

Respondent Thoughts

Jim Hendler

Director, Institute for Data Exploration and Applications
Rensselaer Polytechnic Institute (RPI)

<http://www.cs.rpi.edu/~hendler>, @jahendler

See BRDI slides at

http://sites.nationalacademies.org/pga/brdi/pga_181009

Summary at

http://sites.nationalacademies.org/cs/groups/pgasite/documents/webpage/pga_182845.pdf

Why me

- Lots of history in this stuff
 - Early web development
 - Semantic Web & Linked Data
 - Open Government Data
 - Schema.org dataset/data catalog
 - Way too many committees and such
 - Which meant George knew where to find me

Why me

- Lots of history in this stuff
 - Early web development
 - Semantic Web & Linked Data
 - Open Government Data
 - Schema.org dataset/data catalog
 - Way too many committees and such
 - Which meant George knew where to find me
 - But more importantly: A known curmudgeon*

* Thanks Julie M.

Provocation

- The data world is “inconsistent. Duplicative. Riddled with broken and displaying links”
- The World Wide Web is inconsistent. Duplicative. Riddled with broken and displaying links
 - We live with it globally
 - We fix it locally
- It’s a terrible solution, it’s just that **everything** else has not worked
 - Or been built on top of it (FB, Twitter, etc.)

Provocation

- Scientists spend 75/80% of their time dealing with data issues.
 - And we want to make it even harder??
 - But don't worry, we'll incentivize by getting some funding agents to pay
- Web made it *easier* to do document sharing
 - people used it for many reasons, but part is because it was easy to use and part is “kudos”
 - Entrepreneurs eventually provided the incentives for greater use

Provocations I'll skip...

- The big challenges facing us today are INTERDISCIPLINARY so don't build disciplinary solutions without interoperability
- Human to Human communication is still a critical part of data sharing (get me to the right person is real sharing now)
- If I don't care about your data, why should I put effort into sharing mine?
- Must support the “long tail” community
 - Biomedical/life science is well funded, Geo is sort of well funded, pretty much everyone else is struggling
 - There's as many scientists in the tail as those in the big areas...
- Support 3rd party metadata repair (owl:sameAs, skos:exactmatch, skos:related...)

An example to consider

- Compare Schema.org to every other effort to get metadata out there
 - Used on BILLIONS of web sites
 - Authored by MILLIONS of webmasters
- Guha's talk (106k views)
<https://www.slideshare.net/rvguha/sem-tech2014c>
 - Incentivize usage
 - Simplicity, simplicity, simplicity
 - Incrementality
 - Use URIs
 - Support Collaboration
 - Local and Global

An example to consider

- A horrible way to doing data sharing *per se*
 - But, got things going and keeps growing
 - Provides infrastructure on which to build
 - Grew out of testbeds, interoperability, and incentives
 - works
- What could we do to bootstrap something like this?
 - (several slides removed here – see Carole Goble’s talk)

My (provocative) suggestion

- Build it and they will come
 - We've been talking about this for a long time
 - Standards done before use rarely succeed
 - We are starting to have enough “demonstration” projects
 - And some national-scale proposals
 - Infrastructure needs to be built
 - Grow it from interoperability of platforms
 - Keep interoperability central, supported and simple
 - Metadata is crucial!!
 - Keep metadata central, supported and **simple**
 - Let third parties help