

# Tell me what I need to know: generating reports from your repository

Sébastien François and Justin Bradley  
EPrints Services  
University of Southampton, UK  
{sf2,jb4}@ecs.soton.ac.uk

## **Introduction**

Institutional Repositories (IR) have been a key component of the Open Access initiative over the last ten years. They allow the dissemination of an institution's research output and a myriad of exchange protocols and mechanisms have emerged throughout the years to help with this task. However IR can provide more information beyond *just* storing the full-text documents and their attached metadata.

A key use of the information held within the repository is to generate reports on its holdings and usage. Reports can be used within institutions to help make internal decisions as well as being used to feed external systems which have strict reporting requirements. A growing number of such systems are being set up by research councils and by governments to assess the quality of the research that they fund, such as the Research Outcomes System in the UK and OpenAire in Europe.

We will show that the repository is a key component in making internal and external decisions on an institution's holdings with the use of reports.

## **Usage statistics & internal performance reviews**

IRStats2 will be released in the first quarter of 2013 and will be introducing some new concepts when compared to the first version. Beyond the simple visual make-over, IRStats2 is re-enforcing the notion of reports which become the main component of the package. While with version 1 the repository managers were expected to produce the reports themselves, having to choose what they need to see among many different options, version 2 alleviates this task. The default reports present users with a quick summary of the usage statistics of their repository which they can then share amongst colleagues and research offices. It is also possible to add and edit reports to make sure that the right information, in the right form, is presented to users.

The software has been extended to generate statistics over *any* data source within the repository, it is not exclusively about download figures anymore. For example it is now possible to:

- report on the percentage of open-access and full-text resources within the repository,
- monitor the rate of deposits over time,
- review the number of new users signing-up on open-registration systems,
- analyse the editorial review queue and see the number of items going in and out.

This list is not exhaustive and we expect repository managers to express their own ideas when the software becomes available.

IRStats2 makes it easy to include graphs and figures anywhere on the repository pages and to embed data on other systems, for example on staff profile pages where the researcher's top publications or download graph may be embedded. Furthermore raw data can be exported in the popular formats: XML, JSON and CSV to be ingested by external systems. Work is under way in partnership with the University of Glasgow to automatically include some statistical data into their staff performance reviews.

The basics of statistical reporting have not been forgotten: users may view the data at different levels of granularity, by gathering the figures in sets, for example by author, by item type, by school or, in fact, by *any*

piece of relevant metadata in the repository. Users can then select on which time scale they want to see their stats.

By being able to generate statistical data from any data source and to present it in reports, IRStats2 aims at helping repository managers to review the state and usage of their repository.

### ***Funder reporting***

Reporting to external agencies has stricter requirements than internal reporting. This is clearly the case for the Research Outcomes System (ROS) which collects the outputs, outcomes and impact of research funded by the UK Research Councils. Submission to ROS is mandatory in the UK.

Reporting to such systems raises challenges for institutions: (a) they need to include this in a suitable workflow, (b) they need to make sure they have captured the data in the right way to allow correct reporting.

EPrints Services is involved in two distinct projects where the institutions' Principal Investigators are already adding their outcomes to ROS in addition to the IR. Therefore there is already a duplication of effort and overlap of data between ROS and the IR. We propose to use the IR as a central entry point for the outcomes and to integrate the repository directly with ROS so that manual submission of the outcomes into ROS is no longer required.

Such integration is needed to enhance the repository's functionality to support the ROS requirements. On one hand the IR's metadata schema needs to be reviewed to match ROS's fields and validation process. This will likely involve bringing grants and funders information to the foreground. This data will need to be browsed and searched in the same way that research outputs can currently be consulted in a repository. On the other hand the repository will need to be able to present the captured data in the format(s) expected by ROS. This will be achieved by either importing a CERIF XML document to ROS or by letting ROS harvest the IR via the standard OAI-PMH interface.

Once repositories start using the grants and funders data it will be possible to report on the percentage of resources that meet the ROS data requirements. The report will also be able to highlight missing or invalid data within the resources and point the repository administrators to those items.

### ***Governmental reporting and beyond***

2013 is an important year for the UK higher education institutions: it's the final leg of the Research Excellence Framework (REF) which will assess the research outcomes of all the UK HEIs over a six-year period.

Similarly to ROS reporting, REF defines strict rules for the submission of the research data: institutions are required to submit up to four outputs per researcher as well as extra metadata about the researcher and the outputs. Since the IR is the usual place where the outputs are stored, it is obvious that repositories will play a central part in submitting for REF since most of the information is already available there. However the IR lacks mechanisms for:

- selecting research outputs,
- managing the users who can make selections and who can edit the extra metadata,
- generating REF-compliant reports.

EPrints Services have created a REF extension which tackles the above issues.

The REF plug-in first extends the standard user roles found in a repository. REF introduces the notion of *Champions* who are advanced users in the system and who can manage other people's selections as well as consulting the reports. Even so the Champions may manage the output selection, this task is usually delegated to the researchers themselves who must then be distinguished, in the system, from other staff who

are not making any returns for REF.

New interfaces were added to the repository to help researchers to find their publications easily by matching their names to the publication's list of co-authors during the REF census period. Unique author identifiers were also used for this purpose on repositories which have this functionality enabled.

In terms of reporting, the REF plug-in offers a summary page where errors and warnings are showed for each researcher's output. It is then easy to fill the missing data or to correct any invalid information. By its modular nature the REF plug-in can be extended to incorporate bespoke reports and a number of institutions have asked to present the information in a form which is useful to them for their review meetings. As it is often the case, users find alternative ways to use software to fit their needs and, in the case of reporting, keeping the modularity of the plug-in was an important factor of its design.

Finally, following the REF data requirements, the plug-in can then export all of the submission data in the desired formats. The generated output can then be directly uploaded to the REF submission system for final reviewing.

Although this presentation will focus on specific UK-based examples we intend to highlight general principles and practice for making effective use of your repository's holdings for internal and external reports.

### ***About EPrints Services***

EPrints Services is a UK-based repository software developer and service provider. We help many repository administrators around the world to get the best from their EPrints repository. In 2013 we are focusing on three core challenges -- Validation (*does my repository accurately reflect the published output of my institution?*), Reporting (*covered in this presentation*) and Dataset management (*how can your repository present and disseminate your research data?*) -- and are working closely with our user community to meet these challenges.