



BioRDF: Seeding the Semantic Web

Susie Stephens, Oracle

BioRDF Charter

- Build a life sciences demo that spans from bench to bedside using RDF & OWL to help scientist better understand the value of the Semantic Web
- Explore the effectiveness of current tools for making data available as RDF
- Document our finding to help accelerate the adoption of the Semantic Web by others

BioRDF Participants

John Barkley (NIST), Olivier Bodenreider (NLM), Kyle Bruck (Stanford University), Vinay Chaudhri (SRI), Kei Cheung (Yale University), Tim Clark (MGH), Roger Cutler (Chevron), Don Doherty (Brainstage Research), Yong Gao (MGH), Brian Gilman (Panther Informatics), Carole Goble June Kinoshita (AlzForum), Nikesh Kotecha (Stanford University), Ora Lassila (Nokia), Joanne Luciano (Predictive Medicine), M. Scott Marshall (University of Amsterdam), Robin McEntire (GSK), Brian Osborne (BioPerl), Christophe Poulain (Teranode), Daniel Rubin (Stanford University), Alan Ruttenburg (Millennium Pharmaceuticals), Matthias Samwald (Medical University of Vienna), Andy Seaborne (HP), Matt Shanahan (Teranode), Karen Skinner (NIH), Susie Stephens (Oracle), Charles Tilford (BMS), John Wilbanks (Creative Commons), Elizabeth Wu (AlzForum), Davide Zaccagnini (Language and Computing).

Neuroscience Focus

- Heterogeneous Data
 - Molecules to nervous system
 - Numerous Web resources
 - Effective data sharing and integration needed
- Heterogeneous Interfaces
- Disease Focus
 - Huntingtons
 - Alzheimers
 - Parkinsons

Name	Description	URL
1 ALZForum	Alzheimer forum	http://www.alzforum.org/
2 ASTD	Audimotor Spike Train Database	http://repositories.colib.org/mrcr/1
3 BAMS	Brain Architecture Management System	http://brancusi.usc.edu/bkms
4 BayGenomics	BayGenomics	http://baygenomics.ucsf.edu/
5 BCDB	Brain Collection Database	http://www.brains.tad.msu.edu
6 BIRN	Biomedical Informatics Research Network	http://www.nbim.net/index.htm
7 BIRN-CC	BIRN Coordinating Center	http://www.nbim.net/TestBeds/CoordinatingCenter/index.htm
8 Brain Atlases	Brain atlases	http://brainatlases.ucdavis.edu
9 BrainInfo	BrainInfo	http://braininfo.rprc.washington.edu
10 BrainMap	BrainMap	http://brainmap.org/
11 BrainMaps.org	High Resolution Primate Brain Atlases	http://brainmaps.org
12 BrainML	BrainML	http://brainml.org
13 BrainWeb	BrainWeb	http://www.bic.mni.mcgill.ca/brainweb/
14 Brede	Brede database	http://hendrix.imm.dtu.dk/services/ferme/brede/
15 Catacomb	Modeling Software for Neuroscience	http://www.enorg.org/
16 CCDB	Cell Centered Database	https://ccdb.ucsd.edu/CCDB/
17 CellPropDB	Database for cell properties	http://senselab.med.yale.edu/senselab/CellPropDB
18 ChannelDB	ChannelDB	http://www.modelersworkspace.org/channeldb/ChannelDB.html
19 Chromux	Chromux is a software platform for the analysis of neurobiological time series data.	http://chromux.org
20 CoCoDat	Collation of Cortical (cell and microcircuitry) Data	http://www.cocomac.org/cocodat
21 CoCoMac	Collation of Connectivity data on the Macaque brain	http://cocomac.org/home.htm
22 Comprehensive		http://db-database.org

BioRDF Tasks

ESW Wiki [Login](#)

HCLSIG BioRDF Subgroup Tasks

[FrontPage](#) [RecentChanges](#) [FindPage](#) [HelpContents](#) **Tasks**

[Edit \(Text\)](#) [Edit \(GUI\)](#) [Info](#) [Attachments](#) [More Actions:](#)

Active Tasks

- [/Reagents \(Status\)](#)
- [/SenseLab](#)
- [/Using SW Technologies to Find Small Molecules that Bind to Proteins](#)
- [/Entrez Gene to RDF](#)
- [/OMIM Neural diseases](#)
- [/Natural Language Processing and RDF](#)
- [/Ligand-Receptor Interaction, Molecular Interaction Networks, Ontology Evolution](#)
- [/Vocabulary Requirements](#)
- [/URI Best Practices](#)

Proposed Tasks

- [/Brain Connectivity](#)
- [/Brain Atlas condition to scans](#)
- [/Protein Neural related protein data](#)
- [/Ruby On Rails and ActiveRDF](#)

Converting biological information to the W3C Resource Description Framework (RDF): Experience with Entrez Gene

Report

Lister Hill National Center for Biomedical
Communications (NLM/ NIH)

By: Satya S. Sahoo

Email: sahoo@cs.uga.edu

Under the supervision of: Dr. Olivier Bodenreider

1. Abstract

The Entrez Gene (EG) database stores gene related data from sequenced genomes and of model organisms that are focus of active research [1]. We describe our experience in transforming the EG database into W3C Resource Description Framework (RDF) [2]. Our work is part of a larger effort to create a biomedical repository comprising not only information from structured resources (database and knowledge bases), but also from biomedical text (e.g., MEDLINE), of which information is extracted by SemRep [3]. Using the eXtensible Stylesheet Language Transformation (XSLT) [4] approach, we mapped the element tags of the EG XML representation to more intuitive relationship names manually, and used them during the automatic conversion to RDF. Finally, we store this RDF version of EG in the Oracle 10g [5] relational database with specific support for storing and querying of native RDF data.

Semantic Web Meets e-Neuroscience: An RDF Use Case

Hugo Y.K. Lam¹, Luis Marengo^{2,3}, Tim Clark^{8,9}, Yong Gao⁹, June Kinoshita¹⁰
Gordon Shepherd⁴, Perry Miller^{2,3,5}, Elizabeth Wu¹⁰, Gwen Wong¹⁰, Nian Liu^{2,3}
Chiquito Crasto^{2,4}, Thomas Morse⁴, Susie Stephens¹¹, and Kei-Hoi Cheung^{2,3,6,7}

¹Program in Computational Biology and Bioinformatics, ²Center for Medical Informatics

³Anesthesiology, ⁴Neurobiology, ⁵Molecular, Cellular and Developmental Biology

⁶Genetics, ⁷Computer Science, Yale University, New Haven, CT, USA

⁸Harvard University, Initiative in Innovative Computing, Cambridge, MA, USA

⁹Massachusetts General Hospital, Boston, MA, USA, ¹⁰Alzheimer Research Forum, ¹¹Oracle

{Hugo.YK.Lam, Luis.Marengo, Gordon.Shepherd, Perry.Miller, Chiquito.Crasto, Nian.Liu,

Thomas.Morse, Kei.Cheung}@yale.edu, Tim_Clark@harvard.edu, YGao@partners.org

{EWu, JuneKino}@alzforum.org; Wonglabow@verizon.net; Susie.Stephens@oracle.com

Abstract. Presently, neuroscientists have access to a wide range of neuroscience databases through the Internet. However, most of these databases are neither integrated nor interoperating, which creates a barrier in answering complex neuroscience research questions. Agreement upon a domain ontology is typically useful for querying diverse data sets, but is insufficient for integrating neuroscience data spanning multiple domains. To this end, e-Neuroscience seeks to provide an integrated platform for neuroscientists to discover new knowledge through seamless integration of diverse types and levels of neuroscience data. We present a Semantic Web approach to building this e-Neuroscience data integration framework, which involves using RDF as a standard data model to facilitate representation and integration of data. We have converted a subset of the BrainPharm database into RDF and integrated it with SWAN hypothesis and publication data extracted from Alzforum and made available in RDF as the upper ontology. Our implementation uses the RDF Data Model in Oracle Database 10g for data retrieval, integration, and inference. Our approach should be generalizable across many types of biomedical information.

BioRDF Goals

- Stage 1 (3 month goal)
 - Identify the initial data sources to be used in the demo
 - Explore additional data sources that would be required for the demo to span 'bench to bedside'
 - Learn about GRDDL, SPARQL, OWL, etc.
 - Increase knowledge of neuroscience
 - Set up a Wiki for communication
- Stage 2 (6 month goals)
 - Transform data into RDF from Word, Excel, XML, Relational, etc.
 - Analysis of semantic requirements (connect to ontology sub-group)
 - Move from screen scraping to an API
 - Create documents that describe work undertaken, and observations
- Stage 3 (12 month goals)
 - Use ontologies with the demo
 - Answer scientific questions and hopefully glean new scientific insights through using the demo
 - Validate the effectiveness of the data integration.

Summary

- A BioRDF community now exists
- Converted a number of data sources into RDF
- Reports have been written on a number of projects
- Top priority is to generate a URI document
- Time to start building a demo