

24x7 Presentation Proposal | Islandora Web ARChive Solution Pack

Nick Ruest (ruestn@yorku.ca)

Possible themes:

- Facilitation of reproducible research through access to data, workflows, and code
- Effective re-use of software, services, and infrastructure to support repository development

Islandora Web ARChive Solution Pack

Why?

We are now living in a reality where official records are born and disseminated via the Internet. Many institutions have a strategy in place for transferring official university records that are print or tactile to university archives, but not much exists strategy-wise for websites. What this project aims to do is streamline the curation, dissemination, a discovery of websites using a collection of open source tools and packages.

What is it?

The Islandora Web ARChive Solution Pack is yet another Islandora Solution Pack. This particular solution pack provides the necessary Fedora objects for persisting and disseminating web archive objects; warc files.

What does it do?

The basic iteration of the SP allows a user to upload a warc with an associated MODS form. Once the object is deposited, the associated metadata is displayed along with a download link to the warc file.

Current development work is focused on indexing individual warcs in Solr to allow for keyword searching in a given object (similar to the Islandora Internet Archive Book SP), dissemination of the warc objects in a stand alone version of the Wayback Machine, and an administration panel to aid in the automated accessioning of web sites.

An example of the basic SP is available [here](#) and the source code is available [here](#).