Brief Bio and (PR)²: Problems & Pitches – Rants & Raves by Brian Haugen

Brian Haugen is a Health Science Policy Analyst for the National Institutes of Health, supporting the NIH RePORT and RePORTER projects for the Office of Extramural Research, Office of Data Analysis Tools and Systems (OER DATA Systems). OER DATA Systems is building modular systems to make NIH funding data displays more accessible and relevant and to link NIH funding to investigators, publications, patents, and other products of research. Previously, Brian was a Senior Scientist for Analysis and Evaluation at National Institute for Child Health and Human Development, where he led data management and data analytic projects for the National Children’s Study. Brian first came to the NIH as a Presidential Management Fellow in 2007, and carried out detail assignments across several NIH institutes and Centers in program, policy, review, and information technology. Prior to joining the NIH, he earned his PhD in microbiology from the University of Wisconsin-Madison.

RePORT home page
http://report.nih.gov

Software:
NIH RePORTER http://projectreporter.nih.gov

General Questions

1) Do you consider yourself a developer, user, creator, system evangelist, etc.?
   Creator and system evangelist for our internal and external analytic systems.

2) What are your main interests in attending the workshop?
   - Understanding how biomedical information can be easily shared across institutional and national borders
   - How different research products can be linked to trace the scientific progress
   - Methods for identifying semantic text similarities among documents.

3) What would you like to learn / achieve at the workshop?
   - Identify frameworks and tools to facilitate rapid data integration within an established schema
   - Identify clustering and visualization techniques that are amenable to frequent updates.

4) What are the tools or services you would like to share at the workshop?
   RePORTER, World RePORT, Federal RePORTER

5) Please list three features or functions of your tools or services that are most important for users.
   - Network relationships of disparate entities including patents, publications, funding records, and other products of research.
   - Timely and accessible data
6) What are your major concerns for the software architecture of these tools / services?

- Moving the best of visualizations into the hands of fresh users without overwhelming them.
- Creating maintainable databases of similarity relationships

7) What are some underserved user needs that your systems, idealized versions, or an ecosystem that your systems are integral components of could address?

Understanding the path by which scientific ideas lead to products and improvements in human health.

8) Are you or your group working on any of these challenges? If yes, please explain.

Yes, we are developing tools to generate potential links among different scientific documents, for later assessment by human experts.

9) “Big data” and “cloud computing” get thrown around a lot as terms. How do these concepts and your, your group’s and your users’ interpretation and understanding of them affect your plans for development?

Cloud computing is an attractive solution when computing capabilities need to scale rapidly, but it is not clear whether it is a cost-effective solution(either in dollars or time) when computing demands are stable.

Most of our computing challenges do not enter the “Big Data” realm except in the areas of unstructured text analysis and relationship networking, but we hope that the tools and techniques applied to larger data sets may allow new analyses