

CORE.ac.uk: Sustainability Strategy

17th March 2021

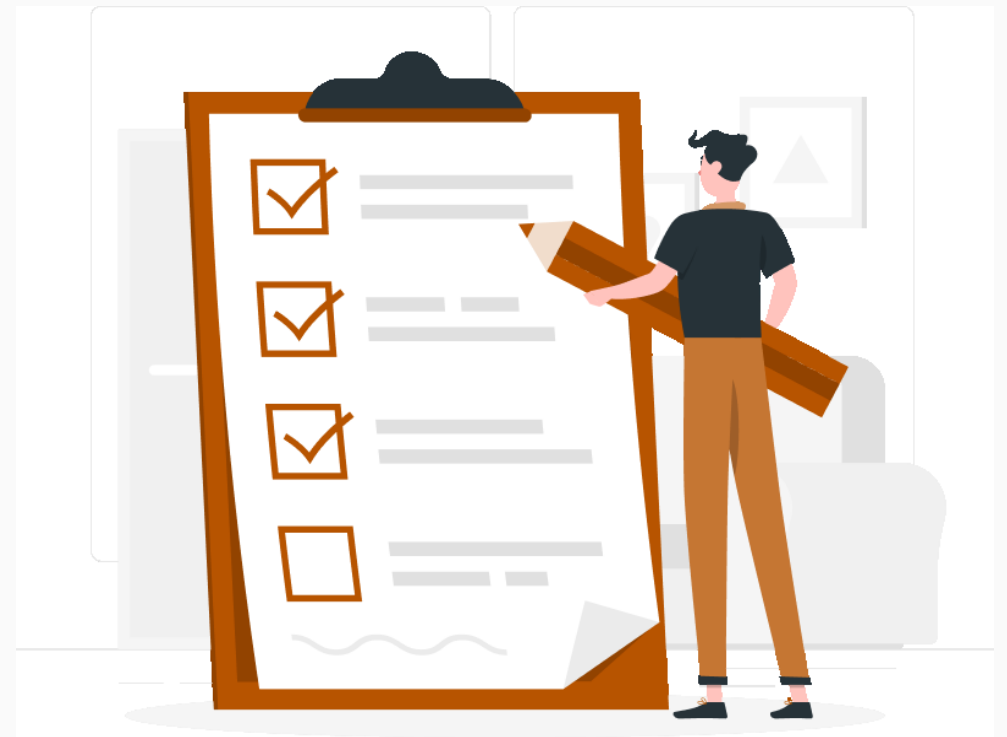
Dr. Petr Knoth

CORE Head & Founder | Senior Researcher in Text and Data Mining
Big Scientific Data and Text Analytics Group
Knowledge Media Institute, The Open University
STEM



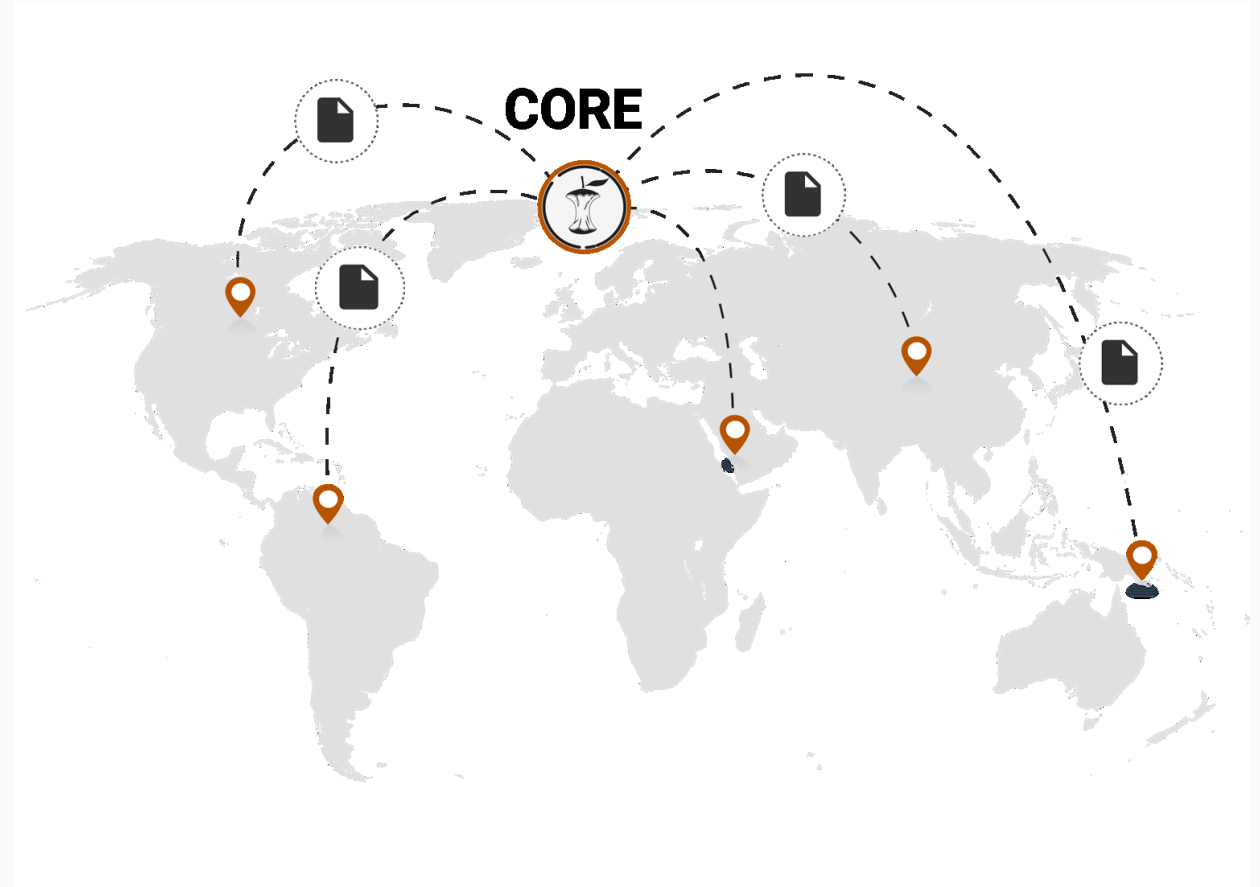
Outline

- 1 What is CORE
- 2 CORE's Sustainability strategy in a nutshell

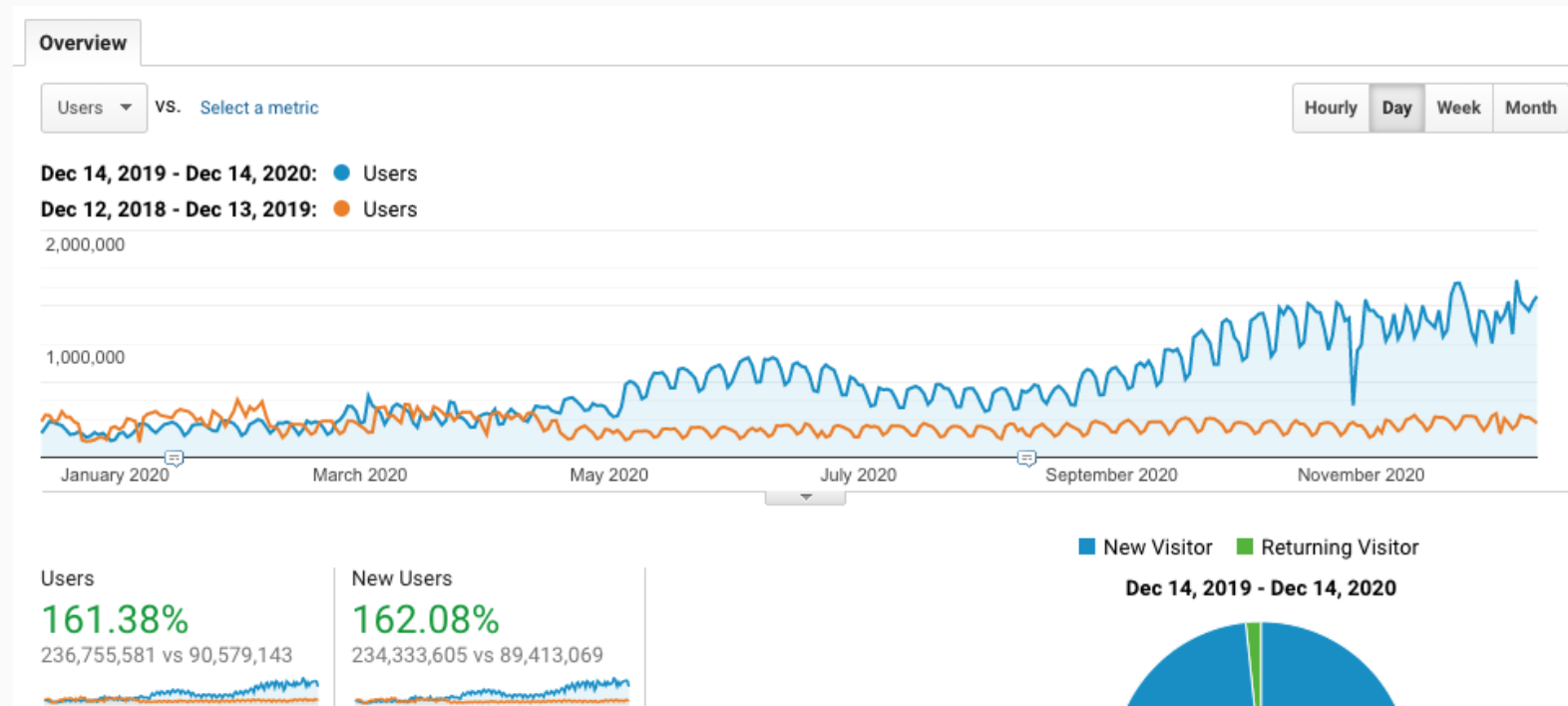


What is CORE

- World's largest aggregator of open access research papers from repositories and journals.
- It is a **not-for-profit service** dedicated to the open access mission.
- We serve the global network of repositories and journals by **increasing the discoverability and reuse of open access content**.



CORE usage



- **236 million** users over the last 12 months
- January 2021: **30 million+ monthly active users**
- June 2020: 20 million monthly active users
- January 2019 – CORE reached over 10M monthly active users for the first time

CORE Usage

Alexa Rank

Estimate ?

This site ranks:

1,373

In global internet traffic and engagement
over the past 90 days

↗ 469

Alexa Rank #



90 days ago
#1,785

Today
#1,373

- February 2021 – CORE within top 1.5k websites globally by user engagement.
- a combination of dail visitors and page views on a website over a 3 month period
- <https://www.alexa.com/siteinfo/core.ac.uk>

Comparison to other key OA infrastructures by Alexa rank

- CORE (1,373)
- Arxiv.org (2,129)
- ORCID (5,474)
- Publons (15,459)
- DOAJ (28,387)
- Jisc (133,724) - services sitting under the jisc.ac.uk domain as an aggregate
- SHERPA (154,862)
- BASE (158,709)
- Unpaywall (192,884)
- OpenAIRE (258,719)
- Open Access Button (709,665)

[Data as of June 2021]



Allows for comparison of worldwide traffic for any domain -

<https://www.alexa.com/siteinfo/>

CORE Services

- Content discovery



Recommender



Discovery



Search

- Raw data services



API



Dataset



FastSync

- Managing content



Repository Dashboard

CORE Search

- Full text search for OA content
- Faceted searching
- What you find is what you get
- Real change of data providers wanting to be included

The screenshot displays the CORE Search website interface. At the top, the CORE logo is visible on the left, and 'Services' and 'About' links are on the right. A search bar contains the term 'allele' with a 'Search' button. Below the search bar, the 'Advanced Search' section is active. On the left, a 'Refine your search' sidebar lists various filters: Publication type (with 'with fulltext only' checked), Year (1999-2018), Languages (English: 49,370, German: 852, Italian: 208, French: 203, Portuguese: 196, Spanish: 72, Hungarian: 42, Indonesian: 32, Croatian: 31, Dutch: 14), Journals (Journal Pengolahan Hasil: 1,856), Repositories (PubMed Central: 37,730, Elsevier - Publisher: 30,360, MUGG (Crossref): 31,814, Springer - Publisher: 28,410, Frontiers - Publisher: 4,480, Edinburgh Research Explorer: 3,085, Harvard University - DASH: 2,034, Raciboud Repository: 2,029, ZORA: 1,891, Erasmus University Digital: 1,780), and Permalink (https://core.ac.uk/search?). The main search results area shows 'Showing results for allele (368,795 articles found)' and a 'Sort by: Relevance' dropdown. Four article results are displayed, each with a thumbnail, title, authors, repository, and a 'Get PDF' button. The articles are: 'The almond Sf allele: an allele in question', '7 GT repetition on 5T allele', 'Allele-Specific p53 Mutant Reactivation', and 'Allele coding in genomic evaluation'. The footer contains 'Useful links' (Blog, Contact us, Services, Cookies, About CORE, Privacy notice), 'Writing about CORE?' (Discover our research outputs and cite our work), and logos for The Open University and Jisc, with a note that CORE is a not-for-profit service delivered by them.

CORE Recommender

The screenshot shows the Apollo Open Content Mining interface. At the top is the University of Cambridge logo and a breadcrumb trail: 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 magnifying glass icon. A link for 'Advanced search' is also present. The main content area is titled 'Open Content Mining' and displays a record for a conference paper. The left sidebar contains navigation links for 'All of Apollo' (Communities & Collections, Authors, Titles, Keywords, Type) and 'This Collection' (Authors, Titles, Keywords, Type). The main content area includes a thumbnail of the document, a 'View / Open Files' section with a link to 'article text (PDF, 9Mb)', and sections for 'Citation', 'Description', 'Abstract', 'Keywords', 'Identifiers', 'Publication Date', 'ISBN', and 'Language'. The 'Citation' section shows the author 'Murray-Rust, P.' and the year '(2012)'. The 'Description' section mentions the 'Conference for the Fellows of OpenForum Academy - 24th September 2012 Brussels'. The 'Abstract' section provides a summary of the paper's content. The 'Keywords' section lists 'Open Content Mining, Index Terms—Open Knowledge, Content mining, Hargreaves process, Text mining, publishers, legal barriers'. The 'Identifiers' section includes the URL 'http://www.dspace.cam.ac.uk/handle/1810/243749'. The 'Publication Date' is '2012-09-24'. The 'ISBN' is 'to be assigned'. The 'Language' is 'English'.

UNIVERSITY OF CAMBRIDGE

Apollo Home / School of the Physical Sciences / Department of Chemistry / Unilever Centre for Molecular Informatics / Panton Discussions / View Item

Apollo

Search Apollo

Advanced search

Browse

All of Apollo

- > Communities & Collections
- > Authors
- > Titles
- > Keywords
- > Type

This Collection

- > Authors
- > Titles
- > Keywords
- > Type

Statistics

- > View Usage Statistics

Open Content Mining

View / Open Files

- article text (PDF, 9Mb)

Authors

Murray-Rust, Peter

Publication Date

2012-09-24

ISBN

to be assigned

Language

English

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 technologists 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>

- Recommending relevant content to users from across all free content
- Recommender plugin for repositories
- <https://core.ac.uk/services/recommender/>

CORE Recommender

Type
Conference Object


Rights
Attribution 2.0 UK: England & Wales
Licence URL:
<http://creativecommons.org/licenses/by/2.0/uk/>


Metadata
[Show full item record](#)


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

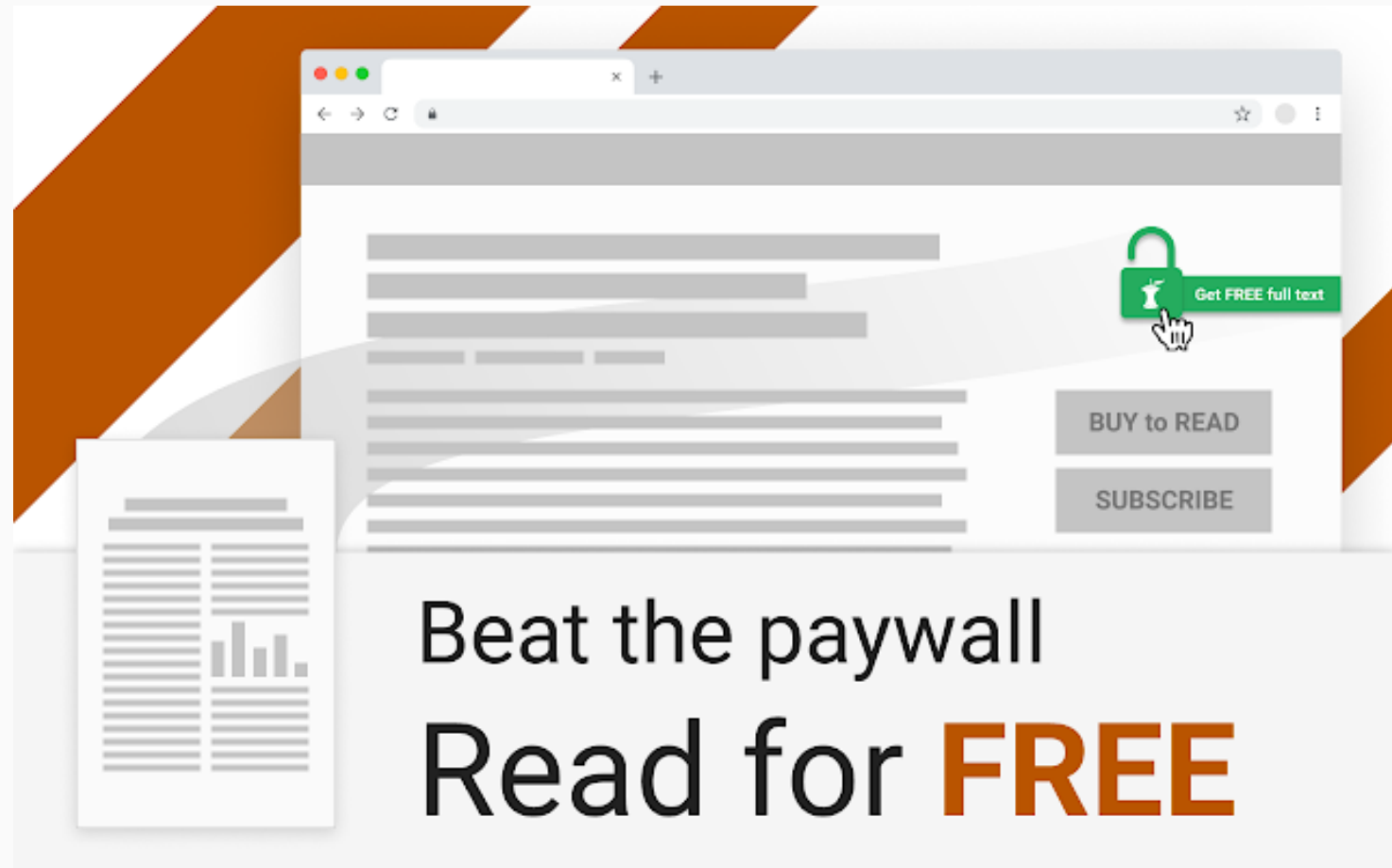
**A Case Study of Data Analysis Process and Tools for a Consulting Company**
Provided by: Aaltodoc Publication Archive | **Year:** 2012
By Gong Peng

Powered by 

- Recommending relevant content to users from across all free content
- Recommender plugin for repositories
- Currently **300+ repositories** have it **actively installed**, including aggregators such as La Referencia.
- <https://core.ac.uk/services/recommender/>

Introducing CORE Discovery

- High coverage of freely available content
- Free service for researchers by researchers. No company controlling the pipes.
- Best grip on open repository content.
- Repository integration
- Discovering documents without a DOI.

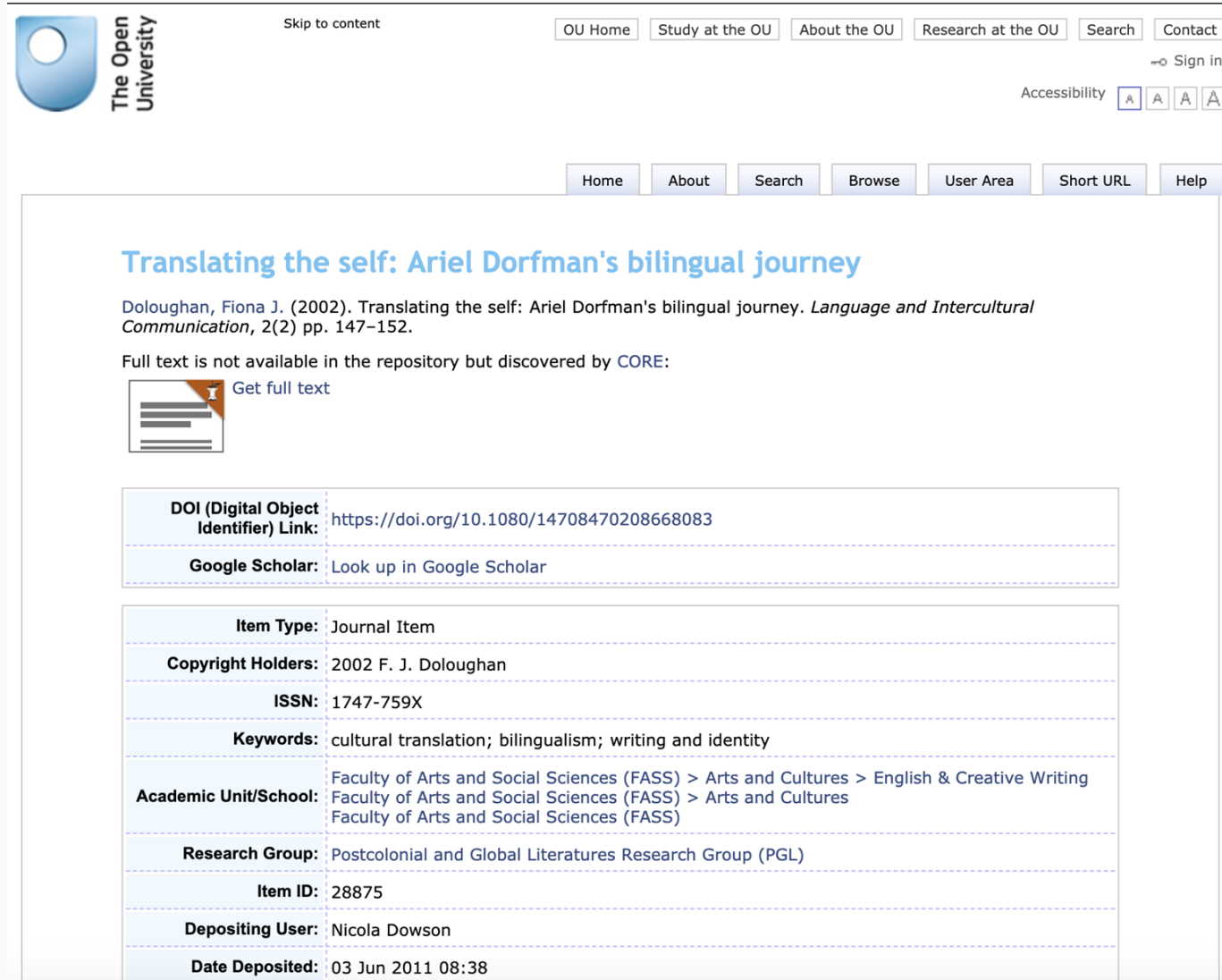


Beat the paywall
Read for **FREE**

<https://core.ac.uk/services/discovery/>

CORE Discovery Repository integration

- Majority of articles in repositories metadata only.
- CORE Discovery repository plugin:
 - turns dead ends of user journeys into journeys fulfilling users' information needs
 - makes repository content more discoverable.




The screenshot displays the Open University website interface. At the top, there is a navigation bar with links for 'OU Home', 'Study at the OU', 'About the OU', 'Research at the OU', 'Search', and 'Contact'. A 'Skip to content' link is also present. Below this, a secondary navigation bar includes 'Home', 'About', 'Search', 'Browse', 'User Area', 'Short URL', and 'Help'. The main content area features a search result for the article 'Translating the self: Ariel Dorfman's bilingual journey' by Fiona J. Doloughan (2002). The article is from the journal 'Language and Intercultural Communication', volume 2(2), pages 147-152. A note indicates that the full text is not available in the repository but was discovered by CORE. A 'Get full text' button is provided. Below the article information, there are two tables. The first table contains links for the DOI (Digital Object Identifier) and Google Scholar. The second table provides detailed metadata for the item, including its type, copyright holders, ISSN, keywords, academic unit, research group, item ID, depositing user, and date deposited.

Translating the self: Ariel Dorfman's bilingual journey

Doloughan, Fiona J. (2002). Translating the self: Ariel Dorfman's bilingual journey. *Language and Intercultural Communication*, 2(2) pp. 147–152.

Full text is not available in the repository but discovered by CORE:

 [Get full text](#)

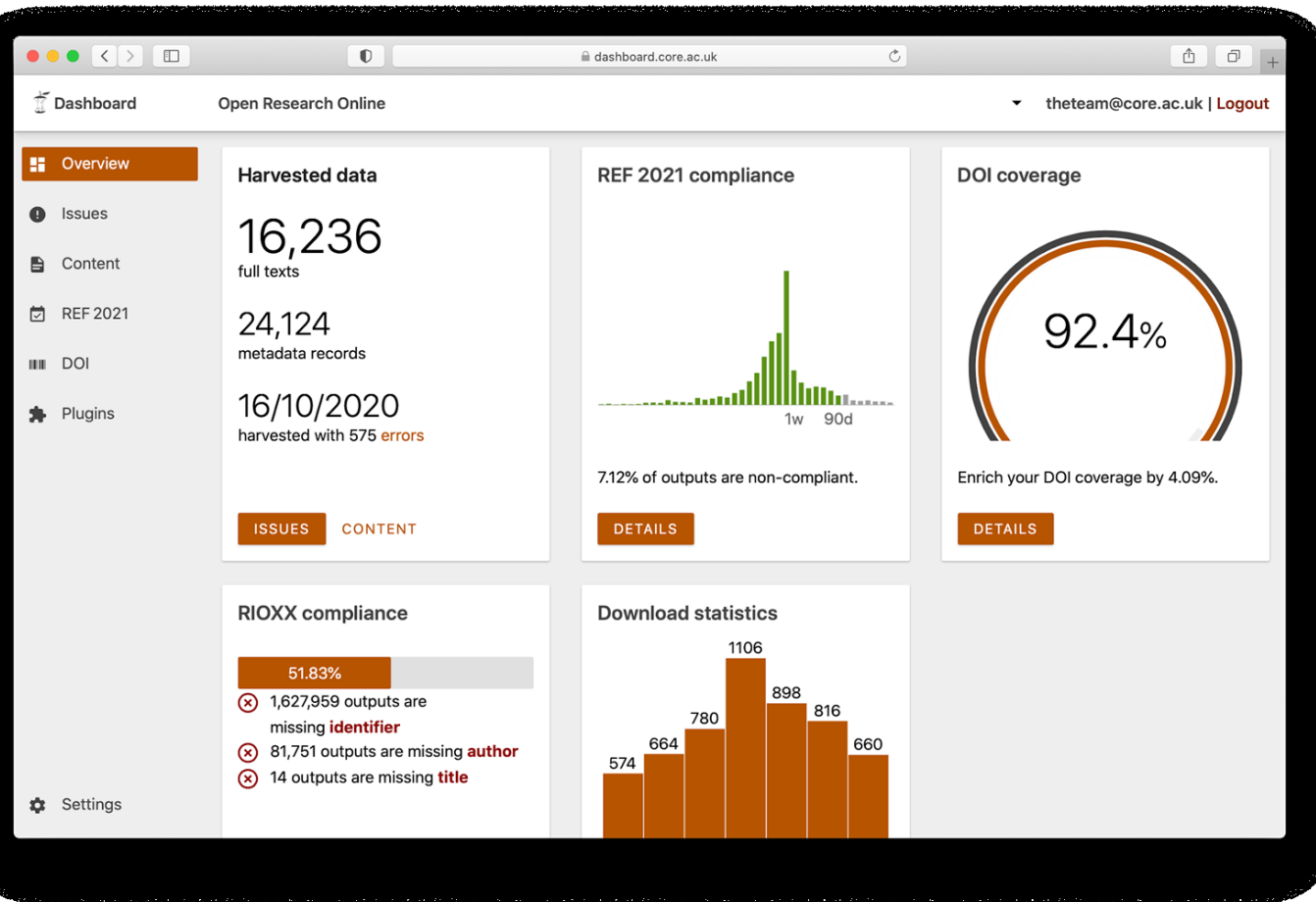
DOI (Digital Object Identifier) Link:	https://doi.org/10.1080/14708470208668083
Google Scholar:	Look up in Google Scholar

Item Type:	Journal Item
Copyright Holders:	2002 F. J. Doloughan
ISSN:	1747-759X
Keywords:	cultural translation; bilingualism; writing and identity
Academic Unit/School:	Faculty of Arts and Social Sciences (FASS) > Arts and Cultures > English & Creative Writing Faculty of Arts and Social Sciences (FASS) > Arts and Cultures Faculty of Arts and Social Sciences (FASS)
Research Group:	Postcolonial and Global Literatures Research Group (PGL)
Item ID:	28875
Depositing User:	Nicola Dowson
Date Deposited:	03 Jun 2011 08:38

CORE's raw data services



CORE Repository Dashboard



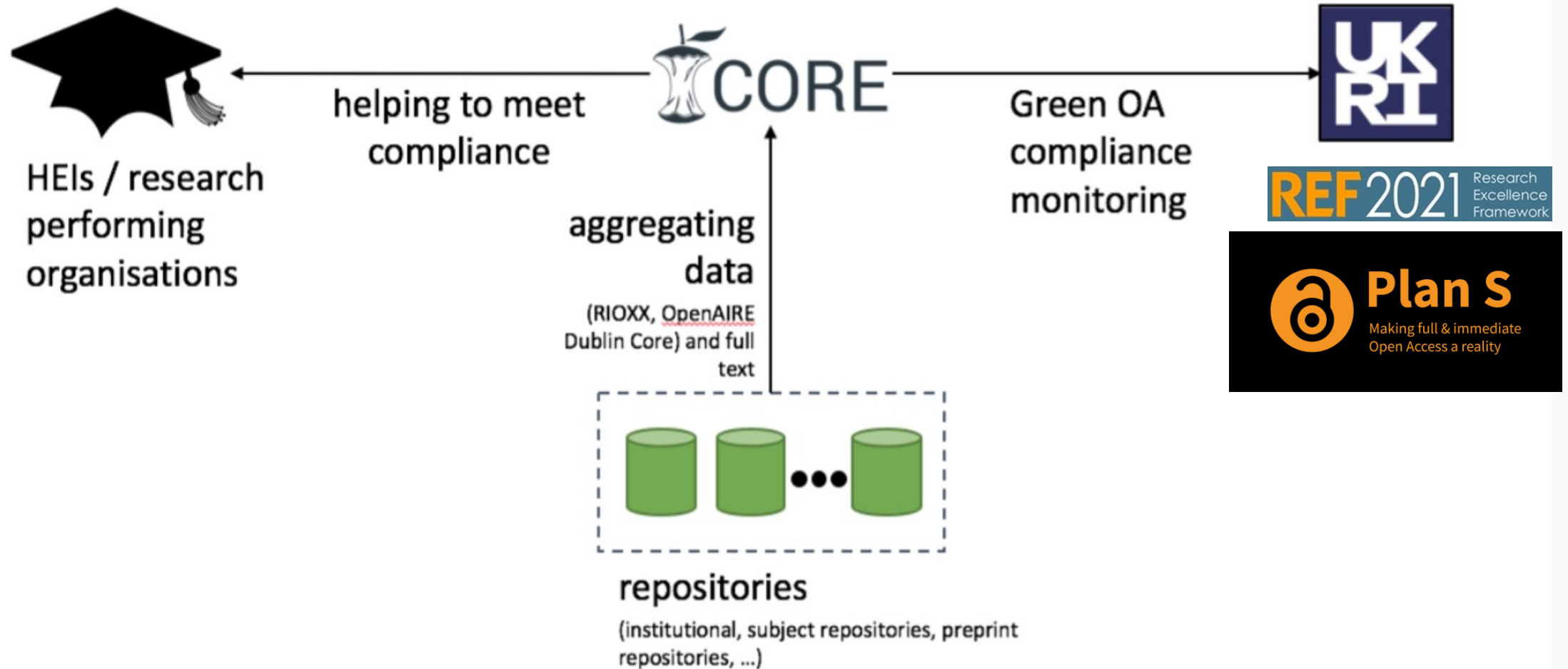
- Download and repository harvesting statistics
- Monitoring, issue reporting and advice to repositories
- Updating of metadata records
- Helps manage access to documents from your repository in CORE
- Data enrichments (DOIs, ORCIDs, ISSNs)
- **Over 100 UK HEIs logged in to the CORE Repository Dashboard in the last 3 months**
- **291 unique registered institutions** and a huge growth potential

The CORE Repository Dashboard provides an excellent level of feedback and control over what gets harvested, aggregated and exposed. It also provides intelligence on harvesting errors which may need troubleshooting, thereby ensuring optimum repository health for text mining as well as discovery of repository content.

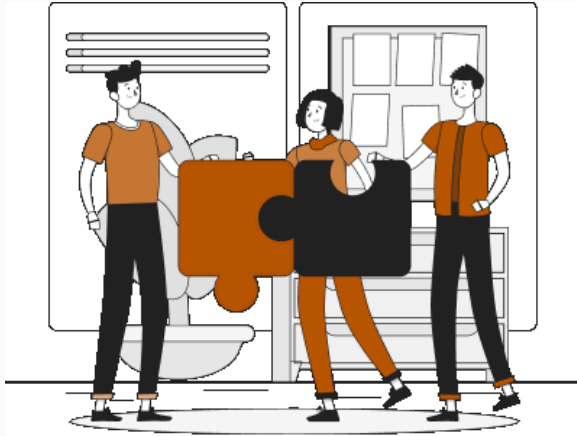


— George Macgregor, Institutional Repository Manager, University of Strathclyde

Green compliance monitoring with CORE

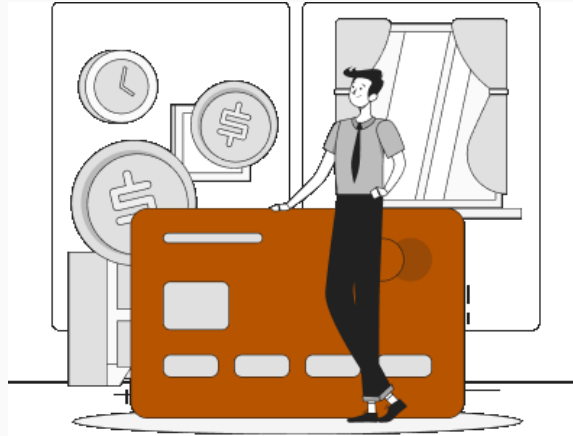


Internal Business Unit



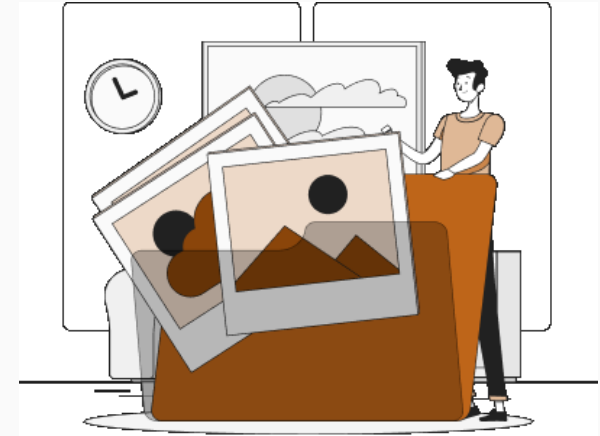
1

Business unit within
The Open University



2

Own budget and slightly
looser admin processes



3

Ability to use the
university infrastructure

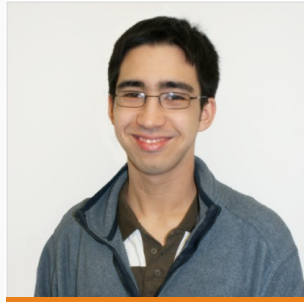
The team



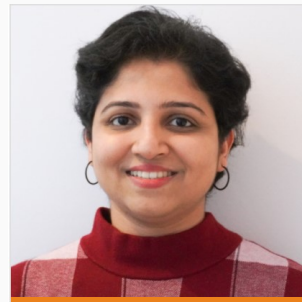
Petr Knoth



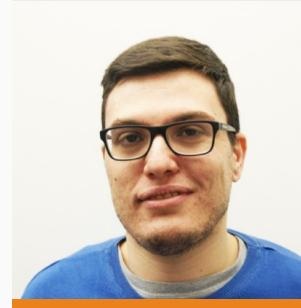
Nancy Pontika



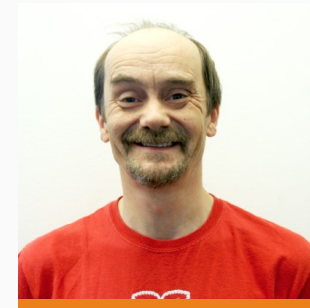
Samuel Pearce



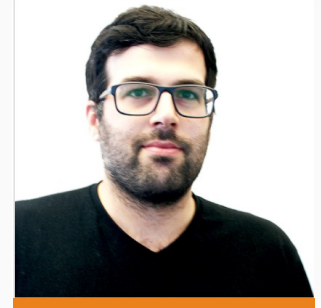
Suchetha
Nambanoor-
Kunnath



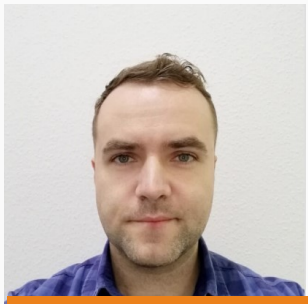
Matteo
Cancellieri



David Pride



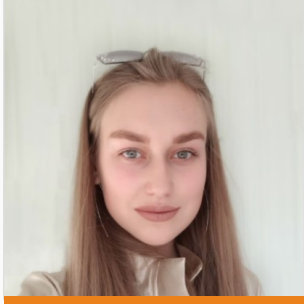
Lucas
Anastasiou



Valeriy Budko



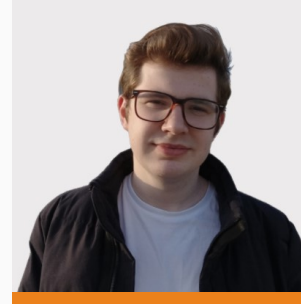
Catherine
Kuliavets



Maria Tarasiuk



Kostiantyn
Voskoboinik



Anton Zhuk



Viktoriia
Pavlenko



Drahomira
Herrmannova

Our locations



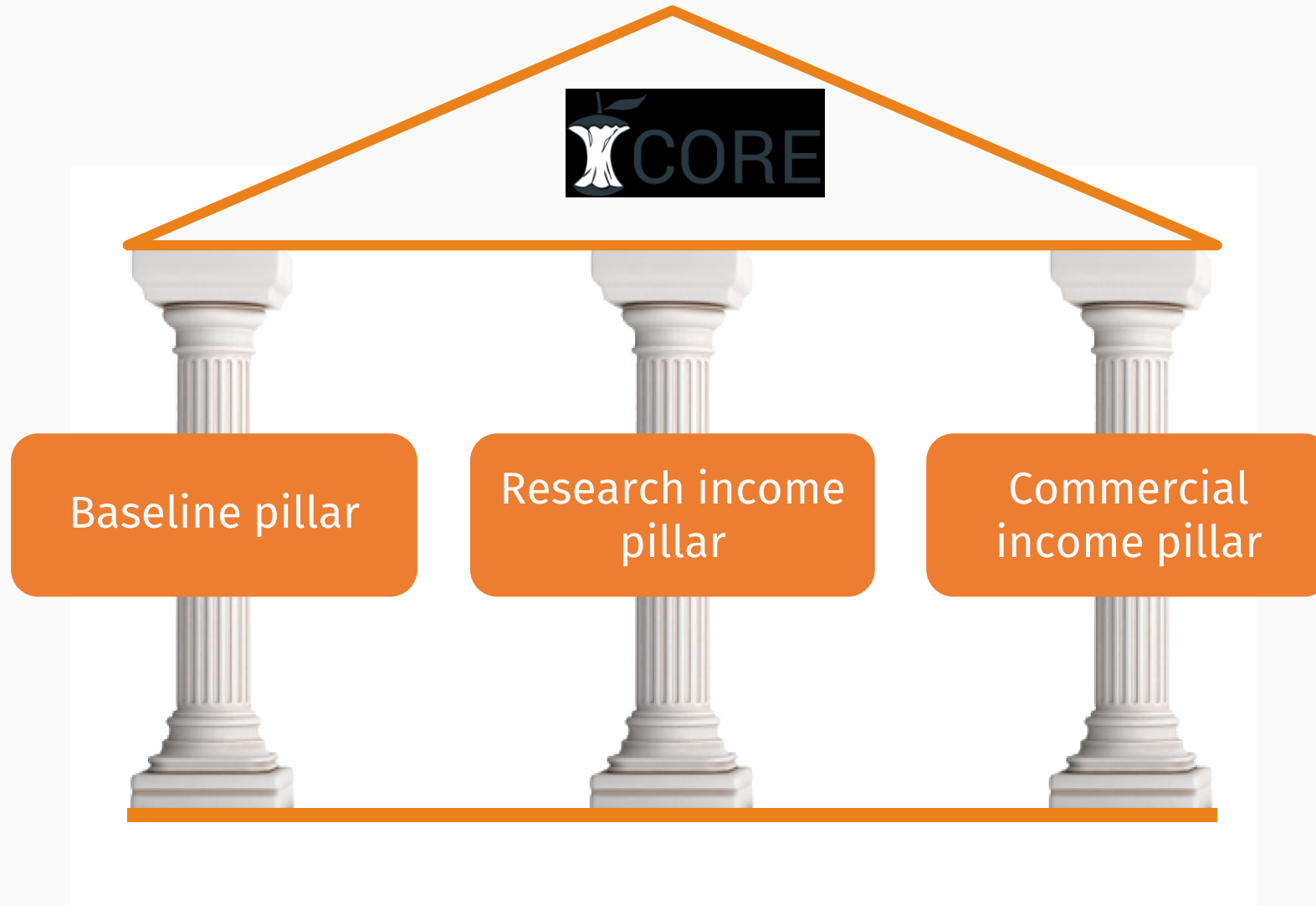
Milton Keynes, UK

- Research
- Development
- Product operation

Kyiv, Ukraine

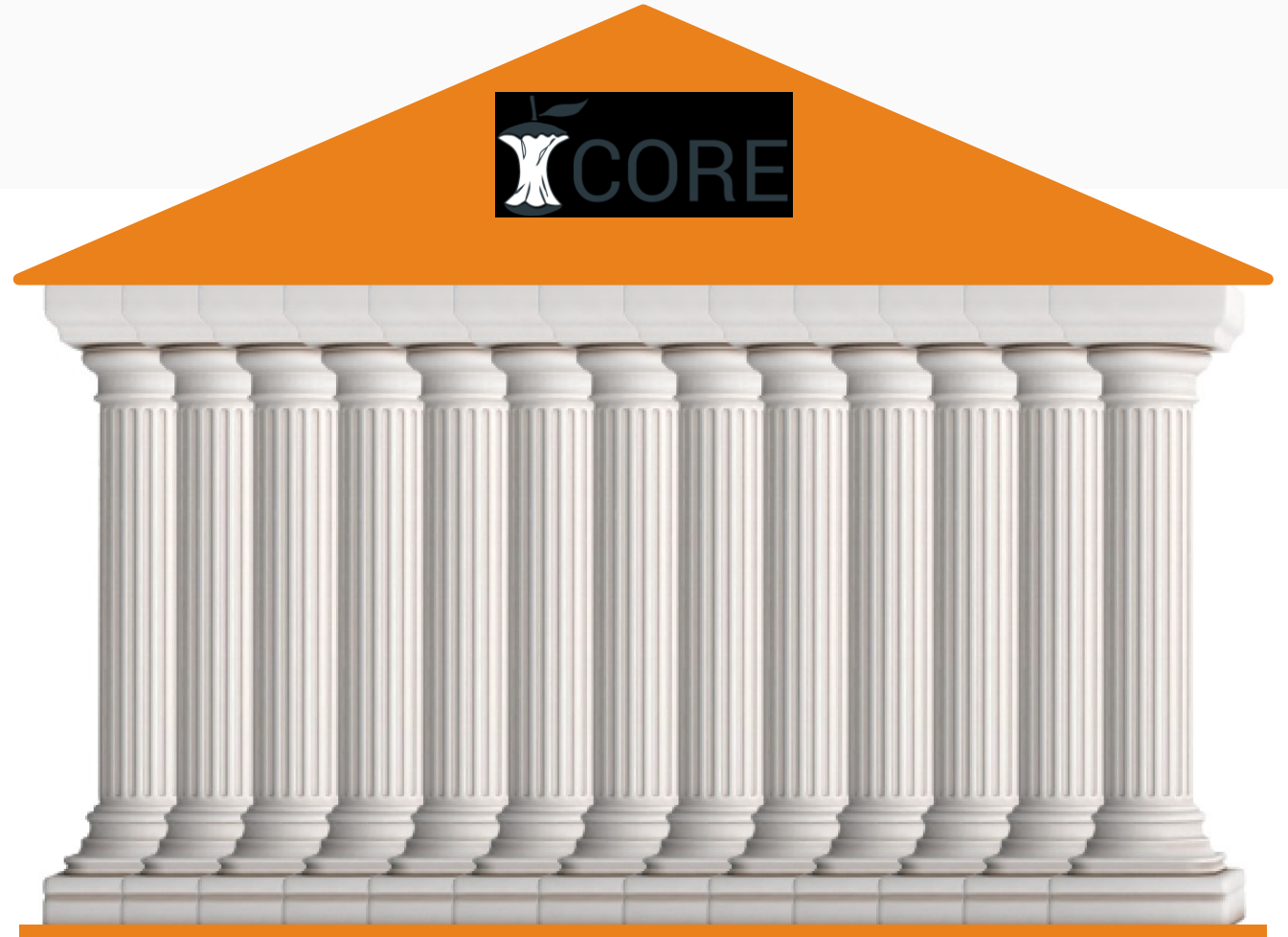
- Development
- Service operation
- Design

The income pillars of CORE



Sustainability strategy

- 1 Sticking to the mission: do not charge the general public and researchers
- 2 Reduce reliance on the baseline and research components and diversify them
- 3 Grow commercial income and use it to sustain the free parts of CORE
- 4 Working towards an even closer alignment with POSI



CORE Sustainability Paths



CORE as a technology and infrastructure open science provider.

Path 1



CORE as part of a UK wide and global institutional support mechanism. (HEIs membership programme)

Path 2



CORE as a provider of Open Access compliance monitoring services in the UK and beyond.

Path 3



Synergies with other services

Path 4

Contributions

- Among the most used open science infrastructures globally
- An enabling infrastructure for open science
- Strictly mission-driven, not-for-profit and value for money
- Products for:
 - B2B – raw data services – text and data mining, service integrations, etc.
 - HEIs
 - compliance monitoring (Research England, Plan S, ...), increased discoverability, metadata quality, support for repository managers
 - CORE Memberships
 - Funders – analytics and compliance
 - Building new products (watch this space)

Thank you!

<https://core.ac.uk>

