India Stewart

Data and Software Preservation for Open Science
Linked Data and the Semantic Web

- Current web is a directed graph of human readable nodes
- Leads to “stupid” querying
Linked Data and the Semantic Web

- Semantic web adds a second layer, a directed graph of machine readable nodes
- Allows intelligent queries

**SPARQL:**

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX : <http://dbpedia.org/resource/>
PREFIX dbpedia: <http://dbpedia.org/>
PREFIX dbpedia-owl: <http://dbpedia.org/ontology/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX ont: <http://dbpedia.org/ontology/>

SELECT ?films ?release
WHERE {?films dbpedia2:director :Quentin_Tarantino .
  OPTIONAL {?films ont:releaseDate ?release .}}
```
Metadata documenting provenance:
- Trust
- Integration with other data
- Credit to content creators

A standard provenance model is crucial to creating a comprehensive Semantic Web
Graphs are constructed in triples
- Subject, predicate, object

- India drinks coffee.
- India knows Bob.

Ontologies are defined to describe the relationships between nodes.
- We are particularly interested in PROV
@prefix : <http://xmlns.com/foaf/0.1/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

</opt/REU/India_foaf.ttl>.
:givenName "India";
:status "Undergraduate";
:birthday "11-28";
:gender "female";
:age "21";
:lastName "Stewart";
:mbox "india.stewart@uky.edu";
:postProject <Mapping the Galactic Disk>;
:currentProject <Data and Software Preservation for Open Science>;
:schoolHomepage <http://uky.edu>;
:knows <Charles Vardeman>;

a :Person.
Work So Far

- Simple FOAF generator
- Prototype of a software provenance generator
  - Dead end due to issues with pulling provenance info from operating system
  - Mining program build information with Signature Matching
- Prototyping dGit, an extension of Git designed to generate provenance
R&Wbase:
- Applies principles of Git to tracking changes in metadata descriptions
- Explores using PROV to document commits

MIT group used Subversion to generate RDF descriptions of tracked data
- Naive handling of rdf data graphs
- Integrates directly with SVN, using hooks to call scripts that grab metadata
Data Git:
- a python wrapper for Git that creates a provenance graph of entities under git control.
- Sophisticated handling of rdf graphs, using rdflib
- Uses PROV to document commits as Entities.
Future Work

- Finish dGit tool
  - Document files as Entities, described by commit Activities
  - Optimize dGit interaction with Git

- Expose URLs to be queriable