The Jorum Paradata Enhancement Project: Integrating Discovery and Usage Analytics for Open Educational Resources

Richard Jones, Cottage Labs, richard@cottagelabs.com
Sarah Currier, Mimas, sarah.currier@manchester.ac.uk

Abstract: Jorum is a national open repository service to UK post-16 education. Funded by Jisc, and run by Mimas at the University of Manchester, Jorum enables sharing and discovery of open educational resources (OER). Mimas and Cottage Labs recently worked together to take two prototype interfaces for Jorum (one providing discovery and the other providing statistics/paradata) and rationalise them into a production-ready system. This produced a novel and flexible user interface which allows end-users to switch between discovery and reporting interfaces with ease, meeting the many and complex needs of the Jorum community.

Jorum [1] is a national open repository service to UK post-16 education. Funded by Jisc [2], and run by Mimas [3] at the University of Manchester, Jorum enables sharing and discovery of open educational resources (OER). Specifically, Jorum provides individuals, communities, organisations and institutions in UK higher and further education with a platform for hosting and showcasing their OER, and gives free global access to these OER, all of which are shared under Creative Commons licences allowing reuse, and - where specified - repurposing.

Since moving from a proprietary learning object repository platform in 2010-2011 [4], Jorum has been built on DSpace [5], which was (and continues to be) modified by the Jorum development team to support the requirements of sharing and hosting OER. Until early 2013, discovery of resources in the repository was provided through a native DSpace interface. However, work began at the end of 2011 [6] to build a new web front end for discovery of Jorum content, using one of the prototype DSpace REST APIs.

In addition, from 2011 on a pressing requirement emerged from Jorum's user community and funders for provision of usage data and analytics to support and encourage a culture of sharing OER for teachers, institutions and national organisations. There was also a growing demand for an API to accommodate both content discovery and usage data reporting. In 2012 Jorum began development on a Dashboard Beta for usage data [7], and worked closely with the Learning Registry [8] [9] [10], a US-based project developing a schema-less database solution for sharing paradata.

Paradata are usage data about learning resources that include not just quantitative metrics (e.g., how many times a piece of content was accessed), but also pedagogic context, as inferred through the actions of educators and learners [11].

The eager response from Jorum's community to this R&D work further emphasised the need for improved access to Jorum paradata, and Jisc provided enhancement funding to extend Jorum's core offer to support these and several other emerging requirements [12]. The Jorum Paradata Enhancement Project brought external developers Cottage Labs [13] to work together with Mimas to take the existing repository infrastructure, and the prototype discovery and paradata interfaces, and rationalise them into a single coherent user experience which fulfilled the many complex use cases of the OER repository.

The project began with a review and definition of a set of user stories for consumers and producers of OER. These user stories are relevant to repository managers, institutional senior management, software developers, teaching and support staff producing and using OERs, and so forth. There were in excess of 40 user stories in total; the following are some samples:

- As a repository manager, I want all paradata for my repository's OERs in Jorum
- As a content seeker, I want to order search results by popularity: views, shares and downloads
- As an institutional manager, I want to filter paradata by institution and by subject area
- As a third-party software developer, I want to connect to an API to extract paradata to present in my local environment(s)
- As an individual OER author, I want all stats for my OERs in Jorum
- As a content seeker, I want to compare stats across N OERs

From our larger set of user stories, we carried out a prioritisation exercise and derived a set of technical requirements which were most critical to Jorum and its community. These requirements were:

- Jorum must register all incoming records (including manually entered ones, OAI-PMH harvested ones, and any other mechanism of ingest) with its origin (e.g. repository) and its affiliation (e.g. University of Manchester). This is so that end-users will ultimately be able to generate reports with appropriate constraints ("show me the paradata for everything which came from my institution's repository").
- The user interface should be able to provide a complete set of statistics for a single item as a standard view (i.e. "these are the stats and paradata for item X"). This is the most basic and essential paradata requirement, and once developed can be used to extend to comparative and aggregate statistics, as required by some of the other user stories.
- Be able to show the aggregated set of statistics for a set of OERs. This builds on the previous requirement, allowing satisfaction of institutional needs to view statistics on, for example, all of their items in Jorum.
- It should be possible to select OERs for reporting via a standard search mechanism. For example, a project manager could search Jorum by the project tag or acronym to discover all OER outputs of the project partners, and get a report on usage of this content. This ties the paradata aspects of the system to the discovery aspects - effectively end-users can use the discovery interface both for discovery and for

building sets of OERs to be reported on. This provides a pleasing symmetry between the two distinct sets of users stories around discovery and paradata while ensuring appropriate re-use of the underlying software.

Based on these requirements, the project delivered a new system architecture to replace the existing prototype systems, providing equivalent and better functionality. This combined both discovery and paradata usages into a novel, unique and user-friendly interface, allowing the end-user to switch seamlessly between use cases, and providing a coherent user experience. The system is also be extensible such that we can go on to develop support for more user stories over time.

Figure 1 shows the target system architecture, combining a search index to provide discovery and paradata information to a flexible front-end, and then utilisation of the DSpace REST API [14] for retrieving details about specific OERs upon request. By combining discovery with paradata and a REST API for content retrieval with an index which supports search and faceted-browse, we can therefore support this novel kind of interface to OER, meet the user stories, and be confident that we can extend the functionality to more advanced user stories in the future.

This presentation will begin by introducing Jorum, the concept of paradata, and the motivations behind the project. We will then present the user stories that motivated us; including those which we did not attempt to support in this initial project, but which are indicative of the direction of the Jorum OER service.

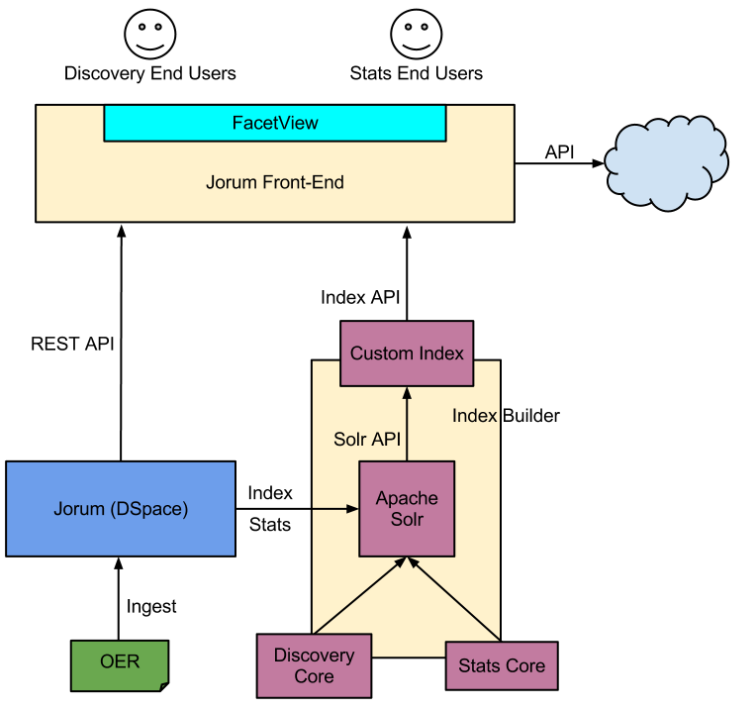

Figure 1: target system architecture

We will then go on to briefly outline the technical approach, how information is extracted from

DSpace and placed into an intermediate indexing layer, and then ultimately on to presentation through a custom user interface based on JavaScript and Ruby-on-Rails.

A brief demonstration of the user experience will be given, including the discovery interface using faceted browse and search. We can then show the construction of paradata reports based on the result set obtained through the discovery interface. This includes tabular, graph-based and map-based presentations of the data, all of which can be downloaded in appropriate formats for inclusion in management reports.

We will conclude the presentation by considering the user stories which were not covered by this first phase of the Jorum Paradata Enhancement Project, the additional side-work that had to be done to support the project (e.g. metadata enhancement), and the vision for the future of Jorum using this technology, which includes offering support for institutions, communities and organisations to create their own branded and tailored “Jorum Powered” [15] repository interfaces.

[1] Jorum: http://www.jorum.ac.uk
[2] Jisc: http://www.jisc.ac.uk
[3] Mimas: http://www.mimas.ac.uk
[4] Jorum History 2002-2011: http://www.jorum.ac.uk/about-us/Developing%20Jorum/2002-2011
[5] DSpace: http://www.dspace.org
[6] Coming improvements to the Jorum user experience [blog post, 12 December 2011]: http://www.jorum.ac.uk/blog/post/20/coming-improvements-to-the-jorum-user-experiece
[7] Collecting statistics just got a whole lot sweeter [blog post, 29 June 2012]: http://www.jorum.ac.uk/blog/post/38/collecting-statistics-just-got-a-whole-lot-sweeter
[8] The Learning Registry: http://www.learningregistry.org/
[9] Jorum OAI-PMH data published [blog post, 14 February 2012]: http://jlernexperiment.wordpress.com/2012/02/14/jorum-oai-pmh-data-published/
[10] Dunking your cake in a tasty source of data [blog post, 13 April 2012]: http://jlernexperiment.wordpress.com/2012/04/13/dunking-your-cake-into-a-tasty-source-of-data/
[11] Paradata: http://en.wikipedia.org/wiki/Paradata_(learning_resource_analytics)
[12] It’s all coming together: Jorum’s 2012 Summer of Enhancements [blog post, 30 August 2012]: http://www.jorum.ac.uk/blog/post/48/it-s-all-coming-together-jorum-s-2012-summer-of-enhancements
[13] Cottage Labs: http://cottagelabs.com
[14] Mimas fork of DSpace REST API: https://github.com/Mimas/dspace-rest
[15] Jorum Powered: http://www.jorum.ac.uk/powered