Reification: Statements about Statements within RDF

Vinod Chachra and Heather Myers, VTLS Inc. (USA)

Abstract

There is a need for visualization tool that allows users to more easily connect to primary sources of information. The idea is to allow users to see for themselves the primary sources of information that validate or justify the assertions made in RDF statements. This will allow users to directly and easily review the primary sources. VTLS has developed a tool that will allow staff to reify RDF assertions with links to the justifying artifact. Once all of the justifications have been established, they are presented via the Web in a graphical interface.

Background

VITAL is an institutional repository system offered by VTLS Inc. It is based on the open source software called Fedora™. This repository system is used in academic, special, and public libraries around the globe. VTLS is involved in development projects with several leading-edge organizations. This paper depicts the outcome of one such project.

The purpose of our development work was to create a new visualization tool that allowed users to more easily connect to primary sources of information related to the statements and assertions made in RDF (Resource Description Framework). The idea is to allow users to see for themselves the primary sources of information that validate or justify the assertions made in the RDF statements. This will allow users to directly and easily review the primary sources. The value of teaching with primary sources is well documented. The recent proliferation of Web-based learning tools that are visual in nature[1,2], led us to create a visual tool that provides access to archived resources and to the contextual information about those resources. In order to provide such a visual tool, a system must exist which stores both the resources and the relationships among them.

This new reification tool adds a layer of depth by creating a new link to the predicate portion of the statement. This new link is to a primary source of information that acts as justification that the predicate properly reflects the relationship between the subject and the object. For example, a user seeing the relationship "Charles Robinson was present at the Free State Party" would have instant access to the digitized copy of the document Charles Robinson signed calling for district elections of delegates to represent the Free State Party. **The reification visualization tool shows not just what we know, but also how we know it to be true.**

The need for such a tool became evident when VTLS was working with Jordan Fields (then at Kansas City Public Library) on a large project to build a Civil War Digital Library presented in an interactive Web site. Thousands of pages of Civil War primary sources from more than twenty-five contributing institutions were digitized as part of this initiative. In addition to these materials, the Web site will include articles written by Civil War scholars, shorter encyclopedia articles, maps, and timelines. The reification visualization tool discussed here will rely on a list of over five hundred people, places, events, and organizations that are relevant to the Civil War on the Missouri-Kansas border. Relationships between the concepts in the content are being documented and will form the backbone of the tool. A user researching one of these concepts will be able to see how it relates to all other concepts from the list. In addition to seeing the relationships, users will have one-click access to the primary source artifact that proves (justifies) the stated connection (predicate) between any two concepts.
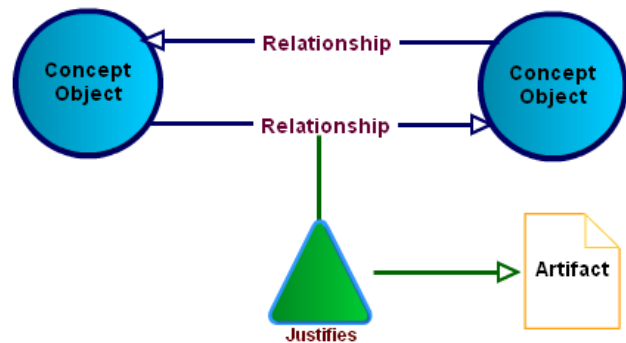

Figure 1. Reification architecture

RDF and Reification

A typical RDF-based statement consists of a subject, a predicate, and an object. RDF is a framework that is used to exemplify data throughout the Internet. Within Fedora, RDF is used to relate digital objects to other digital objects in a datastream called RELS-EXT. Within the RDF syntax, this relationship is structured as:

<subjectFedoraObject><relationshipProperty><targetFedora Object>

Fedora automatically validates the RELS-EXT datastream and enforces the following constraints that limit us from using it to make the justifications for the assertions contained in RDF:

"4. There must be only one `<rdf:Description>` in the RELS-EXT datastream. One description can have as many relationship property assertions as necessary.

5. There must be no nesting of assertions. Specifically, there cannot be an `<rdf:Description>` within an `<rdf:Description>`. In terms of XML "depth," the RDF root is considered at the depth of zero. There must be one `<rdf:Description>` element that must exist at the depth of one. The relationship assertions are RDF properties of the `<rdf:Description>` that exist at a depth of two."[3]

VTLS has determined that the best course of action is to reify the assertions made in RDF. As W3C notes, "RDF applications sometimes need to describe other RDF statements using RDF […] RDF provides a built-in vocabulary intended for describing RDF statements. A description of a statement using this vocabulary is called a *reification* of the statement."[4] Unfortunately, though, standard reification does require the subject to reference an individual instance of a triple statement, as opposed to any random triple with a similar assertion. This means that reification must be done outside of the RDF within the RELS-EXT, through a new reification tool.

Entities and Relationships for Reification

Conceptually, there are three major entities within the ontology: concepts, digital artifacts, and relationships.

Entities	Types	Storage
Concepts	People, Places, Events, Organizations	VITAL/Fedora Repository
Digital Artifacts	Circular, Diary, Letter, Newspaper clips, etc.	VITAL/Fedora Repository
Relationships	Object-Object: livedIn, foughtIn, familyMemberTo, etc. Object=Relationship: justifiedBy, justifies	Occurs within REIFICATION-RDF Datastream in objects Stored within RDF datastore

The Concept Entity

In this project, concepts (people, places, events) are the more than five hundred entities mentioned above and will be stored as "regular" objects within the VITAL Digital Asset Management System. The concepts themselves are described within a MADS datastream (see http://www.loc.gov/standards/mads/mads-outline.html#name). Concept objects will have a Concept Content Model and a Concept Type Content Model (event, person, organization, etc.). In the specific demonstration project we are working on there are approximately five hundred concepts that have associated artifacts and relationships between them.

Datastream ID	Description	Contains
DC	Dublin Core Metadata	Descriptive Metadata
MADS	Metadata Authority Description Schema	Descriptive Metadata

RELS-EXT	Fedora Object Relationships	Content Model declaration (e.g. hasModel) Object-Object relationships (e.g. Collection membership) Concept-Concept relationships (e.g. familyMemberTo)
REIFICATION-RDF	RDF XML with extended reification support	Relationship-Artifact relationships (e.g. isJustifiedBy)

The Digital Artifacts

Digital resources or artifacts are stored as "regular" objects within VITAL. The system supports the linking of multiple datastreams to each object. Minimally, there is the digital content and a descriptive metadata datastream attached to it. Each object in the repository has a Content Model associated with it. Thus, an Artifact Object will have an Artifact Content Model and an Artifact Type Content Model (Diary, Letter, etc.). The structure of the Artifact Object is shown below.

Datastream ID	Description	Contains
DC	Dublin Core Metadata	Descriptive Metadata
OtherMetadata	Descriptive Metadata mapped into Dublin Core	TEI, KCMetadata
ContentDatastreams	Digital Artifacts	pdf, jpg, tiff, etc.
RELS-EXT	Fedora Object Relationships	Content Model declaration (e.g. hasModel) Object-Object relationships (e.g. Collection membership)
REIFICATION-RDF	RDF XML with extended reification support	Relationship-Artifact relationships (e.g. isJustifiedBy)

The Relationships

The typical Fedora ontology as described in the Fedora RELS-EXT ontology consists only of object-object relationships. The ontology described here (the VTLS ontology) goes beyond the traditional Fedora ontology. In order to support the idea of reification the relationships themselves had to be added to the VTLS ontology. Adding the relationships to the ontology will allow Fedora objects to be related to the relationships themselves. In order to talk about relationships, they must be reified. Implementation of reification requires that a unique identifier be

created for an RDF assertion. Once there is such an identifier, it can then be talked about as an RDF resource. This allows the linking of the relationship to a Fedora object. Everything now becomes a standard RDF triple statement. For this specific Civil War Project, there are thirty-five defined relationships and their reciprocals.

The Relationship Editor

In order to facilitate the establishment of relationships VITAL's Relationship Editor was enhanced to include the ability to reify relationships. The Relationship Editor is selected from an object's page. There is a drop-down list box that will contain all of the possible relationships. Administrators will be able to navigate to a Concept Object's Relationship Editor and relate it to other Concepts. VITAL's Relationship Editor is being enhanced to provide the ability to reify these relationships and then to justify them with a relationship to an artifact. REIFICATION-RDF will be automatically populated with the relationships by VITAL's Enhanced Relationship Editor.

An Example

Described below is a series of concepts that are related to each other. In addition, there are artifacts that justify the relationships.

A person: **Charles Dow**, is represented by concept Object, PID 114.

A person: **Franklin Coleman,** is represented by concept Object, PID 157.

An Organization: **Free State Party** is represented by concept Object, PID 192.

A place: **Douglas County, Kansas** is represented by concept Object, PID 287.

A letter: **From James Henry Lane to R.M. Ainsworth** is a scanned TIFF image contained in the system as a datastream within Artifact Object, PID 12674. It is considered to be the first letter of the war and justifies the person Charles Dow was a member of the Organization Free State Party.

A letter: **Statement of Wilson Shannon** is a series of scanned TIFF images contained as datastreams within Artifact Object, PID 12716. It describes the events leading up to the Wakarusa War, including the murder of Charles Dow by Franklin Coleman in Dow's home, Douglas County, Kansas (place).
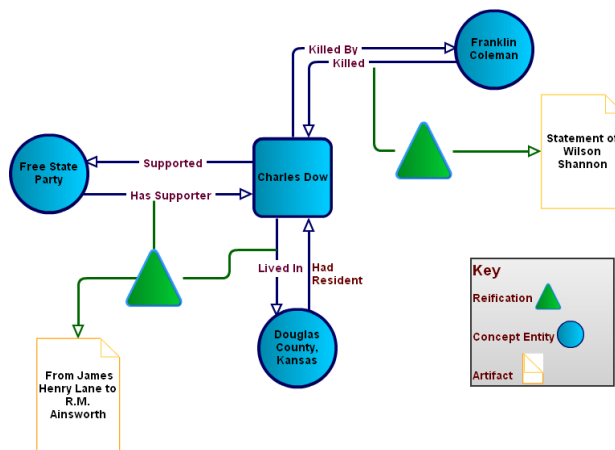

Figure 2: Reification visualization example

Resulting RELS-EXT

In the above example, the concept entity with PID 114, Charles Dow, has the following RELS-EXT datastream:
```
<rdf:RDF>
<rdf:Description rdf:about="info:fedora/authority:114">
< hasModel rdf:resource="info:fedora/vital-system:concept" />
<rel:supported rdf:resource="info:fedora/authority:192"/>
<rel:killed_by rdf:resource="info:fedora/authority:157"/>
<rel:lived_in rdf:resource="info:fedora/authority:287"/>
</rdf:Description>
</rdf:RDF>
```

In the RELS-EXT, the assertions between concepts are made.

Resulting REIFICATION-RDF

In the above example, the concept entity with PID 114, Charles Dow, has the following REIFICATION-RDF datastream:
```
<rdf:Description rdf:about="#authority_114__1">
<history:isJustifiedBy rdf:resource="info:fedora/kcpl:12674"/>
</rdf:Description><rdf:Description
rdf:about="#authority_114__2">
<history:isJustifiedBy rdf:resource="info:fedora/kcpl:12716"/>
rdf:about="#authority_114__3">
<history:isJustifiedBy rdf:resource="info:fedora/kcpl:12716"/>
</rdf:Description>
</rdf:RDF>
```

For each assertion defined in the RELS-EXT, an identifier is provided (e.g. authority_114_1), and the justification is made in a reification statement against that identifier with the identifier of the artifact.

Discovering Primary Sources

Once all of the justifications have been established, they are presented via the Web in a graphical interface. Integrating within the Web site are four tools that will allow users with various learning preferences to find content within the repository. Each of these tools is meant to appeal either graphically, or through text.

Reification Visualization Tool

The reification visualization tool was inspired by recent prototypes built on RDF triples such as the SNAC project (http://socialarchive.iath.virginia.edu/) and the RelFinder (http://www.visualdataweb.org/relfinder.php); however, the tool adds another layer of depth by linking the primary sources to the predicates of the triples as proof that the relationship between the subject and the object exists. **The reification visualization tool shows not just what we know, but also how we know it to be true.**

Timeline

The Web site will contain an interactive timeline that can be navigated with arrows or by dragging the mouse. Clicking a marker will yield a pop-up screen that contains a teaser and links to individual Timeline Event pages. The timeline displays local events in the context of national events, so different markers will be needed in order to differentiate them. Events are then linked to concepts within VITAL. Users will have the opportunity to navigate the timeline to concepts and enter the visualization tool to discover content within the repository.

Figure 3: Timeline screenshot

Digital Gallery

The digital gallery will display all of the artifact objects that are pictorial in nature from VITAL. The pictorial items have been given the theme "Picturing the War" in VITAL. The initial presentation should be of thumbnail images (approximately 200 x 150 pixels) with title directly below. Upon clicking on an image, a lightbox or similar image viewer with the full-size image should pop up and allow forward and backward browsing of the images. The lightbox should provide the image title and a link to the item's individual page in the digital collection. From there, the user can view any relationships that this artifact justifies within the visualization tool, and browse those relationships to other artifacts and concepts.
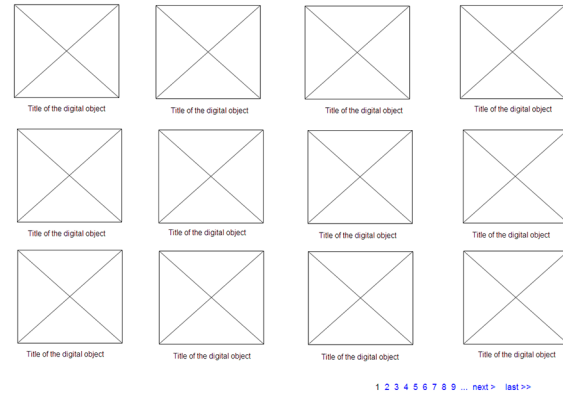

Figure 4

Interactive Map

An interactive map will provide an additional means of accessing the artifacts. Clicking a marker will yield a pop-up screen indicating the title of the location, teaser text and a link to the individual location's page. Each location is a concept within VITAL and from the location page, the user will be able to view any relationships this location has within VITAL, as well as the justifications.
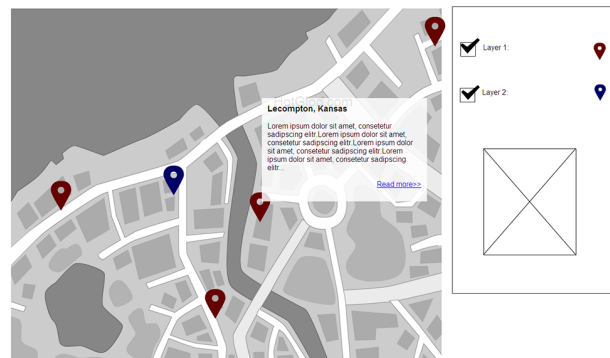

Figure 5: Map screenshot

This tool makes it easier to understand and justify the complex relationships between many objects in the Kansas City Civil War project and has application in other domains.

References

[1] Stripling, Barbara. "Feature Article." *Teacher Resources - Library of Congress*. Library of Congress, Summer 2009. Web. 01 Feb. 2013. <http://www.loc.gov/teachers/tps/quarterly/inquiry_learning/article.html>.

[2] "Primary vs Secondary Sources." *Primary vs Secondary Sources*. Princeton, n.d. Web. 01 Feb. 2013. <http://www.princeton.edu/~refdesk/primary2.html>.

[3] Wilper, Chris. "Digital Object Relationships." *Duraspace Wiki*. Duraspace, 13 Aug. 2011. Web. 01 Feb. 2013. <https://wiki.duraspace.org/display/FEDORA34/Digital Object Relationships>.

[4] "RDF Primer." *RDF Primer*. Ed. Frank Manola and Eric Miller. W3C, 10 Feb. 2004. Web. 04 Feb. 2013. <http://www.w3.org/TR/rdf-primer/>.

Author Biography

Vinod Chachra is the founder and CEO of VTLS Inc., which provides state-of-the-art software systems to more than 1900 libraries and resource centers in 43 countries worldwide. He is the author of several books and articles and in 2012 was inducted as a founding member of the Virginia Tech Faculty Entrepreneur Hall of Fame.

Heather Myers received her BS in Information Systems Management, Systems Analysis and Design from the University of Maryland, Baltimore County. During the course of her career at VTLS, Heather has focused on Digital Repositories. She is currently the Director of Portals and Platforms Development Department.