



Springer

the language of science

Preventing DOI Conflicts

Mike Waters, Process and Content Management

A circular badge with an orange background and a white border. The year '2007' is written in white, bold, sans-serif font in the center.

2007

Springer History and Content

- Plenum Publishers
 - 150 English journals + 100 Russian journals, some hosted on CatchWord, no DOIs
- Kluwer Academic Publishers acquires Plenum
 - plus 400 journals hosted on KluwerOnline + 2000 eBooks on NetLibrary, about 300,000 DOIs
- Bertelsmann acquires Springer-Verlag to form BertelsmannSpringer
 - plus 1100 journals, hosted on LINK
- KAP and BertelsmannSpringer merge under Cinven/Candover to form Springer SBM
 - content migrated to new SpringerLink platform
- Springer acquires journals from Humana Press, Current Medicine Group, Transaction Publishers, Science in China Press, and others
 - **1900 journals, with about 2.9 million DOIs**
 - **21,000 books in 850 book series, with about 500,000 DOIs**
 - **about 5500 new books each year**

Content and Metadata Workflows

- Old Workflows
 - Paper-based routing of manuscripts through editorial and production
 - Electronic files from authors and to vendors via disks, CD-ROM, FTP, HTTP
 - Minimal validation of electronic content before posting to web site
- New Workflows
 - Editorial: Editorial Manager and ScholarOne, capturing metadata and content
 - Production: Documentum production workflow
 - Validation of incoming content: Submission Checker
 - Checks zip file packaging, directory structure, PDFs, images
 - Checks incoming XML with rule-based XML language, similar to Schematron
 - Content delivery: Data repository delivers content to various clients
 - Metadata: CrossRef, PubMed, etc.
 - Metadata + content: SpringerLink, bulk users (consortia)

DOI Conflicts, Where Do They Come From?

- External workflows
 - Sometimes cheap, but not integrated with internal systems
- Platform migrations and parallel workflows
 - Business need to keep things running while migrating
- Introduction of new internal workflows
 - Every new component in the workflow can be a point of failure
- Backfile conversions
 - Sheer size of project complicates production, especially if outsourced
- Journal takeovers
 - Editors more concerned with content, not DOIs
- Upgrades to system software, APIs, and DTDs
 - Can revert to default encodings (Latin-1) from UTF-8
 - Content might not up-translate cleanly to new DTD

How To Prevent DOI Conflicts

- Non-Technical Aspects
 - Promote proper DOI and metadata handling throughout the organization
 - Internal “data bug” tool to keep external partners informed; quality does count
- Technical Aspects
 - Submission Checker is sophisticated Perl script with number of modules, primarily:
 - Content Checker verifies that all DOIs have known Springer prefixes
 - PDF Checker checks consistency of DOIs in PDFs vs. XML metadata
 - DOI Inconsistency Protection (DIP) queries CrossRef
 - queryWithDOI, DOI request to CR Servlet interface
 - queryWithMetadata, metadata request to CR OpenURL interface
 - Log file records results of all tests, giving different classes of messages: Info level, warning level, error level. Error level messages cause submission to be rejected

Using the OpenURL Interface: queryWithMetadata

- Detects if DOI conflict exists, i.e., more than 1 DOI returned in response
- Enables comparison of submission's DOI vs. CrossRef DOI
- Uses uri_escape_utf8 Perl module to escape non-ASCII
- Example Good Request

```
http://www.crossref.org/openurl?pid=USERNAME%3APASSWORD  
&issn=15729559&volume=18&issue=9&page=1663&aulast=Br%C3%A9hat  
&date=1997&noredirect=true&multihit=exact
```

queryWithMetadata (continued)

- Example Good Response

```
<query key="555-555" status="resolved">
<doi type="journal_article">10.1007/BF02678278</doi>
<issn type="print">01959271</issn>
<issn type="electronic">15729559</issn>
<journal_title>International Journal of Infrared and ...</journal_title>
<author>Bréhat</author>
<volume>18</volume>
<issue>9</issue>
<first_page>1663</first_page>
<year>1997</year>
<publication_type>full_text</publication_type>
<article_title>Measurement of the optical constants ...</article_title>
</query>
```

queryWithMetadata (continued)

- Example Bad Request

```
http://www.crossref.org/openurl?pid=USERNAME%3APASSWORD  
&issn=15729559&volume=18&issue=9&spage=1663&aulast=Br%E9hat  
&date=1997&noredirect=true&multihit=exact
```

- Example Bad Response

```
<html>  
<head><title>CrossRef OpenURL Resolver</title></head>  
<body><a href='http://www.crossref.org'><img border=0 alt='CrossRef'  
src=http://www.crossref.org/images/crossref.gif></a><br><br>  
<span style='font-style:bold;font-size:16;color:red'>No DOI found</span>  
<br><br>You may use the  
<a href='http://www.crossref.org/03libraries/37guest_query.html'>  
CrossRef Guest Query </a>form to try and locate a DOI.</body>  
</html>
```

queryWithMetadata (continued)

- Example DOI Conflict Request

```
http://www.crossref.org/openurl?pid=USERNAME%3APASSWORD  
&issn=14320983&volume=29&issue=3&spage=293&aulast=Geiser  
&date=1996&noredirect=true&multihit=exact
```

queryWithMetadata (continued)

- Example DOI Conflict Response

```
<query key="555-555" status="multiresolved">  
<doi type="journal_article">10.1007/s002940050049</doi>  
<issn type="print">01728083</issn>  
<issn type="electronic">14320983</issn>  
<journal_title>Current Genetics</journal_title>  
<author>Geiser</author>  
<volume>29</volume>  
<issue>3</issue>  
<first_page>293</first_page>  
<year>1996</year>  
<publication_type>full_text</publication_type>  
<article_title>Wild chromosomal variants in ...</article_title>  
</query>
```

queryWithMetadata (continued)

- Example DOI Conflict Response (continued)

```
<query key="555-555" status="multiresolved">  
<doi type="journal_article">10.1007/BF02221561</doi>  
<issn type="print">01728083</issn>  
<issn type="electronic">14320983</issn>  
<journal_title>Current Genetics</journal_title>  
<author>Geiser</author>  
<volume>29</volume>  
<issue>3</issue>  
<first_page>293</first_page>  
<year>1996</year>  
<publication_type>full_text</publication_type>  
<article_title>Wild chromosomal variants in ...</article_title>  
</query>
```

Using the Servlet Interface: queryWithDOI

- Detects whether DOI exists at CrossRef
- Enables comparison of submission's metadata vs. CrossRef metadata
- Example Servlet Request

```
http://doi.crossref.org/servlet/query?id=DOI&pid=USERNAME:PASSWORD
```

- Example Servlet Response is parsed using Perl XML module

Summary

- Promote better awareness of DOIs and metadata throughout your organization
- Keep external partners informed and audit for data quality
- Be aware of any changes in systems or software. They can adversely affect quality of data that gets pushed from system to system
- Check incoming data thoroughly against CrossRef, not just against internal systems
- UTF-8 always and everywhere. No exceptions!!

Thanks!