

## Chapter 9

### **An Architecture for Complex Objects and their Relationships**

#### **Preface**

This chapter is based on:

Lagoze, C., Payette, S., Shin, E. and Wilper, C. Fedora: An Architecture for Complex Objects and their Relationships. *International Journal of Digital Libraries*, 6 (2). 124-138 [301].

Fedora, the Flexible Extensible Digital Object Repository Architecture, is an architecture and implementation of a powerful digital object model and repository for storage of and access to those objects. It is perhaps the most successful result of our work on digital libraries at Cornell.

The Fedora work is interesting because of its roots in early digital library concepts, and because of its longevity due to the flexibility of its initial design. Initial work on Fedora began in 1997 as an outgrowth of our earlier work on extensions and implementations of the Kahn/Wilensky Framework [290], Dienst [152], and the Warwick framework [147, 148, 286]. The Kahn/Wilensky Framework, which was described in greater detail in earlier chapters, provided the foundational notion of a *digital object*, an identified container aggregating multiple data streams. The Dienst work contributed the notion of a repository and protocol-based access to digital objects that conform to a document model. Finally, the Warwick Framework work, and its subsequent extension with the notion of *distributed active relationships*, which occurred within the Dublin Core metadata effort, contributed the concept of linking data with services to produce active disseminations of content.

As described in the remainder of this chapter, recent work has extended the original concepts of Fedora into the realm of the semantic Web, service-based architectures, and policy enforcement. As stated in the text the “recent work uniquely integrates advanced content management with semantic web technology.” In this manner, Fedora is both an exemplar of the library meme-based notions described at the beginning of this dissertation blended with the Web 2.0 concepts that developed alongside digital library efforts. Perhaps this is the reason behind the enduring nature of the Fedora Project in the manner in which it is recognized as the technological leader amongst competing projects such as DSpace, Greenstone, and ePrints. Indeed, the words behind the name Fedora – Flexible Extensible Digital Object Repository Architecture – have indeed proven true as it has successfully bridged the digital library and Web 2.0 worlds.

Fedora is still a very active and increasingly influential project, especially in the expanding cyberinfrastructure research area. Under the leadership of Sandy Payette, the Fedora project has now been spun off into a not-for-profit called Duraspace, which focuses on open source projects to support durable digital content<sup>92</sup>.

### **Acknowledgments**

As a project that has spanned more than 10 years, the success of Fedora is based on the efforts of a large number of people. Primary among them is Sandy Payette, the current Chief Executive Officer of DuraSpace, the not-for-profit corporation that is currently responsible for Fedora development and maintenance. Sandy was my chief collaborator on Fedora during the DARPA-funded and NSF-funded research phase. She subsequently led the Fedora project in its transition to widely disseminated open

---

<sup>92</sup> <http://www.duraspace.org/>

source software and architectural integration with developing web frameworks.

Through a unique combination of organizational and technical leadership, Sandy has brought Fedora from its humble beginnings as a research project to a leader in open source technology. Other primary contributors to the work over the years include the rest of the Fedora team at Cornell including Chris Wilper, Eddie Shin, and Dan Davis. In addition, significant contributions have been made by the team at University of Virginia led by Thornton Staples and Ross Wayland. Finally, the large and diverse Fedora community should be recognized for their contributions.

Initial support for the Fedora work came from Digital Library Initiative funding from the National Science Foundation and DARPA. More recent funding is due to the generosity and wisdom of the Andrew W. Mellon foundation, especially Don Waters who as program manager has continued to support Fedora activities. Finally, the transformation from Fedora to Fedora Commons has been possible due to extremely generous support from the Gordon and Betty Moore foundation, and the efforts of Jim Omora, a program officer at Moore.

---

## **Introduction**

At a minimum, technologies for representing digital content should be able to match the richness, and complexity of well-established physical formats. As such, they should allow the representation of a variety of structural organizations, such as chapters and verses; accommodate the flexible combination of different genre of materials, such as text and images; and allow the aggregation of content from multiple sources and the association of metadata with the elements of the aggregation.

However, freed of the constraints of physical media, digital content architectures should do more. Exploiting their networked context, they should allow aggregation of content regardless of its physical location. By leveraging local and remote computing power they should support programmatic and user-directed manipulation of digital content. Finally, they should represent the complex structural, semantic, provenance, and administrative relationships among digital resources.

This paper describes our latest work on Fedora, an open-source digital content repository service, which provides a flexible foundation for managing and delivering complex digital objects. This recent work uniquely integrates advanced content management with semantic web technology. It supports the representation of rich information networks, where the nodes are complex digital objects combining data and metadata with web services and the edges are ontology-based relationships among these digital objects.

The motivation for integrating content management and the semantic web originates from requirements defined by the broader Fedora user community. The most familiar example is the need to express well-known management relationships among digital resources such as the organization of items in a collection and structural relationships such as the part-whole relationships between individual articles and a journal. While the relationships among digital objects in these familiar applications are mainly hierarchical, we are working with other applications where the relationships are more graph-like. For example, in the NSF-funded NSDL (National Science Digital Library) Project<sup>93</sup>, we are using Fedora to implement an information network overlay that represents local and distributed resources and the provenance, managerial, and

---

<sup>93</sup> <http://nsdl.org>

semantic relationships among those resources. We report on the results of this work later in this paper.

While there are a number of schemes for representing these relationships such as conventional relational databases and formalisms like conceptual graphs [440], the products of the semantic web initiative such as RDFS [92], OWL [155], and highly-scalable triple-stores such as Kowari [496] provide extensible open-source solutions for representation, manipulation, and querying these knowledge networks.

The remainder of this paper describes the details of the Fedora architecture that provides the foundation for these rich applications. The structure of this paper is as follows. We begin by summarizing the historical development of Fedora. The next section provides core background on the Fedora digital object model, articulating a graph-based view of the model that is consistent with the semantic web orientation of our latest work. After that, we described the Fedora relationship model that provides a common framework for describing, storing, and querying relationships among objects and their components. The penultimate section describes results of the deployment of Fedora, focusing on applications that exploit features that distinguish Fedora from related work. The final section is a conclusion.

## **Background**

The Fedora Project<sup>94</sup> is an ongoing research and development effort to provide the framework for creation, management, and preservation of existing and evolving forms of digital content. The roots of the project lie in DARPA-funded research in the early 1990's that defined the notion of a *digital object* [255] and implemented Dienst [289], a networked digital library architecture with protocol-based dissemination of digital

---

<sup>94</sup> <http://www.fedora-commons.org/>

objects in multiple formats. Follow-on research extended these initial concepts with the notion of *active digital objects* [148] and *distributed active relationships* [147]. These concepts were refined and prototyped in a CORBA-based Fedora (Flexible Extensible Digital Object Repository Architecture) [403] as part of research with CNRI [401] and in the context of the NSF-funded Prism Project [295]. This prototype provided the context for a variety of research initiatives most notably in the areas of fine-grained policy enforcement [404] and preservation [405].

The transition of Fedora from a research prototype to production repository software began when the University of Virginia Library, seeking a solution for managing increasingly complex digital content, experimented with the Fedora architecture [443]. This experimentation took place in the context of innovations in humanities research. The experimentation proved successful, providing the basis for subsequent funding from the Andrew W. Mellon Foundation to Cornell and Virginia [406] to jointly develop Fedora and make it available as open source software to libraries, museums, archives, and content managers, facing increasing variety and complexity in the digital content that they manage [444]. Mellon-funded development continues through 2009.

The richness of the Fedora digital object model and extensibility of the Fedora service-based architecture has led to its deployment in a variety of domains including digital libraries [279, 298], institutional repositories [457], electronic records archives [488], trusted repositories for digital preservation [248], library systems<sup>95</sup>, educational technologies [279], web publishing<sup>96</sup>, and distributed information networks [298].

---

<sup>95</sup> <http://www.vtls.com/products/vital>

<sup>96</sup> <http://www.encyclopedia.chicagohistory.org/>

Fedora is implemented as a set of web services that provide full programmatic management of digital objects as well as search and access to multiple representations of objects. All Fedora APIs are described using the Web Service Description Language (WSDL) [124]. As such, Fedora is particularly well-suited to exist in a broader web service framework and act as the foundation layer for a variety of multi-tiered systems, service-oriented architectures, and end-user applications. This distinguishes Fedora from other complex object systems that are turn-key, vertical applications for storing and manipulating complex objects through a fixed user interface (e.g., DSpace [439], arXiv<sup>97</sup>, ePrints<sup>98</sup>, Greenstone [494]).

By providing both a model for digital objects and repository services to manage them, Fedora is also distinguished from work focused on defining and promoting standard XML formats for representing and transmitting complex objects (e.g., METS<sup>99</sup>, MPEG-21 DIDL[245], IEEE LOM [4]). However, Fedora is compatible with these efforts since it has the ability to ingest and export digital objects that are encoded in such XML transmission formats<sup>100</sup>. This allows Fedora to comfortably coexist in the archival framework defined by OAIS [418].

As a service-based architecture for complex digital objects, Fedora has some commonality with the aDORe architecture [461] developed at the Los Alamos National Laboratory research library. The aDORe system provides a standards-based

---

<sup>97</sup> <http://arxiv.org>

<sup>98</sup> <http://eprints.org>

<sup>99</sup> <http://standards.loc.gov/mets>

<sup>100</sup> Fedora currently supports ingest/export of digital objects encoded using METS and also the Fedora XML wrapper format (FOXML). Future releases will support MPEG-21 DIDL and possible other formats.

repository for managing and accessing complex digital objects. Objects are encoded in XML using DIDL [47] and a limited set of object relationships can be expressed using RDF. Object dissemination services are available via OAI-PMH [314] and OpenURL [12].

### **Fedora model for complex objects**

The Fedora object model supports the expression of many kinds of complex objects, including documents, images, electronic books, multi-media learning objects, datasets, computer programs, and other compound information entities. Fedora supports aggregation of any combination of media types into complex objects, and allows the association of services with objects that produce dynamic or computed content. The Fedora model also allows the assertion of relationships among objects so that a set of related Fedora objects can represent the items in a managed collection, the components of a structural object like the chapters of a book, or a set of resources that share common characteristics (defined by semantic relationships).

Fedora defines a powerful object model for expressing this variety of complex content and their relationships. This object model can be understood from two perspectives.

- The *representational* perspective defines a simplified abstraction for understanding Fedora objects, where each object is modeled as a uniquely identified resource projecting one or more views, or *representations*. From this perspective the internal structure of a digital object is opaque; however, relationships among objects are observable.
- The *functional* perspective reveals the object components that underlie the representational perspective and provides the basis for understanding how the Fedora object model relates to the management services exposed in the Fedora repository architecture.

### ***Representational View***

The representational perspective of the Fedora object model asserts that each digital object can disseminate one or more representations of itself, and that each object can be related to one or more other objects. A familiar example of a digital object with multiple representations is a document or image where the content is available in multiple formats. All digital objects, and their individual representations, are identified with Uniform Resource Identifiers (URIs). These URIs are specified using the “info” scheme and conform to the syntax described at [462]. This choice frees the architecture from ties with any identifier resolution system (e.g., the Handle System [449]).

This perspective hides complexity and exposes only the access points to content stored in a Fedora repository. Figure 36 depicts the representational view of three inter-related Fedora objects. The diagram shows a directed graph, where the larger nodes are digital objects, and the smaller nodes are representations of the digital objects<sup>101</sup>. These nodes are linked by two types of arcs – relationship arcs connect digital objects, and representation arcs connect digital objects to their respective representations. This graph can be expressed as RDF, stored in a triple store, and queried. This is discussed a later section.

Each digital object in the diagram has at least one representation, related to its originating digital object by a `hasRep` arc. For example, the node labeled `info:fedora/demo:11` is an image digital object with four representations, identified by their respective URIs:

---

<sup>101</sup> This graph-based overlay model can form the basis for interoperability among heterogeneous object models and repositories. This concept was explored as part of an NSF-funded research project, Pathways, a collaboration between the authors of this paper and colleagues at Cornell, LANL, and others [466].

- -Dublin Core record, identified as `info:fedora/demo:11/DC`
- -High-resolution image, identified as `info:fedora/demo:11/HIGH`
- -Thumbnail image, identified as `info:fedora/demo:11/THUMB`
- -Image with zoom/pan utility, as `info:fedora/demo:11/bdef:2/ZPAN`

We have yet to define the underlying source of these representations. In fact, in this view of the architecture such details are hidden from the client application concerned with access to these representations.

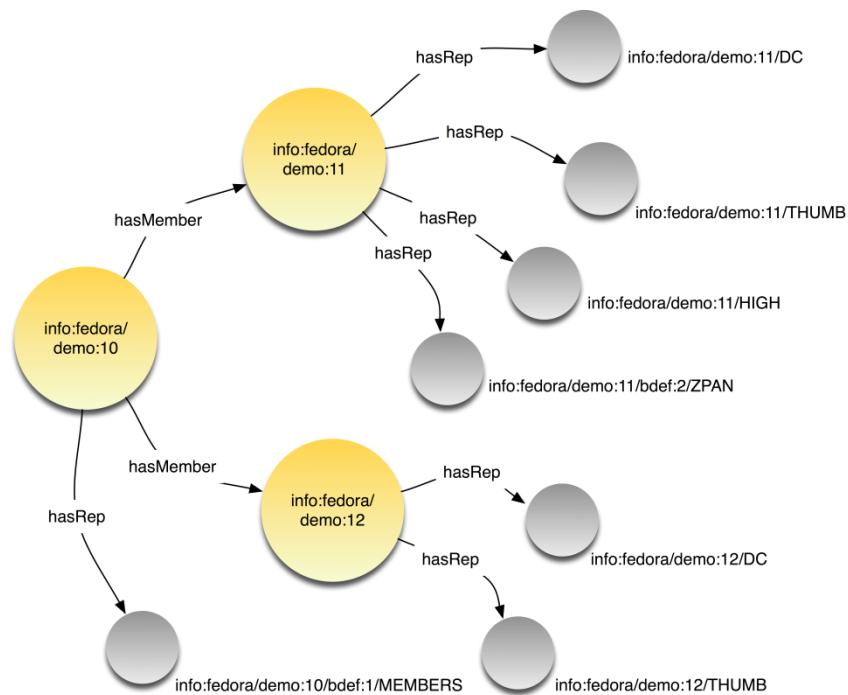


Figure 36 - Representational view of Fedora objects

Figure 36 also demonstrates an example of inter-object relationships. In this example, the node labeled `info:fedora/demo:10` is a “collection” with two “items”, the nodes labeled `info:fedora/demo:11` and `info:fedora/demo:12`. These collection-item relationships are expressed by the `hasMember` arc that emanates from the collection object. The inverse `isMemberOf` relationships are not shown in the diagram for simplification.

This simple representational view forms the basis of Fedora’s REST-based access service (i.e., API-A-LITE), whereby digital object URIs and representation URIs can be easily converted to service request URLs upon Fedora repositories.

### ***Functional View I - Datastreams***

While the representational perspective of the Fedora object model provides a simple, access-oriented overlay for digital resources and collections, the *functional perspective* provides a view of the core underlying data model for Fedora. In the following sections, we take one of the digital object nodes depicted in Figure 36, and drill down to unveil the specific components of a Fedora digital object that enable access to representations. We start with the digital object as a container with a persistent unique identifier (i.e., PID). From there, we unveil the components incrementally, first focusing on components that enable simple content aggregation, then on components that enable dynamic and computed content, and finally on components related to digital object integrity. We note again that these underlying details are invisible to clients concerned only with information access.

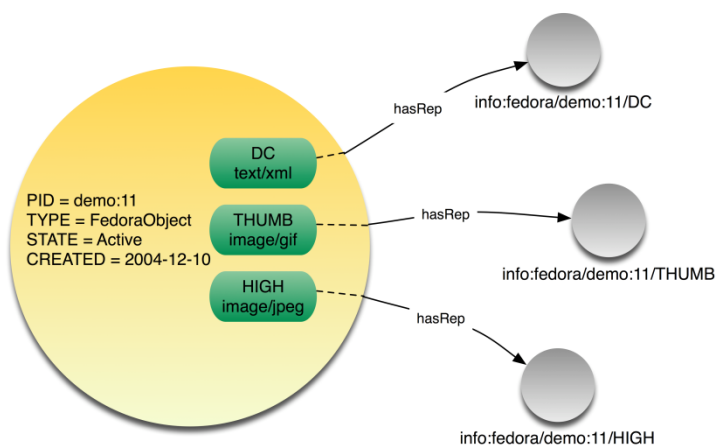
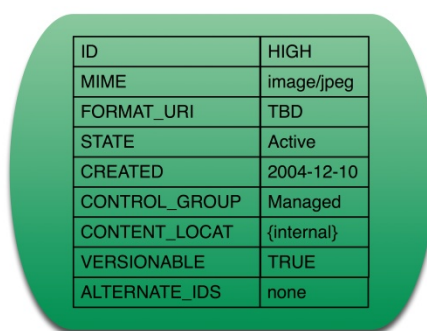


Figure 37 - Fedora object with PID, properties, and datastreams

In its simplest form a Fedora object is an aggregation of content items, where each content item maps to a representation. The Fedora object model defines a component

known as a datastream to represent a content item. A datastream component either encapsulates bytestream content internally or references it externally. In either case that content may be in any media type. Figure 37 shows a digital object as an aggregation of datastreams and the one-to-one correspondence of those datastreams to the representations of the digital object that are exposed to accessing clients. In this simple case, each representation of a Fedora object is a simple transcription of the content that lies behind a datastream component.



ID	HIGH
MIME	image/jpeg
FORMAT_URI	TBD
STATE	Active
CREATED	2004-12-10
CONTROL_GROUP	Managed
CONTENT_LOCAT	{internal}
VERSIONABLE	TRUE
ALTERNATE_IDS	none

Figure 38 - Properties of a datastream component

As seen in Figure 37, a digital object has a unique identifier (PID) and a set of key descriptive properties. Each datastream contains information necessary to manage a content item in a Fedora repository. These are stored as properties of the datastream as shown in Figure 38.

Three datastream properties deserve special attention. The Format URI refines the media type definition and anticipates the emergence of a global digital format registry such as the GDFR<sup>102</sup>. Control group defines whether the datastream represents either local or remote content. Datastreams with a control group of “Managed” are internal content bytestreams that are under the direct custodianship of the Fedora repository.

---

<sup>102</sup> <http://hul.harvard.edu/gdfr/>

Datastreams whose control group is “External” or “Redirected” (the difference between these is outside the scope of this paper) represent content that is stored outside the repository. These datastreams have a content location property that is a URL pointing to a service point outside the repository that is responsible for providing the content. The ability to create digital objects that aggregate locally managed content with external content is a powerful feature of Fedora, and is useful in a variety of contexts. A good example of a hybrid local/remote object is an educational object where local content is an instructor’s syllabus, lecture notes, and exams, and remote content are primary resources included by-reference from other sites.

### ***Functional View II - Disseminators***

In addition to the representations described in the previous section, which are direct transcriptions of datastreams, the Fedora object model enables the definition of *virtual representations* of a digital object. A virtual representation, also known as a *dissemination*, is a view of an object that is produced by a service operation (i.e., a method invocation) that can take as input one or more of the datastreams of the respective digital object. As such, it is a means to deliver dynamic or computed content from a Fedora object.

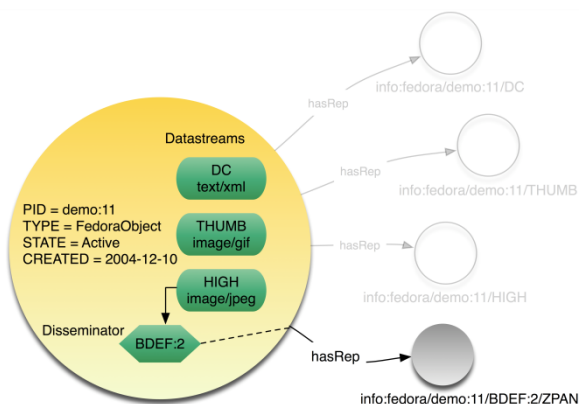


Figure 39 - Fedora object with disseminator added

This is illustrated in Figure 39, where a virtual representation labeled `info:fedora/demo:11/BDEF:2/ZPAN` is highlighted. From the access perspective this representation is an image wrapped in a java application that provides image zoom and pan functions. Note that this representation is not a direct transcription of any Datastream in the object. Instead, it is the result of a service operation defined in the Disseminator component labeled `BDEF:2` inside the object that uses the datastream labeled `HIGH` as input. The light-weight, REST-based interface to Fedora (API-A-LITE) makes it possible for a client application to pass parameters to the invoked service; in this case zoom and pan specifications.

To enable such behavior, a Disseminator must contain three pieces of information: (1) a reference to a description of service operation(s) in an abstract syntax, (2) a reference to a WSDL service description [124] that defines bindings to concrete web service to run operation(s), and (3) the identifiers of any Datastreams in the object that should be used as input to the service operation(s).

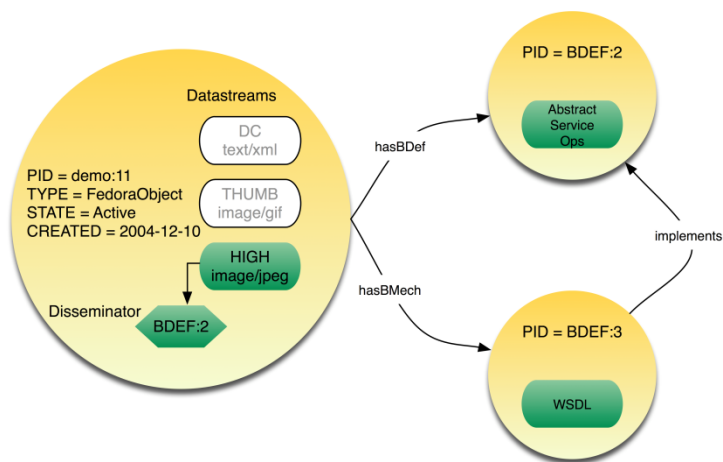


Figure 40 - Disseminators establish relationships to service definition objects

Fedora stores the service operation description and the WSDL service description within special digital objects, respectively known as BDefs (behavior definitions) and

BMechs (behavior mechanisms). Figure 40 depicts a Fedora BDef object and a BMech object along with object-to-object relationships that exist due to the presence of the Disseminator component in the main object (i.e., `demo:11`).

Disseminators are effectively metadata that a Fedora repository uses at run time to construct and dispatch service requests and produce one or more virtual representations of the digital object. From a client perspective this is transparent since virtual representations look just like other representations of the object.

Disseminators are a powerful feature in the Fedora object model. They can be used to create common representational access points for digital objects that have different underlying structure or format. For example, an institutional repository might contain scholarly documents in a variety of root formats (e.g., Word, HTML, TeX), where the root format is stored as a datastream in a Fedora digital object. For interoperability purposes, a virtual representation can be defined on each object that converts the datastream containing the root format to a common format (e.g., PDF). Similarly, a repository manager can decide for archival purposes to convert all documents in a repository to a canonical preservation format without disrupting the manner in which clients access documents for browsing, viewing, etc. Finally, disseminators can add utility operations to digital objects. For example, a Disseminator can be defined for a digital object that provides parameterized query access to the relationships defined for that object. Such a query might return the “members of a collection” or, in the case of an educational digital library such as the NSDL [498], the set of resources that are appropriate for K-12 mathematics education. The implementation of these queries is described later.

### ***Functional View III – Object Integrity Components***

The Fedora object model defines several metadata entities that pertain to managing the integrity of digital objects. These entities are the object’s relationship metadata, access control policy, and audit trail. To keep the Fedora model simple and consistent, integrity entities are modeled as datastream components with reserved identifiers. As such, the integrity entities are stored like other datastreams, however the Fedora Repository system recognizes them as special and asserts constraints over how they are created and modified. Figure 41 depicts these integrity-oriented entities as special datastreams in a digital object, identified as Relations, Policy, and Audit Trail.

A Relations datastream is used to assert object-to-object relationships such as collection/member, part/whole, equivalence, “aboutness,” and more. The previously discussed “hasMember” relationship is an example of the type of assertion that can be managed via the Relations datastream, described later in this chapter.

A Policy datastream is used to express authorization policies for digital objects, both to protect the integrity of an object and to enable fine-grained access controls on an object’s content. In Fedora objects, a policy is expressed using the eXtensible Access Control Markup Language (XACML)<sup>103</sup>, which is a flexible XML-based language used to assert statements about who can do what with an object, and when they can do it. Object policies are enforced by the authorization module (i.e., AuthZ) implemented within the Fedora repository service.

---

<sup>103</sup> [http://www.oasis-open.org/committees/tc\\_home.php?wg\\_abbrev=xacml](http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xacml)

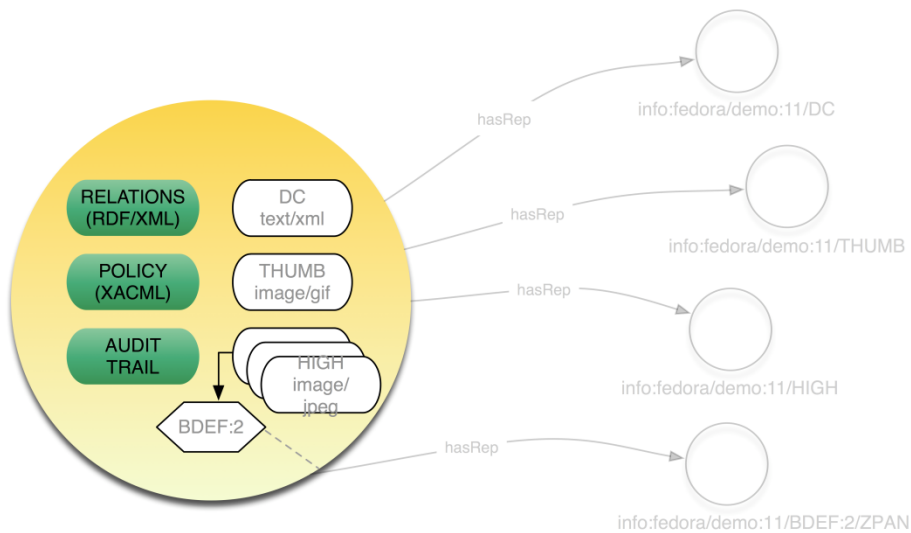


Figure 41 - Integrity datastreams - relationships, policy, and audit trail

The Audit Trail is a system-controlled datastream that keeps a record of all changes to an object during its lifetime. The Fedora repository service automatically creates an audit record for every operation upon an object, detailing who, what, when, where, and why an object was changed. This information is important to support preservation and archiving of digital objects.

Another feature for managing the lifecycle of objects is versioning. Versioning is important for applications where change tracking is essential, as well as for preservation and archiving systems that must be able to recover historical views of digital objects. The Fedora object model supports component-level versioning, meaning that datastreams and disseminators can be changed without losing their former instantiations. Fedora automatically creates a new version of these components whenever they are modified.

This is depicted in Figure 41, which shows a digital object with multiple versions of a datastream (see component labeled HIGH). Also, the versioned datastream is input to the disseminator labeled BDEF:2. Requests for representations of this digital object

can be date-time stamped and the Fedora repository service will ensure that the appropriate component version is returned. This feature applies for representations that are direct transcriptions of datastream content, as well as for virtual representation where datastream content is mediated via a Disseminator.

### **Relationships in Fedora**

As described earlier, the Fedora object model can be abstractly viewed as a directed graph, consisting of *internal* arcs that relate digital object nodes to their representation nodes and *external* arcs between digital objects. In this section we focus on that relationship graph and describe the Fedora *Resource Index* module, which allows storage and query of the graph. This module builds on RDF (Resource Description Framework) [273] primitives developed within the semantic web community. The Fedora system defines a base relationship ontology that, in the fashion of any RDF properties, can co-exist with domain-specific ontologies from other namespaces. Each digital object's relationships to other digital objects are expressed in RDF/XML [44] within a reserved datastream in the object. The Resource Index is a relationship graph over all digital objects in the repository that is derived by merging the relationships implied by the Fedora object model itself with the relationships explicitly stated in an object's relationship datastream. The triples representing this graph are then stored in a triple-store providing the capability for searching over the graph.

The combination of representing explicit relationships as RDF/XML in a datastream of a digital object and then mapping them to the triple store offers the "best of both worlds". The explicit representation provides the basis for exporting, transporting, and archiving of the digital objects with their asserted relationships to other objects. The mapping to the triple store provides a graph-based index of an entire repository and the basis for high-performance queries over the relationships. An added advantage

of the dual representation is that the entire triple store can be rebuilt by importing and parsing the XML-based digital objects.

### ***Representing object-to-object relationships***

There has been a significant amount of work in the area of structural metadata for complex objects. These efforts have been focused on developing XML schema for expressing structural relationships with individual digital objects. One early example was the Making of America [161] project that formalized structural metadata and defined a set of templates that correspond to well-known physical artifacts such as a book composed of chapters and diaries consisting of entries. The current exemplar of this encoding of structural metadata is in METS [347].

Our focus in Fedora has been to decompose these structural units into separate digital objects. The motivation for this is that the units can then be reused in a variety of structural compositions. In addition, this lays the basis for expressing other types of non-structural relationships among digital objects such as:

- The organization of individual resources into larger *collection* units, for the purpose of management, OAI-PMH harvesting [314], user browsing, and other uses.
- The relationships among *bibliographic* entities such as those described in the Functional Requirements for Bibliographic Relationships [3].
- *Semantic* relationships among resources such as their relevance to state educational standards or curricula in an educational digital library like the National Science Digital Library [498].

- Modeling more complex forms of *network overlays* over the resources in a content repository such as citation links [202, 236], link structure, friend of a friend<sup>104</sup>, etc.

All of these relationships, including structural relationships, should be expressible both within individual digital objects and among multiple digital objects. For example breaking the components of a structural entity, such as the chapters of a book, into separate digital objects provides the flexibility for reuse of those individual components into other structural units. This is even more important for the other forms of relationships. For example, a single resource may be part of multiple collections or may be relevant for multiple state standards.

The remainder of this section describes a relatively simple example of inter-object relationships to demonstrate how these are expressed in Fedora. The simple techniques illustrated here can be used to express more complex inter-object relationships. In a later section, we will describe a more complex example in the context of our NSDL work.

The expression of arbitrary, inter-object relationships in Fedora is enabled by a reserved datastream known as the Relations datastream. This datastream allows for a restricted subset of RDF/XML where the subject of each statement must be the digital object within which the datastream is defined.

Since predicates from any vocabulary can be used in Relations, the repository manager has considerable flexibility in the kinds of relationships that can be asserted. Table 4 shows an example Relations datastream in a Fedora digital object identified by the URI, `info:fedora/demo:11`. The RDF/XML refers to three different relationship

---

<sup>104</sup> <http://www.foaf-project.org/>

vocabularies (hypothetical for the purpose of this example) and asserts the following relationships:

- `demo:11` is a member of the collection represented by the object `demo:10`,
- `demo:11` fulfills the state educational standard represented by the object `demo:Standard5`,
- `demo:11` is a manifestation of the expression represented by the object `demo:Expression2`.

Table 4 - Example relations datastream

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:nsdl="http://nsdl.org/std#"
  xmlns:rel="http://example.org/rel#"
  xmlns:frbr="http://example.org/frbr#">
  <rdf:Description rdf:about="info:fedora/demo:11">
    <rel:isMemberOf rdf:resource="info:fedora/demo:10"/>
    <std:fulfillsStandard rdf:resource="info:fedora/demo:Standard5"/>
    <frbr:isManifestionOf rdf:resource="
  info:fedora/demo:Expression2"/>
  </rdf:Description>
</rdf:RDF>
```

### ***Object representations and properties in the Resource Index***

As described earlier, a Fedora digital object consists of a number of core components such as datastreams and disseminators, which bind to BDefs and BMechs. In addition each Fedora digital object has system metadata or properties. The architecture provides a system-defined ontology to represent the relationships among these core components. For example, the relationship of an object to its representations is expressed using the `<fedora-model:disseminates>` predicate as shown in the triple in Table 5.

Table 5 - Object-representation relationship

```
<info:fedora/demo:11>  
<fedora-model:disseminates>  
<info:fedora/demo:11/HIGH>
```

In addition to these relationships, the system-defined ontology also represents object data properties whose range contains date and boolean datatypes, as shown in the triple in Table 6.

Table 6 - Data type properties

```
<info:fedora/demo:11/HIGH>  
<fedora-view:lastModifiedDate>  
"2004-12-12T00:22:00"^^xsd:dateTime
```

Unlike the relationships expressed in the Relations datastream, these relationships are not explicitly asserted within the digital object. Instead they are derived from the structure of the object itself and mapped into the Resource Index, alongside the relationships represented in the Relations datastreams. This is described in the next section.

### ***Storing and querying the relationship graph***

All these relationships – the relationships explicitly stated in the Relations datastream, the relationships implied by the object structure, and the data relationships contained in the object properties – are stored in the Resource Index. This index is automatically updated by the repository service whenever an object structure is modified or its Relations datastream is changed.

The Resource Index handles queries over these relationships. The combination of all relationships into a single graph, and the automated management of that combined graph, enables a powerful and flexible service model. External services may issue queries combining relationships from different name spaces, since they are all RDF properties. For example,

Table 7 shows a query listing all the representations of all objects that are members of a particular collection.

Table 7 - Sample RDF query using iTQL

```
select $dissemination
from<#ri>
where ($object <fedora-view:disseminates> $dissemination)
and $object <rel:isMemberOf><demo:10>
```

An early design goal of the Resource Index was to allow the use of different triplestores and thus permit the Fedora repository administrator to choose the most appropriate underlying store. To that end, the Resource Index employs a triplestore API similar in spirit to JDBC, to provide a consistent update and query interface to a variety of triplestores. Extensive testing of both query performance time and query language features ultimately led to the selection of Kowari as the default triplestore<sup>105</sup>.

The query interface to the relationship graph currently supports three RDF query languages, RDQL<sup>106</sup>, iTQL<sup>107</sup>, and SPARQL [410]. Both RDQL and iTQL share a superficially similar syntax to SQL, with RDQL enjoying broader implementation support, but iTQL providing a richer feature set [223].

The RDF query results naturally take the form of rows of key-value pairs, again similar to the result sets returned by a SQL query. However, it is often useful to work with a sub-graph or a constructed graph based on the original. To this end, the query API may also return *triples* instead of *tuples*.

---

105 Our preliminary report on query performance of various triple store technologies is available at <http://tripletest.sourceforge.net/2005-06-08/index.html>.

106 <http://www.w3.org/Submission/2004/SUBM-RDQL-20040109/>

107 <http://kowari.org/271.htm>.

### *Using the relationship graph*

The Resource Index is exposed as one of the interfaces of the core Fedora repository service. This facilitates the development of other services in the Fedora Service Framework<sup>108</sup>. The Resource Index interface is exposed in a REST architectural style to provide a stateless query interface that accepts queries by value or by reference. The service has been implemented with an eye toward eventual conformance to the W3C Data Access Working Group's SPARQL protocol for RDF [127], as it matures. One example of a service exploiting the Resource Index is the OAI Provider Service that exposes metadata about resources in a Fedora repository. This OAI Provider Service is quite flexible in that it can be configured to allow harvesting not only of static metadata formats, but those that are dynamically produced via service-based disseminations of Fedora objects.

Table 8 - A query to build an OAI response

```
select $member $collection $dissemination
from<#ri>
where $member <rel:isMemberOf><info:fedora/demo:10>
and $member <rel:isMemberOf> $collection
and $member <rel:isMemberOf> $dissemination
and $member <fedora-view:disseminates> $dissemination
and $dissemination <fedora-
view:disseminationType><info:fedora/*/bdef:OAI/getQualifi
edDC>
```

An example of the interaction of this service with the Resource Index is as follows. An external OAI harvester requests qualified Dublin Core records for a particular set

---

<sup>108</sup> <http://fedora.info/download/2.1/userdocs/server/features/serviceframework.htm>

of resources from the repository. The OAI Provider service processes this by issuing the query to the Resource Index listed in Table 8. This query effectively requests “all *disseminations* of qualified Dublin Core records of resources that are members of the collection identified as `demo:10`. The significance of requesting disseminations is that the Dublin Core records may not statically exist as datastreams within the object, but they may be derived from another metadata format such as MARC. The Resource Index query would return the tuples shown in Table 9 that can provide the basis of an OAI response. Note that the OAI representations were not shown earlier in Figure 36.

Table 9 - The query response as triples

member	collection	dissemination
<code>info:fedora/ demo:11</code>	<code>info:fedora/ demo:10</code>	<code>info:fedora/ demo:11/ bdef:OAI/getDC</code>
<code>info:fedora/ demo:12</code>	<code>info:fedora/ demo:10</code>	<code>info:fedora/ demo:12/ bdef:OAI/getDC</code>

## Results

Fedora has been tested in the field with the real-world collections of our collaborators. These applications demonstrate the flexibility of the Fedora object model and repository service to accommodate a diverse set of information management problems. They distinguish Fedora from seemingly similar architectures such as DSpace, arXiv, and ePrints, whose focus is primarily on institutional repositories for scholarly publications. The applications supported by Fedora include not only complex digital library collections<sup>109</sup> [279] and institutional document repositories [457], but also

---

<sup>109</sup> <http://www.lib.virginia.edu/digital/collections/>

electronic records archives [488], trusted repositories for digital preservation [248], and distributed information networks such as the NSDL. This section describes results of Fedora deployment in three of these contexts: the University of Virginia digital library collections; the Encyclopedia of Chicago, a multimedia cultural heritage resource; and the National Science Digital Library, a distributed information network of educational resources and contextual information about those resources.

The core functionality of Fedora has proven effective for integrating rich digital collections at the University of Virginia Library (UVA). The UVA digital repository is built upon well-defined “content models” for digital content, where a content model specifies the number and types of datastreams and disseminators for particular genre of complex digital objects, including images, books, letters, archival finding aids, and data sets. The result is a seamless integration of content in a repository that enables consistent management of digital objects, consistent interfaces to access digital objects, and easy re-use of digital materials in different contexts. The architecture provides a means to easily aggregate materials from different collections and create new views on content using both Fedora relationship metadata and custom disseminators. One example is a cross-collection object that builds upon multiple objects: one that disseminates architectural drawings about historical buildings, another that contains recent photographs of those buildings from art collections, and another that contains historical letters that mention the buildings from the electronic text collections.

Northwestern University’s use of Fedora demonstrates how Fedora’s flexibility allows the management and publication of rich multimedia objects. Most compelling is the

electronic version of the Encyclopedia of Chicago<sup>110</sup> produced in collaboration with the Chicago Historical Society. The encyclopedia is a multi-media resource that exploits the Fedora disseminator capability in novel ways. A notable feature of this application is the design of digital objects and disseminators to create rich, clickable maps. These maps are linked to disseminators that provide multi-dimensional views and contextual information about a location in Chicago. For example a street map of Chicago highlights sites of labor unrest. By clicking on the map, a user can discover and launch numerous disseminations that link to population statistics, newspaper articles, and historical data that relate to a particular place on the map. This is all done using well-designed digital object content models and rich, service-based disseminators to produce dynamic transformations of digital object content. Nearly every piece of content on the web site is a dissemination of a Fedora digital object, and interestingly, the entire web site itself is published via a single dissemination of a master collection object.

Our work in the context of the NSF-funded NSDL (National Science Digital Library) Project [498] is perhaps the most interesting example of the power of Fedora's relationship architecture. Our goal in the NSDL is not only to provide a digital library allowing search and access to distributed resources, but also to augment NSDL resources with context that defines their usability and reusability in different learning and teaching environments. By "context", we mean information such as the provenance of the resources, the manner in which resources have been used, comments by users that annotate and explain primary resources, and linkages between the resources and relevant state educational standards. While the NSDL work is

---

<sup>110</sup> <http://www.encyclopedia.chicagohistory.org/>

specifically targeted at the education domain, we argue that the notion of *contextualization* is increasingly important as a means of adding value to digital content and defining its quality based on provenance, utility, and other factors.

Using the content management and semantic web tools in Fedora we have implemented an *information network overlay* [298]. This architecture represents the data underlying the NSDL as a graph of typed nodes, corresponding to the information entities in the NSDL, and semantic edges representing the contextual relationships among those entities. The nature and variety of these relationships will evolve over time and, thus, any fixed schema approach for representing the network overlay would be too restrictive. Our results thus far indicate that the semantic web approach of Fedora is particularly well-suited for this application.

Figure 42 illustrates a fragment of the network overlay. The nodes in the overlay graph correspond to Fedora digital objects – each shape corresponding to an information entity in the NSDL. These entities include agents, resources, metadata, and the like. The edges are relationships among these entities, which are represented in the Resource Index. For example, Figure 42 shows the grouping of resources in collections, and the provenance trail of who originally recommended those resources and who manages them. Relationships from other ontologies, such as state education standards, are overlaid on this base graph. These are similarly represented in the Resource Index alongside the base ontology relationships. The entire knowledge base can then be queried by external services to build rich portals for users and tools for inferring quality, usability, and educational value.



foundation for the deployment of extended services and user/client applications that apply Fedora in a variety of domains.

Increasingly rich digital content is placing greater demands on the institutions responsible for the creation, storage, management, and preservation of that content.

Fedora is well positioned to meet those demands and its open architecture provides the basis for meeting new requirements as they develop in the future.