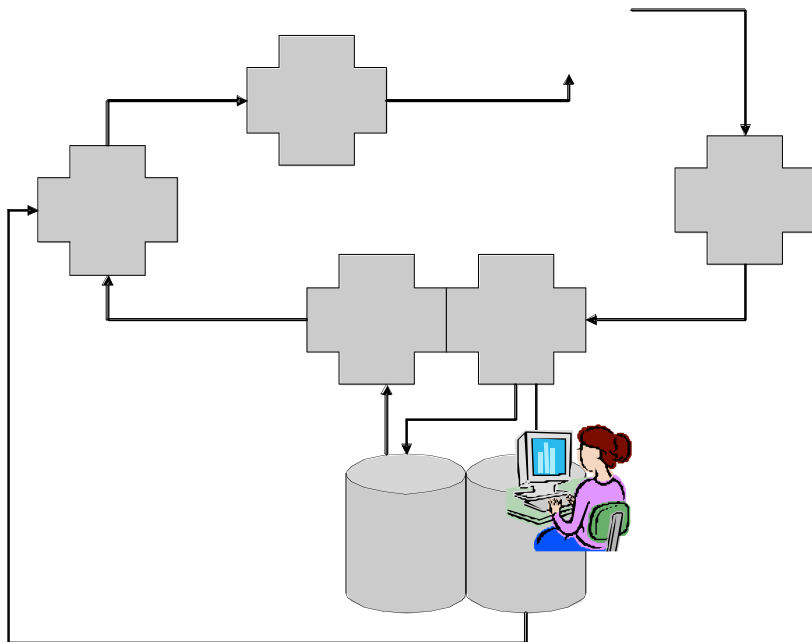


# WikiD (Wiki/Data)



**Jeffrey A. Young**  
**OCLC Office of Research**  
[jyoung@oclc.org](mailto:jyoung@oclc.org)

**code4lib 2006**  
**Oregon State University, Corvallis,**  
**Oregon**  
**15 February 2006**

## WikiD Overview

- WikiD is an open-source online database for servicing multiple data collections involving arbitrary schemas of XML
- J2EE webapp
- Customizable and extensible
  - Displays (XSL)
  - Crosswalks (XSL)
  - Services (Java abstractions)
- Built using modular standards-based APIs
  - SRU/SRU Update
  - OpenURL 1.0
  - OAI-PMH 2.0
  - RSS 2.0

## OAI/WikiD Comparisons

### OAI-PMH

- Identifiable distributed “repositories”
- oai:**gsafd.oclc.org**:localID
  
- Identifiable “items”
- oai:gsafd.oclc.org:GSAFD000001
- Multiple *metadata*Formats
  
- OAI Registry at UIUC
- oai:url-registry.uiuc.edu:gsafd.oclc.org
  
- Fixed OAI “verbs”
- Repository-level “branding”
- Read-only
- Access latest records only

### WikiD

- Identifiable local “collections”
- **Collection**Gsafd
  
- Identifiable “items”
- CollectionGsafd:GSAFD000001
- Multiple *record*Formats
  
- “CollectionCollection”
- CollectionCollection:CollectionGsafd
  
- Extensible wiki-style “actions”
- Collection-level XSL
- Read/write
- Access historical records

## Primary Bootstrap Collections

- **CollectionCollection**
  - Collection items: descriptions of collections
- **CollectionExternalSchemas**
  - Collection items: descriptions of XML Schemas that are referenced by items in CollectionCollection
    - Schema title (for use in pull-down menus)
    - Namespace URI
    - Schema URL
- **CollectionWikiPages**
  - Collection items: wiki markup pages
  - Provides the user interface for WikiD

## Persistent URIs

- All items in all collections are referred to internally by URIs
  - info:wikid/**CollectionCollection:CollectionGsafd**
  - info:wikid/**CollectionGsafd**
  - info:wikid/**CollectionGsafd:GSAFD000001**
- The WikiD API can recognize and service shorthand forms
  - CollectionCollection:CollectionGsafd
  - CollectionGsafd
  - CollectionGsafd:GSAFD000001
- Persistent in the same sense that OAI identifiers persist



Message: Action complete: Update

[To the FrontPage](#)

[RSS](#) Collection: [Wiki Pages](#) ([config](#))

Item: CollectionWikiPages:WikiDSandbox

## WikiD Sandbox

---

### GSAFD Thesaurus

- [?CollectionGsafd](#)



[OAI-PMH Identify](#) (CollectionWikiPages)

<b>Collection Name</b>	<input type="text" value="GSAFD Thesaurus"/>		
<b>Description</b>	<input type="text" value="Guidelines on Subject Access to Individual Works of Fiction, Drama, etc. (GSAFD)"/> <a href="http://www.oclc.org/research/projects/termservices/resources/gsafd.htm">http://www.oclc.org/research/projects/termservices/resources/gsafd.htm</a>		
<b>Local Identifier Type</b>	<input checked="" type="radio"/> User-assigned <input type="radio"/> System-assigned <input type="radio"/> External ( <a href="#">Instructions</a> )		
<b>Administrator's email</b>	<input type="text" value="mailto:jyoung@oclc.org"/>		
<b>XSL Stylesheet</b>	<input type="radio"/> System-created <input checked="" type="radio"/> Manually-created ( <a href="#">Instructions</a> )		
<b>Native Record Schema</b>	default <input type="radio"/>	<input type="text" value="marc: MARC 21 XML Schema"/> <a href="#">Add new</a>	MetadataPrefix: <input type="text" value="marc"/>
<b>Crosswalks</b>	See the <a href="#">instructions</a> to learn how to configure crosswalks		
<b>Crosswalk Schema</b>	default <input checked="" type="radio"/>	<input type="text" value="zthes: A Profile for Thesaurus Navigation"/> <a href="#">Add new</a>	MetadataPrefix: <input type="text" value="zthes"/>

## Content Storage Overview

- Updates to items in WikiD are written to the fileserver according to the collection's "native format"
- But WikiD can't intelligently index arbitrary XML, so it creates and indexes metadata instead
- Metadata fields: (with their approximate OAI equivalents)
  - Persistent item identifier ("local-identifier")
  - Collection identifier ("namespace-identifier")
  - Sort/brief display value (e.g. dc:title or MARC 245)
  - Date created
  - Date last updated (oai:datestamp)
  - Relative URL containing the native record (dc:identifier)
- The metadata records for all items in all collections get stored in a single SRU database (and also to the fileserver)



## Content Storage Detail: WebEdit

- A lightweight open-source web service to transform HTML form submittals into a target XML format
  1. HTML form fields get encoded in a linear XML structure
    - `<document>`  
`<fieldName1>fieldValue1</fieldName1>`  
`<fieldName2>fieldValue2</fieldName2>`  
...  
`</document>`
  2. This document gets transformed into a *metadata* record by an XSL Stylesheet indicated by a hidden form field
  3. Included in this stylesheet are templates to construct the native record and Base64-encode it as an element in the metadata record

## Content Storage Detail: SRU Update

- A standards-based web service protocol developed by the SRU community to update XML records in an SRU database
- Ralph Levan's open-source implementation automatically handles the strange version-controlled native/metadata bifurcation described earlier
  - Yanks the Base64-encoded native record from the metadata record, decodes it, and stores it on the fileserver
  - Inserts the relative URL of the native record back into the metadata record
  - Touches the metadata record's date fields
  - Stores/indexes the metadata record in the database

## GSAFD Collection Description (Metadata)

- Persistent item identifier
  - **CollectionGsafd**
- Collection identifier
  - **CollectionCollection**
- Sort/brief display value
  - **GSAFD Thesaurus**
- Relative URL containing the native record
  - **2006/01/27/20/inf\_3awikid\_2fCollectionCollection\_3aCollectionGsafd\_5f200601272016502020**



<a href="#">To the FrontPage</a>	<b>RSS</b> Collection: GSAFD Thesaurus <a href="#">(config)</a>	New item identifier: <input type="text"/> <input type="button" value="Create Item"/>							
<table border="1"><thead><tr><th>Item identifier</th><th>Item label</th><th>Edit link</th><th>Deposit record link</th><th>Record content link</th></tr></thead></table>					Item identifier	Item label	Edit link	Deposit record link	Record content link
Item identifier	Item label	Edit link	Deposit record link	Record content link					
<p>There are no records in this collection</p>									
<b>Import records:</b> <a href="#">(marc)</a>	<input type="text" value="p.org/~vazine/GSAFD/MARC-XML/2004-07-20/gsafd-lcsh-lcshac.2004-07-20.marc21a.xml"/> <input type="button" value="Import"/>								
	<a href="#">OAI-PMH Identify</a> (CollectionGsafd)								

## GSAFD Item (Metadata)

- Persistent item identifier
  - **GSAFD000001**
- Collection identifier
  - **CollectionGsafd**
- Sort/brief display value (taken from the MARC 245 field)
  - **Adventure fiction**
- Relative URL for the native file
  - **2005/06/10/23/inf\_3awikid\_2fCollectionGsafd\_3aGSAFD000001\_5f20050610232847520**

[To the FrontPage](#)

**RSS** Collection: GSAFD Thesaurus  
([config](#))

New item identifier:

Create Item

Item identifier	Item label	Edit link	Deposit record link	Record content link
GSAFD000153	<a href="#">Western television programs</a>	<a href="#">Edit item</a>	<a href="#">View deposit</a>	<a href="#">View raw data</a>
GSAFD000152	<a href="#">Western stories</a>	<a href="#">Edit item</a>	<a href="#">View deposit</a>	<a href="#">View raw data</a>
GSAFD000151	<a href="#">Western radio programs</a>	<a href="#">Edit item</a>	<a href="#">View deposit</a>	<a href="#">View raw data</a>
GSAFD000150	<a href="#">Western films</a>	<a href="#">Edit item</a>	<a href="#">View deposit</a>	<a href="#">View raw data</a>
GSAFD000149	<a href="#">Western comic books, strips, etc.</a>	<a href="#">Edit item</a>	<a href="#">View deposit</a>	<a href="#">View raw data</a>
GSAFD000148	<a href="#">War television programs</a>	<a href="#">Edit item</a>	<a href="#">View deposit</a>	<a href="#">View raw data</a>
GSAFD000147	<a href="#">War stories</a>	<a href="#">Edit item</a>	<a href="#">View deposit</a>	<a href="#">View raw data</a>
GSAFD000146	<a href="#">War radio programs</a>	<a href="#">Edit</a>	<a href="#">View deposit</a>	<a href="#">View raw data</a>



## The Need for OpenURL 1.0

- SRU is great for interactive searching, but sometimes we want to provide services involving more complex operations
  - Dereference the native items in the metadata records
  - Perform complex query chains
    - Include a pull-down menu of alternative record formats obtained from the collection description record on an item-level record display
  - Result transformations
    - Crosswalks
    - SRU to OAI-PMH
    - SRU to RSS
- OpenURL 1.0 gives us a single consistent API for performing any and all services that reference these items

- 
- 
- 
- 
- 

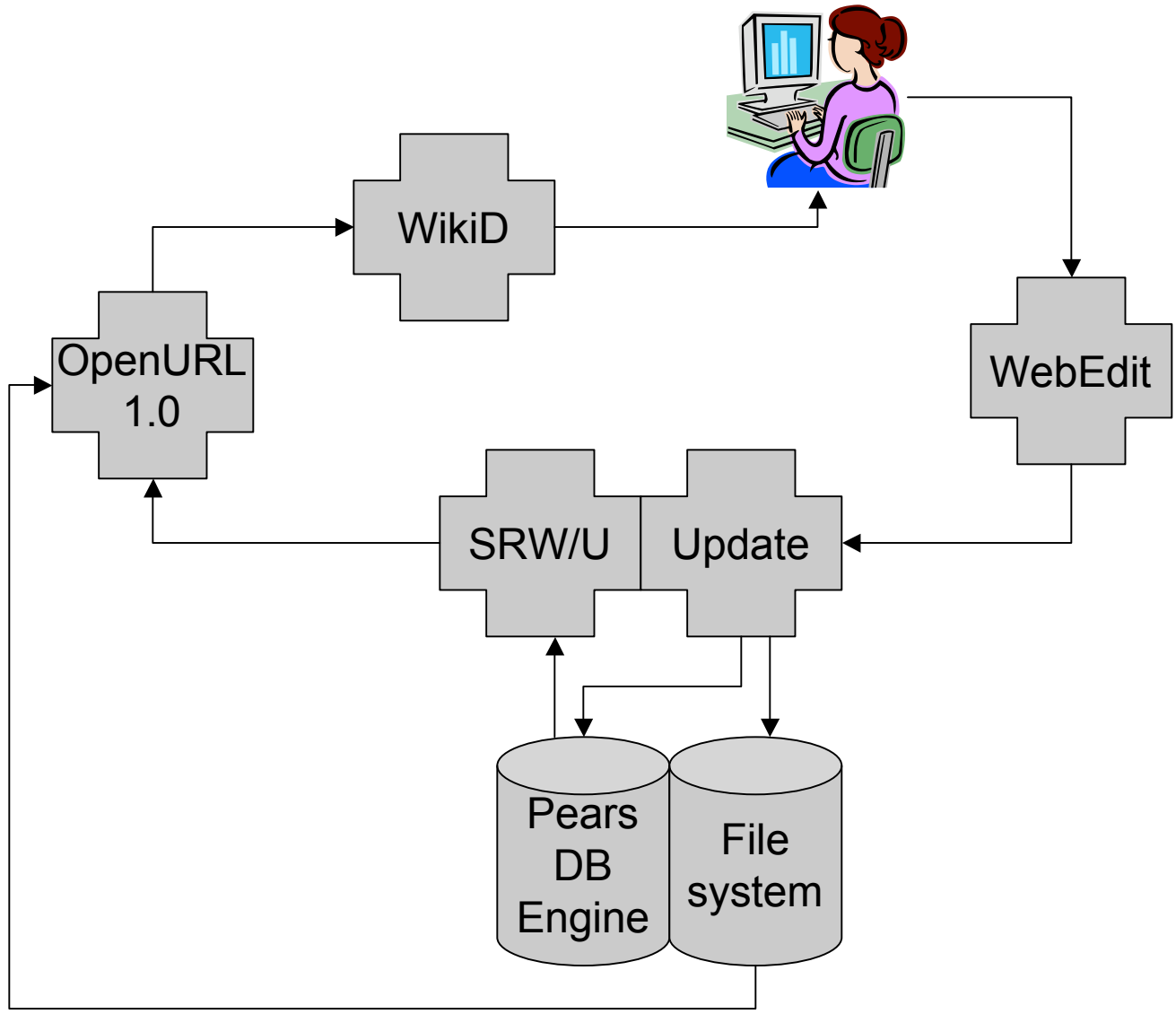
<b>Term</b>	Adventure fiction
<b>SN</b>	Use for works characterized by an emphasis on physical and often violent action, exotic locales, and danger, generally with little character development.
<b>UF</b>	Adventure stories
<b>UF</b>	Swashbucklers
<b>UF</b>	Thrillers
<b>RT</b>	<a href="#">Sea stories</a>
<b>NT</b>	<a href="#">Picaresque literature</a>
<b>NT</b>	<a href="#">Robinsonades</a>
<b>NT</b>	<a href="#">Western stories</a>
<b>MT</b>	Adventure and adventurers — Fiction [ <a href="#">lcsnac</a> ]
<b>MT</b>	Adventure stories [ <a href="#">lcsn</a> ]



## WikiD API

- Although the OpenURL 1.0 API can mediate all conceivable functionality, the URLs involved are hideous
  - [http://localhost:8080/wikid/resolver?url\\_ver=Z39.88-2004&url\\_ctx\\_fmt=info%3Aofi/fmt%3Akev%3Amtx%3Actx&ctx\\_enc=info%3Aofi/enc%3AUTF-8&rft\\_id=info%3Awikid/CollectionWikiPages%3AFrontPage&svc\\_dat=action%3Dedit&rft\\_id=info%3Asid/oclc.org%3Areferred/WikiD](http://localhost:8080/wikid/resolver?url_ver=Z39.88-2004&url_ctx_fmt=info%3Aofi/fmt%3Akev%3Amtx%3Actx&ctx_enc=info%3Aofi/enc%3AUTF-8&rft_id=info%3Awikid/CollectionWikiPages%3AFrontPage&svc_dat=action%3Dedit&rft_id=info%3Asid/oclc.org%3Areferred/WikiD)
- In contrast, the WikiD Servlet accepts wiki-style URLs
  - <http://localhost:8080/wikid/FrontPage?action=edit>
  - WikiD performs a simple mechanical mapping into the OpenURL equivalent and hands it off to the OpenURL resolver
- Human users will generally prefer the WikiD API, but automated clients might prefer to access the OpenURL, SRU, WebEdit, and SRU Update APIs directly

# Servlets and data flows



## Reference Links

- WikiD project page
  - <http://www.oclc.org/research/projects/wikid/>
- WikiD demo
  - <http://alcme.oclc.org/wikid/>