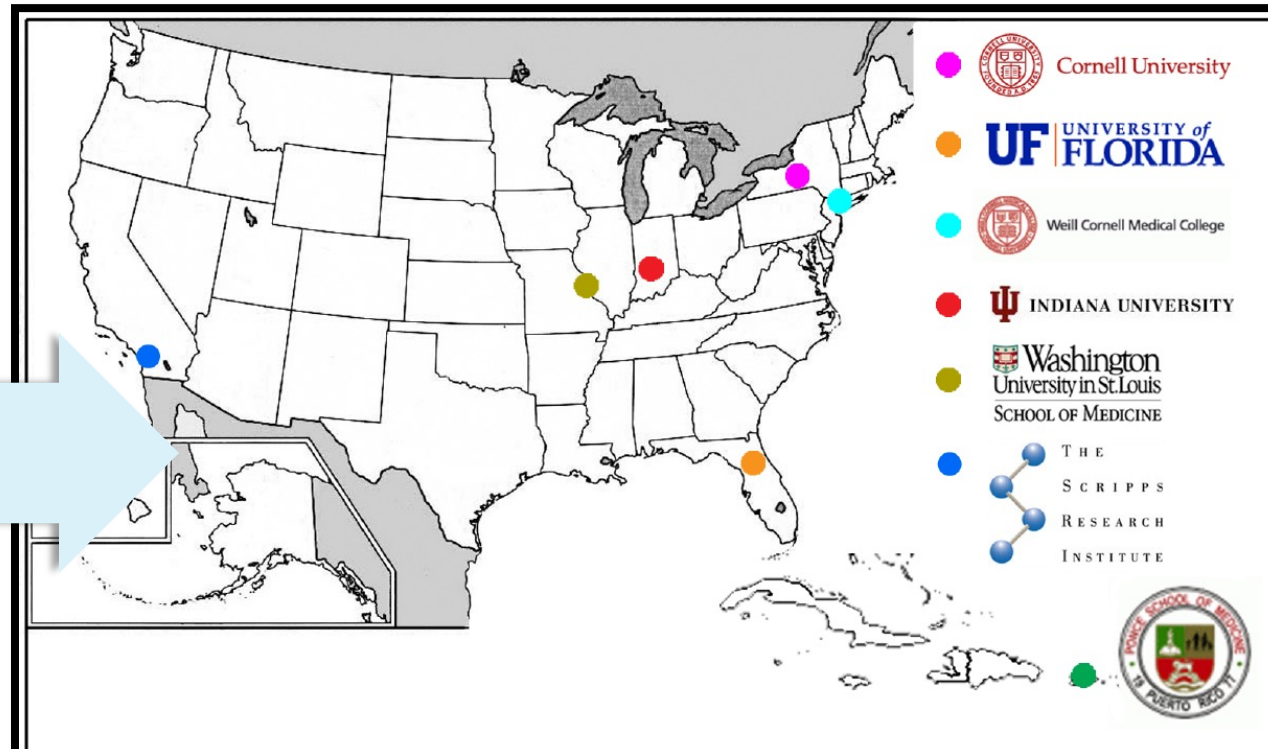# VIVO

Enabling National Networking of Scientists

VIVO and Linked Open Data
December 13, 2010

Dean B. Krafft
Chief Technology Strategist and Director of IT
Cornell University Library

VIVO Origins and Current Status

- Originally developed at Cornell University in 2004 to support Life Sciences.
- Reimplemented using RDF, OWL, Jena and SPARQL in 2007.
- Now covers all faculty, researchers and disciplines at Cornell.
- Implemented at University of Florida in 2007.
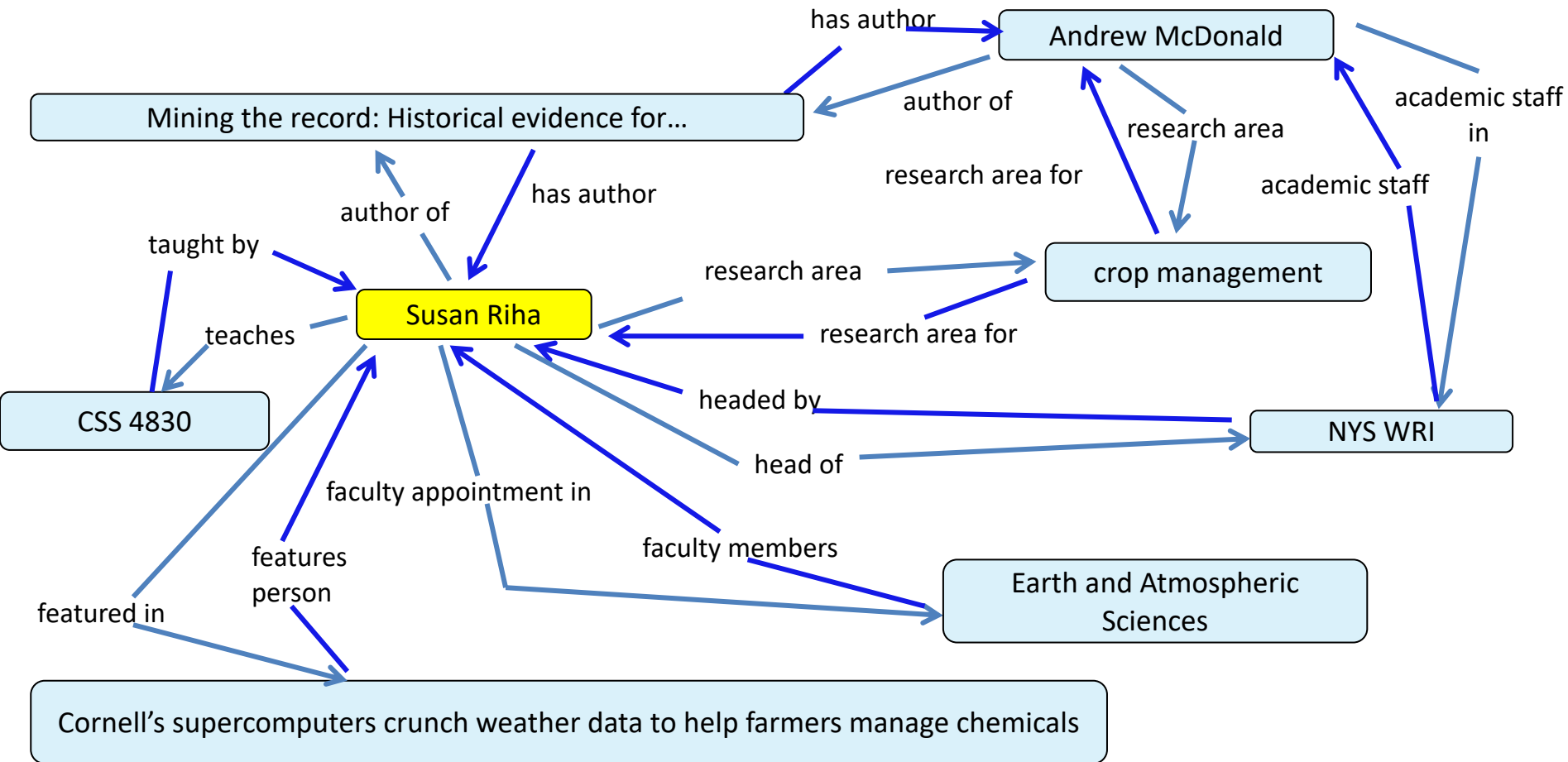- Underlying system in use at Chinese Academy of Sciences and Australian Universities.

In September 2009, seven institutions received $12.2 million in funding from the National Center for Research Resources of the NIH to to enable **National Networking with VIVO**

Cornell University

UF | UNIVERSITY of FLORIDA

Weill Cornell Medical College

INDIANA UNIVERSITY

Washington University in St.Louis
SCHOOL OF MEDICINE

THE SCRIPPS RESEARCH INSTITUTE

Data in VIVO

- Stored in **Resource Description Framework (RDF)** triples.

- Uses the **shared VIVO Core Ontology** to describe people, organizations, activities, publications, events, interests, grants, and other relationships

- VIVO Core Ontology extends Friend-of-a-Friend (FOAF) and Bibliographic Ontology (BIBO)

- Supports local ontology extensions for institution-specific needs.

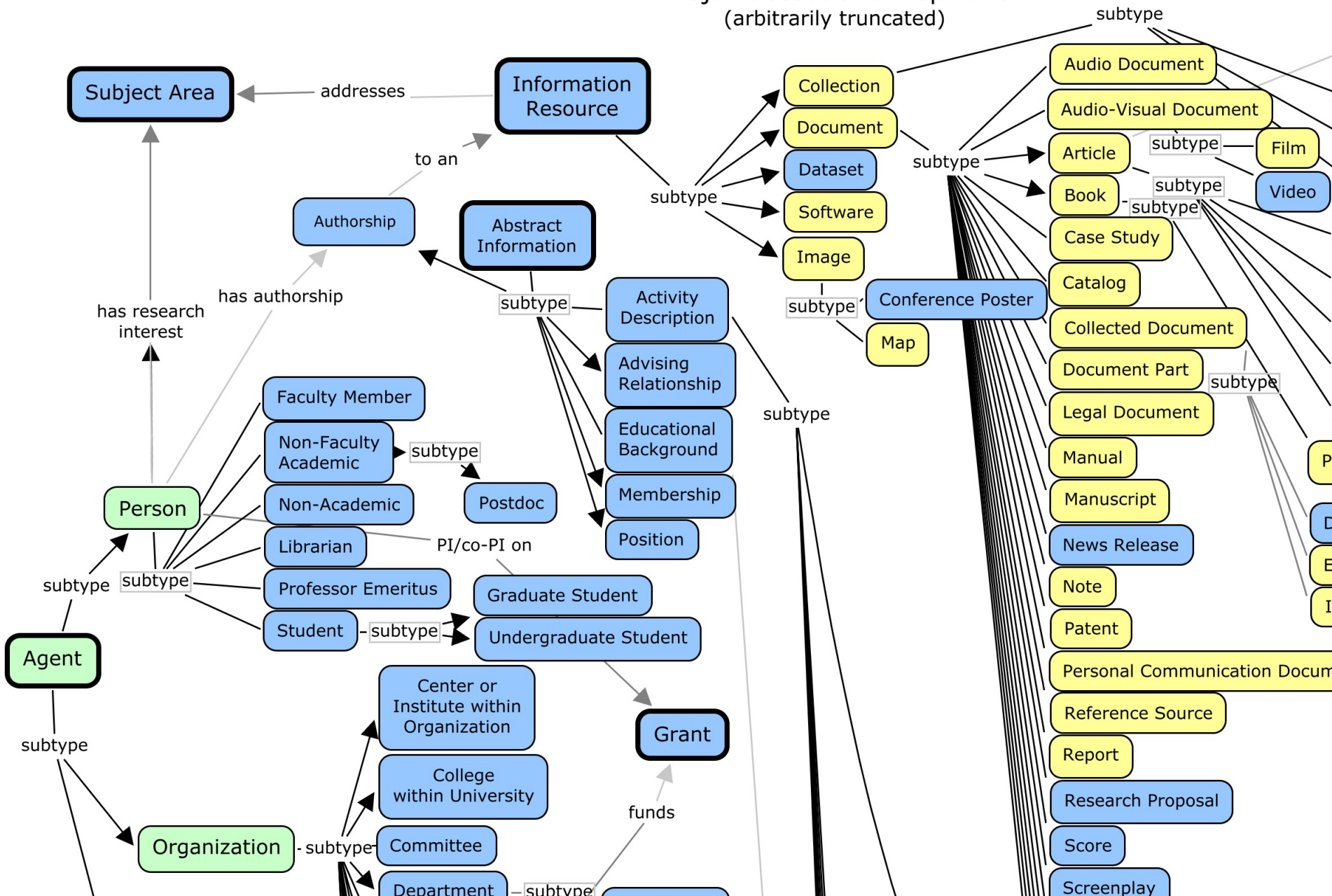- Project will develop mappings to other standard ontologies

Detailed relationships for a researcher

VIVO Core Ontology
Major Classes and Properties
(arbitrarily truncated)

FOAF | VIVO Core | BIBO | Event

- Subject Area
- Information Resource
- Authorship
- Abstract Information
- Person
- Agent
- Organization
- Subject Area ← addresses ← Information Resource
- to an
- has authorship
- has research interest

Subtypes / Classes:

- Collection
- Document
- Dataset
- Software
- Image
- Audio Document
- Audio-Visual Document
- Article
- Film
- Book
- Video
- Case Study
- Catalog
- Collected Document
- Conference Poster
- Map
- Document Part
- Legal Document
- Manual
- Manuscript
- News Release
- Note
- Patent
- Personal Communication Docum...
- Reference Source
- Report
- Research Proposal
- Score
- Screenplay

Person subtypes:

- Faculty Member
- Non-Faculty Academic
- Non-Academic
- Librarian
- Professor Emeritus
- Student
- Postdoc
- Graduate Student
- Undergraduate Student

Abstract Information subtypes:

- Activity Description
- Advising Relationship
- Educational Background
- Membership
- Position

- PI/co-PI on
- Grant
- funds

Organization subtypes:

- Center or Institute within Organization
- College within University
- Committee
- Department

How can you get our RDF?

- Currently get RDF associated with a single individual with that individual's URI

- Next release will have an RDF-enabled index page to allow crawling of all instances and access to all RDF

- Should we make available a site map in some form? A download of all the triples?

- We do not make available a public SPARQL endpoint – not reliable and too vulnerable to misuse

http://vivo.library.cornell.edu/ns/0.1#:CornellAcademicEmployee

foaf:Agent

http://vivo.library.cornell.edu/ns/0.1#:CornellAffiliatedPerson

http://vivoweb.org/ontology/core#:NonFacultyAcademic

foaf:Person

http://vivo.library.cornell.edu/ns/0.1#:CornellEmployee

http://vivo.library.cornell.edu/ns/0.1#:CornellAcademicStaff

owl:Thing

http://vivo.library.cornell.edu/ns/0.1#:affiliatedWithAsAcademicStaffMember

http://vivo.cornell.edu/individual/individual390
http://vivoweb.org/ontology/core#:authorInAuthorship

http://vivo.cornell.edu/individual/n9217
http://vivoweb.org/ontology/core#:featuredIn

http://vivo.cornell.edu/individual/CornellsVIVOconceptwillexpandtoconnectresearchersnationwide
rdfs:label
      Krafft, Dean B.
http://vivoweb.org/ontology/core#:hasPrincipalInvestigatorRole

http://vivo.cornell.edu/individual/n2907098
http://vivoweb.org/ontology/core#:hasPrincipalInvestigatorRole

http://vivo.cornell.edu/individual/n3482598
http://vivo.library.cornell.edu/ns/0.1#:presentations
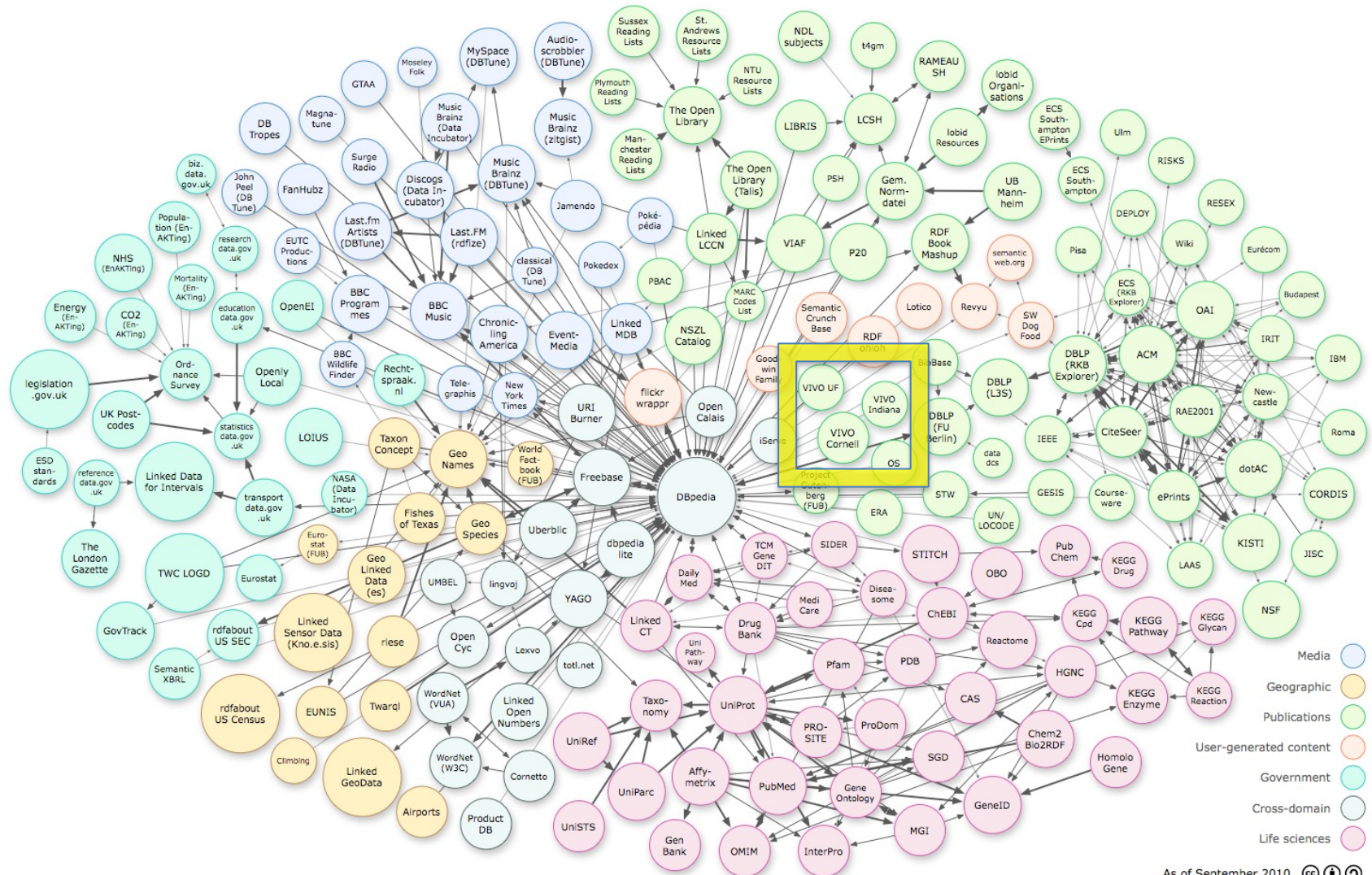   1. Wrenching Change: Why a Research Library Needs a Chief Technology Strategist, Institute for Computer Policy and Law, Ithaca, NY
      2. NSDL: Creating an Extensible, Collaborative, Community-based Digital Library for STEM Education, NSF EHR Cyberinfrastructure for
      3. Organizing the Digital Library, Joint DELOS/NSDL Summer School on Digital Libraries, Florence, Italy, (May 2007).
      4. Semantic Digital Libraries, WWW 2007, Banff, Alberta, Canada (with Sandy Payette, et al.), (May 2007).
      5. NSDL 2.0: Creating a Collaborative Digital Library, Coalition for Networked Information Spring 2007 Task Force Meeting, Phoenix, A

VIVO enables authoritative data about researchers to join the Linked Data cloud

But if we're authoritative, what do we link to?

- Currently link to Geographic URIs in DBpedia
- Reluctant to link to uncontrolled information about VIVO researchers (e.g. Wikipedia entries)
- Will assert Same-As relationship to author identifiers (e.g. ORCID)
- Willing to link publications to authoritative repositories (e.g. PubMed)
- Willing to link grants to authoritative sources (e.g. NIH Reporter), but they'll need to provide permanent URIs
- Willing to link to Medical Subject Headings (MESH) ontology terms (currently fixed URIs, but not RDF) and other large public ontologies/controlled vocabularies

VIVO/LOD Challenges

- **Privacy:** VIVO aggregates a lot of information about people and institutions – it's all nominally public, but …

- **Using LOD ontologies:** our approach is to define what we need, use what is most obvious, and let others do the mapping

- **Presentation:** Faculty may want to present a subset of publications and grants – how does the presentation version differ from the LOD version?

- **Dirty data:** We've just spent six months cleaning and disambiguating people, terms, publications, etc. from Activity Insight

- **Provenance:** Internally, we use private graphs to associate data with sources, but hard to expose this information as LOD

- **Temporality:** Easy to state a fact – harder to state *when* it was true

VIVO/LOD Opportunities

- **Usability:** The VIVO Ontology makes all of an institution's faculty/researcher information available in a self-describing, structured format

- **Reusability:** Information can be repurposed within the institution: department portals, entrepreneurship portal, faculty impact statements, graduate education portal

- **Extensibility:** University of Melbourne has created a data registry system based on VIVO with an extended ontology

- **Integration:** Used OpenCalais developer libraries to annotate Cornell news release with VIVO URIs as a test

- **Delivery:** VIVO is a source for faculty/researcher profiles and context for other systems: data curation; research facilities/resources; publications