# Choosing Technology That Can Evolve With User Needs: A service-oriented approach to e-research, e-scholarship and advanced scholarly publication

Sandy Payette
Co-Director of the Fedora Project
Researcher at Cornell Information Science
payette@cs.cornell.edu

***Abstract:***
*Emerging trends in the processes of scholarly research and scholarly communication present significant challenges to libraries and other information-oriented institutions in terms of choosing technologies that best position for the future. The challenge of "connecting with users" has typically been associated with the most visible layer of an information system: the user interface itself. At the same time, it is essential that we understand how the choice of technologies that lie beneath this visible layer will reverberate throughout a system, ultimately enabling or disabling users in their creation, manipulation, and use of information resources. This is especially true over time as user needs change, and information must be re-used and re-purposed for new or evolving contexts. The significance of service-oriented architectures, with Fedora as an example, is discussed in the context of meeting user requirements in the areas of institutional repositories, scholarly publication, e-research, and e-scholarship applications.*

The challenge of "connecting with users" is usually interpreted as attending to user needs at the most visible layer of an information system: the user interface itself. While recognising the importance of this tangible component of the user experience, it is also essential that we understand how the choice of technologies that lie beneath this visible layer will reverberate throughout a system, ultimately enabling or disabling users in their creation, manipulation, and use of information resources. This is especially true over time as user needs change, and information must be re-used and re-purposed for new or evolving contexts.

Currently, there is increasing investment by libraries and academic institutions in technologies to support institutional repositories, scholarly communication, and the evolving processes often characterised as e-research and e-scholarship. Among these technologies are applications such as DSpace[i], EPrints[ii], and Greenstone[iii], as well as generic digital object architectures such as Fedora[iv] and aDORE[v]. In Australia, there is growing interest in Fedora as evidenced by its use in several prominent projects, including ARROW[vi], eScholarshipUQ[vii], and DART[viii]. Fedora is an open-source, web-service-based repository system that is being adopted worldwide as a foundation for digital libraries, institutional repositories, scholarly publication, digital archives, and more. In this plenary session, I will discuss how the architectural philosophy embodied in systems like Fedora can be beneficial to institutions whose mission is serving the evolving information needs of diverse user communities.

Fedora is deployed as a set of web services, and is not a turn-key, end-user application. In the short run, adoption of turn-key applications for institutional repositories, scholarly publishing, and the like are attractive due to minimal investment in programming/technical resources. However, from a long-term perspective, institutions must be wary of applications that cannot be easily disaggregated, whose components are not easily replaced or extended, and whose overall application design is too tightly integrated. Some may wonder how a system like Fedora can connect with users, when it is not deployed with its own out-of-the-box end-user experience. Paradoxically, the lack of end-user functionality is actually one of Fedora's greatest strengths in terms of connecting with users. A deeper analysis reveals that by offering flexible, *service-oriented* repository architecture, Fedora is well positioned as an enabling technology for accommodating a diverse set of user needs across different communities. Fedora can also be understood as an evolutionary technology since its design enables it to be adapted to fulfil new user requirements over time.

To appreciate Fedora, one must first understand a few basics of service-oriented architectures. Generally speaking, a service-oriented architecture promotes flexibility and extensibility through the design of well-defined services that are exposed with formally-defined program interfaces, each embodying a clear functional purpose. Services act as building blocks that can be composed within clients or middleware to create customised systems that fulfil different user needs, at different institutions, at different points in time. Services that can be re-used in different contexts are conducive to "orchestration" where they can be sequenced together to support workflows and processes that fulfil current and future end-user needs. All of this suggests that services should be *loosely coupled*, meaning they should not be laden with hard-wired interdependencies to other services or to particular end-user application interfaces. Figure 1 depicts such an environment built upon a Fedora service-oriented foundation. The Fedora-based services at the bottom of the diagram provide generic functionality upon which multiple

custom services can be built to support the work processes of scholars, archivists, library patrons, and other constituencies. The more modular and flexible the foundation layer of the environment, the more opportunity there is for institutions to create specialised services and applications that can make a compelling impact on the end user.

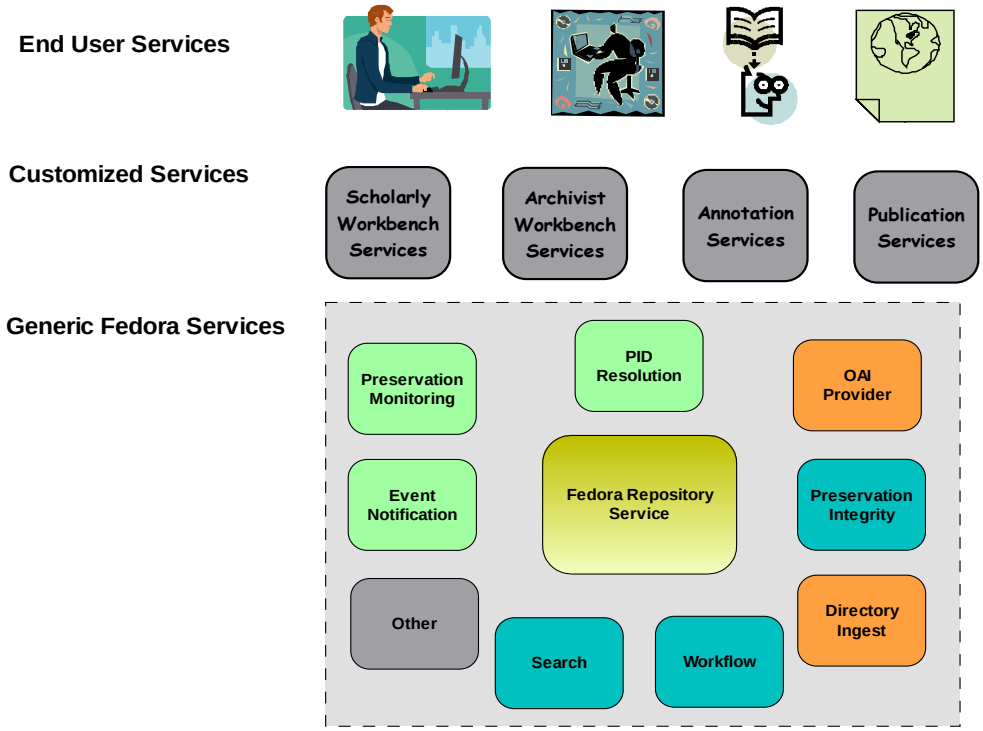


**Figure 1: Enabling user needs with service-oriented architecture**

Fedora has been designed in accordance with basic principles of service-oriented architecture, and is implemented with standard web services technologies (e.g., HTTP, SOAP/REST, WSDL, XML). Below is a brief summary of key features that enable Fedora to support diverse and evolving user requirements:

1. *Expandable service framework*: One of the most significant ways a system can enable users is to provide the capability for new services to be plugged into a system to address new requirements. The Fedora Service Framework allows a set of cooperating services to be built around the core Fedora repository. Currently, Fedora provides services for repository management, digital object access, OAI provider, ingest, and search. Development is under way for workflow and advanced search services, with preservation, event messaging, and advanced ingest planned for 2006 and 2007. Additionally, institutions can build customised services to fulfil community-specific needs.
2. *Flexible XML-based digital object model:* A flexible data model allows institutions to create, manage, and publish objects that reflect the diverse forms of content that is produced and consumed by their user communities. Digital objects, the primary information entities in Fedora, can aggregate any number and any variety of content items (known as “datastreams”). These datastreams can be local to the repository or may

reference content anywhere on the web. Fedora's XML-based object wrapper format can be used to create different "content models" such as books, articles, learning objects, datasets, and more.

3. *Extensible metadata:* Systems that are tied to a fixed metadata schema are not accommodating of various forms of community-specific metadata. Since metadata and data are treated uniformly in Fedora's digital object model, any number and variety of metadata formats may be stored as datastreams, alongside content, in a digital object.
4. *Expressive inter-object relationships:* Fedora digital objects can contain RDF metadata that can express any type of relationships such as membership in collections, structural associations like articles in journals, pictures in albums, or taxonomic relationships. Relationship metadata is indexed and can be searched using semantic web query languages.
5. *Interoperable Ingest/Export:* To promote the exchange of digital objects with other systems or services, Fedora supports ingest and export of whole digital objects in several XML-based formats (METS, FOXML, forthcoming MPEG21-DIDL). Fedora's modular software design makes it straightforward to expand this list, ensuring the utility of the software over the longer term.
6. *Versioning and auditing*: The complete history of changes to Fedora digital objects is stored within the object XML wrapper. This makes Fedora well suited for a dynamic scholarly environment where object editing occurs frequently, yet specific object versions need to be exposed as published documents.
7. *Advanced policy-based access control:* Fedora includes XML-based access control built upon a standard known as the eXtensible Access Control Markup Language (XACML). This enables flexible and fine-grained access controls to be placed on the Fedora repository service (down to specific service operations), as well as on individual digital objects or constituent parts of digital objects.

The above capabilities have been major motivating factors for institutions that have selected Fedora as the foundation for systems to target both current and future user needs. At the University of Virginia Library[ix], Fedora's flexible object model has enabled consistent management of heterogeneous digital collections, plus the ability to customise and evolve the end-user experiences for diverse digital library collection including images, documents, letters, books, and scholarly editions. In the area of institutional repositories, the ARROW project is a leader in adopting service-oriented architecture principles. The use of Fedora as a generic repository service has enabled the creation of custom user interface applications (e.g., VITAL, VALET, Fez) and specialised services above the foundational repository services. In the area of cultural heritage, Northwestern University (with the Chicago Historical Society) has made clever use of Fedora's object model, especially the ability to associate services with objects. The showpiece of this effort is the electronic edition of the Encyclopedia of Chicago[x], a rich multimedia publication that integrates scholarly essays, historical newspaper articles, population data, and GIS. The National Science Digital Library (NSDL) is transitioning from a relational-database-driven metadata catalog system to a new Fedora-based architecture for digital libraries known as an *information network overlay*[xi]. The NSDL uses Fedora to enable the storage and exposure of educationally-relevant relationships among NSDL resources in the repository, as well as relationships between NSDL resources and external entities (e.g., national educational standards). Other projects such as OhioLINK's Digital Resource Commons[xii] and the University

of Hull's RepoMMan[xiii] project are demonstrating how workflow and orchestration services can interact with Fedora repository services to create a system that better reflects the requirements of scholarly research processes.

These innovative projects attest to the flexibility of Fedora, but more importantly, they speak to the benefits of service-oriented architectures in the domains of digital libraries, institutional repositories, e-scholarship, and related initiatives. While Fedora is a system that exhibits modularity, extensible service-orientation, and flexible XML-based data models, these same principles can be applied in heterogeneous computing environments. The need for scholarly systems that are flexible and can evolve is becoming increasing evident in new nationally and internationally-funded initiatives such as E-Research by Australian Research Counsel[xiv], the JISC Framework Program[xv], as well as activities in the areas of e-Science and grid computing. At Cornell, the NSF-funded Pathways[xvi] project (with LANL) is investigating new service-oriented architectures for enabling interoperability among heterogeneous repository systems (e.g., DSpace, Fedora, arXiv, aDORE), as well as ontology-based approaches for matching services to digital objects in a manner that transcends the boundaries of each individual system. All of these efforts explicitly or implicitly acknowledge that investment in new service-oriented architectures is necessary to support the emerging requirements of scientific and scholarly communication processes in the digital age.

## Endnotes

[i] DSpace, http://www.dspace.org/

[ii] EPrints, http://www.eprints.org/

[iii] Greenstone, http://www.greenstone.org

[iv] Flexible Extensible Digital Object Repository Architecture (Fedora), http://www.fedora.info

[v] aDORe, http://comjnl.oxfordjournals.org/cgi/content/abstract/48/5/514

[vi] Australian Research Repositories Online to the World (ARROW), http://arrow.edu.au/

[vii] eScholarshipUQ, http://www.library.uq.edu.au/escholarship/

[viii] Dataset Acquisition Accessibility & Annotation e-Research Technologies (DART), http://dart.edu.au/

[ix] University of Virginia Digital Collections, http://www.lib.virginia.edu/digital/

[x] Encyclopedia of Chicago, http://www.encyclopedia.chicagohistory.org

[xi] NSDL and Fedora, http://www.dlib.org/dlib/november05/lagoze/11lagoze.html

[xii] OhioLINK DRC, http://info.drc.ohiolink.edu/

[xiii] RepoMMan, http://www.hull.ac.uk/esig/repomman/

[xiv] ARC E-Research, http://www.arc.gov.au/grant_programs/special_research_initiatives.htm

[xv] JISC Framework, http://www.jisc.ac.uk/index.cfm?name=programme_frameworks

[xvi] Pathways, http://www.infosci.cornell.edu/pathways/