





Book Analytics Dashboard (BAD) Demonstration Project:

Updated Technical
Development Roadmap
2022-2023

Contact Details

	join our mailing list here email us here: info@book-analytics.org
	follow us on twitter
	visit the BAD project website here
	visit the BAD project Zenodo community here

Revision History

Version	Date	Author	Comment
0.1	28/06/2022	COKI team: Cameron Neylon, Kathryn Napier, Lucy Montgomery, Rebecca Handcock	Cameron Neylon wrote and edited this draft version with feedback and editing from Kathryn Napier, Lucy Montgomery and Rebecca Handcock.
0.2	30/06/2022	COKI team: Kathryn Napier	Links to the University of Michigan Press dashboard and documentation included.
0.3	29/08/2022	COKI team: Cameron Neylon, Kathryn Napier	Community feedback incorporated and summarised.
0.4	28/09/2022	COKI team: Kathryn Napier	Appendix B removed and further feedback incorporated.
1	11/01/2023	COKI team: Cameron Neylon, Kathryn Napier	Incorporated feedback from the project team and updated technical priorities (user experience improvement).



This project deliverable by the Book Analytics Dashboard Project is licensed under a Creative Commons Attribution 4.0 International License.

Contents

Contact Details	2
Revision History	2
Contents	3
Preface	4
Purpose of this Document	4
How to Provide Feedback	4
Outline and Summary	5
1. Main Development Areas and Project Management	7
1.1 Project Management	7
1.2 Technical Roadmap Consultation (WP1)	7
1.3 Workflow Standardisation (WP2)	7
1.4 New Data Sources	7
1.5 Maintenance and System Upgrades	8
1.6 User Experience Improvement	8
2. Issues for Community Consultation and Input	9
2.1 Technical Roadmap Consultation (WP1)	9
2.2 Access Control Systems (WP2)	10
2.3 New Data Source Prioritisation (WP1)	11
2.4 Standardisation and Partner Onboarding (WP2, WP3)	12
2.4.1 Standardising Partner Onboarding (WP3)	12
2.4.2 Standardising Data Ingest Pipelines	13
3. Detailed Technical Work Plans April 2022 - March 2023	14
3.1 Work Package 1	14
3.1.1 BAD-2: Technical Roadmap Consultation	14
3.2 Work Package 2	14
3.2.1 BAD-3: Workflow Standardisation	14
3.3 Work Package 3	14
3.3.1 BAD-4: Partner onboarding Phase 1	14
Appendix A	16

Preface

Purpose of this Document

The draft technical roadmap was provided as a commentable public document that we shared with the publishers and stakeholder communities engaged with the Book Analytics Dashboard (BAD) demonstration project that is now underway. The draft technical roadmap was intended to allow us to capture feedback on technical development priorities for the BAD project.

The draft technical roadmap was also a first step in a community consultation process that will be tested and developed over the course of the BAD project. Our aim is to develop a robust process for community consultation that will feed into an annual cycle of technical development of the Book Analytics Dashboard service capabilities; as well as reporting, governance and accountability.

For more information and the latest announcements about the project please visit our [project page](#) or sign up to the project's [mailing list](#).

How to Provide Feedback

The deadline for community feedback on the draft technical roadmap document was **31 July 2022**.

There were 4 ways that people could provide us with their feedback:

1. Commenting was enabled and people were welcome to make comments anywhere throughout the document.
2. We provided prompts in the form of specific questions that we requested input on. These questions were flagged with the heading 'Questions for Community Feedback' and appear towards the end of each section of the document. People were able to leave responses to these questions by leaving a comment.
3. Responses to our questions could also be provided using [this online form](#).
4. Our email address was provided and people invited to send us an email with their comments and suggestions. Comments via email can be sent to: coki@curtin.edu.au

Once feedback was received, the draft technical roadmap was revised to this updated document that will be shared with the community. This updated technical roadmap will form the basis of the technical work that we do over the next 12 months.

Additional opportunities to help shape the BAD project will occur throughout the year; and in 2023 the community will once again be invited to help us to shape a technical development roadmap.

Outline and Summary

The work-plan for technical development in the [BAD project](#) is focused on delivering an efficient, stable, and scalable set of systems to support book community members to gain access to useful analytics systems to guide their work and decision making. This work-plan involves three main strands of activity across the three years of the project:

- **Year 1:** standardising and simplifying onboarding of new partners;
- **Year 2:** testing and refining these onboarding systems;
- **Year 3:** scaling up the set of partners.

This technical roadmap focuses on decisions that we need to make about the activities that the project will prioritise during Year 1.

The [previous project \(Developing a Pilot Data Trust for Open Access Ebook Usage \(2020-2022\)\)](#) developed a [prototype infrastructure](#) for capturing, integrating and displaying open access book usage data for [publisher partners](#). There are several areas where critical improvements to this prototype infrastructure are required for enabling scaling, particularly in the area of control systems for user access. Additional work is also required to standardise and simplify the partner onboarding process, to make this more accessible to small and medium partners, and to reduce the ongoing costs of the system as far as is feasible.

The existing infrastructure from the previous project is capable of ingesting usage data from multiple platforms, such as [JSTOR](#), [Google Books](#), Google Analytics, [OAPEN IRUS](#) (COUNTER v4 and v5 reports) and demonstrated the ingest of usage data from an eprints repository ([UCL Discovery](#)), along with general bibliographic data from public datasets ([Crossref](#), [OAPEN](#)). The data from these sources are integrated with publisher data via the book-specific metadata standard, [ONIX](#). The University of Michigan Press have kindly made their [open access book usage dashboard publicly available](#), providing a detailed example of the usage statistics that are currently aggregated and visualised. Detailed user guides for this dashboard [are available here](#).

Expanding the set of usage data sources is important and there is scope for selecting a small number of additional usage data sources to be added in Year 1. Options for community prioritisation include [EBSCO](#), [Project Muse](#), [Fulcrum](#) (IRUS), [SciELO books](#), and [ProQuest](#) as

Official project title: OA Book Usage Analytics for Diverse Communities - A Demonstration Project

Short project title: Book Analytics Dashboard (BAD) Project

1 April 2022 - 31 March 2025

some of the possibilities. Other platforms which have been raised for potential future investigation include [Amazon](#), the [Internet Archive](#) and [Science Open](#).

1. Main Development Areas and Project Management

1.1 Project Management

The BAD project uses Jira for project management. Work packages from the [Grant Proposal](#) are captured as Epics. The overall project Gantt chart can be seen in Appendix A. The main technical work packages are [summarised here](#). The core technical packages of work for Year 1 are summarised below. Where these are a current distinct work package in our project management system (which aligns currently with the proposal text) the Work Package number is noted for reference.

1.2 Technical Roadmap Consultation (WP1)

This Technical Roadmap consultation is a distinct work package and analysis task in Year 1. It is both a consultation and prioritisation exercise as well as an iterative process to develop appropriate systems for ongoing technical planning and consultations with community input.

1.3 Workflow Standardisation (WP2)

This work package focuses on an initial standardisation of the process of bringing on new partners. It will develop and implement updated Data Sharing Agreements with existing partners and refine processes for onboarding of new partner presses. It will ensure that the technical and legal capacity needed to scale new members in WP3, WP7 and WP9 are in place.

1.4 New Data Sources

Our goal is to support a regular cycle of incorporating new sources of data as prioritised by the community. The work to ingest a new data source (aka a “telescope”) involves scoping (both technical and legal), design of software systems, development, testing, and user validation. Depending on the complexity of the data source this can take from a few weeks to a few months of effort.

As part of our overall system design we are actively working to simplify and standardise these systems to make the development of new “telescopes” quicker and easier. As the entire system is open source it is also possible for others to contribute to the code base to ingest new data sources. Such community contributions would require testing and validation prior to deployment so still has resourcing implications, but in the long term we see this as a good way to support new requirements.

1.5 Maintenance and System Upgrades

Maintenance is an ongoing process and a significant part of the costs of operations. This work includes monitoring workflows for issues, updates to the data ingest pipelines (e.g. when data source schemas change or an API is modified), as well as ongoing system updates to ensure infrastructure and system stability (e.g. upgrades to new versions of our core underpinning infrastructures including Terraform, Airflow, Google Cloud Platform and Elastic/Kibana).

1.6 User Experience Improvement

As the first year of the BAD project commenced and after several project related workshops were held in Europe in October 2022 with [BAD project and OPERAS-PLUS project partners](#), we realised that user experience improvement of the current dashboards needed to become a higher priority. As the work package Sustainability Model and Community Cultivation (WP5) of the project came online, we realised that the standardised dashboard design task in WP2 needed to incorporate a different dashboarding solution to produce dashboards suitable for scalable and sustainable service. We therefore produced a [template dashboard](#) using [Looker Studio](#), a dashboarding solution offered by Google. This template dashboard will form the basis of focus group community consultations to take place in the first quarter of 2023.

2. Issues for Community Consultation and Input

2.1 Technical Roadmap Consultation (WP1)

The BAD project proposal envisages an annual cycle of publishing a draft roadmap and consulting with the community on priorities. Our goal is to have as simple and efficient a system for this consultation as possible, supporting input from both technical and non-technical stakeholders which can easily be incorporated into the project's project management systems. In the medium term we would expect new developments to be proposed and prioritised by the community.

On an annual cycle we would provide a document similar to this current document detailing proposed work, options, and specific issues requiring community input. This document would be accompanied by an input mechanism (such as an online form or survey), and a publicly available representation of the previous, current, and proposed state of the project management system.

Many other organisations have opted to use a Kanban-style project board such as via Trello to provide a public facing version of the technical roadmap. Our current thinking is that this risks duplication of effort or a disconnect between internal project management systems and the public facing information. Our goal is simplicity and transparency so we initially opted in this cycle to provide direct screenshots from our internal project management system.

The Questions Posed for Community Feedback Were:

1. Are there processes and systems that have worked in other contexts that our team should be aware of? What are they?
2. What is the best way for us to engage the community in helping us to identify key technical priorities for the BAD project? We are especially interested in suggestions about how we can make this process meaningful for stakeholders with a high level of technical knowledge; as well as stakeholders who care about what the service can deliver but who may not feel confident about commenting on technical details.

There was limited specific feedback on the approach for community engagement, although interest was expressed in being involved in the development of documentation. One aspect that will require further work is understanding the different audiences for technical documentation and community feedback. In particular the original draft roadmap assumed some prior knowledge, that may not be appropriate for those commenting from the perspective of optimising the ecosystem (as opposed to partners or potential partners with specific functional needs).

Similarly there were questions about how the BAD project relates to other projects, suggesting a value in including more contextual documentation in future rounds of consultation. This naturally has to be balanced with respect to the density of information.

With respect to modes of community engagement there was interest in community discussion and groups as a forum for prioritising:

In the previous project there were some group discussions with the community, where we gave our ideas/approach about specific topics. Might it be interesting to organise something like that again once a year or so? Where you could also show the progress of what you have developed so far and explain it in a not too technical way ;)

In parallel with the consultation we also investigated ways to provide more directly live information from our Jira Project Management system. This has led to us adopting functionality that will allow the [live version of tasks in the roadmap](#) to be shared publicly for those with that level of interest, ensuring that this info is kept up to date.

2.2 Access Control Systems (WP2)

The BAD technical team strongly recommends that resources be allocated to implementing an industry standard access control system. The existing manual systems for user access management used in the Pilot project are not scalable and do not provide for a high quality user experience.

Specific issues with the current system (Kibana dashboards)

1. New users must be centrally created, which is not scalable or cost effective in the medium term.
2. Password resets (for forgotten passwords) need to be actioned manually, leading to delays in access and loss of productivity for users.

The Questions Posed for Community Feedback Were:

We request community support to prioritise the access control system work. This will involve

- standardising user roles to enable partners to manage system users and control access to their dashboards, and
- implementing a third party “identity provider” and integrating it into Elastic Search to enable a password reset option.

3. Do you support the prioritisation of the access control system work? If not, why?

There were no objections raised to prioritising this work, though the project team have since prioritised development of a BAD template dashboard in Looker Studio, which will now be the project used to produce BAD project dashboards. The Kibana dashboards for existing project partners will be migrated to Looker Studio (see the University of Michigan Press Looker Studio [dashboard here](#), compared to their [Kibana dashboard](#) developed during the pilot project), and new partners will be onboarded to Looker Studio Dashboards.

2.3 New Data Source Prioritisation (WP1)

Each data source (generally a hosting platform) requires that a “telescope” be built to ingest usage data specific to the platform. Each new telescope requires development, testing, validation, and integration to bring data together, plus revisions to standardised dashboards to incorporate the new data. We expect to be able to add two new data sources in the 2022-2023 technical cycle of the BAD project.

Potential data sources that have been identified are (in alphabetical order with no particular prioritisation):

- [EBSCO](#)
- [Fulcrum](#) (IRUS)
- [Project Muse](#)
- [ProQuest](#)
- [SciELO books](#)

The Questions Posed for Community Feedback Were:

4. Are there other important data sources that have been missed?

5. How would you prioritise the incorporation of these data sources?

One issue that was raised with respect to these questions was the need for more information, specifically on how many OA scholarly monographs does each platform host and whether they make books immediately available to readers or through intermediaries such as libraries. Our initial focus is on platforms that provide free and immediate access to readers but this was not entirely clear.

Only one comment was received thus far on specific prioritisations and that was that Project Muse is important to the commenter. Arguably Muse is one of the most common platforms

offering direct access to users for small presses that we do not already cover, so it does make sense to prioritise.

The project team and BAD project Advisory Board have since confirmed the two new data sources to be incorporated in year 1 are Fulcrum (IRUS) and MUSE.

2.4 Standardisation and Partner Onboarding (WP2, WP3)

A key part of the current project is to provide standardised and simple systems that enable the user community to scale rapidly. This involves both simplifying and clarifying the requirements for partners to onboard into the analytics service, and ensuring that the underlying systems are robust and consistent to reduce maintenance requirements and costs.

2.4.1 Standardising Partner Onboarding (WP3)

The primary output from this package of work will be a documentation guide for new partners that details:

- the Dashboarding Service's requirements for data quality and completeness;
- provides detailed and practical information about the steps that need to be taken to make data available to the Dashboard Service;
- what usage data sources are currently supported and offered;
- the choices partners may make in fields to visualise, and dashboard preferences.

The Questions Posed for Community Feedback Were:

6. Do you want to be involved in helping us to develop and improve the documentation and systems that are used to onboard new partners?

There was interest expressed in helping to develop and improve the documentation. This is planned as part of the next cycle of partner onboarding and will be a continual process with opportunities for feedback and criticism as the systems evolve. The partner onboarding documentation may be accessed from the BAD project Zenodo community: [Data Onboarding Background](#) and [Data Onboarding Questionnaire](#).

2.4.2 Standardising Data Ingest Pipelines

Through the development of the pilot systems some inconsistencies in process and naming have arisen. This package of work will ensure that data ingest points (such as email addresses) are consistent, naming conventions are standardised and overall ingest pipelines are simplified and made more robust. Stakeholders with technical experience are invited to examine the [code base](#) and [technical documentation](#) and are encouraged to lodge an issue in [Github](#) for suggestions for improvement or recommendations.

3. Detailed Technical Work Plans April 2022 - March 2023

This section details the work plans documented in the BAD project's project management system, Jira. Epics and tasks are labelled as ('BAD1', 'BAD2' etc). The current status of the work plans (e.g. 'to do', 'in progress', or 'done') are captured in the live [Book Analytics Dashboard - Technical Priorities 2022 - 2023](#).

3.1 Work Package 1

3.1.1 BAD-2: Technical Roadmap Consultation

- BAD37: WP1 - Technical roadmap - publish draft technical roadmap
- BAD38: WP1 - Technical roadmap - capture community feedback
- BAD39: WP1 - Technical roadmap - finalise initial priorities and share with community

3.2 Work Package 2

3.2.1 BAD-3: Workflow Standardisation

- BAD41: WP2 - Collate/standardise data collection processes
- BAD43: WP2 - Access control systems - integrate identity provider to ES
- BAD163: WP2 - Access control systems - standardise partner management role
- BAD397: WP2 - BAD Template Redesign - DataStudio/Looker
- BAD137: WP2 - Standardise data ingest pipelines - naming
- BAD210: WP2 - Standardise data ingest pipelines - workflows
- BAD162: WP2 - New data sources - develop telescopes
- BAD125: WP2 - System maintenance - Telescope updates
- BAD161: WP2 - System maintenance - Version upgrades

3.3 Work Package 3

3.3.1 BAD-4: Partner onboarding Phase 1

- BAD48: WP3 - Create partner onboarding handbook and documentation
- BAD147: WP3 - Partner onboarding - Identify potential new partners
- BAD45: WP3 - Partner onboarding - Onboard partner data
- BAD15: WP3 - Partner onboarding - Setup account

Official project title: OA Book Usage Analytics for Diverse Communities - A Demonstration Project

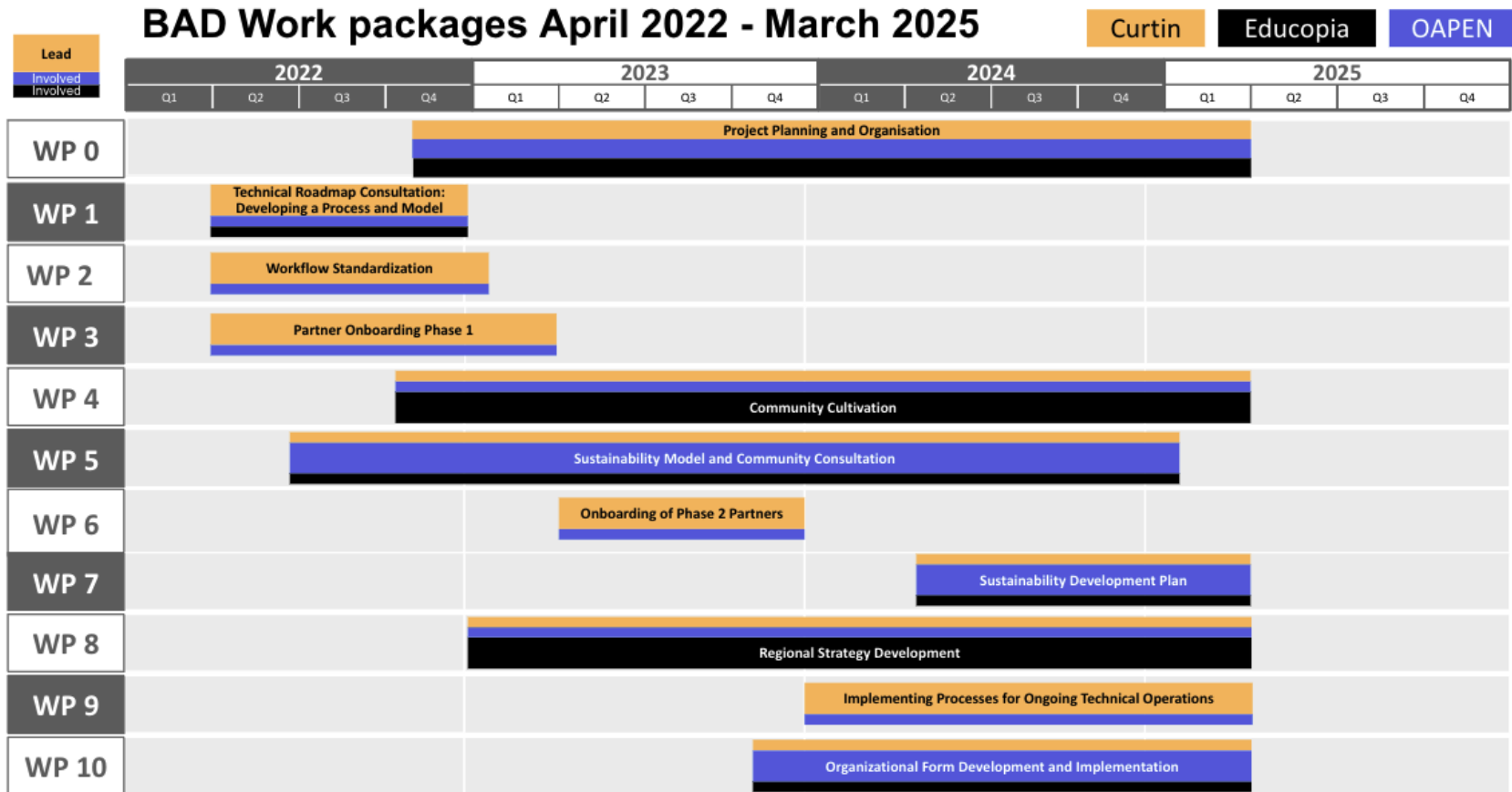
Short project title: Book Analytics Dashboard (BAD) Project

1 April 2022 - 31 March 2025

- **BAD46: WP3 - Partner onboarding - Deploy dashboards and access**
- **BAD47: WP3 - Develop workflow to coordinate onboarding**

Appendix A

BAD Project Gantt Chart April 2022 - March 2025



Updated: 16 June 2022