

The Way Ahead.[™]

Semantic Web for Data Interpretation & Integration: Lessons Learned from Scientific Publishing and the Distributed Annotation System

Stephen A. Chervitz, Ph.D. Software Development, Affymetrix, Inc.

W3C Semantic Web for Life Sciences Workshop, 27 October 2004



Summary

- Scientific publishing
 - RDF-enhanced publications
- Distributed Annotation System (DAS)

- Semantic web issues for DAS 2.0
- Tool vendors and data producers
 - Request for guidance



Scientific publishing RDF-enhanced publications

Cline et al. PSB paper on alt splicing & protein motifs

- Data released in RDF and N3 formats
- Ontology relating splicing and protein annotation
- Advantages of RDF/N3 format
 - Ontology + data + rules in same file
 - Easy to extend, merge with external data
 - Tools available

Scientific publishing, continued

The Way Ahead.™

AFFYMETRIX.

N3 rule relating splice variation to disease/tissue specificity

```
## Define gene and splice variants
```

```
:Nox4 a affx:Gene ;
        affx:strand "+" :
        affx:chr "chr7" ;
        affx:hasTxVariant [affx:representedBy :gi14582296] ;
        affx:hasTxVariant [affx:representedBy :gi18204713] .
## Add new facts about a disease and tissue specifity
:HepatitisB a bg:Disease ;
        bg:manifestedIn :Liver .
 bg:associatedWithDisease :HepatitisB .
:Liver a affx:Tissue .
{ affx:hasTxVariant ?v . ?v affx:representedBy .}
=>{?v affx:foundIn :Liver } .
## Rule to link transcript to disease through tissue specificity
{ ?gene affx:hasTxVariant [affx:representedBy ?tx ; affx:foundIn ?tissue] .
  ?gene bg:associatedWithDisease ?disease .
  ?disease bg:manifestedIn ?tissue . }
=>
{ ?tx bg:expressedIn ?tissue .
 ?tx bg:conveys ?disease . } .
```



Scientific publishing, continued

RDF-enhanced publications

- Drawbacks/barriers to adoption
 - Verbosity, scalability issues for current tools
 - Storage, accessibility
- Community response extremely positive

Distributed Annotation System (DAS)

- DAS 2.0 development is officially underway
 - Multi-center grant
 - July 2004

AFFYMETRIX

The Way Ahead.™

- Numerous community-requested enhancements
 - Editing, not just viewing
 - Hierarchical features
 - Alternative data payload formats
 - Non-genomic coordinate spaces
 - Server registry and discovery
- Integration with semantic web systems



Distributed Annotation System (DAS)

DAS challenges for the Semantic web

- Using LSIDs within DAS
 - All IDs will be URIs
 - Permits URLs or LSIDs
 - Issues with LSID
 - Layer of complexity at odds with RESTful architecture
 - Data constancy
 - Content negotiation
 - Metadata handling
 - Resolver requirement



Distributed Annotation System (DAS)

DAS challenges for the Semantic web, continued

- Ontologies within DAS
 - Sequence Ontology term replaces category & type
 - Extensible
 - Rich RDF annotations

<link href="some-url" type="x-application/xml+rdf"

- Query over DAS servers
 - Feature attributes and relationships
 - RDF-based query protocol?

Tool vendors and data producers Request for guidance

Omics tool vendors want SW-friendly data

- Interpreting, interrelating mountains of data
- Guidelines for tools and data
 - Assign LSIDs for core objects
 - Provide LSID resolution services
 - Publish RDF-OWL for inter/extra-relations
- Practical questions arise
 - LSID assignment "best practices"
 - Standard ontologies to reference
- Ethical, Legal, Social implications

The Way Ahead.™

AFFYMETRIX