Current state of TDM licensing developments

Brussels, 14 October 2013

Wim van der Stelt, Springer
Agenda

- Introduction
- Sample Licensing Clauses
- PLSclear
- Prospect
- CCC
Licensing Clauses

Sample clauses are the first step for any solution

- Sample clause for licencees (librarians) available since early 2013
- Click through license for the individual researcher
- PDR for commercial use (since 2012)
PLS’s TDM permissions wizard

Simple 3-step process
TDM CHALLENGES 1. Not knowing what information publishers need to grant permission

Questions defined by a PLS Publishers working group and refined with a researcher.
TDM CHALLENGES  2. Not knowing what publications to mine

Requests can be made to mine all of a Publisher’s content sets or specific sets only.
TDM CHALLENGES

3. Not knowing who to contact at the relevant publisher

4. Having to make multiple requests

Requests will be routed to contacts on PLS’s database.

Multiple requests can be processed at the same time.
Not-for-profit association of scholarly publishers

4,000+ organizations from all over the world

All subjects, all business models

83 non-publisher affiliates, 2000 library affiliates

62.7 million content items
Services

- Cross-publisher reference linking
- Cross-publisher Cited-by linking
- Cross-publisher metadata feeds
- Cross-publisher plagiarism screening
- Cross-publisher update identification
- Cross-publisher funder identification
- Cross-publisher text and data mining
prospect
by Crossref

Text and Data Mining Hub for Researchers
1. Researchers want to get full text content from publishers’ sites for OA or subscribed content.

Solution:

Simple web form to get access key for subscribed full text content from many different publishers.
2. Researchers want to know whether text and data mining is allowed, and if not, get permission.

Solution:

Licensing information embedded in article metadata and a registry for supplemental text and data mining terms and conditions (licenses).
Researcher has content (s)he wants to mine and gets unique identifiers for them (DOIs)

"10.1177/107769909507200314",
"10.1016/S0022-2828(85)80098-5",
"10.1114/B:ABME0000036643.45452.6d",
"10.1097/prs.0b013e31826d9c8d",
"10.1115/1.3446498",
"10.1111/1467-9620.00198",
"10.1111/j.1440-1703.2003.00624.x",
"10.1021/jp036721c",
"10.5264/eiyogakuzashi.70.294",
"10.1515/cclm-2012-0408",
"10.1098/rsta.1927.0004",
"10.1007/s11665-012-0211-6",
"10.1007/bf02739852",
"10.2307/1240591",
"10.1371/journal.ppat.1003333.g006",
"10.1007/978-0-387-79418-1_28",
"10.2310/6650.2004.11943"
How do the researchers know if they can TDM a particular article or book?
The researcher checks the metadata for the license reference provided by the publisher

<lic_ref>http://creativecommons.org/licenses/by/3.0</lic_ref>

If the license is recognized (e.g. CC) then he/she is good to go

Or the publisher provides supplementary TDM Terms
Given Names
Josiah

Family Name
Carberry

Email
gbilder@breakawayrepublic.com

API Token
_Vm1TkAA3DyKzJ_X8fPz4A

<table>
<thead>
<tr>
<th>License</th>
<th>Status</th>
<th>Review Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>CrossRef Generally Nice License</td>
<td>unread</td>
<td></td>
</tr>
<tr>
<td>Way Too Complex License</td>
<td>unread</td>
<td></td>
</tr>
<tr>
<td>Annals of PyschoCeramics B Prohibition Of TDM</td>
<td>unread</td>
<td></td>
</tr>
</tbody>
</table>

Copyright 2013, CrossRef.
Transfer of a complete and conditions for use of dealing. (c) Counsel. (c) Counsel. The agency or indirectly harmful to copyright, trade shows and description of the Board or retailers for issuance under any duty assignment for purposes of mutilated, Destroyed, Lost or governmental agencies and deliver a national securities and acknowledgments that within thirty (30) days following MUTUAL COVENANTS AND allowances from time to benefit obligations under the transfer, deliver any interest shall have the Committee. (b) Any of a loan and authority to hold such Holder thereof, exercised by such commissions shall not be imposed in a longterm disability for such Warrant certificates and shall become payable to be
<table>
<thead>
<tr>
<th>License</th>
<th>Status</th>
<th>Review Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>CrossRef Generally Nice License</td>
<td>accepted</td>
<td>2013-5-18</td>
</tr>
<tr>
<td>Way Too Complex License</td>
<td>unread</td>
<td></td>
</tr>
<tr>
<td>Annals of PsychoCeramics B Prohibition Of TDM</td>
<td>unread</td>
<td></td>
</tr>
</tbody>
</table>

Copyright 2013, CrossRef.
Researcher fill out request in simple web interface

Publisher verifies API token with Prospect
(frequency at publisher discretion)

If token verified AND access control allows, publisher returns fulltext
Prospect Benefits

- Streamlines researcher direct access to distributed full text for TDM
- Enables machine-to-machine, automated access for recognized TDM (i.e. researchers won’t be locked out of publisher sites)
- Enables article-level licensing info and easy mechanism for supplemental T&Cs for TDM (publishers encouraged to use model STM license)
Progress to date

- DOI Content Negotiation
- CrossRef support for recording links to fulltext
- CrossRef metadata support for ORCID
- CrossRef metadata support for FundRef
- CrossRef metadata Search for Discovery
- CrossRef metadata support for license URIs
- Click-through TDM license registry
- Prospect publisher API for verifying, managing tokens

* being extended to support mime-types
Licensing Service for Text and Data Mining (TDM)

CCC’s TDM Pilot Service Overview
October 2013

John Billington
Product Manager Corporate Products and Services
Problems for TDM Researchers:

• No single source for commercial TDM licenses. Negotiating with hundreds of publishers not feasible.

• Accessing full text XML articles is difficult and time consuming.

• Unsubscribed content is not accessible for Textmining research.

• Very few sources of structured XML which is needed for textmining.
The service is designed to be used in conjunction with existing enterprise text mining systems.
CCC’s Text and Data Mining Service: Features

• Search and access multiple STM publishers’ full-text archives via web interface or RESTful API

• Obtain a standard license for all content

• Create result sets including articles from existing subscriptions and unsubscribed articles

• Download result sets in normalized XML format (not PDF)
Pilot Status

• We plan to operate the TDM pilot program until the end of 2013. Pending feedback from customers we will work toward a commercial release in 2014 and expand publisher participation.

• CCC is working with 6 major STM rightsholders who have agreed to provide full text content for the purposes of a pilot.
  – We do not plan to add more publishers or content until we create a commercial service.
### TDM Pilot Roadmap

<table>
<thead>
<tr>
<th>Event</th>
<th>Timeframe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Pilot Release</td>
<td>May - Web UI</td>
</tr>
<tr>
<td></td>
<td>August - Enhancements</td>
</tr>
<tr>
<td>Pilot End</td>
<td>December</td>
</tr>
<tr>
<td></td>
<td>Decision to begin publisher outreach.</td>
</tr>
<tr>
<td>API Release</td>
<td>October 2013</td>
</tr>
<tr>
<td></td>
<td>Additional user testing</td>
</tr>
<tr>
<td>Commercial Release</td>
<td>Content Purchase capability</td>
</tr>
<tr>
<td></td>
<td>Timing 2014 TBD</td>
</tr>
</tbody>
</table>
Business Model Overview

• The business model includes an annual fee to download and mine full-text subscribed content.
  – A portion of this annual fee will be distributed to publishers based on downloads.
  – Users can purchase unsubscribed content with tokens on a per article basis.
Set Criteria for Your Project

Users enter keyword search or other criteria to define a set of documents.
Refine and Filter Content Sets

CCC returns article results with subscription status into a customized index for each organization.
Users can then download the XML versions of the articles for loading into their text mining tools.
Contact us for more information

John Billington
*Product Manager, CCC Corporate Products and Services Department*

+1 978.646.2633
jbillington@copyright.com