THE BOOK ANALYTICS DASHBOARD PROJECT

CREATING A SUSTAINABLE ANALYTICS SERVICE TO SUPPORT DIVERSE OPEN ACCESS BOOK PUBLISHERS

Kathryn R Napier[1*], Rebecca N Handcock[1], Keegan R Smith[1], Lucy Montgomery[2], Cameron Neylon[2], Katherine Skinner[3], Ronald Snijder[4], Niels Stern[4]

[1]Curtin Institute for Data Science, Curtin University, AU; [2]School of Media, Creative Arts and Social Inquiry, Curtin University, AU; [3]Educopia Institute, US; [4]OAPEN Foundation, NL
*kathryn.napier@curtin.edu.au

OBJECTIVE

The Book Analytics Dashboard (BAD)[1] project is transforming the capability of small and medium size scholarly book publishers to understand the global impact and reach of digital open access (OA) books. By putting leading-edge analytics capabilities in the hands of diverse publishers we are helping to support the continued diversity of the publishing ecosystems that translate Humanities research into formal outputs; and supporting a shift towards publishing approaches that maximise their positive impacts.

The BAD project's goal is to provide publishers with user-friendly tools to navigate complex data about how their books are being used, by developing a sustainable service[2] that will safeguard and support diversity in the voices, perspectives, geographies, topics, and languages made visible through OA books.
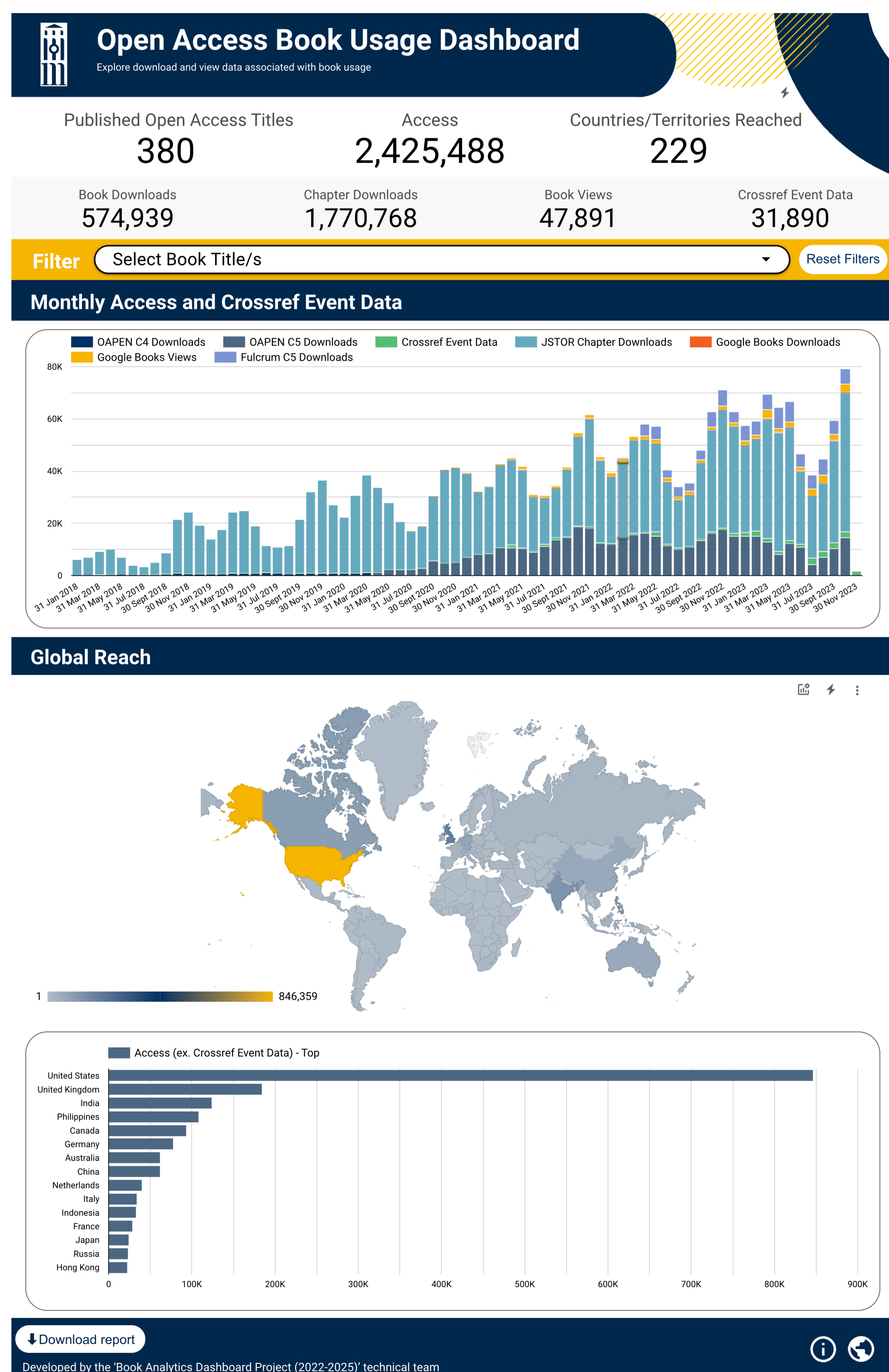

METHODOLOGY AND RESULTS

Usage reports are provided for many platforms where OA books are hosted (e.g. JSTOR, the OAPEN library, publishers websites, Google Books), and aggregating this usage information can be a considerable challenge for publishers. The BAD project has developed automated workflows that collect, link and aggregate usage data and books metadata from a number of different platforms, and present this information in a range of visualisations on dashboards (Fig. 1).

The book usage data workflows code base[3] is built on an open-source workflow system written in Python, and is designed to run using Apache Airflow workflows as a manager/scheduler in conjunction with the Observatory Platform[4], an environment for fetching, processing and analysing data. The workflows transfer, process and transform, disambiguate and link identifiers, and aggregate usage data about OA books from multiple sources via the Google Cloud Platform. The workflows run automatically every week, with usage data aggregated over time by month of usage.

The books metadata and aggregated usage data is then visualised in dashboards provided by Looker Studio (Fig. 2), and are automatically updated as the workflows run. The Overview dashboards are designed to give publishers an overview of the global usage of their OA books longitudinally over time (Fig. 2). Additional dashboards allow publishers to drill down further and assess usage by country or territory, institution, author or editor, subject, or mentions online and citations in datasets. The BAD project dashboards are currently used by 9 publishers and projects from around the globe.
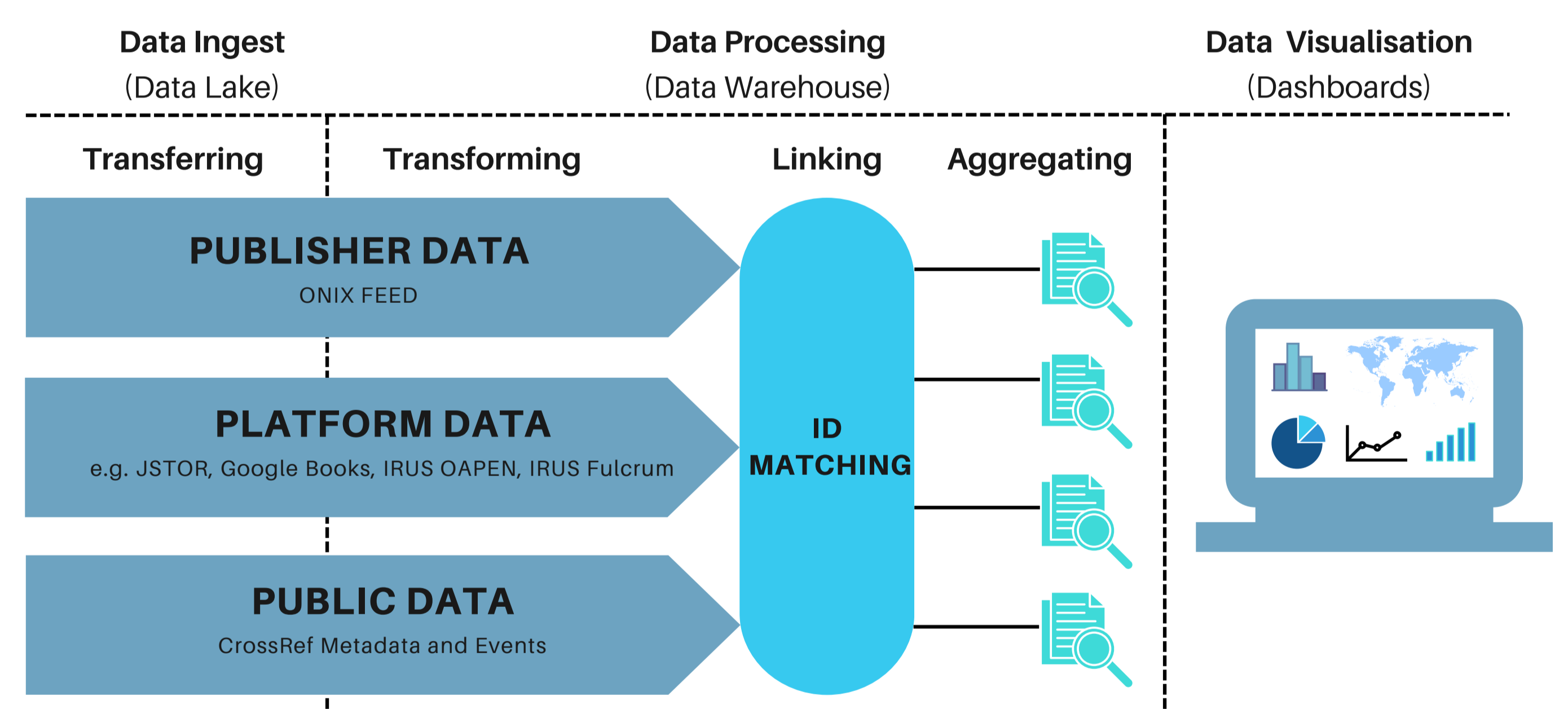

Figure 1. A high level overview of the book usage workflows that transfer data into Google Cloud Storage, transform, link and aggregate data in BigQuery, with dashboards visualised using Looker Studio.

IMPLICATIONS

The BAD project is creating a ground-breaking OA book focused analytics service, the first of its kind in OA publishing. It provides publishers with vital information about the global usage of their OA books, acting as a tool for advocacy for OA publishing and infrastructures.

FOR MORE INFORMATION

1 Visit the BAD project page: http://book-analytics.org/
2 Visit the Book Analytics Service page: https://oapen.org/article/book-analytics-service
3 Book usage data workflows: https://github.com/The-Academic-Observatory/oaebu-workflows
4 The Observatory Platform: https://github.com/The-Academic-Observatory/observatory-platform
5 The University of Michigan Press Dashboard: https://ebc.press.umich.edu/impact/#oa-book-usage

To see the BAD template dashboard, view the QR code:

Figure 2. The Overview dashboard for the University of Michigan Press[5]