



Seamless access to the world's
open access research papers

Nancy Pontika



The Open
University

core.ac.uk

In this talk

- What is CORE?
- Who is CORE for?
- What services does CORE offer?

Mission

*Aggregate all open access articles across relevant data sources worldwide, enrich this content and **provide seamless access to it through a set of data services ...***



World's largest open access articles collection

- > 135,116,215 Metadata
- > 105,290,852 Abstracts
- > 11,817,841 Hosted full texts
- > 78,781,218 Links to full text

60TB of data
3,788 Data providers
~ 50k Journals



CORE API



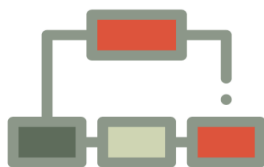
CORE Dataset



CORE Search



CORE Repository Dashboard



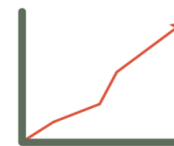
CORE Publisher Connector



CORE SDKs

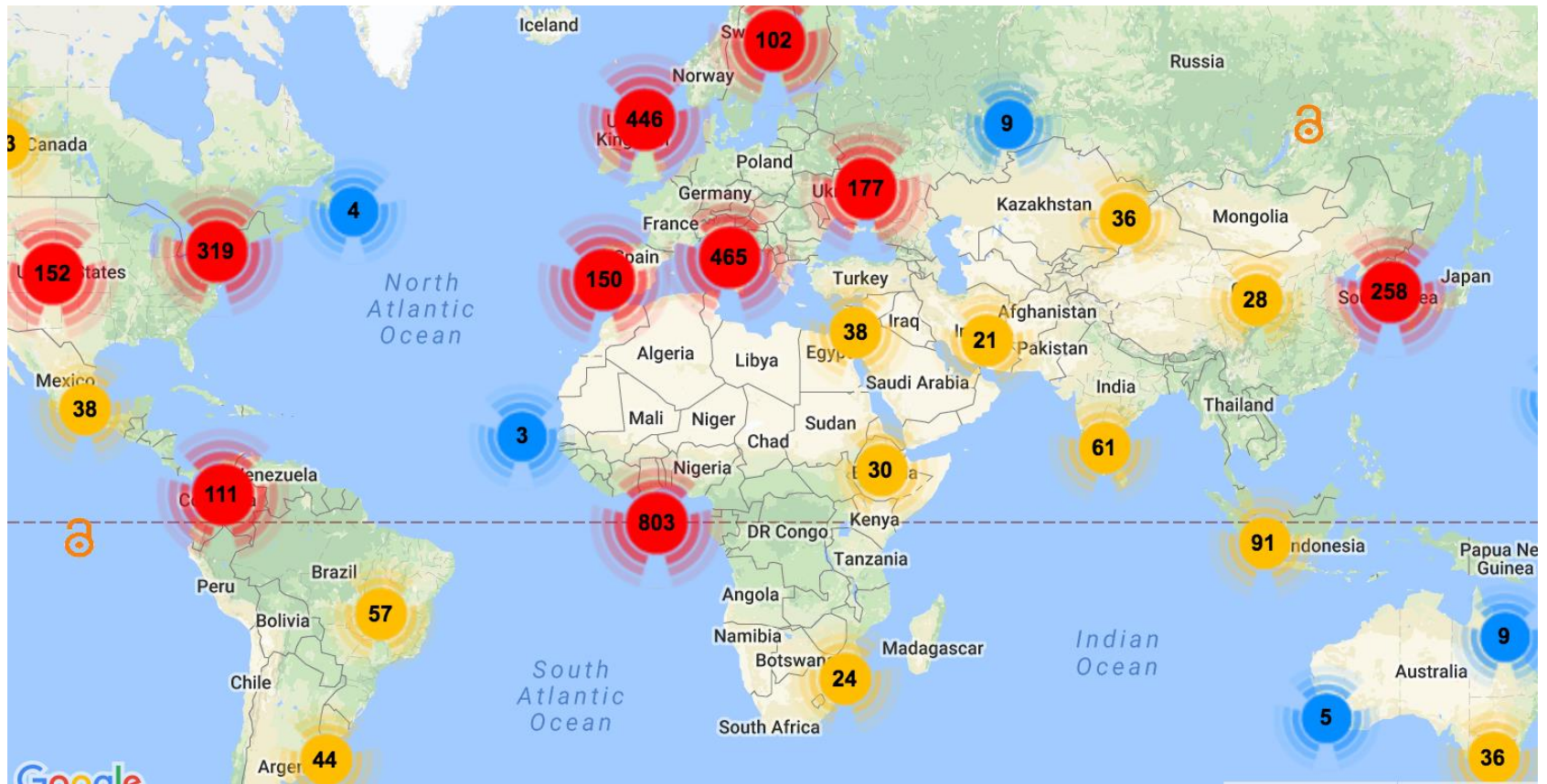


CORE Recommender



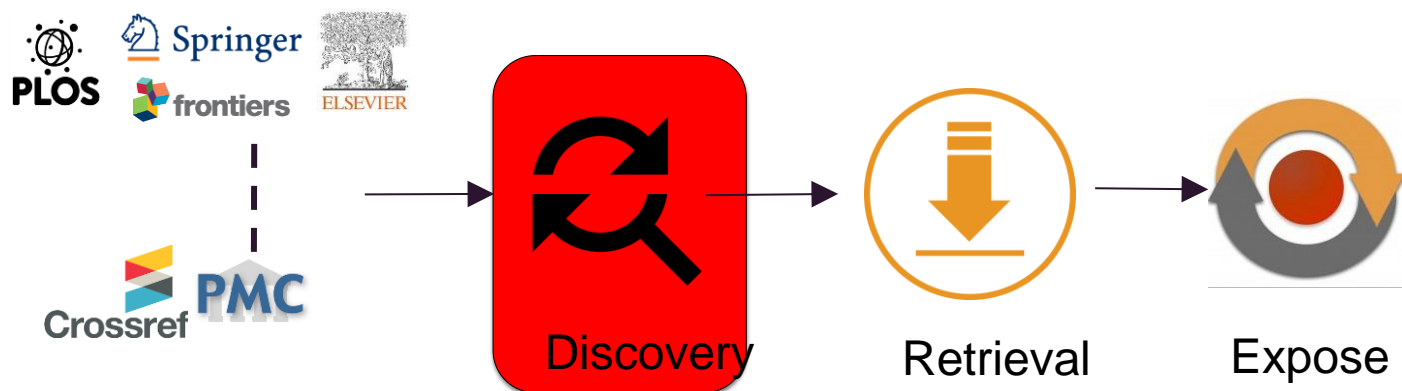
CORE Analytics Dashboard

Global reach



Harvesting

1. OAI-PMH
2. ResourceSync
3. Bespoke APIs of Publishers
4. Custom discovery service





CORE Pipeline

- Metadata download, extraction and harmonisation
- Full text download
- Text extractions and sections extraction
- Metadata validation
- Thumbnails generation
- References and citation contexts extraction
- API enrichment (e.g. finding DOIs, linking to other systems)
- Document type classification
- Deduplication
- Indexing
- Exposing (data dumps, API, ResourceSync)

CORE Search



CORE

Search 135,363,324 open access articles

Search

Aggregating the world's open access research papers

We offer seamless access to millions of open access research papers, enrich the collected data for text-mining and provide unique services to the research community.

[Read more](#)

[Join us](#)

CORE Search



Refine your search

Publication type
 with fulltext only

Year

Languages

<input type="checkbox"/> English	1,229,632
<input type="checkbox"/> Indonesian	48,325
<input type="checkbox"/> Portuguese	28,231
<input type="checkbox"/> Czech	13,902
<input type="checkbox"/> Italian	12,365
<input type="checkbox"/> German	9,571
<input type="checkbox"/> Spanish	7,832
<input type="checkbox"/> French	5,685
<input type="checkbox"/> Dutch	2,765
<input type="checkbox"/> Slovene	1,237

+ Show more

Journals

Showing results for **data mining** (4,482,672 articles found) Sort by: Relevance ▾

Energy efficient data mining scheme for high dimensional data
By **Mohammad Alwadi** and **Girija Chetty**
Repository: University of Canberra Research Repository | 2015
...Communication Technologies (ICICT 2014) Energy Efficient **Data Mining** Scheme for High Dimensional Data, Mohammad Alwadi, Girija Chetty, The University of Canberra, 2015
[Get PDF \(468 KB\)](#) [Similar articles](#)



Smart phone based data mining for human activity recognition
By **Girija Chetty**, **Matthew White** and **Farnaz Akther**
Repository: University of Canberra Research Repository | 2015
...committee of the International Conference on Information and Communication Technologies (ICICT 2014), 14-16 Oct 2014, 2015, 01, 2015, School of Information and Communication Technologies
[Get PDF \(661 KB\)](#) [Similar articles](#)



Analysing qualitative data from virtual worlds: using images and text mining
By **Brad James Austin McKenna**, **Michael Myers** and **Lesley...**



CORE Search

- Faceted searching
- What you find is what you can download
- We really host the full text content
- We are not here to replace Google Scholar, we serve a different community
- >22 million yearly unique visitors

CORE Recommender



CORE Recommender

The screenshot displays the Apollo repository interface. At the top, it shows the University of Cambridge logo and navigation links: Apollo Home / School of the Physical Sciences / Department of Chemistry / Unilever Centre for Molecular Informatics / Panton Discussions / View Item. Below this is the Apollo logo and a search bar with the text 'Search Apollo' and a search icon. A link for 'Advanced search' is also visible. The main content area is titled 'Open Content Mining' and includes a thumbnail of the document cover. The record details are as follows:

- Citation:** Murray-Rust, P. (2012). Open Content Mining.
- Description:** Conference for the Fellows of OpenForum Academy - 24th September 2012 Brussels
- Abstract:** Abstract— We present evidence that content-mining of scholarly articles is now technically feasible and highly valuable both. However researchers and information technologist are blocked by legal and contractual barriers from using it and developing the methodologies. We review the problems and propose changes in legal policy which we have already submitted to the UK's Hargreaves report on intellectual property reform. We put forward the fundamental rights of scholars and embed them in a manifesto: "The right to read is the right to mine", "Users and providers should encourage machine processing, and "Facts don't belong to anyone".
- Keywords:** Open Content Mining, Index Terms—Open Knowledge, Content mining, Hargreaves process, Text mining, publishers, legal barriers
- Identifiers:** This record's URL: <http://www.dspace.cam.ac.uk/handle/1810/243749>
- View / Open Files:** article text (PDF, 9Mb)
- Authors:** Murray-Rust, Peter
- Publication Date:** 2012-09-24
- ISBN:** to be assigned
- Language:** English

The left sidebar contains navigation options: 'Browse' (All of Apollo, Communities & Collections, Authors, Titles, Keywords, Type), 'This Collection' (Authors, Titles, Keywords, Type), and 'Statistics' (View Usage Statistics).

- Recommending relevant content to users from across the CORE's content
- Recommender plugin for repositories
- 23 repositories using it, e.g. Cambridge University

CORE Recommender



CORE Recommender

Type
Conference Object

Rights
Attribution 2.0 UK: England & Wales

Metadata
[Show full item record](#)

Licence URL:
<http://creativecommons.org/licenses/by/2.0/uk/>

Recommended or similar items

Suggested articles | [Suggested articles in Apollo](#)

Effectively and Efficiently Mining Frequent Patterns from Dense Graph Streams on Disk
Provided by: Elsevier - Publisher Connector | **Publisher:** The Authors. Published by Elsevier B.V. | **Year:** 2014
By Braun Peter, Cameron Juan J., Cuzzocrea Alfredo, Jiang Fan, Leung Carson K.

The right to read is the right to mine: Text and data mining copyright exceptions introduced in the UK.
Provided by: LSE Research Online | **Publisher:** London School of Economics and Political Science | **Year:** 2014
By Mounce Ross

Global boom, local impacts: Mining revenues and subnational outcomes in Peru 2007-2011
Provided by: EconStor | **Publisher:** Washington, DC: Inter-American Development Bank (IDB) | **Year:** 2014
By Zambrano Omar, Robles Marcos, Laos Denisse

Environmental security, mining and good governance : mining regulation in the Kyrgyz region. A review
Provided by: UEF Electronic Publications | **Publisher:** University of Eastern Finland
By Honkonen T

A Case Study of Data Analysis Process and Tools for a Consulting Company
Provided by: Aaltodoc Publication Archive | **Year:** 2012
By Gong Peng

Powered by CORE

- Recommending relevant content to users from across the CORE's content
- Recommender plugin for repositories
- 23 repositories using it, e.g. Cambridge University

CORE API, data dumps, SDKs



CORE Dataset



CORE API



CORE SDKs

- Enable others to programmatically interact with CORE data.
- Essential for developing applications that make use of information about research papers
- Supporting a variety of use cases
- >500 registered users, including companies for CORE API
- Use:
 - CORE API to interact with CORE in real time
 - CORE Data Dumps to run batch process, develop prototype applications and do research
 - CORE SDKs to gain easy access to CORE API from several programming languages
 - CORE FastSync

CORE API, data dumps, SDKs



CORE Dataset



CORE API



CORE SDKs

- It is beyond human capacities to read all scientific literature
- Example use cases in which CORE is applied:
 - Improving discovery
 - Plagiarism detection
 - Question answering in science
 - Literature based discovery
 - Fact checking and detection of misinformation
 - Analysing research trends
 - Finding experts in a particular domain
 - Research evaluation and scientometrics
 - Exploratory and visual search
 - ...

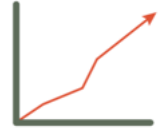
CORE Repository Dashboard



CORE Repository Dashboard

- Helps repositories to:
 - Manage content aggregated by CORE
 - Identify and resolve issues in the configuration of their repository
- >60 registered repositories

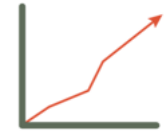
CORE Analytics Dashboard



CORE Analytics Dashboard

- A tool for analytical evaluation of universities research outputs based primarily on CORE data.
- Research Impact insights on institutional level, comparison with other institutions or group of institutions
- Based on citation data, social impact data
- Data sources:
 - CORE dataset (primary pivot)
 - MAG dataset
 - Crossref Event Data
 - Mendeley

CORE Analytics Dashboard



CORE Analytics Dashboard

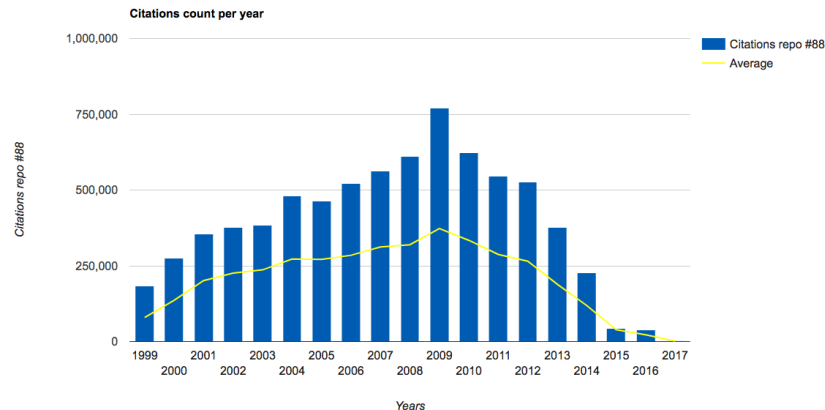


Find in CORE

Oxford University Research Archive

University of Oxford
ora@bodleian.ox.ac.uk

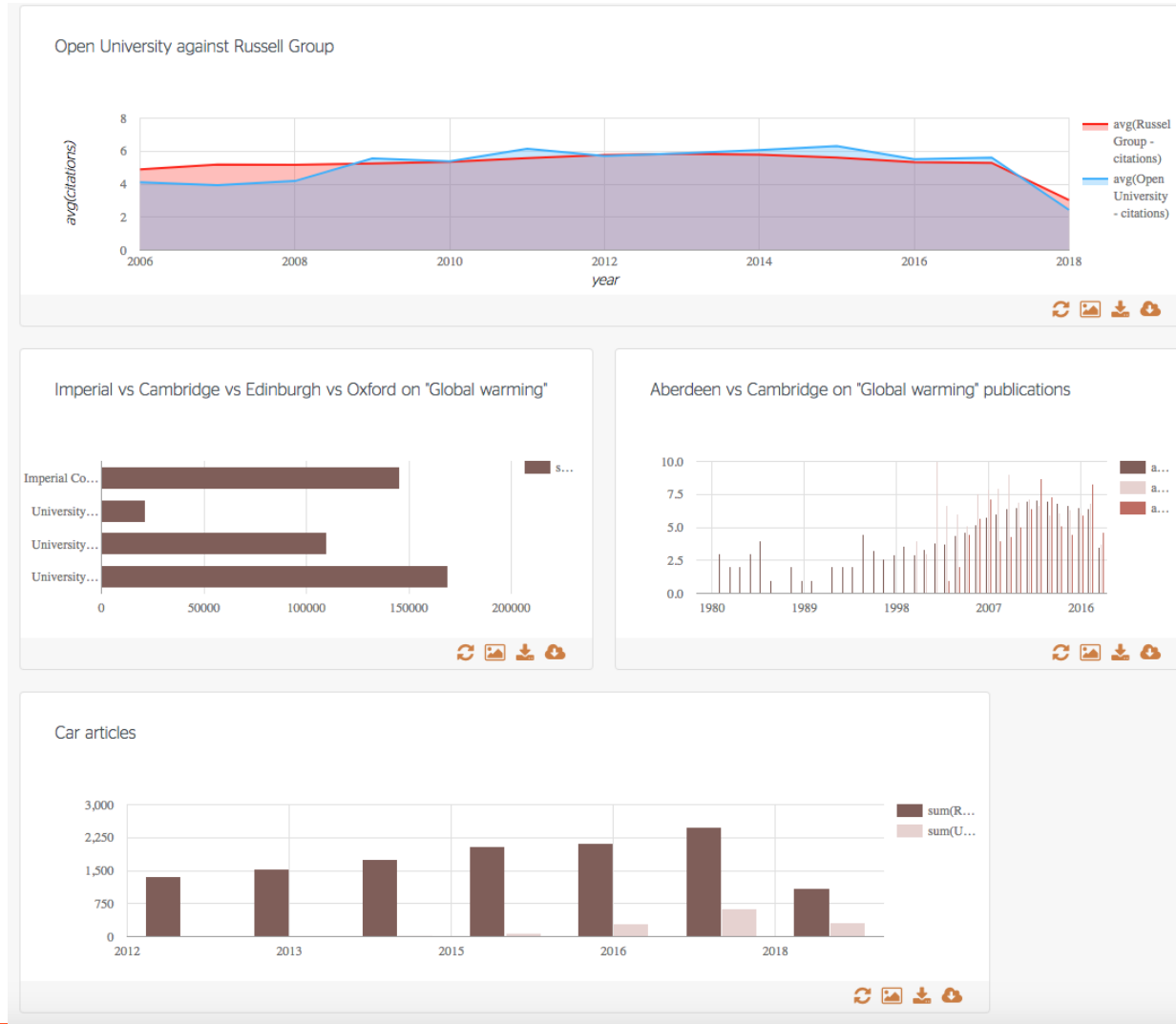
Count toggle Avg toggle Norm toggle



- Helps you analyse the impact of research at your institution.
- See how you perform against competitors.
- Slice and dice by discipline, field, any kind of a query

CORE Analytics Dashboard: sneak peek


- Customizable area
- Pre-defined widgets of standard data reports



Key customers/partners

- Main source of data for **Naver Academic**.
- **IRIS.AI**: A start up of 20 people entirely developing tools to help researchers be more productive. Highly dependent on CORE API.
- **ProQuest**: CORE supplies data about OA publications to improve discovery of OA content in Primo and Summon (new) additional discovery collaborations on the way.
- Providing a key API for **OA Button**.

Our contribution to the community (examples)

- Harvesting and delivering data from UK repositories supporting **RIOXX** to **OpenAIRE** (every 2 weeks) on behalf of Jisc.
- **OpenAIRE** to use CORE Publisher Connector 
- Delivering repository statistics data (including historical information) to **OpenDOAR** (new).
- Providing statistics on the growth of OA and disciplinary trends (in particular to British organisations).
- **IRUS-UK**: IRUS has created in collaboration with CORE the IRUS-CORE service which integrates the capturing of download data from repositories with CORE.
- **OpenMinTeD**: EU eInfrastructure project for TDM of research data. CORE the largest supplier of data.
- Driving innovation in this area (**EOSC/ResourceSync**), **NGR**.

Conclusions

- Global world's largest aggregator of open access papers
- A suite of data services for a variety of stakeholders
- Large user base
- Moving fast from research to a product. Always innovate.

Thank you!



CORE

Bringing science to all



@oacore



theteam@core.ac.uk



core.ac.uk/team



The Open
University

