Book Analytics Dashboard Demonstration (BAD) Project (2022-2025)

OA Book Usage Analytics for Diverse Communities - A Demonstration Project: 2022-2025

EDUCOPIA INSTITUTE

open Open Access

Curtin University

Mellon Foundation

Project overview

Project partners and people

Partner Organisations

- **Curtin University**
- **COARD**
- **OAPEN**
- **Educopia**

Technical Team

- Rebecca Handcock
- Kathryn Napier
- Keegan Smith

Principal Investigators:

- Lucy Montgomery
- Cameron Neylon
- Katherine Skinner
- Ronald Snijder
- Niels Stern

Project Management and Support:

- Laura J. Wilkinson
- Kathryn Barker
- Isabelle Perger

Description of Proposed Work

Formal Project Title: **OA Book Usage Analytics for Diverse Communities – A Demonstration Project**

Short Title: Book Analytics Dashboard Project

Duration: April 2022 – April 2025

**Funders:** Mellon Foundation, COARD

The Book Analytics Dashboard Project (2022-2025) is focused on creating a sustainable OA Book focused analytics service.

Our goal in creating this service is to help safeguard and support diversity in the voices, perspectives, geographies, topics and languages made visible through OA Books.

EDUCOPIA INSTITUTE

open Open Access

Curtin University

Mellon Foundation

Background and Context

The Book Analytics Dashboard project is building on an earlier Mellon-funded initiative: Developing a Pilot Data Trust for OA eBook Usage (2020 – 2022).

2020-22 Usage Data Trust Pilot Project

Dashboard Service (BAD)

OAeBU Data Trust
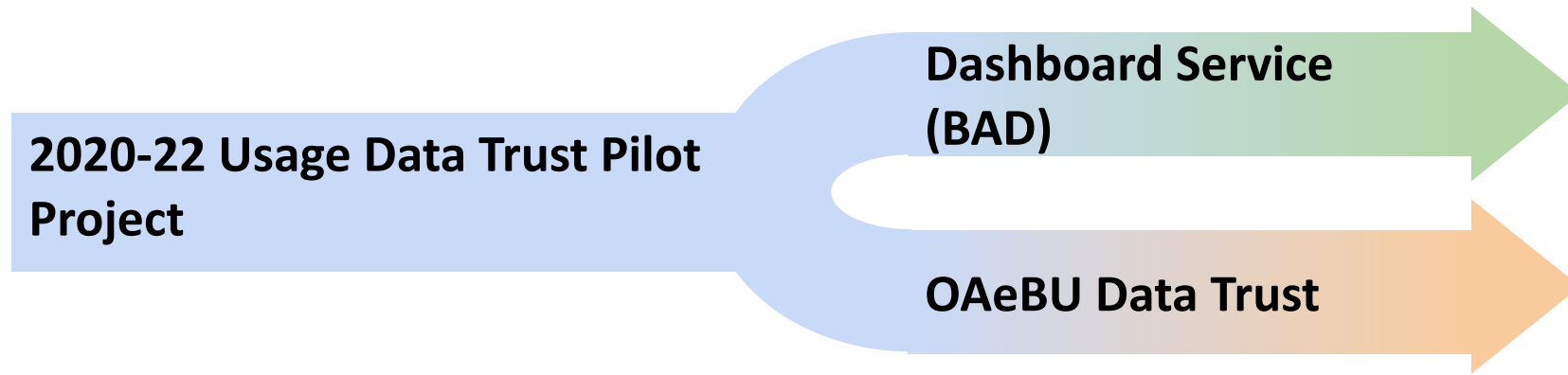
The Book Analytics Dashboard project is:

- Scaling dashboard infrastructure and workflows
- Developing a service model, including customer support
- Identifying and implementing a long-term plan for housing, maintenance and funding
- Providing an analytics service as a sustainable community infrastructure for OA book stakeholders

Mellon Foundation

2020-2022 (OAeBU) PROJECT GOALS

- 2-year Mellon Foundation funded project

- Develop & pilot data trust infrastructure

    - 6 dashboard partners

    - Platform/data partners

    - Implement and extend relevant open-source technologies

- Develop personas and use cases

- Document the OA eBook supply chain

- Define sustainability, governance, and policy models

Mellon
Foundation

2020-2022 (OAeBU) Technical work

1. **Open Access Book usage** (from JSTOR, OAPEN IRUS-UK, Google Books, ANU Press Repository, Google Analytics, UCL Discovery Repository)

2. **6 dashboard partners across a range of sectors and sizes** (5 Presses and 1 Repository/Platform)

3. **Benchmarking of usage data across Pilot partners**

4. **Open Source re-hostable codebase -** state of the art workflow, data-warehouse and dashboard systems deploying Airflow, Google Cloud Platform, ElasticSearch and Kibana

2022-2025
BAD Project work

Mellon
Foundation

BAD Work packages April 2022 - March 2025

- **WP1:** Technical Roadmap Consultation: Developing a Process and Model
- **WP2:** Workflow Standardization
- **WP3:** Partner Onboarding Phase 1
- **WP4:** Community Cultivation
- **WP5:** Sustainability Model and Community Consultation
- **WP6:** Onboarding of Phase 2 Partners
- **WP7:** Sustainability Development Plan
- **WP8** Regional Strategy Development
- **WP9:** Implementing Processes for Ongoing Technical Operations
- **WP10:** Organizational Form Development and Implementation

Activities and rationale - Full document: https://doi.org/10.5281/zenodo.6618791

Mellon Foundation

Useful links

Code and documentation

- [Book Usage Workflows Read the Docs](#)
- [Book Usage Workflows GitHub Repository](#)
- [University of Michigan Press website with book usage dashboards](#)
- Directly to the University of Michigan Press's public book usage dashboards [here](#)
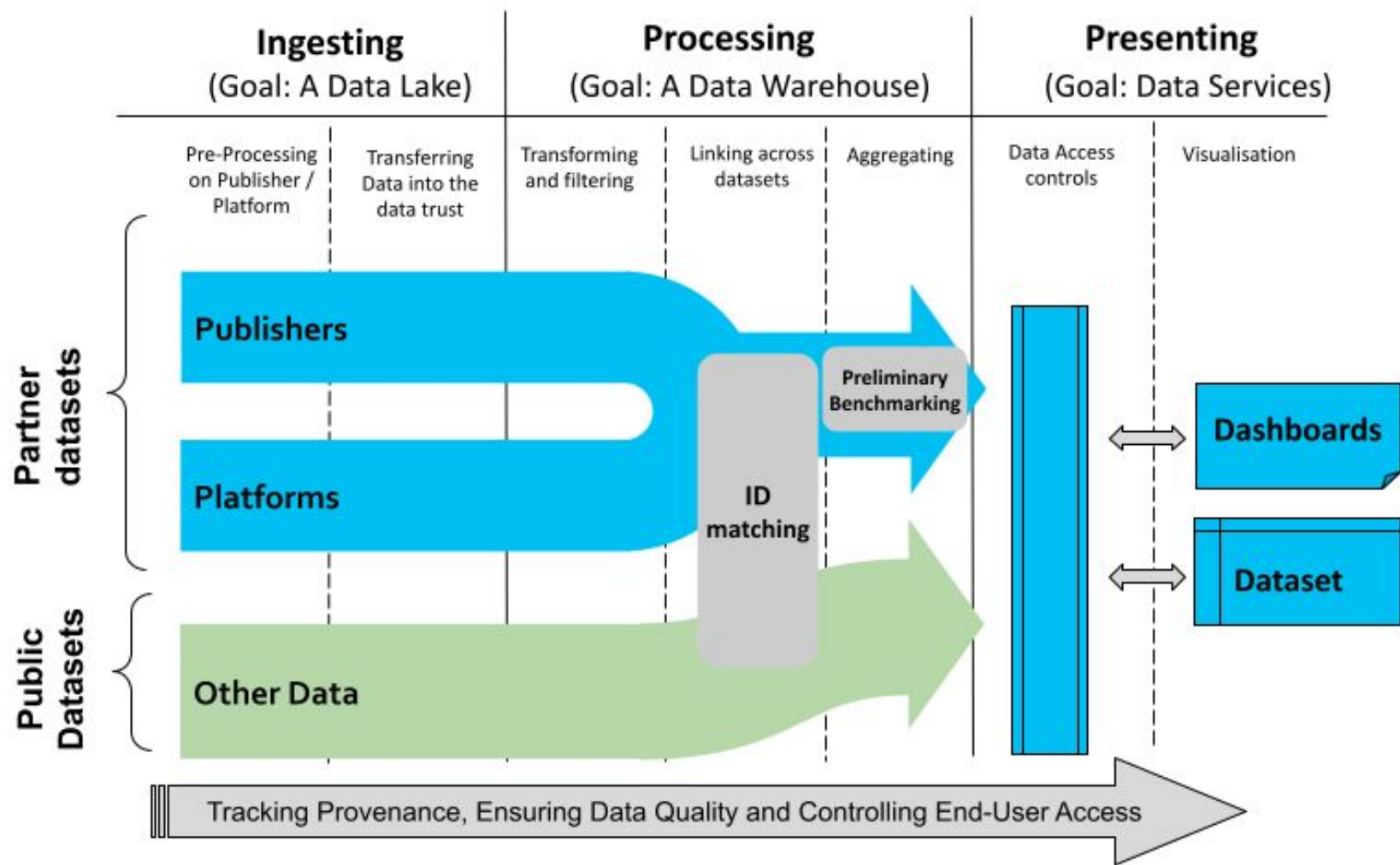
2023 - 2025 Project page

- [2023 - 2025 project page](#)
- BAD [Zenodo Community](#) and [Info Landing Page](#)
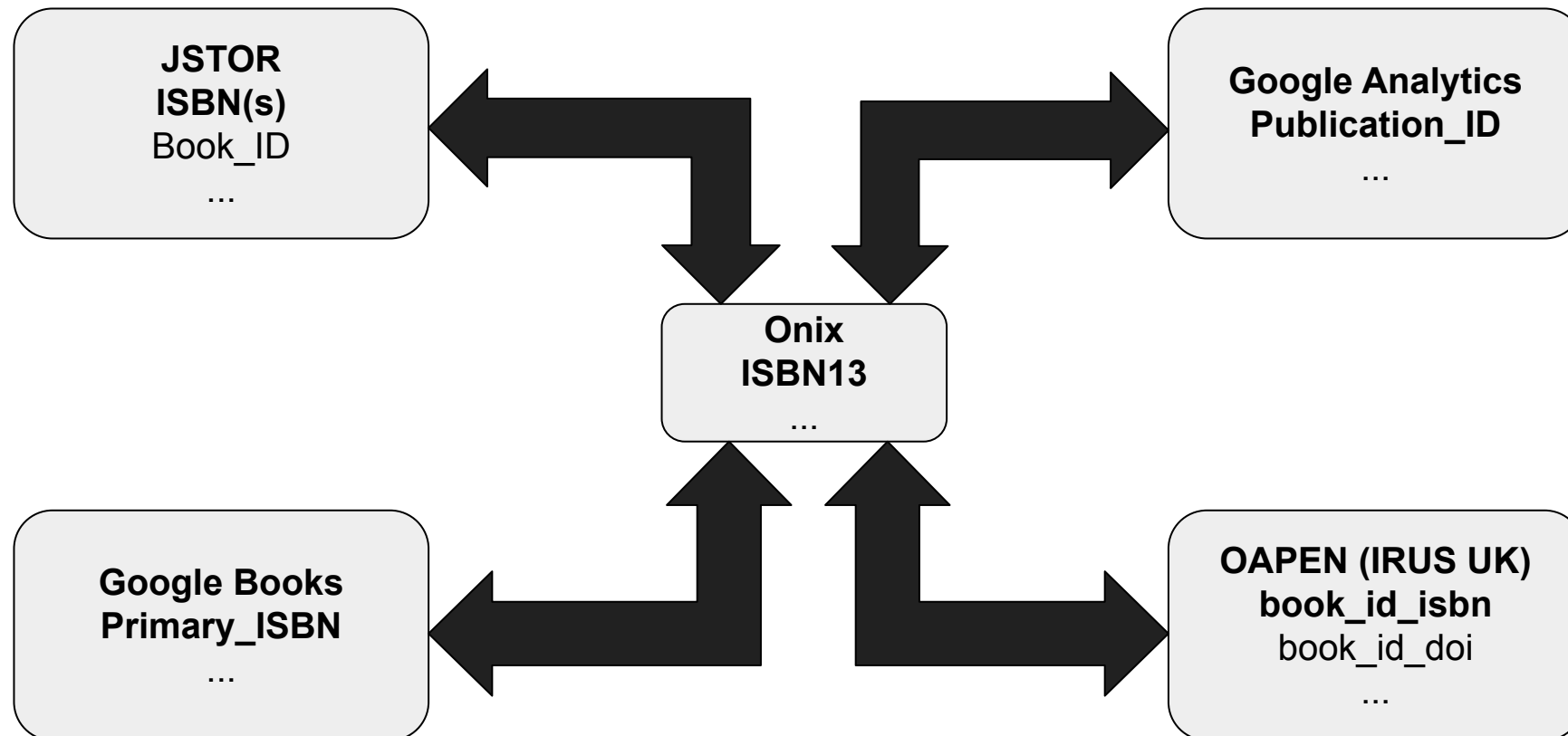
2020-2022 Project

- [2020 - 2022 project page](#)
- [OAeBU Zenodo repository community](#)
- The [OAeBU project page](#) on the Educopia Institute website

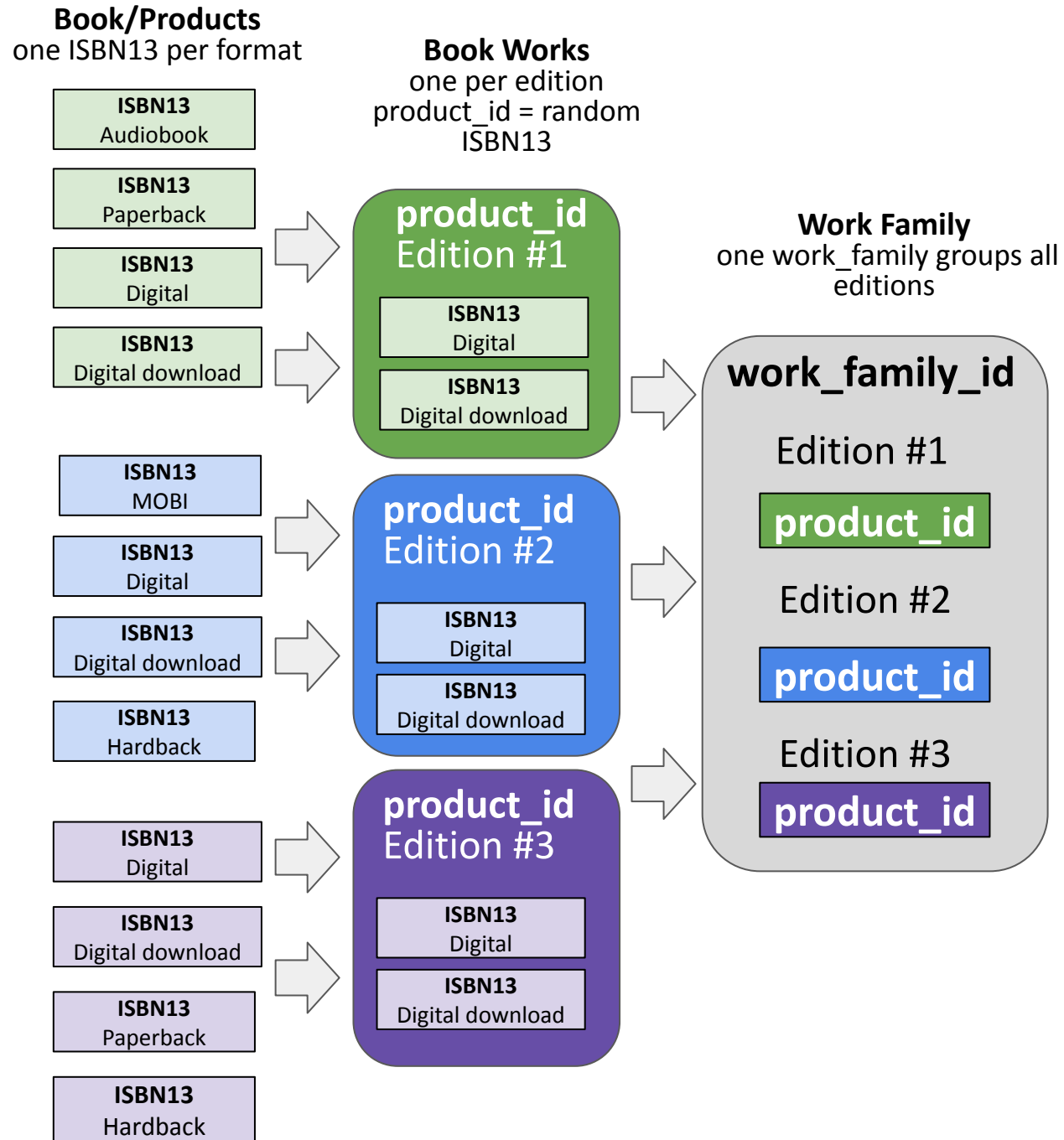# BAD Technical System Outline

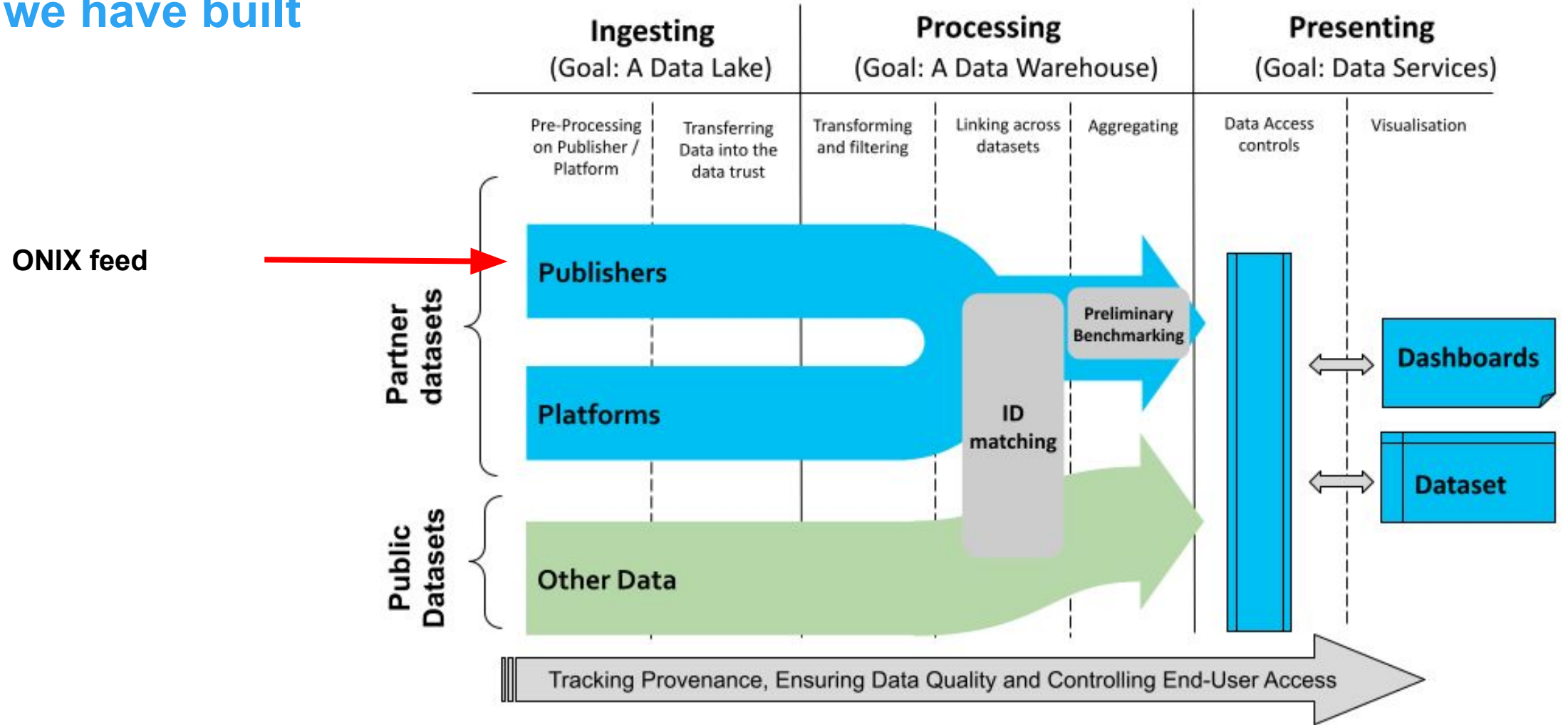# What we have built

ISBN-13 and ONIX as the coordinating ground truth

Book ID relationships

ONIX 3.0 standard requires ISBN13 to be transmitted in the record

The publisher can optionally disclose relationships via the "RelatedWorks" field

Book/Products
one ISBN13 per format

| ISBN13 Audiobook |
| ISBN13 Paperback |
| ISBN13 Digital |
| ISBN13 Digital download |

| ISBN13 MOBI |
| ISBN13 Digital |
| ISBN13 Digital download |
| ISBN13 Hardback |

| ISBN13 Digital |
| ISBN13 Digital download |
| ISBN13 Paperback |
| ISBN13 Hardback |

Book Works
one per edition
product_id = random
ISBN13

product_id
Edition #1
- ISBN13 Digital
- ISBN13 Digital download

product_id
Edition #2
- ISBN13 Digital
- ISBN13 Digital download

product_id
Edition #3
- ISBN13 Digital
- ISBN13 Digital download

Work Family
one work_family groups all editions

work_family_id

Edition #1
- **product_id**

Edition #2
- **product_id**

Edition #3
- **product_id**

Mellon Foundation

What we have built

ONIX feed

How it works
Combine, collate, & aggregate view

1

Publisher ONIX feed
ISBN13 identifier
Published-year, DOI, Title,
Authors & affiliations ...
Keywords, Subjects,

ISBN

ISBN

ISBN

ISBN

ISBN

ISBN

ISBN   ISBN   ISBN   ONIX Metadata

Crossref Metadata   Crossref Metadata

2
Usage data providers e.g.

Google Books

ISBN

ISBN

ISBN

etc ...

JSTOR

ISBN

ISBN

ISBN

etc ...

1. Metadata
2. Usage data by Country, Title, Institution etc.

3 **Bigquery Book Usage Dataset**
Example: one ISBN from one publisher with usage data by country for one month, from each data provider

ISBN **Metadata collated for ISBN**	
Onix feed	Source of "truth"
Usage Providers	Used for QC
Crossref Metadata	Used for publisher info

Usage data - JSTOR	
Usage from Country A	◆
Usage from Country B	⬤
Usage from Country C	✖

Usage data - GoogleBooks	
Usage from Country A	◆
Usage from Country B	⬤
Usage from Country C	✖

4 **Assemble pre "aggregations" for ElasticSearch / Kibana**
Example: ISBN, Country, Month

Country Usage data - JSTOR		
ISBN	Usage A from JSTOR GoogleBooks etc.	◆
ISBN	Usage B from JSTOR GoogleBooks etc.	⬤
ISBN	Usage C from JSTOR GoogleBooks etc.	✖
ISBN	Usage A from JSTOR GoogleBooks etc.	◆
ISBN	Usage B from JSTOR GoogleBooks etc.	⬤
ISBN	Usage C from JSTOR GoogleBooks etc.	✖

Aggregation types

ISBN, Usage Data Provider, Month
ISBN, Country, Month
ISBN, Institution, Month
ISBN, Publisher, Month
ISBN, Event, Month
ISBN, CIty, Month
Author, Month
Publisher, Month
Subject, Month

Data is not "aggregated" but tables of "aggregatable" elements are collated.

Standardisation is necessary as usage data providers have different usage aggregation formats for data
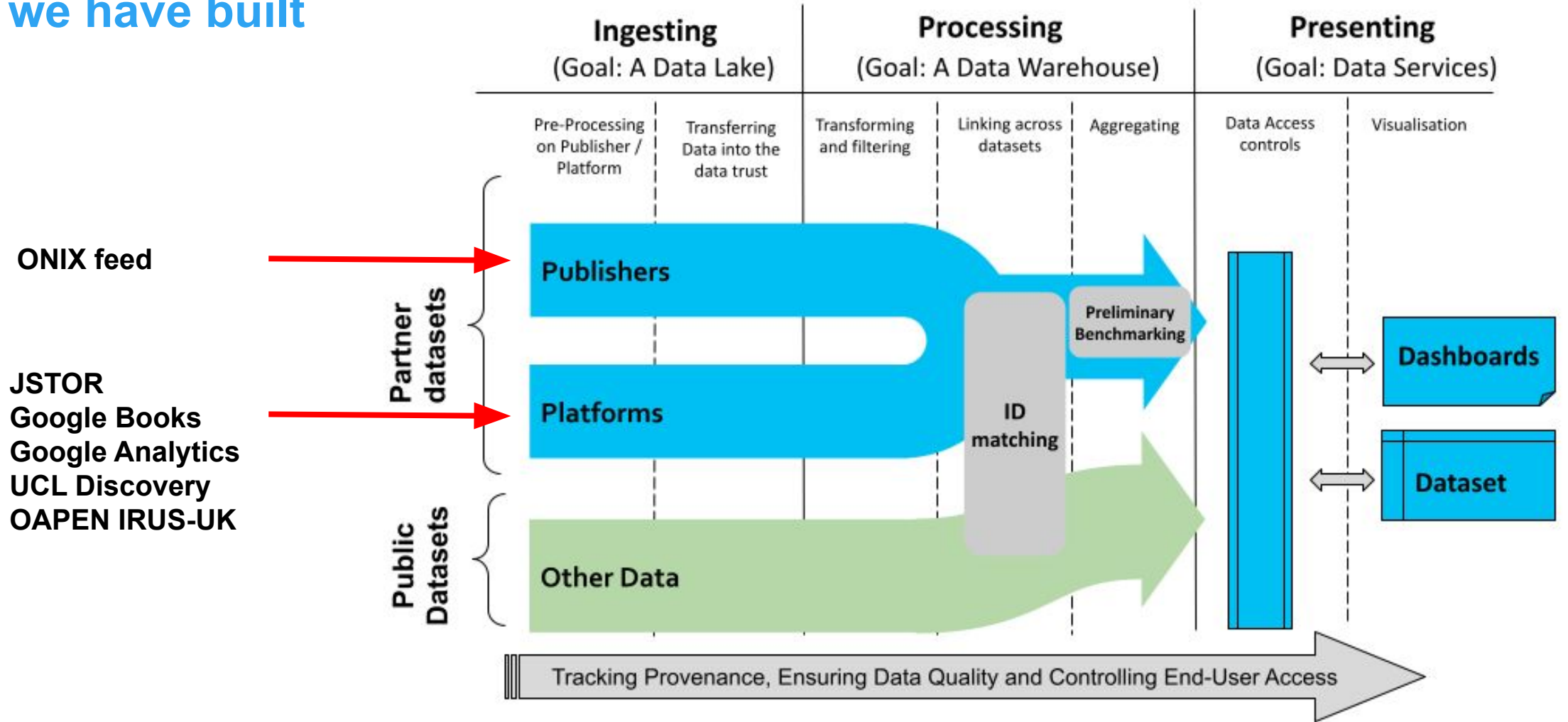
5
Dashboards in ElasticSearch / Kibana

Data elements are "aggregated" as part of the Kibana visualizations

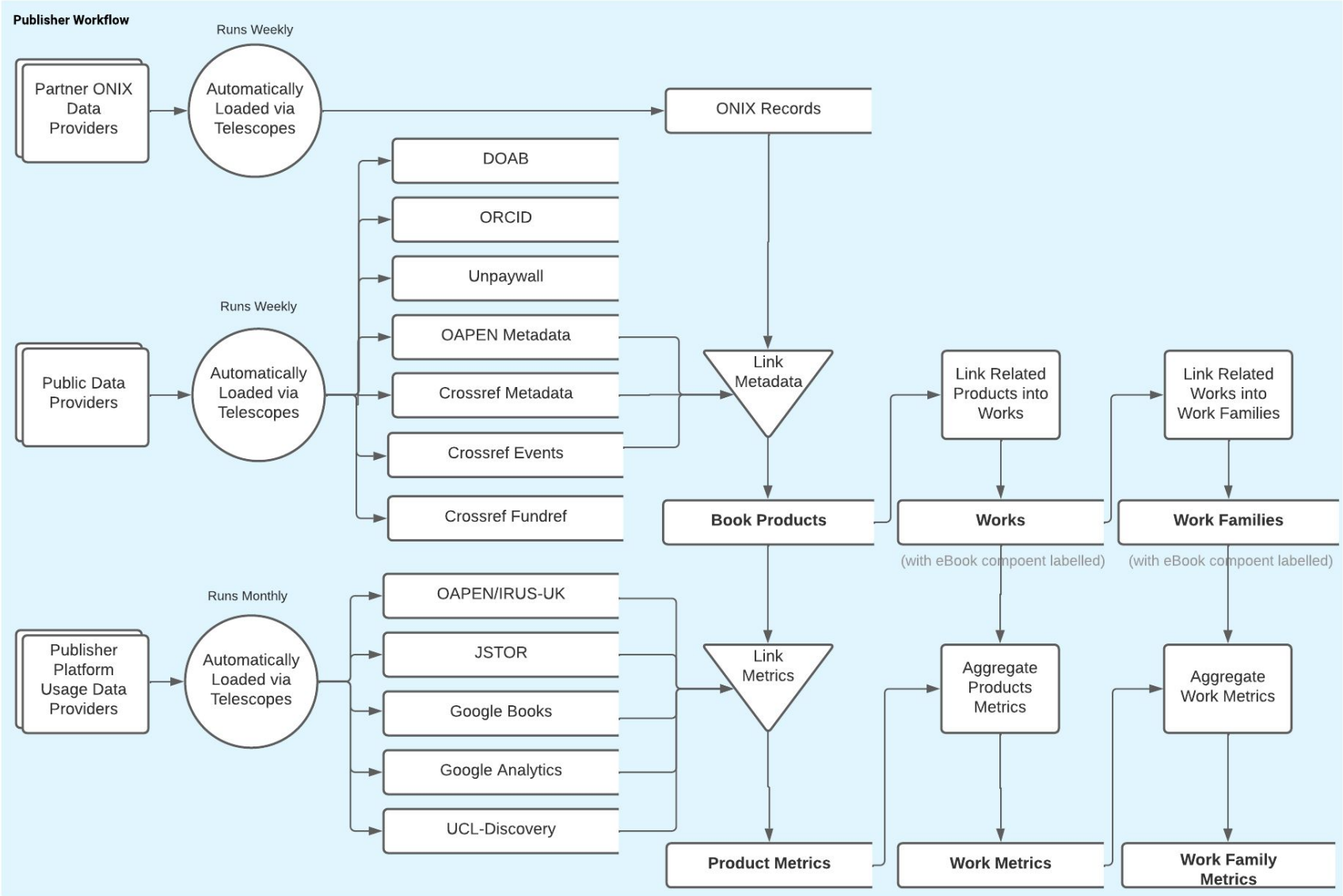**1** → Source metadata → **2** Combine & collate usage data → **3** Standardize aggregation variables → **4** Assemble aggregations and export for ElasticSearch → **5**

What we have built

ONIX feed

JSTOR
Google Books
Google Analytics
UCL Discovery
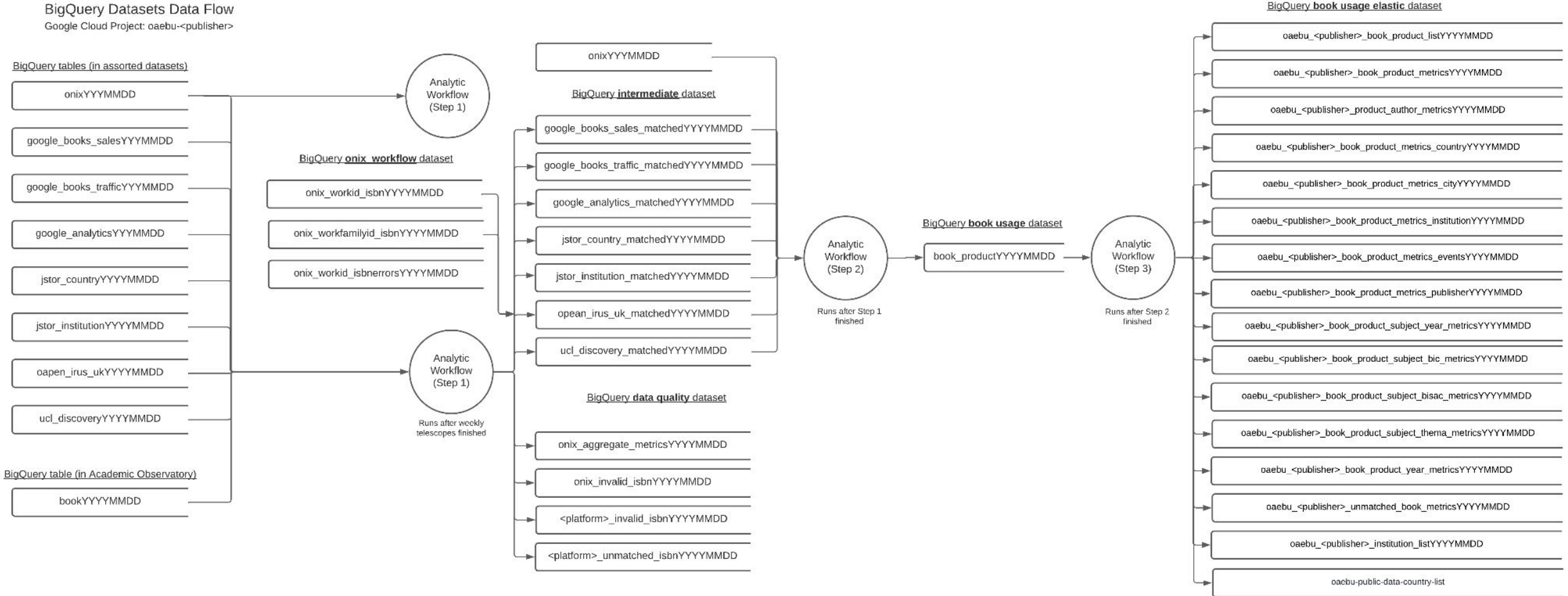OAPEN IRUS-UK

Crossref events
Crossref metadata
OAPEN metadata

What we have built

What we have built

Useful links

Code and documentation

- [Book Usage Workflows Read the Docs](#)
- [Book Usage Workflows GitHub Repository](#)
- [University of Michigan Press website with book usage dashboards](#)
- Directly to the University of Michigan Press's public BAD dashboard [here](#)

2023 - 2025 Project page

- [2023 - 2025 project page](#)
- BAD [Zenodo Community](#) and [Info Landing Page](#)

2020-2022 Project

- [2020 - 2022 project page](#)
- [OAeBU Zenodo repository community](#)
- The [OAeBU project page](#) on the Educopia Institute website