Improve interoperability across publisher platforms to support text and data mining

Nancy Pontica (1), Petr Knoth (2), Lucas Anastasiou (3), Giorgio Basile (4), Samuel Pearce (5), and Giulia Dore (6)

(1) CORE, KMi, The Open University, UK, Orcid 0000-0002-2091-0402, (2) CORE, KMi, The Open University, UK, Orcid 0000-0003-1161-7359, (3) CORE, KMi, The Open University, UK, 0000-0002-1587-5104, (4) CORE, KMi, The Open University, UK, (5) CORE, KMi, The Open University, UK, 0000-0001-5616-7000, (6) University of Glasgow

We propose a two hours workshop addressed mainly to publishers, but also to other text and data mining stakeholders. In this workshop we will make interoperability suggestions across publishers platforms and suggest best practices to enable text and data miners achieve harmonised access to hybrid gold open access content. We will outline the technical difficulties we experienced in achieving this and present how we succeeded in harvesting metadata records and full text content of millions of OA articles from publisher APIs. We will also show our documentation with suggestions on what publishers could do to improve the situation and will demonstrate how we have managed to provide an interoperable layer over these data using ResourceSync.
By the end of this workshop we aim to:
- have established a common ground of understanding between publishers and the community of what is considered best practice in implementing programmable APIs for consuming or harvesting publishers' content.
- list common challenges and pitfalls while attempting to collect data via these APIs
- best practice of API usage from the scholarly publications related stakeholders