Apps and Tools Call 20140826

Date	2014-08-26
Topic	Data Visualization with Michael Bales

Agenda

2014 conference recap

Upcoming hackathon at Cornell in October

https://wiki.duraspace.org/display/VIVO/Hackathon+October+2014+at+Cornell

 Registration: https://www.surveymonkey.com/s/9L5ZN3M

Duraspace Wiki Accounts - now able to create accounts by completing forms

 https://wiki.duraspace.org/signup.action

Data Visualization discussion - Michael Bales from Weill Cornell Medical College

      Follow at https://twitter.com/bale0019

Michael Bales's presentation on visualizations

Slides - http://goo.gl/X3aE4D

Survey of what's available

- 2011 review – 43 sites
- 8 no longer publicly available
- Most had no analytic or visualization elements
- 9 of 43 had at least one analytic or visualization element

Sites reviewed in 2011:

- academia.edu
- BibApp
- BioCrowd
- BiomedExperts
- Research Expertise at UT Arlington
- Community Academic Profiles - CAP
- Digital|Vita
- Epernicus Solutions & Epernicus Network
- Expertise @ Maryland
- Faculty Profile System
- Faculty Research Interests Database
- GENIUS
- Google Scholar
- iamResearcher
- iAMscientist
- INDURE
- LabRoots
- labslink
- LatticeGrid
- Loki
- Mendeley
- myExperiment
- MyNetResearch
- MySDScience
- Nanopaprika
- Nature Network

- OpenWetWare
- Pivot
- Profiles Research Networking Software
- Research Crossroads
- Research in View
- ResearcherID
- ResearchGate
- RKBExplorer
- Science Stage
- SciTable
- SciVal Experts
- sequilab
- The Research Cooperative
- VIVO
- Columbia University Scientific Profiles
- Yaffle

Sites with the most extensive analytic and visualization content:

- Harvard Catalyst Profiles
- LatticeGrid

Other sites with some analytic and/or visualization content:

- academia.edu
- BibApp
- Google Scholar
- INDURE
- myExperiment
- ResearcherID
- VIVO

The good

- Clean interfaces; visually appealing
- Well indexed – what interactive elements are available on the site?
- Provide sufficient explanatory detail
- Good visual organization
- Size, color, visual arrangement used to convey hierarchy (e.g., ranking or popularity)
- Adherence to cognitive model of what users want to accomplish

The bad

- Visualizations only available for selected profiles
- Visual chaos
- Difficult navigation; e.g., >20 infographic choices displayed on one screen
- Broken links; error messages

To do

- Review research on what users have said they want
- Inspiration from visualization and analytics beyond researcher profiling systems
- Consider how to incorporate "latest and greatest" in analytics and visualization
- Prototype, evaluate, refine

For discussion

- Other visualization paradigms of interest?
- Topics, collaboration, metrics – other ways of categorizing? (Grants could be an additional category; metrics could apply to individuals, departments, organizations)
- What sites have I missed – review of vendor offerings?

Questions asked

- Don: do we have Jira tickets or space to track use cases? A: No, but we can get started.

- Chris: Data journalism… Mike Bostock (creator of D3? Don: "centered visualizations on the DOM lending itself more to styling")... NYT and WSJ information graphics good examples (http://www.amazon.com/Street-Journal-Guide-Information-Graphics/dp/0393347281)... Miles used D3 for his work on VIVO Dashboard
- Chris: where do geographic visualizations fit? where do grads come from? where are funding offices based? A: think about the what and the how as 2 different things…
- Chris: how did you get started down this path? A: researching NLP, …, social network analysis, team science -- interest in viz for a long time… looking forward to hackathon, getting hands dirty with data, using D3
- Chris: would be interesting to visualize the scope/coverage/other of VIVO data itself (?)
- Alexandre: thanks to new query API, viz should be an application outside of VIVO. A: Chris we agree -- pulls out JSON. The trick is getting Don, Paul, others to put up public SPARQL endpoints ;) Don: you have to open up the linked data. Alex: could viz help administrators make informed decisions to support opening up their institutions' data? Michael: when you put data out in support of "metrics" it can become sensitive.
- PUBLIC DATA sameas PUBLIC DATA => bumper sticker?
- Jim: There's a line there that can be walked… don't want to require the heavy duty programming required by the first round (NIH grant era) of VIVO visualizations… thinks that loosely coupled paths are available now. API and AJAX services… Michael: observation that when an ontology defines relationships in a well defined way… the VIVO data are curated separately at each institution… so there might be missing/mismatched data… only a problem to extent that people are encouraged to fill in the data…
- Chris: better way to visualize the VIVO-ISF ontology… Don: how to keep D3 viz decoupled from ontology details and SPARQL/Solr expertise… Shahim (via chat): "yes, and as part of the future ISF work, these issues will be addressed as much as possible"... Chris: increase VIVO value prop by reducing cost to building viz
- Paul: viz of search results would be great improvement over default search result usability
- Alex: How to approach functional and usability testing of visualizations?
  - How do you know if the visualization is functioning properly (or how to "debug")? Michael: not sure of a set of heuristics for that… PaulF: LatticeGrid… tests that data is in correct format, and that the page renders… expect that the viz library is functioning… currently porting to work with VIVO instead of relational DB
  - How to test a visualizations usability/UX? Michael: usability testing with video recording, screen/click capture and "think aloud" technique
  - Alex: I like how the Profiles RNS viz includes "Why?" explanations...
- Alexandre: having some competitions to stimulate contributors (including students) to develop open source viz… Chris: discussing these types of ideas with DuraSpace, e.g. mini-grants

Notes

- Known "issues" described of overview:
  - haven't updated list of scientific social networking sites that have emerged since 2011
  - haven't covered VIVO visualizations yet, or related (VIVO Dashboard, etc)
- Two sites with most comprehensive visualizations: Harvard Profiles and LatticeGrid
- Most popular visual element was the tag cloud
- People like to see faces in visualizations
- Harvard Profiles: force-directed node layout
- Northwestern has a wheel graph which is an alternative to Harvard's
- Chord diagram: good for sharing collaboration among a group
- Harvard study: looked at the effect of co-location on the success of a publication
- Harvard Profiles: collaboration timeline shows how collaborations have changed over time; concept timelines also does this
- Have you looked at the interactive capability map viz at Find an Expert (http://findanexpert.unimelb.edu.au/)? For example, network viz on the term "disaster" -- http://115.146.84.177/search/#disaster%7C20%7C1
- Collaboration systems: how do you find people who have researched similar topics but have not yet co-authored
- Harvard Profiles: has similarity scores of people
- Pros
  - providing sufficient explanatory detail, but the level of details depends on the audience and how in depth they want to go
  - using size and color to indicate hierarchy
- What do people say they want with visualization and analytic elements?
- Good to look at other domains to get ideas