



TWC



↔ <http://bit.ly/lebo-cold-2013>

Content-Preserving Graphics



Timothy Lebo, Alvaro Graves, Deborah McGuinness
(presented by **Patrice Seyed**)
Tetherless World Constellation
Rensselaer Polytechnic Institute

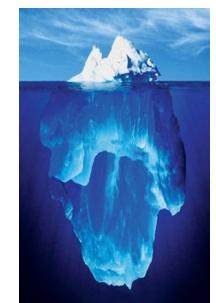
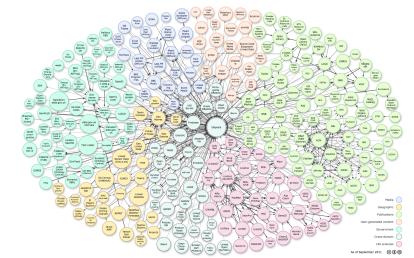
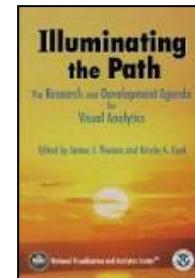


Rensselaer



Outline

- Visual Analytics
 - Sensemaking loops
 - 5 Real-world problems
- Can Linked Data Help?
 - Easy access, combination
 - Vocabulary reuse, inference foundations
- Helping Linked Data
 - Practical, sharable visual handles for LOD
 - Interrogating and re-purposing analytical results
- Future work: Pieces that LOD is missing





Visual Analytics: Sensemaking Loops

Analytics is *iterative* and *ad hoc*:

- *Data Foraging loop*
seeks, searches, and filters information
- *Sense making loop*
develops mental model to suit objective

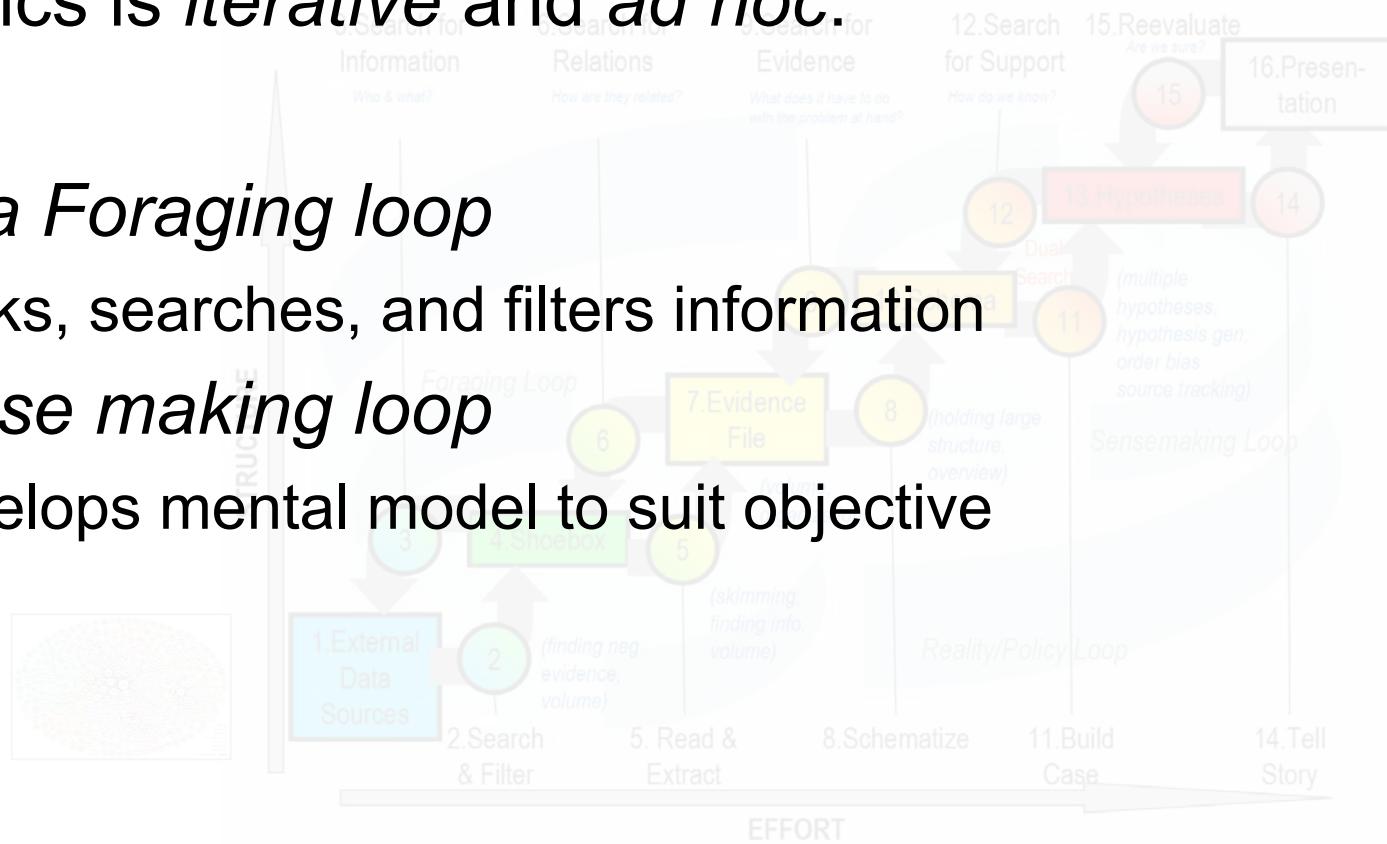


Figure 2. Notional model of sensemaking loop for intelligence analysis derived from CTA.

P. Pirolli and S. Card The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of International Conference on Intelligence Analysis*, 2005.



Visual Analytics: Real-World Problems

Among 35 data analysts from 25 commercial organizations:

- Most tedious and time-consuming task is **discovering** and **wrangling** data
- Analytical results are **static**
- Analytical results are **shared** via email, a shared file system, or during group meetings
- Difficulties discovering when **relevant** data becomes available
- ***Visualizations avoided because considered a barrier to underlying data***

S. Kandel, A. Paepcke, J. Hellerstein, and J. Heer. Enterprise data analysis and visualization: An interview study. *Visualization and Computer Graphics, IEEE Transactions on*, 18(12):2917–2926, Dec. 2012.



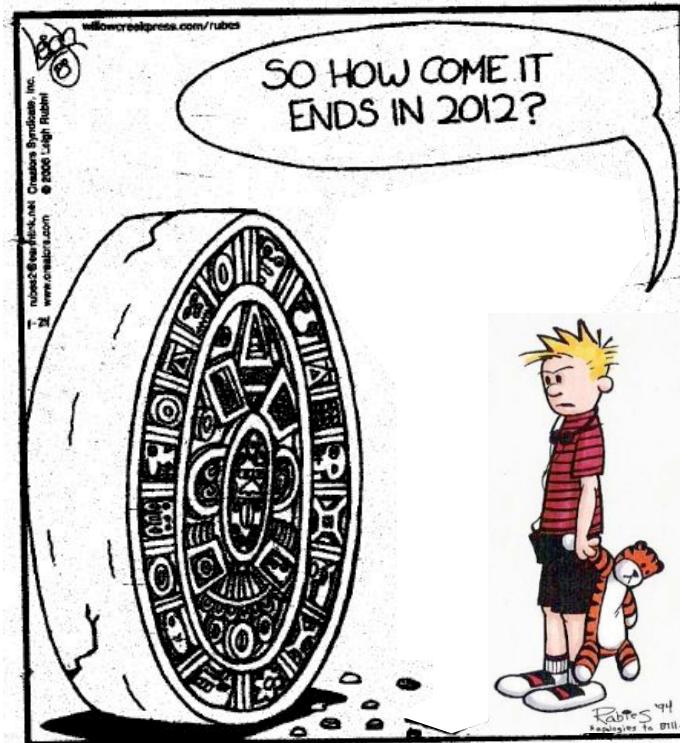
Avoiding Archaeological Endeavors

*What are the **things** that you're talking about?
How do those things **relate**?*

RUBES by Leigh Rubin



At last, the mystery of the
Mayan calendar revealed.



A few years later, or the **same moment**
somewhere else on the web.



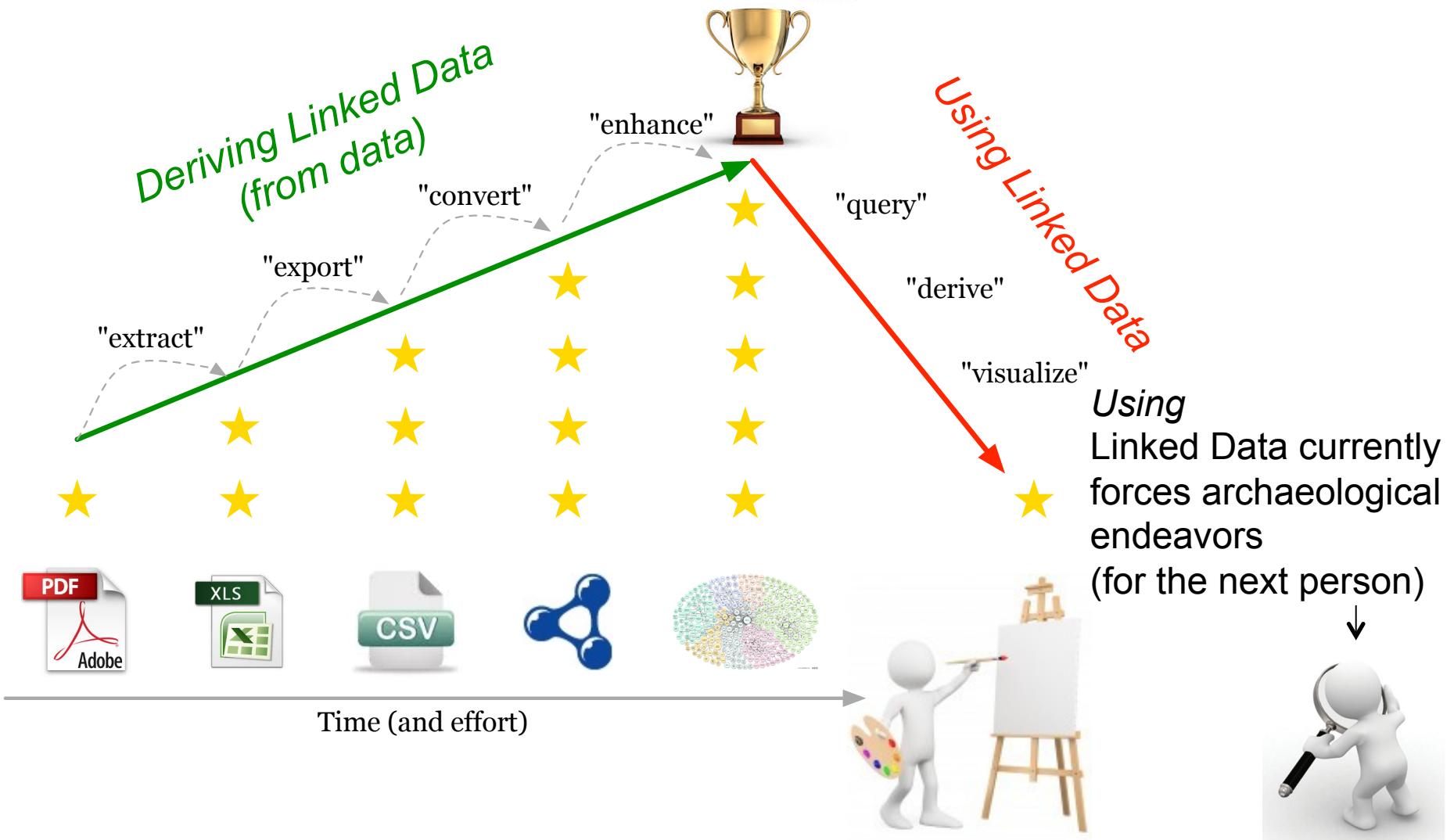
Avoiding Archaeological Endeavors

- ★ make your stuff available on the Web (whatever format) under an open license¹
- ★★ make it available as structured data (e.g., Excel instead of image scan of a table)²
- ★★★ use non-proprietary formats (e.g., CSV instead of Excel)³
- ★★★★ use URIs to denote things, so that people can point at your stuff⁴
- ★★★★★ link your data to other data to provide context⁵

Michael Hausenblas' <http://5stardata.info>



The Irony of *using* Linked Data





If We Don't Use Linked Data, Who Will?

{

```
"license_title": "Creative Commons A",
"maintainer": "",
"private": false,
"maintainer_email": "",
"num_tags": 14,
"id": "91d2c0de-75a4-4bb6-b260-bc2946e1be8b",
"metadata_created": "2011-11-29T14:11:14.282333",
"relationships": [
```

]

},
"lic
"me
"au
"au
"do
"st
"ve

4383

datahub

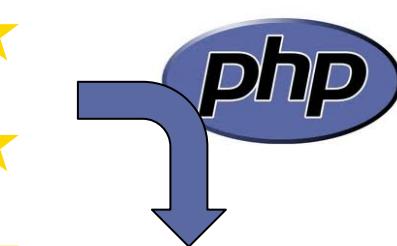
Farmers Markets Geographic Data (United States)

Followers 0

Organization Tetherless World Constellation

Social Google+ Twitter

JSON



lodcloud.graffle

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE plist PUBLIC "-//Apple//DTD PLIST 1.0//EN"
<plist version="1.0">
<dict>
 <key>ActiveLayerIndex</key>
 <integer>0</integer>
 <key>ApplicationVersion</key>
 <array>
 <string>com.omnigroup.OmniGrafflePro</string>
 <string>139.18.0.187838</string>
 </array>

As of October 2013



Why Linked Data Is Not Used

“It’s *easier* to do something else.”

“It’s *more useful* to do something else.”

Content-preserving graphics enables two new kinds of utility:

Interrogate visual results

Re-purpose visual results



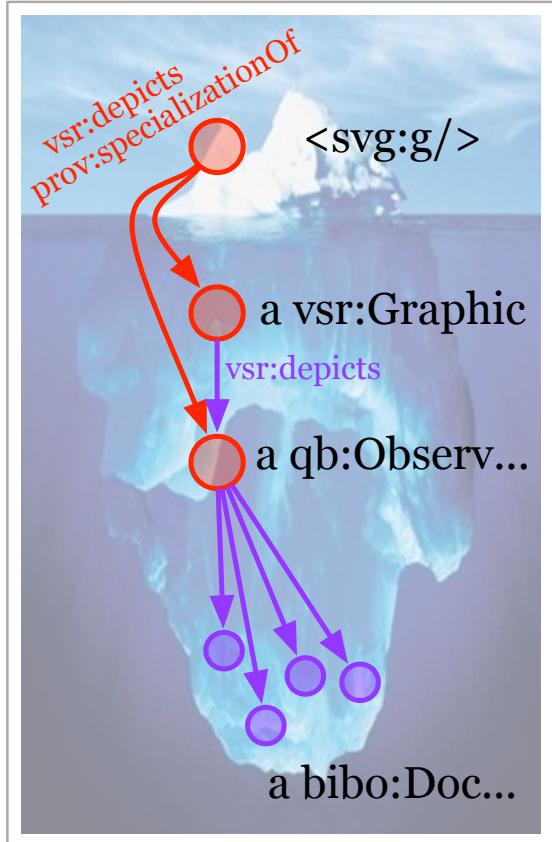
Content-Preserving Graphics

Useful Visual Handles on Useful Portions of LOD

Traditional Linked Data Applications



Content-Preserving Graphics



available by GRDDL

available by dereference

} Concrete Visual Information

} Abstract Visual Information

} Derived Information

} Existing Linked Data

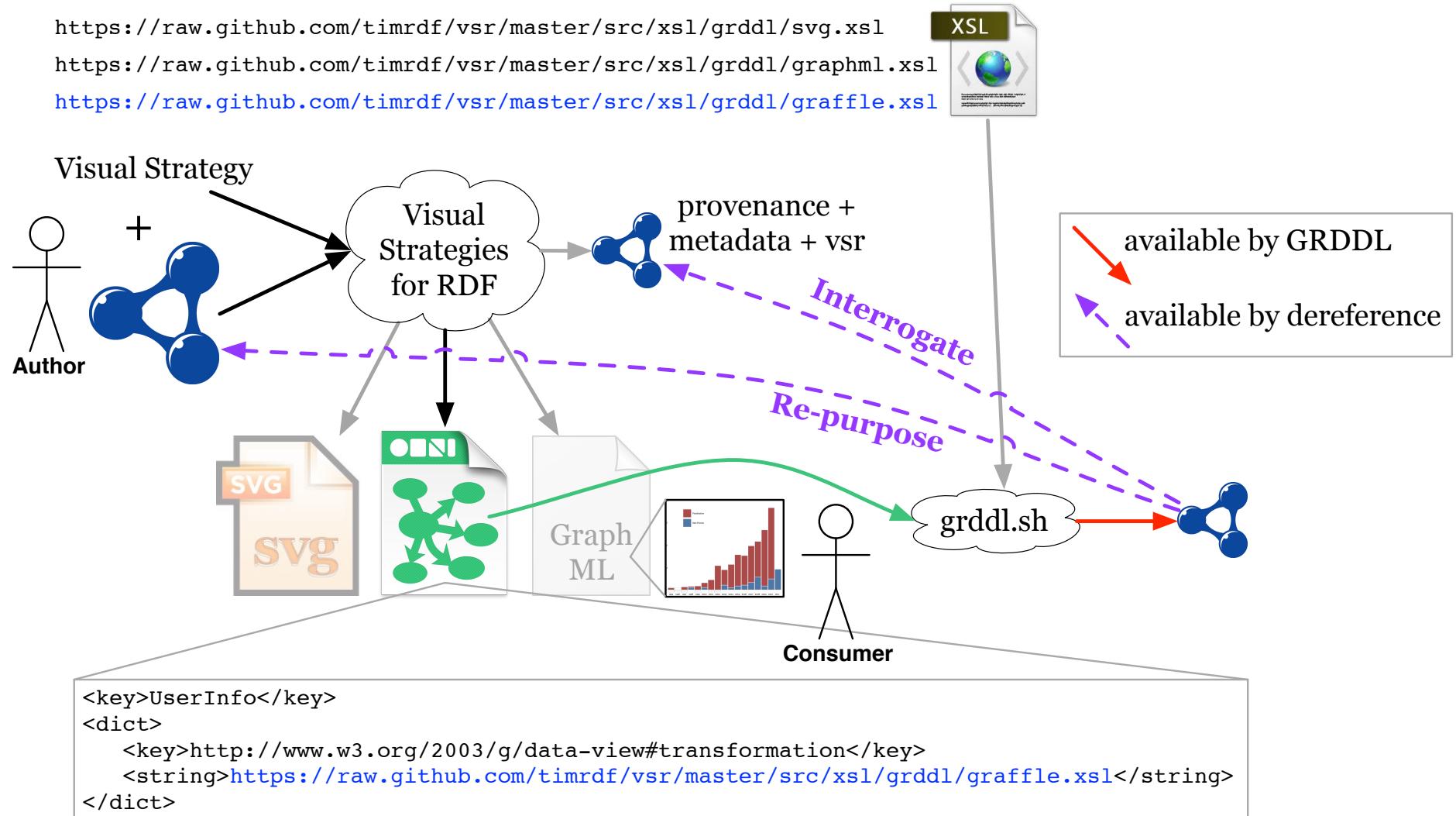


Creating and Using a Content-Preserved Graphic

<https://raw.github.com/timrdf/vsr/master/src/xsl/grddl/svg.xsl>

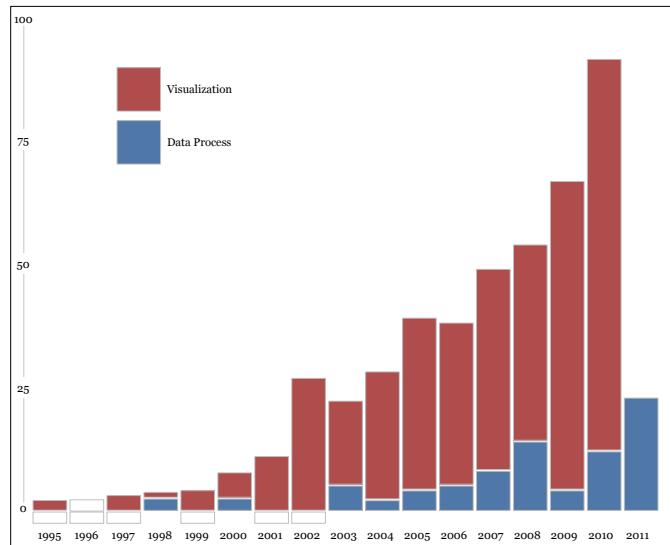
<https://raw.github.com/timrdf/vsr/master/src/xsl/grddl/graphml.xsl>

<https://raw.github.com/timrdf/vsr/master/src/xsl/grddl/graffle.xsl>



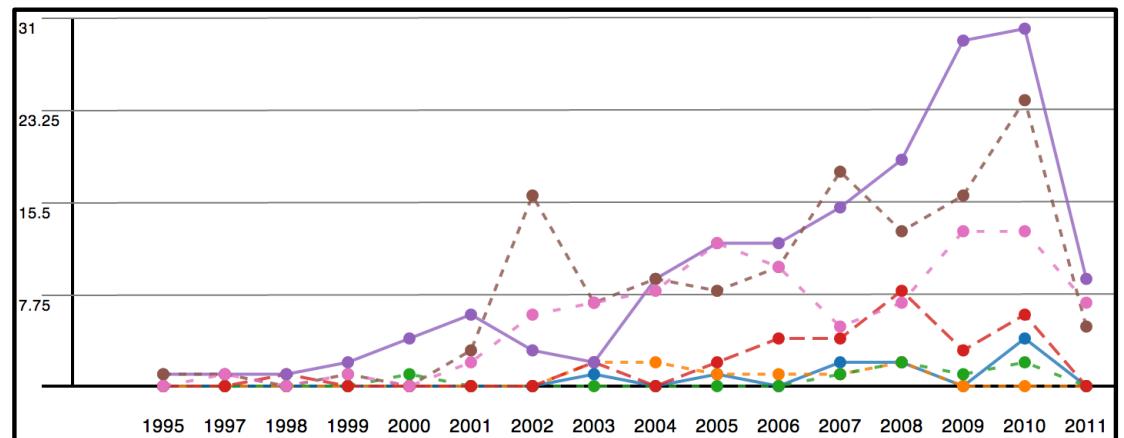


Re-purposing a Content-Preserved Graphic



← Categorizing publications from 1995 to 2011 into **two** categories

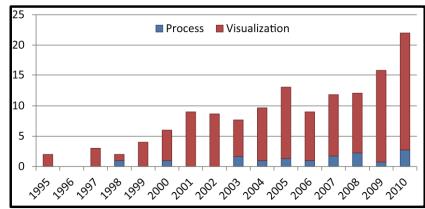
Categorizing publications from 1995 to 2011 into **seven** categories ↓



H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical studies in information visualization: Seven scenarios. *Visualization and Computer Graphics, IEEE Transactions on*, 18(9):1520–1536, Sept. 2012.

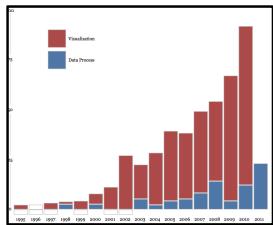


Seeding a Linked Data Crawl from a Content-Preserved Graphic

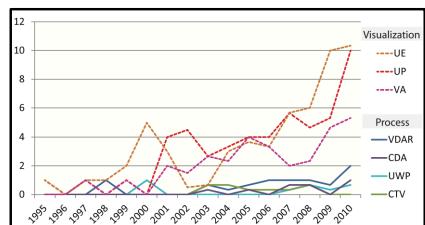
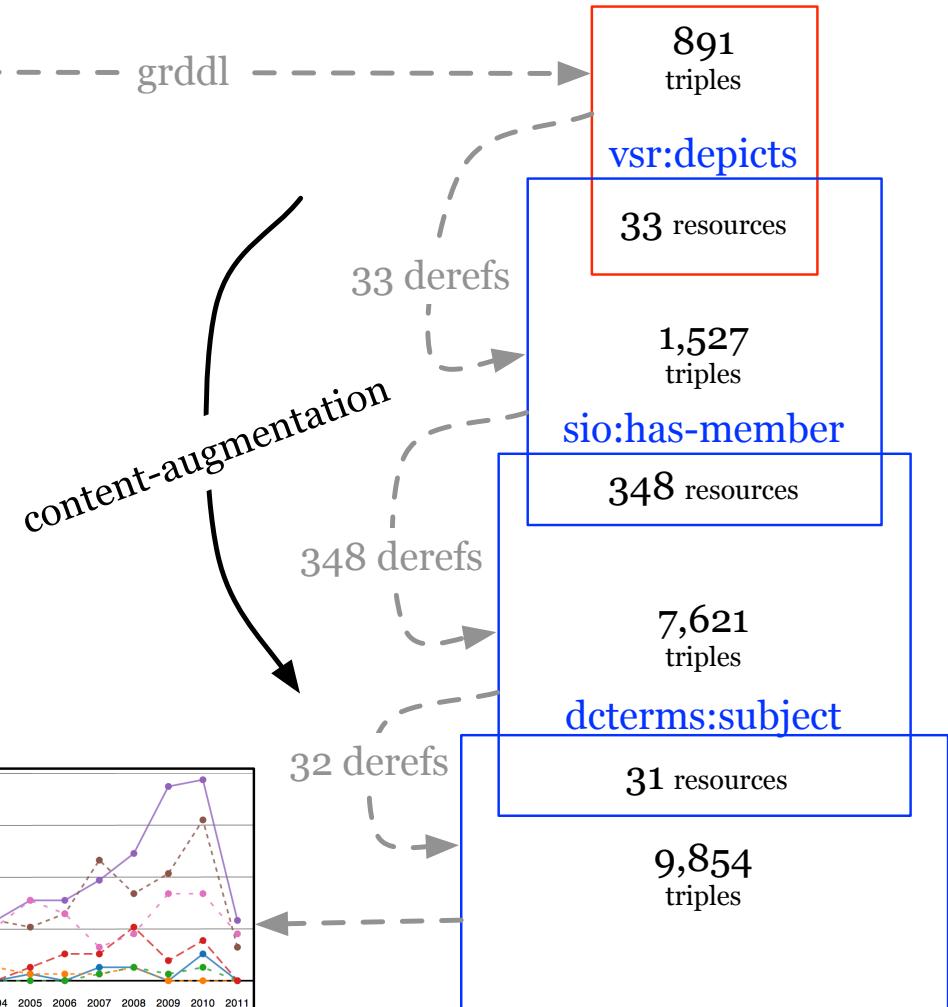


Lam's Figure 1
(original)

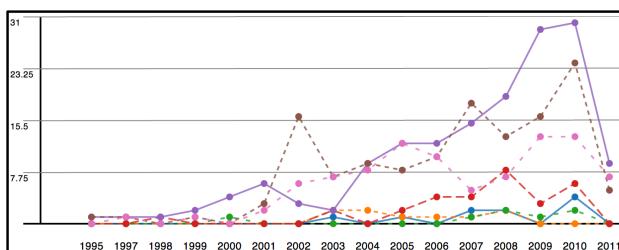
NOT
derivable
↓
(original)



Replica 1



Replica 2





Conclusion

- Content-Preserving Graphics
 - Adds **computations** as additional Linked Data
 - Adds **graphics** as additional Linked Data
 - Adds links from graphics **back to underlying** Linked Data
 - Enables **interrogation** and **re-purposing** of results
- Addresses current challenges in the analytic cycle
 - **Discovery** and **wrangling** of source data
 - Receiving **static** results **shared** through ad hoc channels
 - Finding **relevant** supplemental data



Future Work: What LOD Needs

- Linked Data Composition User Interfaces
 - We need more/better tools to let users pull together certain portions of Linked Data
 - e.g. *crawling the vsr:depicts, sio:has-member dcterms:subject chain to create our timeline*
 - Many crawlers exist, but can't be steered and managed by users at a visual level
 - Many tools assume fixed sources and are tailored to specific content (vocabularies)
 - We're starting a *Data Sculptor*

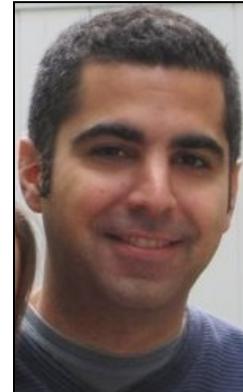


Future Work: What LOD Needs

- Linked ~~Data~~ → Linked *Computation*
 - We need more/better frameworks to allow analysts to derive new RDF from existing RDF
 - e.g. computing *qb:Observations* of *bibo:Documents*
 - SADI (Semantic Automated Discovery and Integration) is a good step, but limited to per-resource
 - csv2rdf4lod’s “SDV” dataset organization, but depends on “black box” computations



Thanks!



- Tim Lebo
- Alvaro Graves
- Patrice Seyed



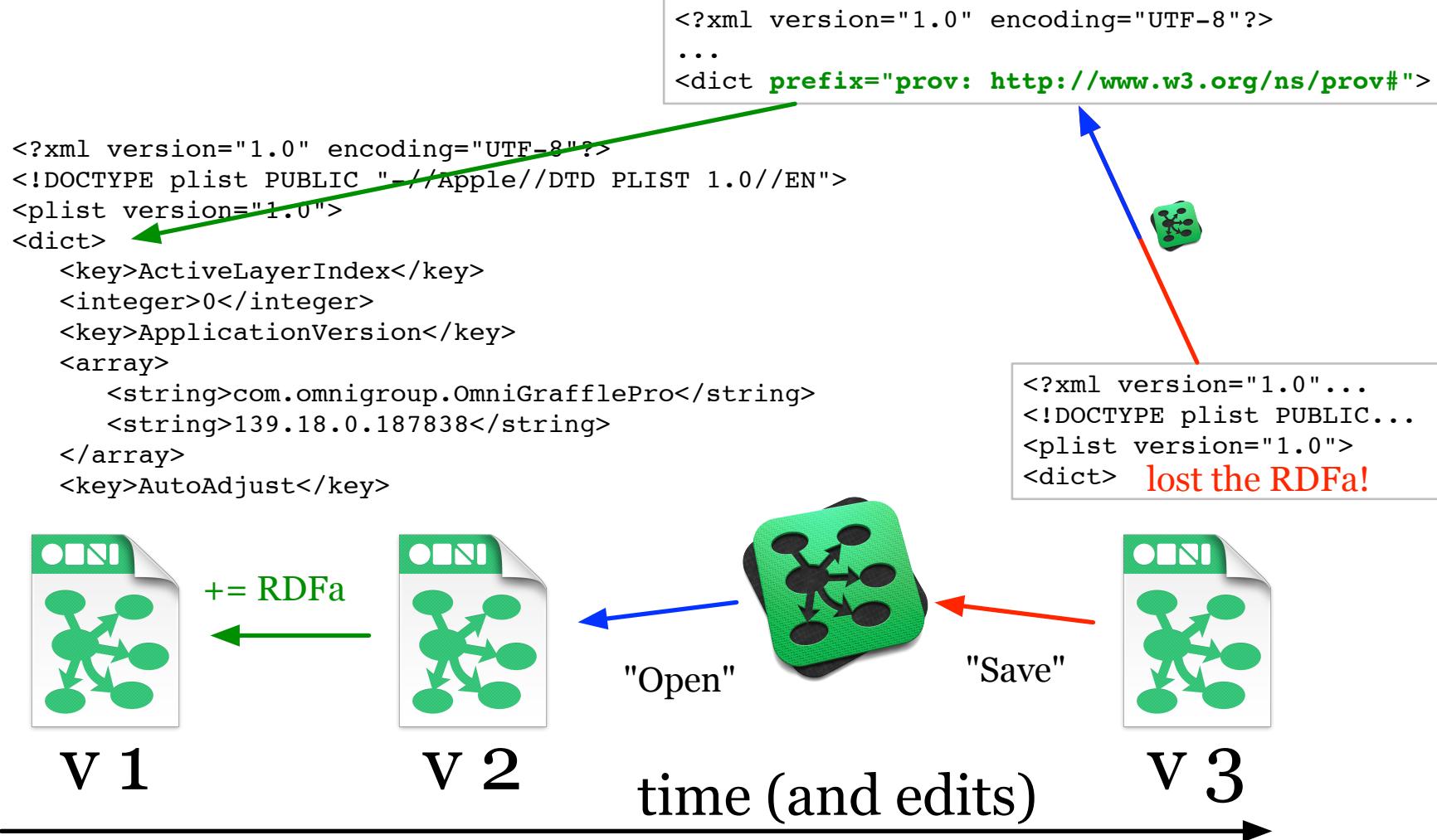
Deborah McGuinness



- Backup slides



Why not just use RDFa?





Bootstrapping

- Bootstrapping the *Content-Preserved Graphic*

