

The value of aggregating Open Access research

Dr Nancy Pontika
Open Access Aggregation Officer
CORE

Email: nancy.pontika@open.ac.uk

Twitter: @nancypontika

Agenda

- What is Text and Data Mining (TDM):
 - benefits,
 - why repositories and libraries should support TDM
- CORE and TDM
- The OpenMinTeD project
- Publishers' connector for open access research papers
- How you can use CORE
- Ethiopia videos

What is text and data mining (TDM)

The process of extracting high quality and meaningful information from text to answer unknown questions.

An illustration on a light green background. At the top, a network of white circles connected by lines is overlaid on a grid of binary code (0s and 1s). Below the network, a large stack of white papers with horizontal lines representing text is shown. Two red stick figures are positioned on either side of the papers, each holding a red scythe and appearing to be cutting through the stack. Below the papers, the text "Millions of research papers ready to text-mine" is written in black. At the bottom left is the CORE logo, and at the bottom right is the URL "core.ac.uk" in red.

Millions of research papers
ready to text-mine

CORE core.ac.uk

TDM benefits

- Reduces costs of conducting research (access, staff & infrastructure)
- Unlocks hidden information and develops new knowledge
- Explores new horizons
- Improves research and evidence base
- Improve research process quality

Why should repositories support TDM

The list of essential tools will evolve over time, but includes OA repositories and journals, free and open-source repository software, free and open-source journal management software, tools for **text- and data-mining**, directories of OA journals and repositories, directories of university and funder policies, providers of open licenses, digital preservation services, current awareness services, services for cross-linking and persistent URLs, and search engines.

BOAI10 Recommendations

<http://www.budapestopenaccessinitiative.org/boai-10-recommendations>

Researchers can make copies of any copyright material under fair dealing provisions (and conduct text and data mining). They will be able to do this

HEFCE FAQs

<http://www.hefce.ac.uk/rsrch/REFimpact/faqs/>

The aggregation of repository content can offer the foundation for a whole host of text mining activities to be developed on top of the content. Text and data mining are becoming valuable analytical methods that allow researcher to discover interesting patterns and extract new knowledge from a corpus of content. Repository collections

COAR

<http://www.rcuk.ac.uk/documents/international/coarstateofrepositories-pdf/>

Repositories and TDM



Institutional
Repositories

Subject
Repositories

Publishers/
OA journals



TDM & Libraries (Repository Managers)

- Established and maintain a close collaboration with researchers
- Extensive experience in advocacy, i.e. open access
- Knowledgeable about the repository's collection
- Participate in the Academic Institution's Research Committees
- Knowledgeable of your repository's collection
- Familiarity with Copyright issues and Creative Commons Licenses

Open Access research content in CORE

- Discover content in the CORE search engine



Search 70,477,809 open access articles

Search

- Use the CORE API
<https://core.ac.uk/services#api> and Dataset to
apply TDM practices
<https://core.ac.uk/services#dataset>

OpenMinTeD project

ACCESSIBLE CONTENT

Via standardised programmatic interfaces and access rules

DISCOVERABLE SERVICES

Well-documented easily discoverable text mining services and workflows which process, analyse and annotate text

EFFICIENT PROCESSING

Operate on public e-Infrastructures via standardized APIs

TDM COMMUNITIES

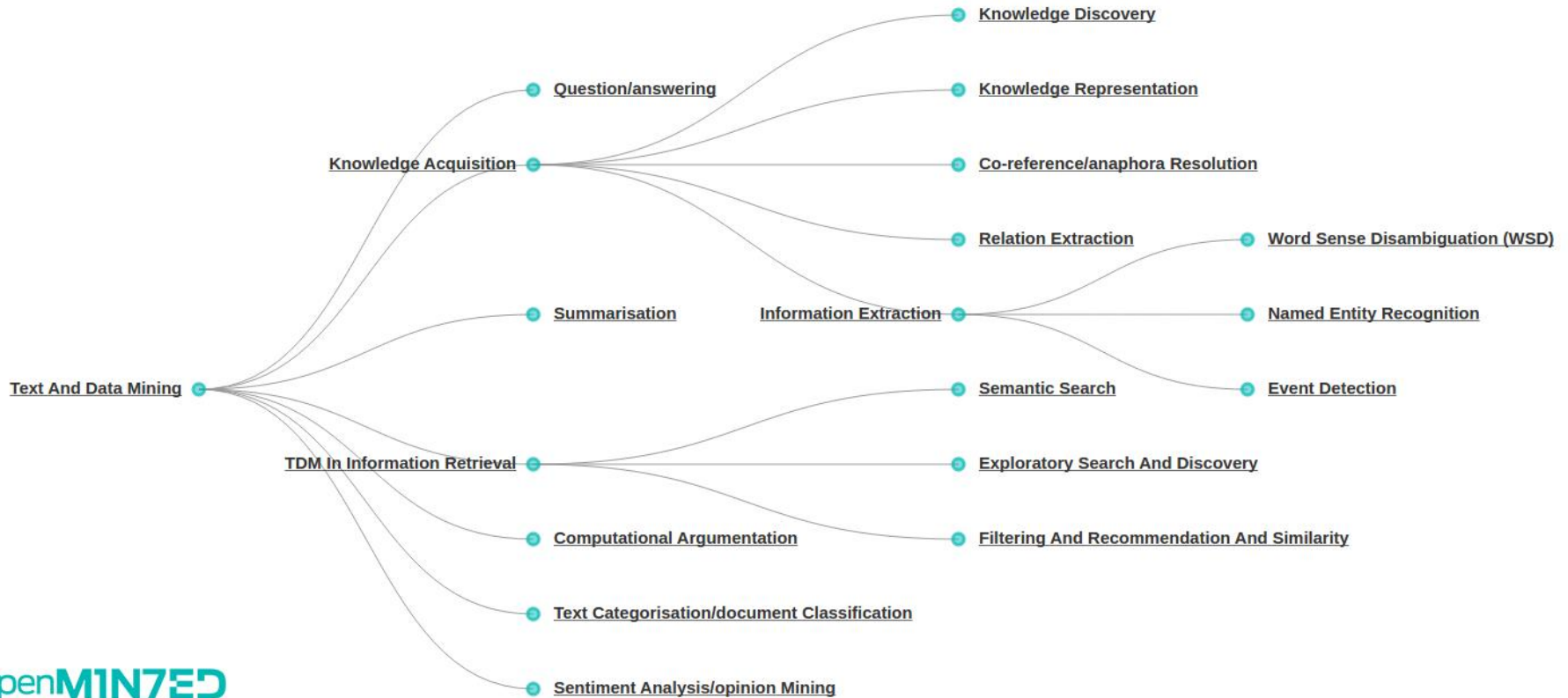
Different scientific communities have different challenges

VALUE ADDED APPS

Community-driven applications to illustrate the value of the infrastructure. Engage with industry.

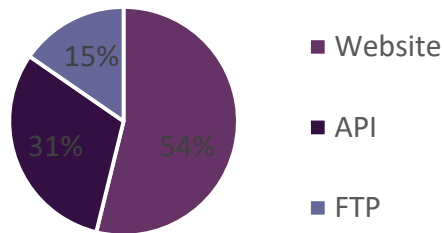
TDM taxonomy

Text and Data Mining

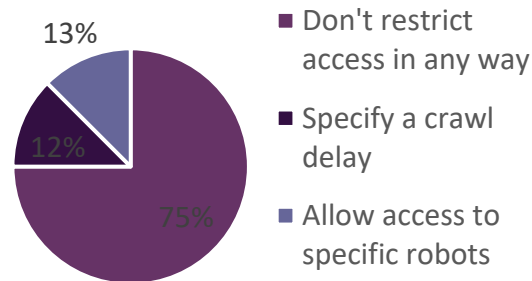


Key players' interoperability for machine access to metadata

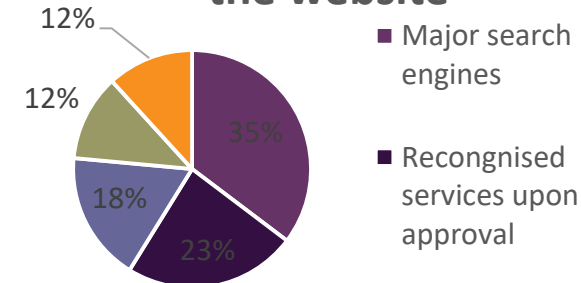
Automated downloads of OA full-text



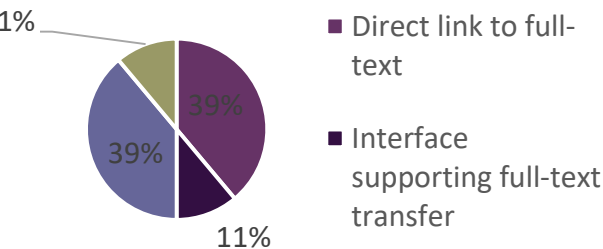
Restricting access to full-text



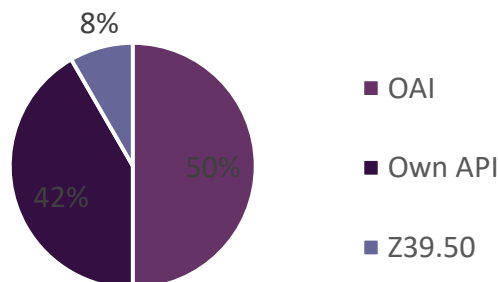
Accessing full-text by harvesting the website



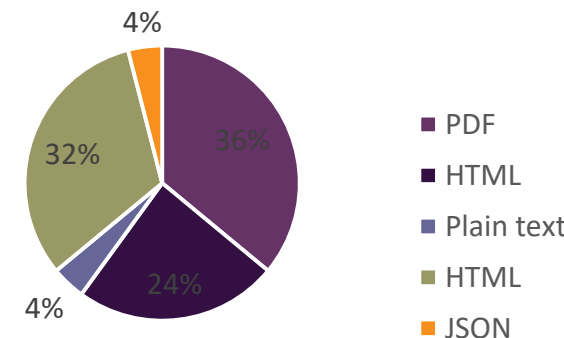
Reference of an article's full-text on metadata



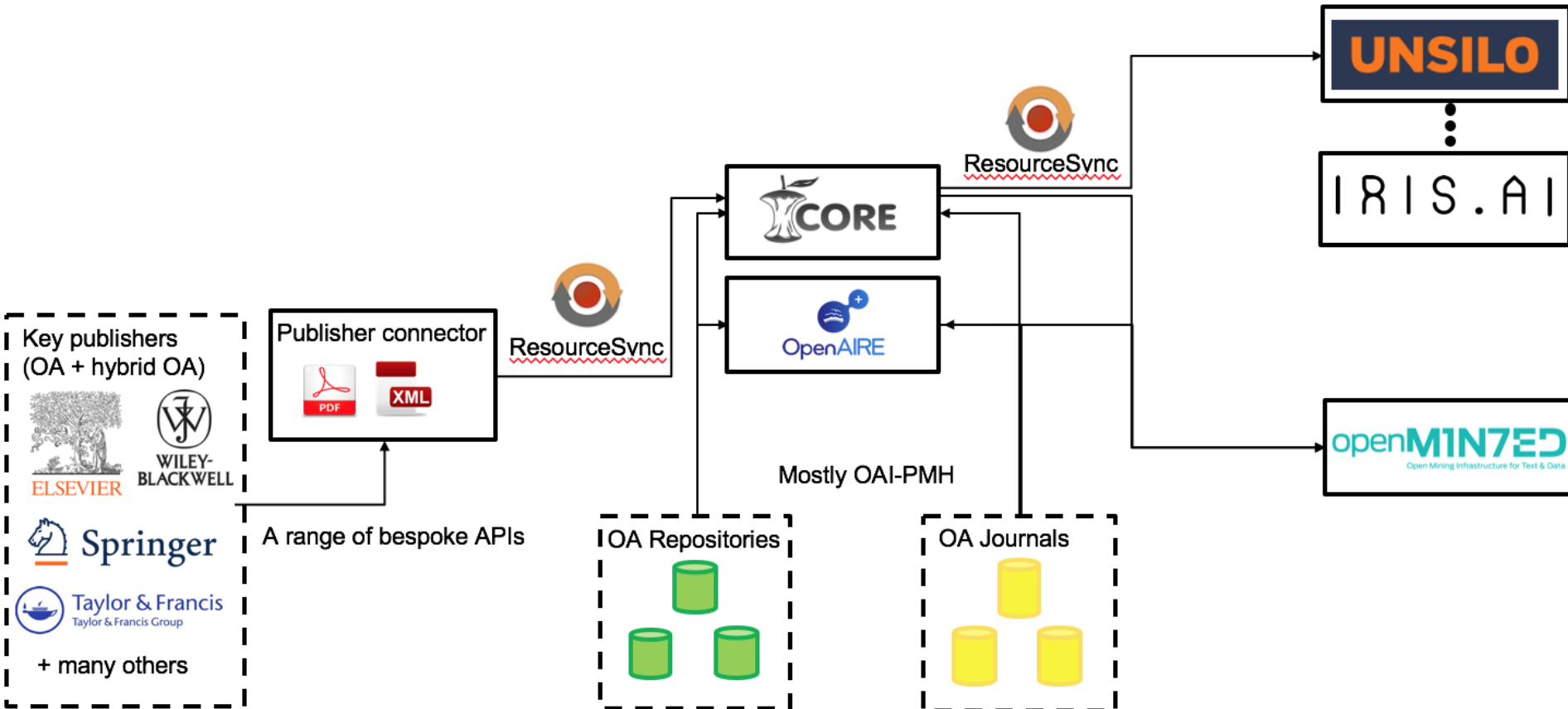
Accessing content standards



Files format



Exposing enriched data for TDM



How repository managers can use CORE

CORE has already more than 1,000 data providers

<https://core.ac.uk/dataproviders>

- Join CORE as a data provider <https://core.ac.uk/join>
- Ask your library to list the CORE search engine in the library pages
- When you become a CORE data provider request access to the CORE Dashboard <https://core.ac.uk/services#dashboard>
- Install the CORE recommender in your repository <https://core.ac.uk/services#recommender>
- **Become a CORE Ambassador**
<https://core.ac.uk/about/ambassadors>

Videos at the Ethiopia meeting

We want to hear from you!

- Value of CORE
 - Librarians
 - Research support administrators
 - Researchers
 - Students
 - Life long learners

Thank you!

CORE

Website: <http://core.ac.uk>
Email: theteam@core.ac.uk
Twitter: @oacore

OpenMinTeD

Website: <http://openminded.eu/>
Twitter: @openminded_eu
LinkedIn: <https://www.linkedin.com/groups/8402199>
Slideshare: https://www.slideshare.net/openminded_eu
YouTube Channel: <https://www.youtube.com/channel/UCjAgOP2ALi4IHAL-8sWTG8Q>
Flickr: <https://www.flickr.com/photos/146359736@N07/>