nucleic Acids Research. Submitted by Philippe Thomas.

Call for proposals for Biocuration 2015. Get Ready!

The Executive Board of the International Society for Biocuration would like to invite tenders to host the 8th International Biocuration Conference during the Northern Spring or Summer of 2015.

Individuals and organizations may apply by sending a proposal to the ISB Executive Committee (intsocbio@gmail.com) on or before August 1st of 2013. We would very much appreciate an email indicating early intentions to bid. The successful bidder will be notified by October 1st, 2013. The ISB Executive Committee will publicly announce the selected organization or individual during the 7th International Biocuration Conference, to be held in April 6-9, 2014 in Toronto, Canada.

For more information about the ISB and our previous conferences, please visit: http://www.biocurator.org

- Proposals should be short:
  - One side of an A4 or US letter sheet, 11 point.

- The proposal should contain:
  - The name and institution of the local organizer
  - Details of the proposed venue for up to 350 participants
  - The range of dates available for the conference
  - A brief outline of a strategic plan to attract a broad range of participants from the Biocuration community

Upcoming Conferences

1st International Workshop on Semantics for Biodiversity (S4BIODIV)
http://semantic-biodiversity.mpl.ird.fr/
When: May 26-27, 2013
Where: Montpellier, France
What: The workshop aims to identify the key challenges faced by the bioinformatics community, discuss potential solutions and identify the opportunities emerging from the trans-disciplinary interactions between Plant Science and Informatics experts. Therefore, they expect the bioinformatics experts to explain how they apply semantic web standards and tools to their scientific topic, from biology, agriculture, agro-ecology, genomics, environmental studies, to social sciences, citizen sciences.

10th Extended Semantic Web Conference (ESWC 2013)
http://2013.eswc-conferences.org/
When: May 26-30, 2013
Where: Montpellier, France
What: The ESWC is a major venue for discussing the latest scientific results and technology innovations around semantic technologies. Building on its past success, ESWC is seeking to broaden its focus to span other relevant research areas in which Web semantics plays an important role.

4th International Conference on Biomedical Ontologies 2013 (ICBO2013)
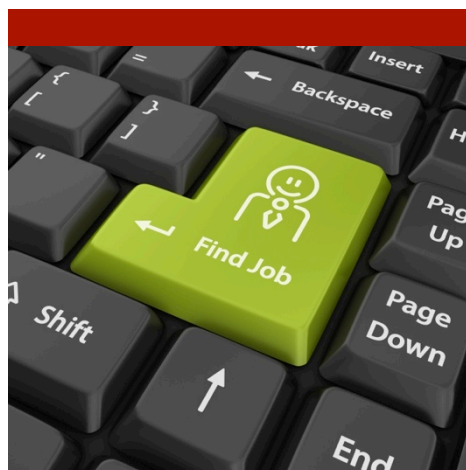http://goo.gl/PGUVX and http://goo.gl/K4eUC
When: July 7-9 2013
Where: Concordia University, Montreal, Quebec, Canada.
What: The conference aim is to foster discussion, exchange, and innovation in research and development in the areas Biomedical Ontology. Researchers and professionals from

Job Opportunities

Scientific Data, Editorial Biocurator, Nature Publishing Group, UK. Posted April 19, 2013. More information is available at http://goo.gl/ZcgSg

Senior Biocurator, Saccharomyces Genome Database, Stanford University. Posted April 19, 2013. Find more about this position at http://goo.gl/HwRGy

Data Wrangler, ENCODE, Stanford University. Posted April 19, 2013. Details available at http://goo.gl/ZZA80

Python Software Engineer for UI construction. Saccharomyces Genome Database and ENCODE Data Coordination Center. Posted March 30, 2013. Please see the complete job advertisements at http://goo.gl/9M1HF

Research Assistant Position in EU project Linked2Safety, Digital Enterprise Research Institute (http://www.deri.ie/about/jobs), National University of Ireland, Galway. Posted March 23, 2013. Find more about this position at http://goo.gl/uHwlm

Scientific Curator, Eugene, OR, USA. Posted March 23, 2013. Details available at http://goo.gl/z6aMX

Scientific Biocurator for Swiss-Lipids, Geneva, Switzerland. Posted March 15, 2013. More information at http://goo.gl/Jxigq

Upcoming Conferences (Continued)

biology, medicine, computer science and engineering are invited to share their knowledge and experience. The event is part of the Semantic Trilogy 2013 featuring: i) International Conference on Biomedical Ontologies (ICBO 2013), ii) Canadian Semantic Web Symposium (CSWS 2013), and iii) Data Integration in the Life Sciences (DILS 2013).

Bio-Ontologies SIG 2013
http://www.bio-ontologies.org.uk
When: July 20, 2013
Where: Berlin, Germany.
What: The Bio-Ontologies SIG provides a forum for discussion of the latest and most innovative research in the application of ontologies and more generally the organization, presentation and dissemination of knowledge in biomedicine and the life sciences.
Notes: Co-located with ISMB/ECCB 2013. www.bio-ontologies.org.uk/submissions

BioNLP Shared Task 2013 at the BioNLP'11 Workshop
http://2013.bionlp-st.org/participation
When: Aug. 8-9 2013
Where: Sofia, Bulgaria
What: The BioNLP Shared Task (BioNLP-ST) series represents a community-wide trend in text-mining for biology toward fine-grained information extraction (IE).
Notes: Co-hosted with the ACL/HLT 2013 BioNLP workshop. Registration for participation to the BioNLP Shared Task is required. Registration open at the BioNLP Shared Task 2013 website. Training and development data sets of BioNLP shared task 2013 are publicly available on the task Web pages of the BioNLP Shared Task 2013 website: http://2013.bionlp-st.org/tasks

The 12th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE 2013)
http://www.onthemove-conferences.org/index.php/odbase13
When: Sept 10-11, 2013
Where: Graz, Austria
What: ODBASE'13 provides a forum on the use of ontologies and data semantics in novel applications. Of particular relevance to ODBASE 2013 are papers that bridge traditional boundaries between disciplines such as databases, social networks, mobile systems, artificial intelligence, information retrieval, and computational linguistics. Submission of research and practical experience papers concerning scale issues in ontology management, information integration, and information extraction, as well as papers that examine the information needs of various applications is encouraged.
Notes: Deadlines: Abstract Submission: May 11, 2013. Paper Submission: May 18, 2013. Author Registration: July 16, 2013. Springer Verlag will publish the final proceedings as LNCS (Lecture Notes in Computer Science). Author instructions can be found at: http://goo.gl/f96OI

7th Int'l Workshop on Modular Ontologies (WoMO)
http://lat.inf.tu-dresden.de/~dwalther/womo2013/
When: September 15, 2013
Where: Corunna, Spain
What: Topics include, but are not limited to: What is modularity?; Logical/foundational studies; Algorithmic approaches; Evaluation of modularizations; Applications: semantic web; life sciences; earth sciences; bio-ontologies; natural language processing; collaborative ontology development, ontology versioning, and others.
Notes: Paper submission deadline: July 5, 2013

Job Opportunities (Continued)
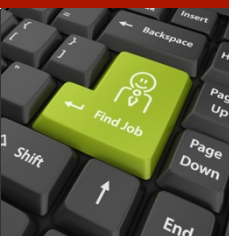
Software Engineer, Zebrafish Model Organism Database, Eugene, OR, USA. Posted March 15, 2013. Find more about this position at http://goo.gl/p4byB

Postdoctoral and Staff Researchers, BindingDB (http://www.bindingdb.org). Posted February 22, 2013. For more information visit http://goo.gl/OvG9t

Software developer, NISEED (http://www.aniseed.cnrs.fr) team at CRBM, Montpellier, France. Posted February 20, 2013. Find more about this job at http://goo.gl/bIVI9

Upcoming Conferences (Continued)

Workshop on Ontologies and Data in Life Sciences
http://tinyurl.com/odls-2013-en
When: Sep 19-20
Where: Koblenz, Germany
What: This workshop integrates workshops of two series: the workshop series Data in Life Sciences and Ontologies in Biomedicine and Life Sciences (OBML), thereby connecting two important areas of biomedical research. This connection covers the overall spectrum of biomedical information processing, from experimental data acquisition and data management, across analysis, structuring and interpretation of data, up to the development of structures of knowledge in the form of ontologies with their various applications. Notes: Paper submission ended Apr 22.

12th International Conference on Bioinformatics - InCoB2013
http://ictbi.imed-cn.org/incob2013
When: Sep 20-22, 2013.
Where: Taicang (China).
What: InCoB2013 has overall objective to provide a platform for experts and budding bioinformaticians to discuss and exchange ideas and thoughts on the development of bioinformatics in the Asia Pacific region.
Notes: Deadline for paper submissions (BMC Bioinformatics, BMC Genomics and BMC Systems Biology) ended April 30, 2013.

BioCreative IV Critical Assessment of Information Extraction in Biology Challenge & Workshop
http://www.biocreative.org/events/biocreative-iv/CFP/
When: October 7-9, 2013
Where: NCBI, National Institutes of Health, Bethesda, Maryland.
What: BioCreative is a community-wide effort for evaluating text mining and information extraction systems applied to the biological domain. One key goal of BioCreative is the active involvement of the text mining user community in the design of the tracks, preparation of corpus and the testing of interactive systems.

NETTAB 2013: Semantic, Social, and Mobile Applications for Bioinformatics and Biomedical Laboratories
http://www.nettab.org/2013/
When: Oct 16-18, 2013
Where: Lido of Venice, Italy
What: NETTAB 2013 will explore mobile, social, and semantic solutions for bioinformatics and laboratory informatics. A savvy combination of these technologies could greatly enhance the research outcome of life scientists and markedly simplify workflows in biomedical laboratories.
Notes: Deadlines: July 5: Abstract submission for Oral communications. July 31: Abstract submission for Posters. September 13: Early registration.

Discovery Informatics Symposium: AI Takes a Science-Centered View on Big Data
http://www.discoveryinformaticsinitiative.org/dis2013
When: November 15–17, 2013
Where: Arlington, VA (USA)
What: Part of the AAAI Fall Symposium Series, this symposium will provide a forum for researchers interested in understanding the role of AI techniques in improving or innovating scientific processes.
Notes: To receive further notifications about this and related meetings you may subscribe to:
http://discoveryinformaticsinitiative.org/mailman/listinfo/dii

Your ISB Executive Committee

Left to right: Back: Claire O'Donovan, J. Michael Cherry, Marc Robinson-Rechavi (Treasurer), Renate Kania, Alex Bateman. Front: Teresa Attwood, Pascale Gaudet (Chairperson), Chisato Yamasaki, Monica Munoz-Torres (Secretary).

Executive Committee Meetings
The ISB Executive Committee meets monthly. Minutes from the meetings are posted on the ISB website http://www.biocurator.org/executive_committee_minutes.html

ISB Newsletter Archive
Previous issues of this Newsletter can be found on the ISB website at http://biocurator.org/newsletter.shtml
M. Munoz-Torres edits the ISB Newsletter.

Share your Ideas!

Share your ideas with members of ISB! Prepare a news article written to biocurators as the audience. In 260 words or less spread the word about new tools in your site, or tools for distribution to the wider community. Broadcast announcements and advances from your team to all the members of ISB by sending an email to intsocbio@gmail.com. Submission deadline for the June Newsletter is Friday, May 31st at midnight (GMT).

Kind Regards,
The ISB Executive Committee