

EPrints as a data repository at UWE

We intend to extend our successful EPrints research-output repository to include research data.

We chose EPrints because

- We use it already
- It won't cost us any more than we are already committed to
- We have the skills needed to extend and repurpose it quickly
- Other institutions have already successfully used it to publish research data

Things to think about

The most important choice for the implementation is whether to include research data in the research-output repository itself or whether to segregate research data in its own EPrints instance. The most straightforward option to implement is to simply include research data with research output. However, it may not be desirable to mix the two, and one could argue that keeping the two separate helps with managing processes for REF and the like. Also, research data may benefit from slightly different, say, upload configuration, from research outputs, so keeping it in a separate instance would give us more flexibility. It would also allow us to customise and simplify the metadata-entry screens, which would benefit users.

The metadata schema already used in UWE EPrints is comprehensive and is unlikely to need much work to be fit for research data. We've already added to the default set of EPrints fields, including UWE-specific ones such as research funder and research project, which are likely to be needed for the research data project.

EPrints is likely to need more storage space to be able to handle research data sets for the project. Our EPrints storage is currently specified for research outputs, which tend to be documents of a few megabytes. Research data is likely to be larger, and there is a limit on what EPrints can be expected to hold, probably around few gigabytes per project. For data which won't reasonably fit into EPrints, we intend to create a metadata-only record linked to our offline access procedures.

We have a small amount of project money to buy more storage. EPrints storage is held on UWE's storage area network, so expanding should be straightforward with no new hardware needed.

Implementation

Assuming we choose to create a separate instance for research data, we would need to

- Create the new EPrints instance
- Create a new web site address
- Install a new SSL certificate
- Clone the metadata, plug-ins and screens from the research output repository
- Customise the screens

- Test with users and iterate the customisation

We can expand the storage as a separate job.