THE UNIVERSITY OF STRATHCLYDE

Repository aggregation : CORE as essential open scholarly infrastructure & service platform

Dr George Macgregor [a, b, c]

[a] Scholarly Publications & Research Data (SPRD), IS Information Management, University of Strathclyde
[b] iSchool, Department of Computer & Information Sciences, University of Strathclyde
[c] CORE Ambassador

https://purl.org/g3om4c | @g3om4c | @g3om4c@mastodon.social | 0000-0002-8482-3973

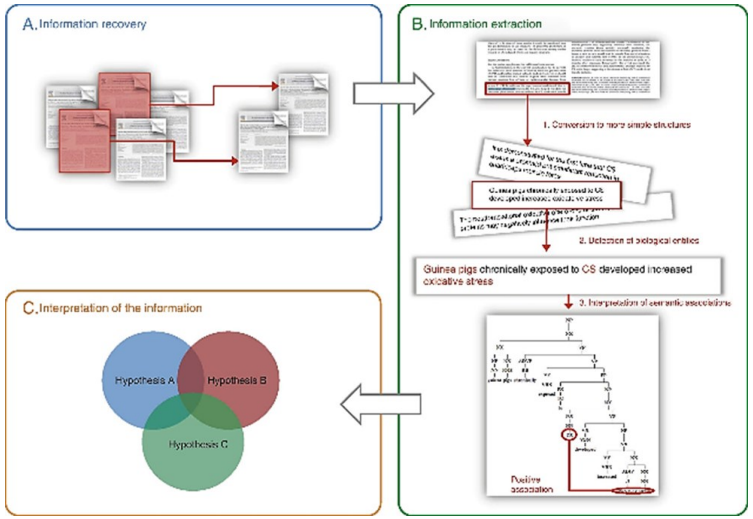
This deck available: https://strathprints.strath.ac.uk/80959/

Thursday 09 June 2022

Quick background…

- The forgotten origin of repositories…?
- OAI-PMH and distributed global repository network
- Harvesting structured data – and beyond…
- Harvesters = Aggregators
- Aggregation brings huge scientific benefits & opportunities
- Contributing to the knowledge graph and knowledge commons
- Enabling creation of new scientific knowledge and discoveries
- ‘Big data’ = patterns
- OA beyond human readable access…
- CORE, BASE, OpenAIRE, OAIster
- CORE – Commitment to the Principles of Open Scholarly Infrastructure (POSI)

Don Swanson ≠ Don Johnson



THE UNIVERSITY OF STRATHCLYDE

The COnnecting REpositories (CORE) project aims to facilitate the access and navigation across relevant scientific papers, stored in Open Access repositories across the UK.

This is being achieved by:

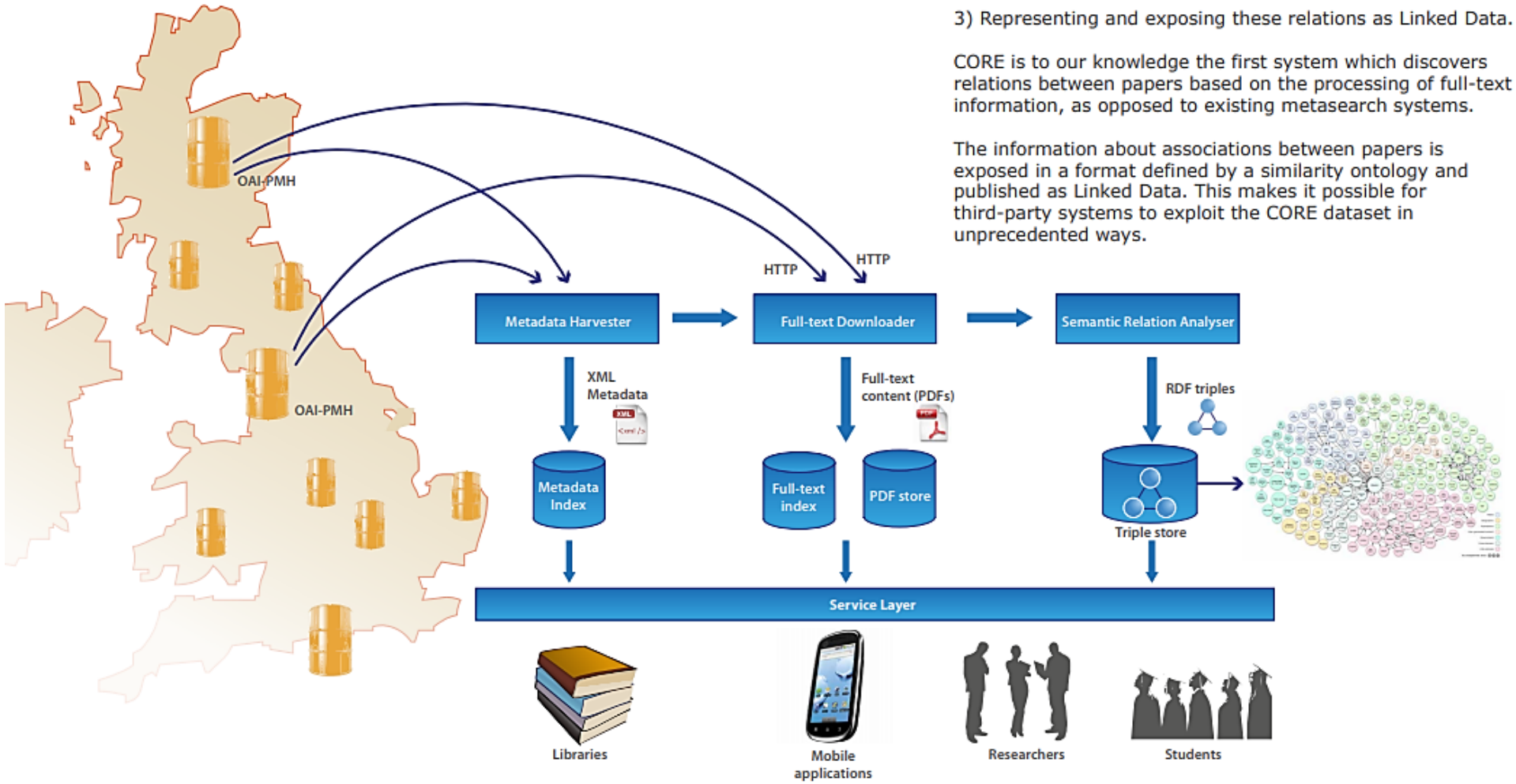
1) Harvesting metadata and full-text content from Open Access repositories.

2) Applying text mining techniques to discover semantically related articles.

3) Representing and exposing these relations as Linked Data.

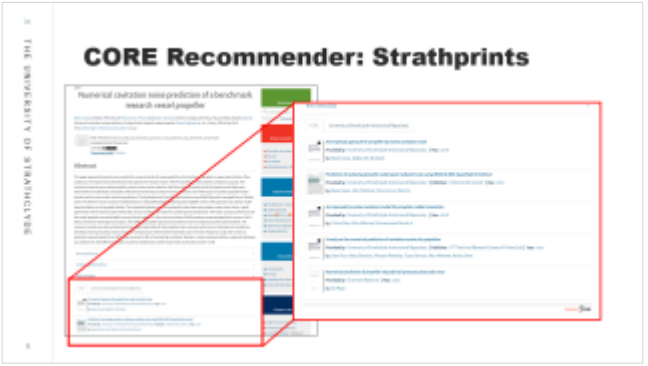
CORE is to our knowledge the first system which discovers relations between papers based on the processing of full-text information, as opposed to existing metasearch systems.

The information about associations between papers is exposed in a format defined by a similarity ontology and published as Linked Data. This makes it possible for third-party systems to exploit the CORE dataset in unprecedented ways.



Knoth, P. et al. (2017) CORE: Connecting Repositories, OA17 Workshop,CERN, Geneva.
https://blog.core.ac.uk/2011/06/17/core-oai7-poster/

Data aggregation enables…



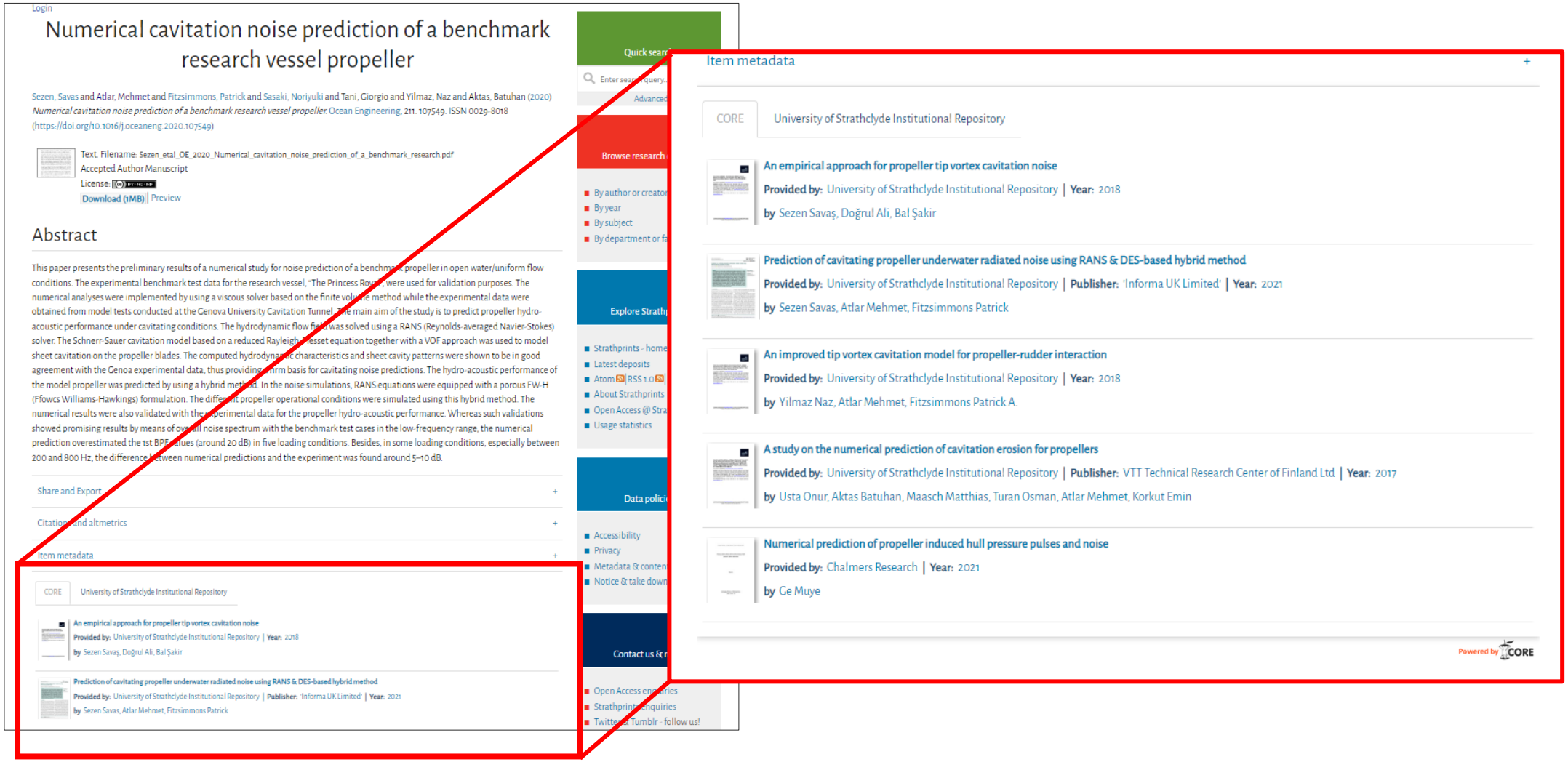
\ Data \

- **CORE dataset**: download millions of outputs for TDM (largest dataset…)
- **CORE API**: harness platform to perform TDM or build novel applications
- **'FastSync'**: Does what it says on the tin!
- And **search** circa 250 million items…

\ Discovery \

- **CORE Discovery**: Unpaywall but using CORE dataset (Mozilla; Chrome & Chromium; Opera)
- **CORE Recommender**: Recommendation engine for repos
- **OAI resolver**: New! Does what it says on the tin - again!
- **Repository Dashboard**: Maintain, analyse & validate your repository data within CORE

CORE Recommender: Strathprints



Critical open scholarly infrastructure

- CORE = essential open scholarly infrastructure
- Aligns with the raison d'être of the repository cooperative & network
- Repositories are not islands
- Importance of the ‘network effect’;
- More repositories delivering more content = greater knowledge creation
- This is *critical*
- But also delivers greater platform potential;
- Better recommendations, more paywalls thwarted, new services built on the platform in future
- CORE sustainability & the future…

Further reading

- \ CORE \ https://core.ac.uk/
- \ CORE API \ https://core.ac.uk/services/api
- \ CORE Dataset \ https://core.ac.uk/services/dataset

- Introducing CORE (2019) [video]

CORE