Stories, that Persuade With Data: Some Thoughts on Making Networks of Knowledge.

Anita de Waard
VP Research Data Collaborations
a.dewaard@elsevier.com

Elsevier
Research Data Services

Carnegie Mellon University

IEDA
INTEGRATED EARTH DATA APPLICATIONS

Outline:

- Stories, that persuade with data
- Networks of claims and data
- Research Data Management: some thoughts.

Scientific articles are stories...

The Story of Goldilocks and the Three Bears	***Story***	***Grammar***
Once upon a time	*Time*	*Setting*
a little girl named Goldilocks	*Characters*	
She went for a walk in the forest. Pretty soon, she came upon a house.	*Location*	
She knocked and, when no one answered,	*Goal*	*Theme*
she walked right in.	*Attempt*	
At the table in the kitchen, there were three bowls of porridge.	*Name*	*Episode 1*
Goldilocks was hungry.	*Subgoal*	
She tasted the porridge from the first bowl.	*Attempt*	
This porridge is too hot! she exclaimed.	*Outcome*	
So, she tasted the porridge from the second bowl.	*Activity*	
This porridge is too cold, she said	*Outcome*	

...that persuade (editors/authors/readers!)...

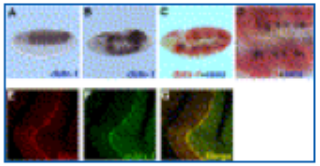
Aristotle	**Quintilian**	
prooimion	Introduction / *exordium*	The introduction of a speech, where one announces the subject and purpose of the discourse, and where one usually employs the persuasive **appeal to ethos** in order to establish credibility with the audience.
prothesis	Statement of Facts/*narratio*	The speaker here provides a narrative account of what has happened and generally explains the nature of the case.
	Summary/ *propostitio*	The propositio provides a brief summary of what one is about to speak on, or concisely puts forth the charges or accusation.
pistis	Proof/ *confirmatio*	The main body of the speech where one offers logical arguments as proof. The **appeal to logos** is emphasized here.
	Refutation/ *refutatio*	As the name connotes, this section of a speech was devoted to answering the counterarguments of one's opponent.
epilogos	*peroratio*	Following the refutatio and concluding the classical oration, the peroratio conventionally employed **appeals through pathos**, and often included a summing up.

… with data.

***Drosophila* and Mammalian Ataxin-1 Physically Interact with Sens and Gfi-1**

Sens and Gfi-1 Are Coexpressed with Atx-1 Homologs in *Drosophila* and Mice

The findings that fly Atx-1 and Sens as well as mammalian Atx-1 and Gfi-1 physically interact prompted us to examine if Atx-1 and Sens/Gfi-1 are coexpressed in vivo. In situ hybridization and Northern analyses show that *datx-1* is expressed in embryonic stages (Figures 3A–3D and data not shown). The expression of *datx-1* is first observed in the dorsolateral region in the stage 5 embryos (Figure 3A). During gastrulation, *datx-1* is expressed in the dorsolateral ectoderm that encompasses the peripheral neuroectoderm (Figure 3B). *sens* mRNA is first expressed in presumptive sensory organ precursor (SOP) cells at stage 10 (Nolo et al., 2000). We found that *sens* is expressed in a subset of cells within the region of *datx-1* expression (Figures 3C and 3D). In mice, Gfi-1 is expressed in many areas that give rise to neuronal cells during embryonic development (Wallis et al., 2003). However, our data show that, in the adult cerebellum, Gfi-1 expression is mainly confined to PCs, where Atx-1 is most abundant (Figures 3E–3G) (Banfi et al., 1996).



Full-size image (90K)
High-quality image (1040K)

Figure 3. Fly and Mouse Atx-1 Colocalize with Sens and Gfi-1 in Certain Cell Types

Figure 2. dAtx-1 and hAtx-1 Interact with Sens and Gfi-1, Respectively

As claims get cited, they become facts:

Voorhoeve et al, Cell, 2006:

To investigate the possibility that miR-372 and miR-373 suppress the expression of LATS2, we... Hypothesis

Therefore, **these results point to** LATS2 as a mediator of the miR-372 and miR-373 effects on cell proliferation and tumorigenicity, Implication

Raver-Shapira et.al, JMolCell 2007 Cited Implication

... two miRNAs, miRNA-372 and-373, function as potential novel oncogenes in testicular germ cell tumors by inhibition of LATS2 expression, **which suggests that** Lats2 is an important tumor suppressor (Voorhoeve et al., 2006).

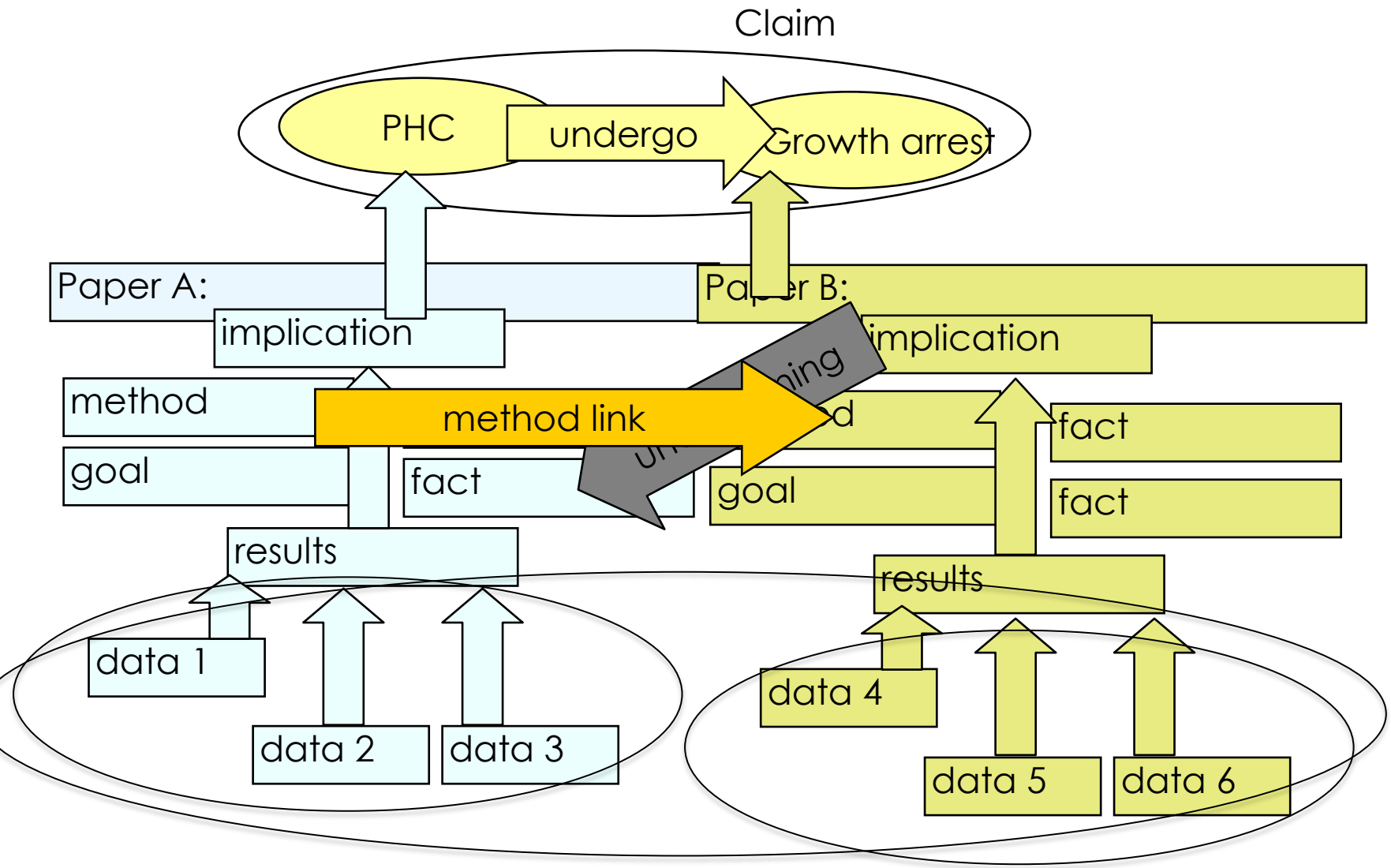
Yabuta, JBioChem 2007: Fact

miR-372 and miR-373 **target** the Lats2 tumor suppressor (Voorhoeve et al., 2006)

There are many problems with this system!

- There are too many papers to read – for people.
- Papers are not written in a way that makes reading easy – for computers.
- Issues with reproducibility.
- Hard to access or assess the data.
- And how do we know how many data points a claim was based on?

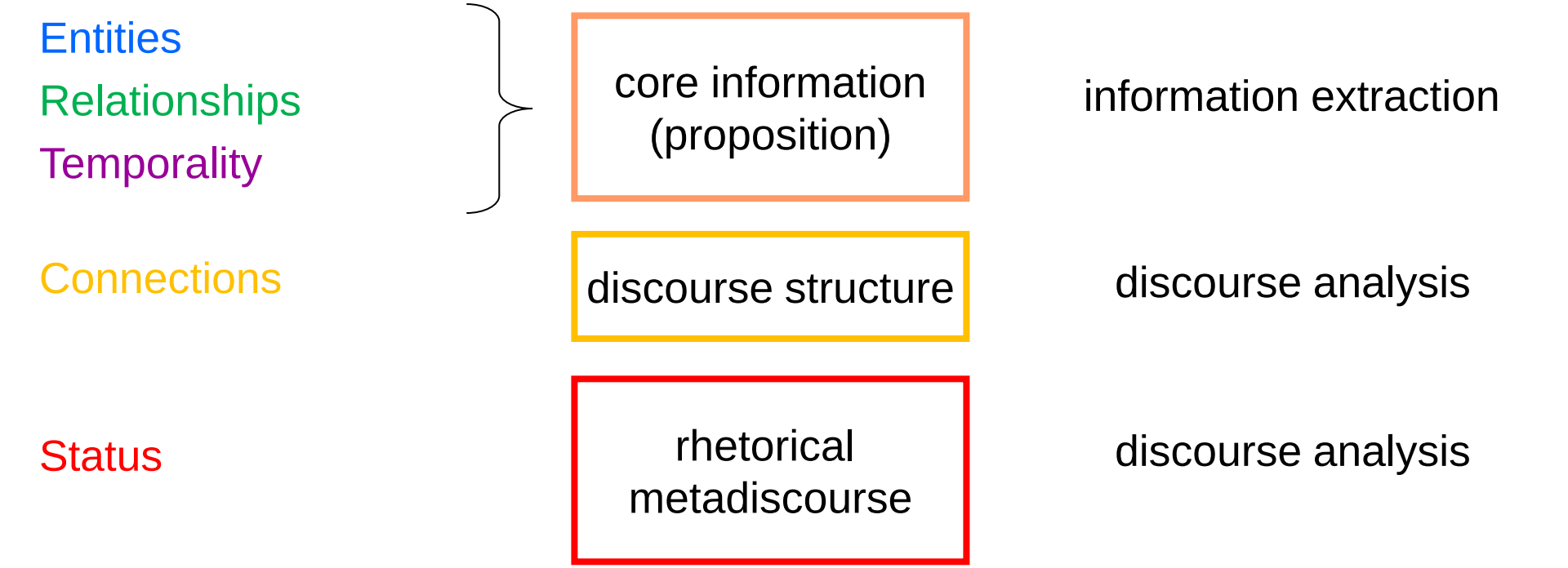
What might work: networks of claims and data:



How do we get there? Find claims:

E.g., scientific discourse analysis:

In contrast with previous hypotheses compact plaques form before significant deposition of diffuse A beta, suggesting that different mechanisms are involved in the deposition of diffuse amyloid and the aggregation into plaques.



Sándor, Àgnes and de Waard, Anita, (2012).

Turn claims into formal representations:

Biological statement with BEL/ epistemic markup	BEL representation:	Epistemic evaluation
These miRNAs neutralize p53-mediated CDK inhibition, possibly through *direct inhibition of the expression of the tumor-suppressor LATS2*.	***r(MIR:miR-372) -|(tscript(p(HUGO:Trp53)) -| kin(p(PFH:"CDK Family")))*** *Increased abundance of miR-372 decreases abundance of LATS2* ***r(MIR:miR-372) -| r(HUGO:LATS2)***	Value = Possible Source = Unknown Basis = Unknown

Biological statement with Medscan/epistemic markup	MedScan Representation:	Epistemic evaluation
Furthermore, we present evidence that the *secretion* of ***nesfatin-1*** into the culture media was dramatically *increased* during the differentiation of *3T3-L1* preadipocytes into *adipocytes* (P < 0.001) and after treatments with TNF-alpha, ***IL-6***, insulin, and dexamethasone (P < 0.01).	**IL-6 → *NUCB2*** (*nesfatin-1*) Relation: *MolTransport* Effect: Positive CellType: *Adipocytes* Cell Line: *3T3-L1*	Value = Probable Source = Author Basis = Data

Use Linked Data to point to claims, and connect them:

the xml is fixed, but the structure is open!

```
<!DOCTYPE chapter
  PUBLIC "-//ES//DTD ehs book DTD version 5.1.1//EN//XML"
  "ehs_book511.dtd" []>
<chapter id="ch1">
  <info>
    <ce:pii>B0-323-01679-0/10003-4</ce:pii>
    <ce:isbn>0-323-01679-0</ce:isbn>
    <ce:copyright type="full-transfer"
      year="2003">Mosby, Inc.</ce:copyright>
  </info>
  <ce:floats>
    ...
  </ce:floats>
  <ce:label>Chapter 1</ce:label>
  <ce:title>Core Issues in Primary Care</ce:title>
  <ce:author-group>
    ...
  </ce:author-group>
  <ce:intro>
    <ce:para>Text of opening paragraph...</ce:para>
  </ce:intro>
  <ce:sections>
    <ce:section id=#123>
      <ce:section-title>Core Issues in Primary Care
        Today</ce:section-title>
      <ce:para>Text of opening paragraph...</ce:para>
      <ce:para>Text second paragraph...</ce:para>
    </ce:section>
    <ce:section>
      <ce:label>1.2</ce:label>
      <ce:section-title>Core Issues</ce:section-title>
      <ce:para>Text of opening paragraph...</ce:para>
      <ce:para>Text second paragraph...</ce:para>
    </ce:section>
  </ce:sections>
  <ce:bibliography>
    ...
  </ce:bibliography>
</chapter>
```

allows for layers of annotation

but we all know
she was deluded then

said @anitawaard
on January 9, 2014

this says

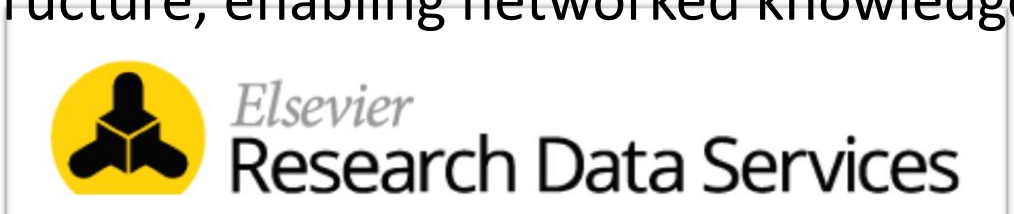
mice like cheese

What about the data?

- Can we see it?
- How can we evaluate it?
- Can it be reproduced?
- Can we combine or compare data points?

Elsevier Research Data Services

- Goals:
  - Increase data preservation: quantity and quality
  - Improve data use: by and for authors, readers, and lay people
  - Enhance interoperability: between systems, journals, databases
  - Help develop a sustainable data infrastructure.
- Guiding principles:
  - In principle, all data stays open
  - Work with existing repositories and tools (so URLs, front end etc stay where they are)
  - 2013/2014: Series of pilots and questionnaires to drive data strategy/data policy and contribute optimally to an integrated data infrastructure, enabling networked knowledge.



Research data management today:

Using antibodies

and squishy bits

Grad Students experiment

and enter details into their lab notebook.

The PI then tries to make sense of their slides,

and writes a paper.

End of story.

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

A simple method of *in vitro* electroporation allows visualization, recording, and calcium imaging of local neuronal circuits☆

Kenneth R. Hovis[a,b], Krishnan Padmanabhan[a,b], Nathaniel N. Urban[a,b,c,*]

[a] Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, United States
[b] Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA 15213, United States
[c] Center for Neuroscience, University of Pittsburgh, Pittsburgh, PA 15260, United States

ARTICLE INFO

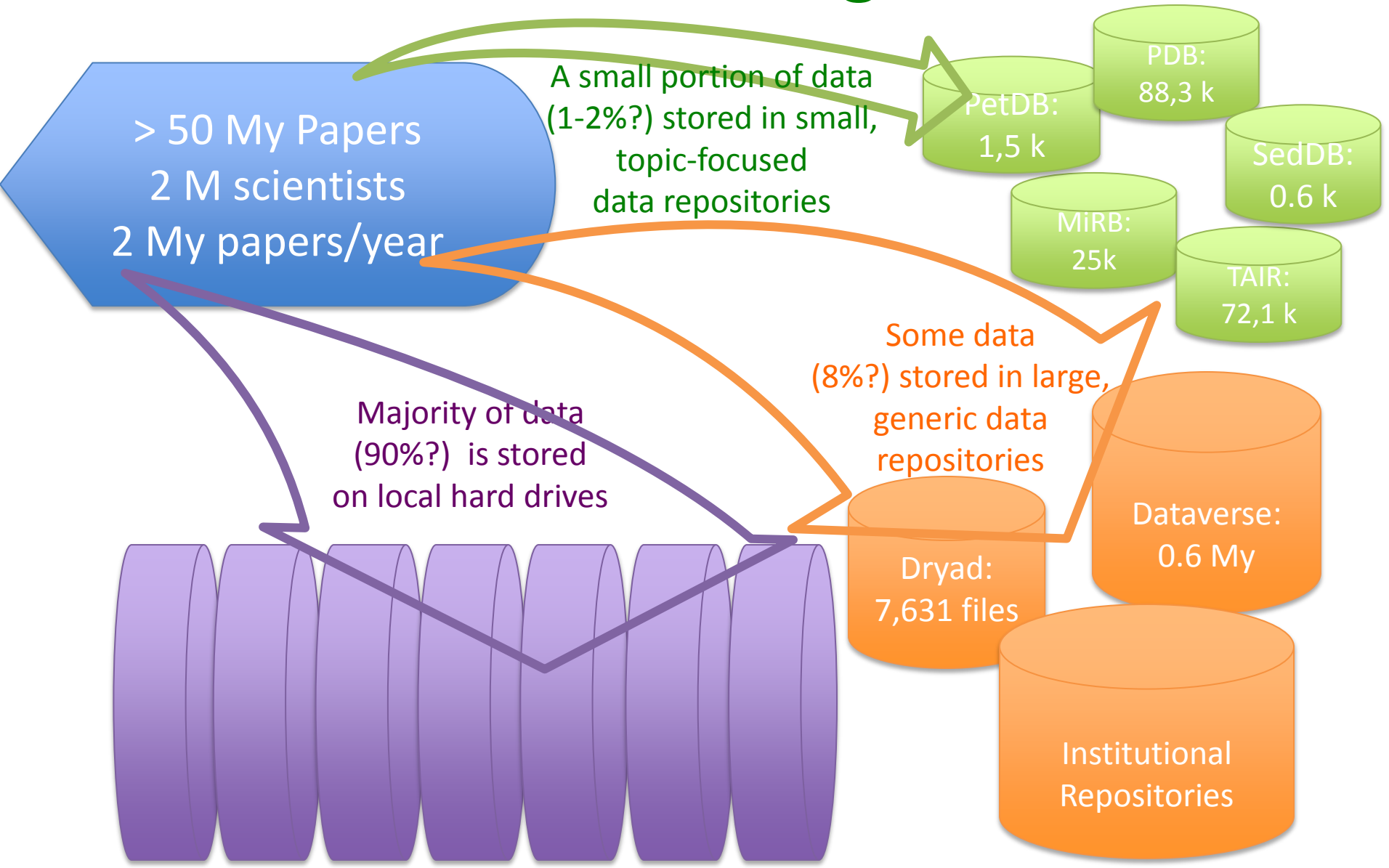
Article history:
Received 3 March 2010
Received in revised form 17 May 2010
Accepted 24 May 2010

Keywords:

ABSTRACT

Since Cajal's early drawings, the characterization of neuronal architecture has been paramount in u
standing neuronal function. With the development of electrophysiological techniques that pr
unprecedented access to the physiology of these cells, experimental questions of neuronal function
also become more tractable. Fluorescent tracers that can label the anatomy of individual or popula
of neurons have opened the door to linking anatomy with physiology. Experimentally however, cu
techniques for bulk labeling of cells *in vitro* often affect neuronal function creating a barrier for ex

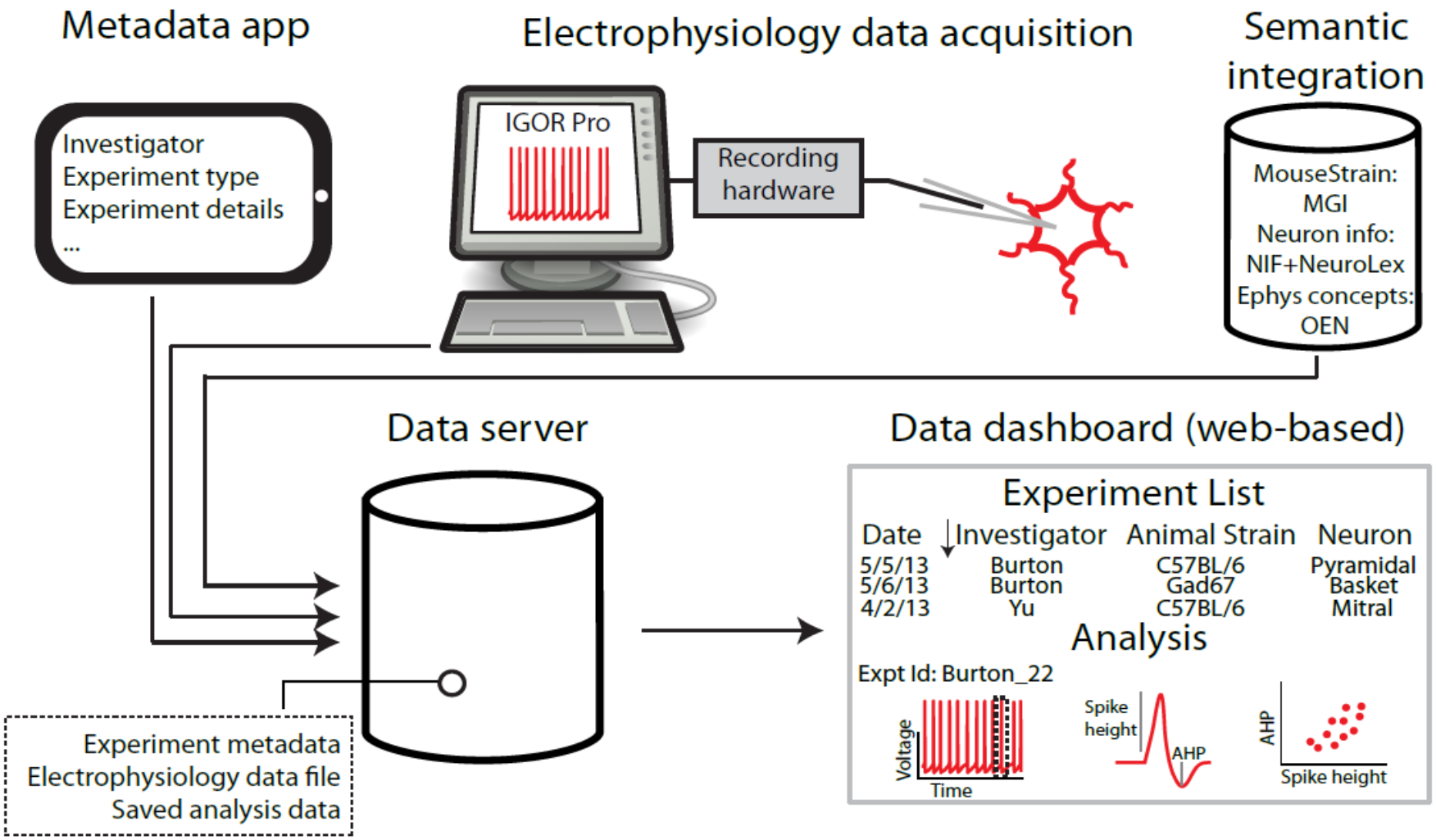
Where research data goes now:



An Urban Legend is born:

- How can we make a standard neuroscience wet lab more data-sharing savvy?
- Incorporate structured workflows into the daily practice of a typical electrophysiology lab (the Urban Lab at CMU)
  - What does it take?
  - Where are points of conflict?
- 1-year pilot, funded by Elsevier RDS:
  - CMU: Shreejoy Tripathy, manage/user test
  - Elsevier: development, UI, project management

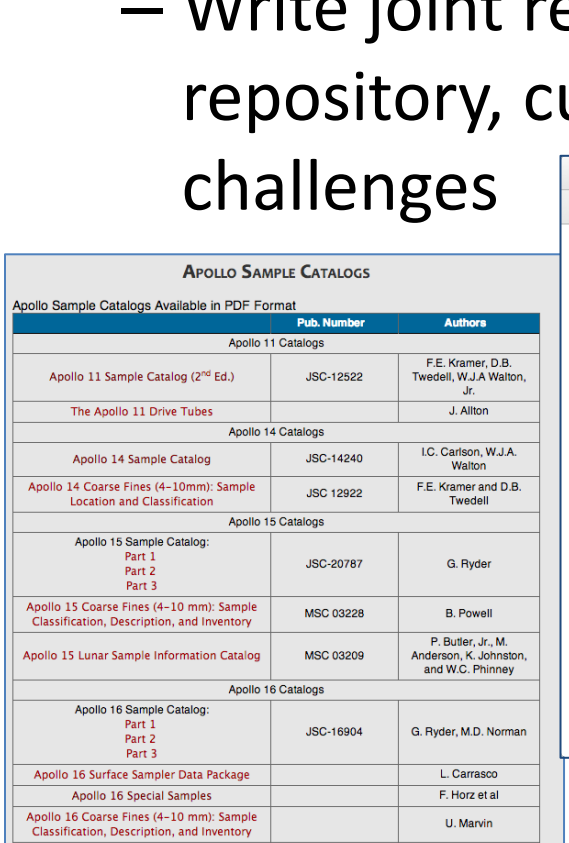
Annotating data during experimentation:



What does high-quality data curation take?

Pilot project with IEDA:

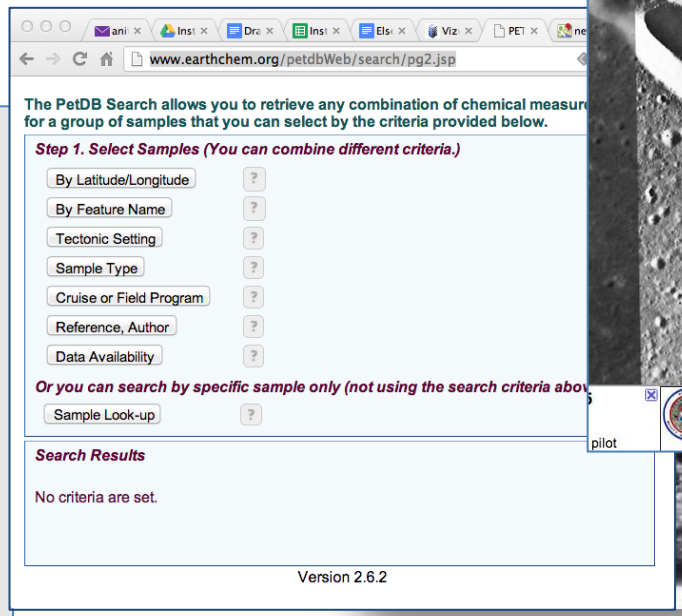
- Build a database for lunar geochemistry
- Write joint report on building repository, curation, costs and challenges

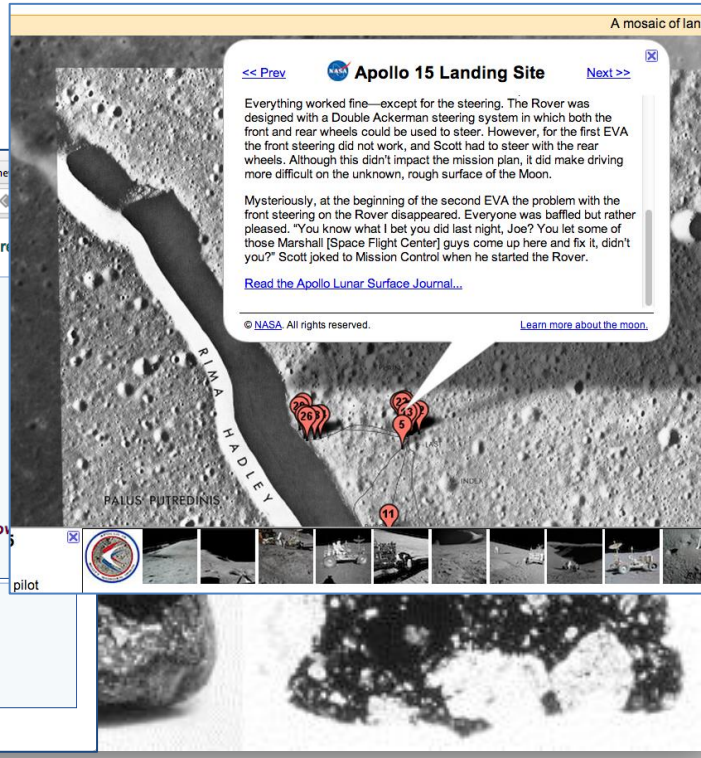


Apollo Sample Catalogs

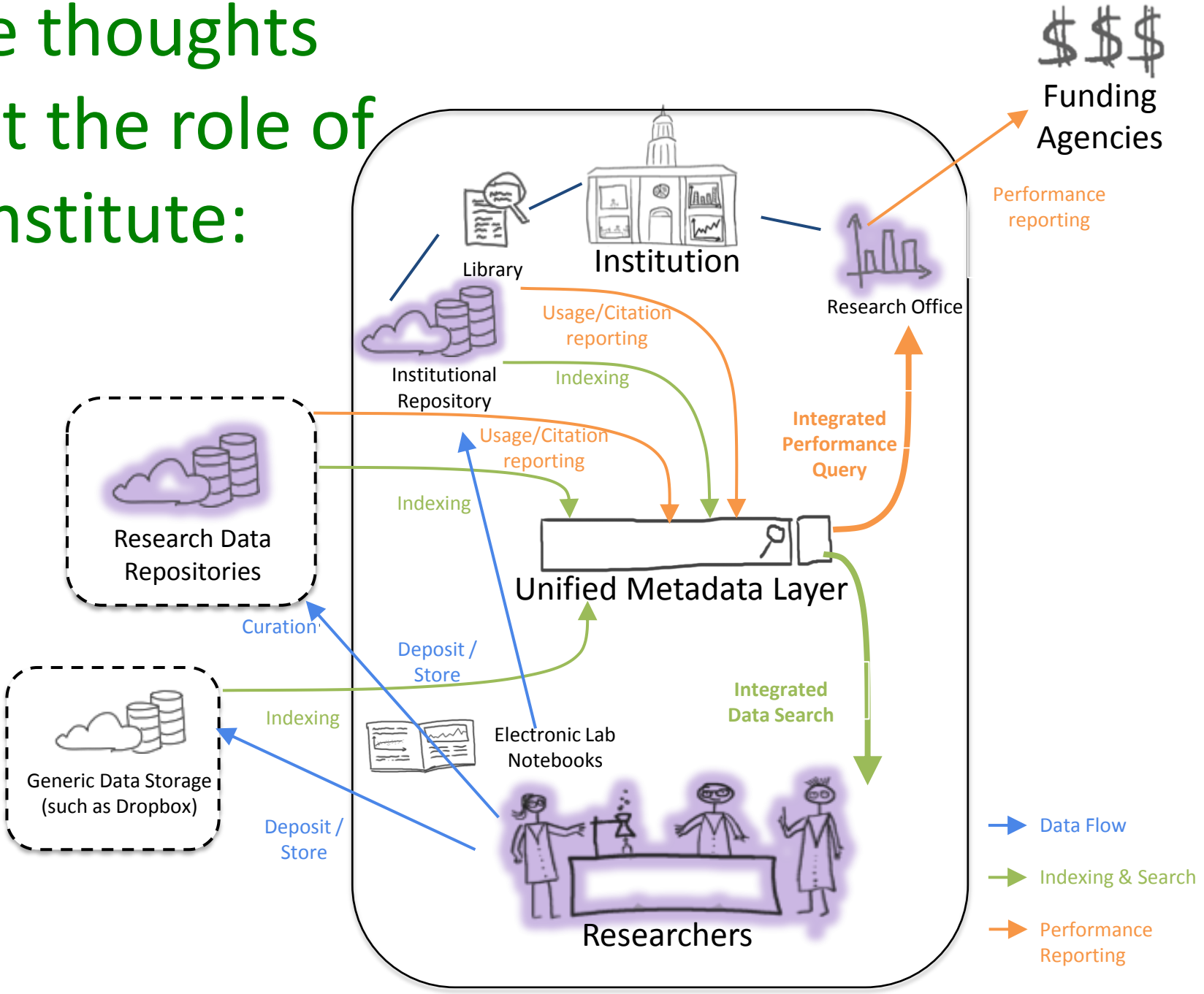
Apollo Sample Catalogs Available in PDF Format

	Pub. Number	Authors
Apollo 11 Catalogs		
Apollo 11 Sample Catalog (2nd Ed.)	JSC-12522	F.E. Kramer, D.B. Twedell, W.J.A Walton, Jr.
The Apollo 11 Drive Tubes		J. Allton
Apollo 14 Catalogs		
Apollo 14 Sample Catalog	JSC-14240	I.C. Carlson, W.J.A. Walton
Apollo 14 Coarse Fines (4–10mm): Sample Location and Classification	JSC 12922	F.E. Kramer and D.B. Twedell
Apollo 15 Catalogs		
Apollo 15 Sample Catalog: Part 1 Part 2 Part 3	JSC-20787	G. Ryder
Apollo 15 Coarse Fines (4–10 mm): Sample Classification, Description, and Inventory	MSC 03228	B. Powell
Apollo 15 Lunar Sample Information Catalog	MSC 03209	P. Butler, Jr., M. Anderson, K. Johnston, and W.C. Phinney
Apollo 16 Catalogs		
Apollo 16 Sample Catalog: Part 1 Part 2 Part 3	JSC-16904	G. Ryder, M.D. Norman
Apollo 16 Surface Sampler Data Package		L. Carrasco
Apollo 16 Special Samples		F. Horz et al
Apollo 16 Coarse Fines (4–10 mm): Sample Classification, Description, and Inventory		U. Marvin





Some thoughts about the role of the institute:



FORCE11
The Future of Research Communications and e-Scholarship

Data Initiatives:

- Data Citation group:
  - Synthesize principles of proper data citation
  - 'Declaration of Data Citation Principles', 8 principles of successful data citation -http://www.force11.org/datacitation
- Resource Identification Initiative:
  - Promote research resource identification, discovery, and reuse
  - Resource Identification Portal http://scicrunch.com/resources
  - Central location for obtaining research resource identifiers (RRIDs) for materials and software used in biomedical research
    - Antibody: Abgent Cat# AP7251E, ABR:AB_2140114
    - Tool: CellProfiler Image Analysis Software, NIFRegistry:nif-0000-00280
    - Organism: MGI:MGI:3840442

Resource Identification Initiative

In summary:

- Stories, that persuade with data:
  - We need better ways to communicate science!
- Networks of claims and data:
  - Promising steps towards identifying claims
  - Entity identification and Linked Data helps
  - Problem: access to data
- Research Data Management:
  - Key issue: get data in up-front
  - Evaluate and scale up role of repositories
  - Codevelop view of role of institutions/libraries

Questions?

Anita de Waard
VP Research Data Collaborations
a.dewaard@elsevier.com

http://researchdata.elsevier.com/

Thank you!

Collaborations and discussions gratefully acknowledged:

- CMU: Nathan Urban, Shreejoy Tripathy, Shawn Burton, Rick Gerkin, Santosh Chandrasekaran, Matthew Geramita, Eduard Hovy
- UCSD: Phil Bourne, Brian Shoettlander, David Minor, Declan Fleming, Ilya Zaslavsky
- NIF/Force11: Maryann Martone, Anita Bandrowski
- OHSU: Melissa Haendel, Nicole Vasilevsky
- California Digital Library: Carly Strasser, John Kunze, Stephen Abrams
- Elsevier: Mark Harviston, Jez Alder, David Marques