

CCC's Text and Data Mining Pilot Service

Overview of 2013 Pilot Program

John Billington, *Product Manager Corporate Products and Services*
Copyright Clearance Center



Agenda

- Introduction to CCC
- CCC's activity in Text and Data Mining
- Market and use cases for TDM
- Challenges: Voice of the TDM user
- TDM Pilot features

Making Copyright Work

Rightsholders



600+ million rights from:

- Publishers
- Authors
- Agents
- Creators



- **Licensing Solutions**
- **Rights Management**
- **Content Delivery**
- **Copyright Education**



A Copyright Clearance Center Subsidiary



Content Users



- 35,000 companies
- 1,200 colleges and universities
- Publishers and Authors

CCC Activities in Text and Data Mining

- CCC has developed a pilot service that provides licensed, compliant access and retrieval of full text XML and metadata from multiple scientific publishers for the purposes of text mining.
- CCC's role is to provide an authorized means to **access and retrieve** published content in a standard format.
- The initial pilot is focused on corporate users.



Our Corporate Use Cases

- Corporate R&D users
 - Mid-Large Pharmaceutical companies.
 - Computational bioinformatics departments.
 - Project teams focusing on drug discovery.
- Future Markets:
 - Corporate marketing users.
 - Academic uses.



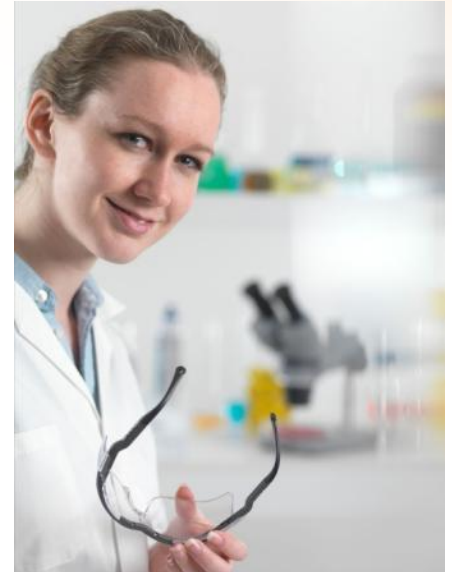
Research Challenges

Access to XML

“Up till now, it’s been very challenging to extract full text from different publishers and convert to a normalized format that is usable in text mining technologies.”

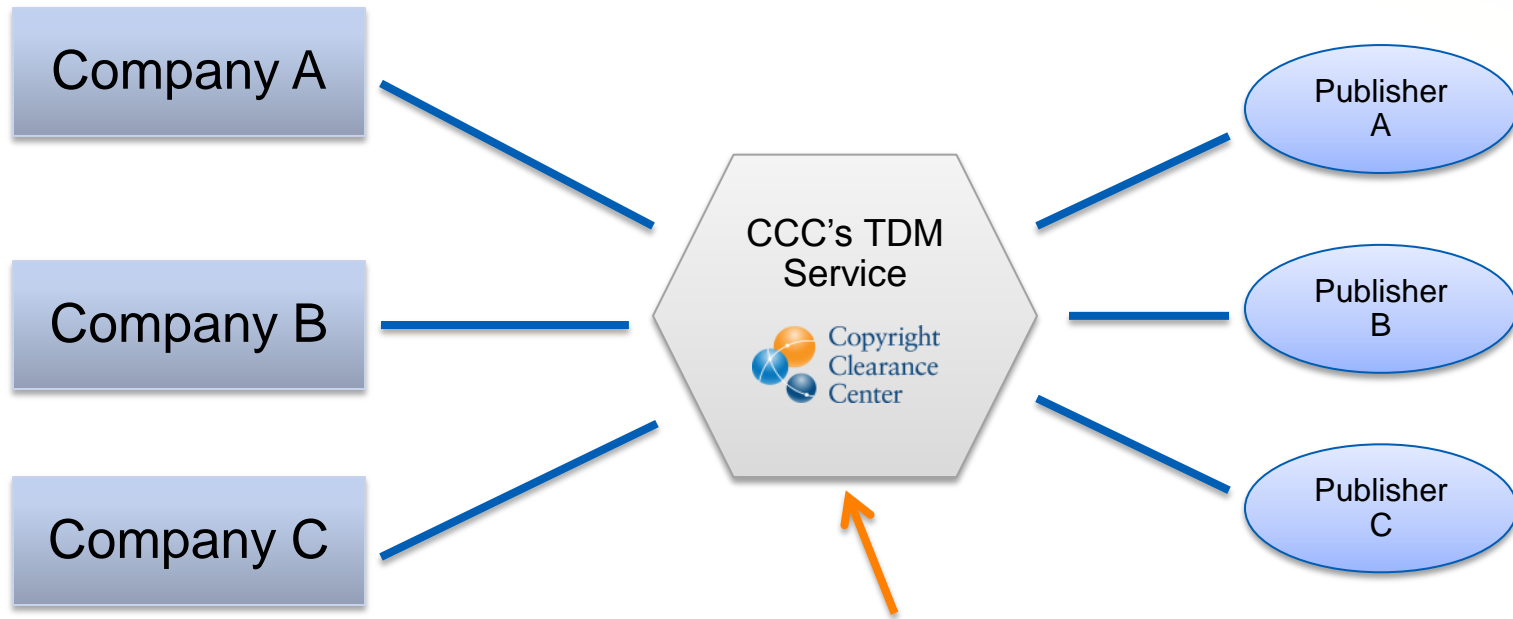
Lack of Federation

“[existing tools are] still very difficult. The only other approach we’ve had is to go to PubMed Central and download open access articles, and there isn’t anything else other than scraping.”



CCC's Text and Data Mining Pilot

The service will help researchers access and download XML content in a standardized way. We are working with users and publishers to design a business model and an access method that works for both sides.



The service is designed to be used in conjunction with existing enterprise text mining software and methods.

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 from existing subscriptions **and** from unsubscribed articles.
- Purchase mechanism for full text unsubscribed articles.
- Download result sets in normalized XML format.

Questions?

- John Billington, *Product Manager*, CCC Corporate Products and Services.

jbillington@copyright.com
+1 978 646 2633

