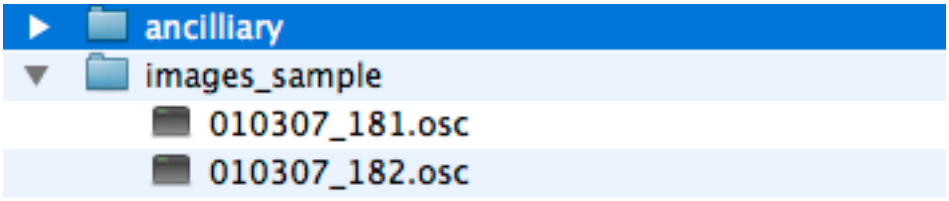
Steve Androulakis - 24 August 2009

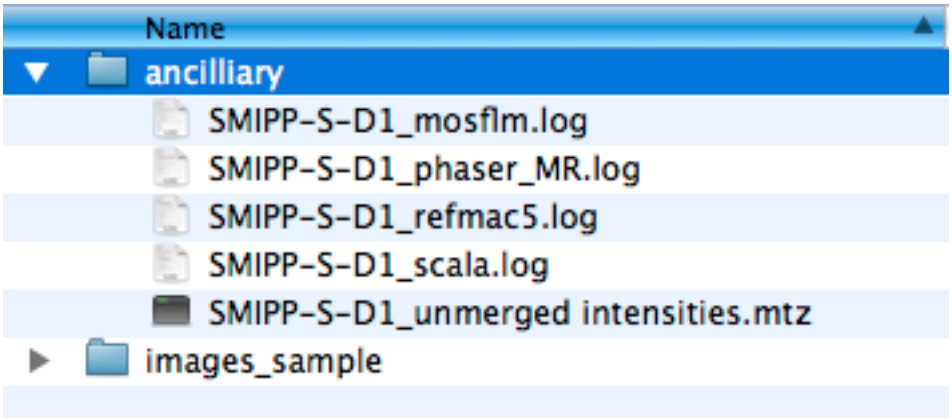
TARDIS Data Deposition Guide

Organise Data for Submission

There are two types of Datasets able to be published, raw and ancillary. Each set of raw diffraction images (img/osc/adsc etc..) should be in its own directory. Ancillary files are for log files, processed files, and anything else to be included with the published experiment.



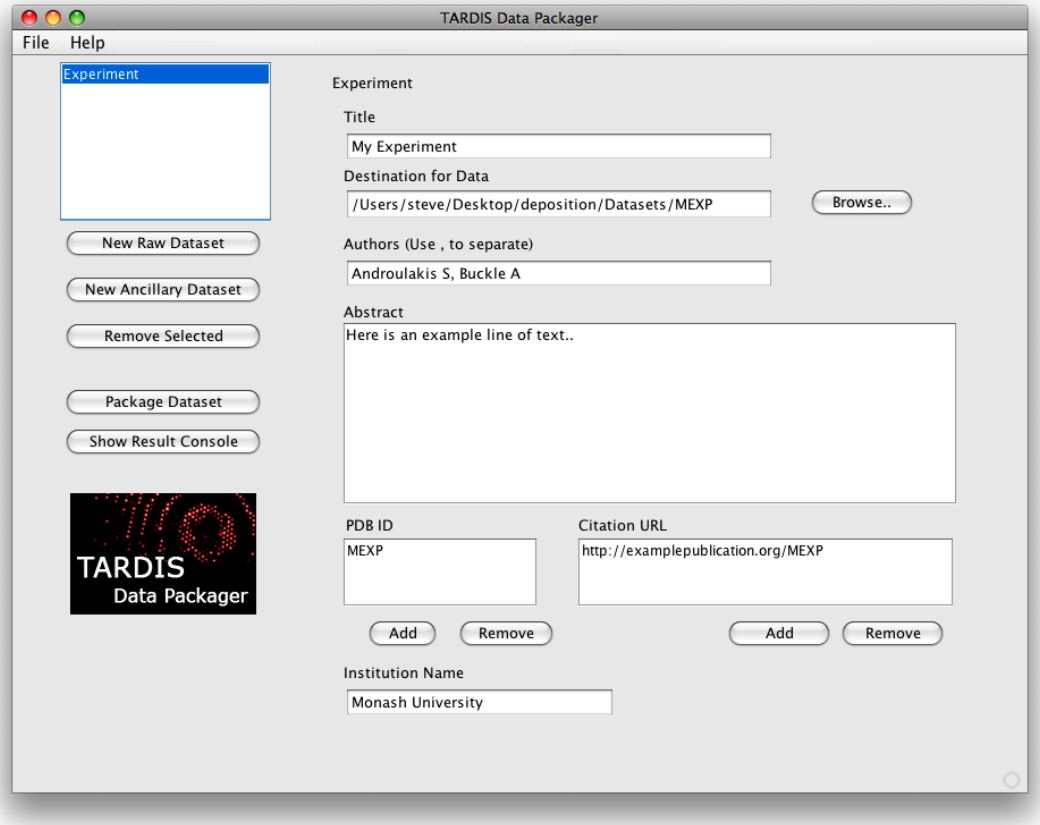
'images_sample' – a directory containing raw diffraction images



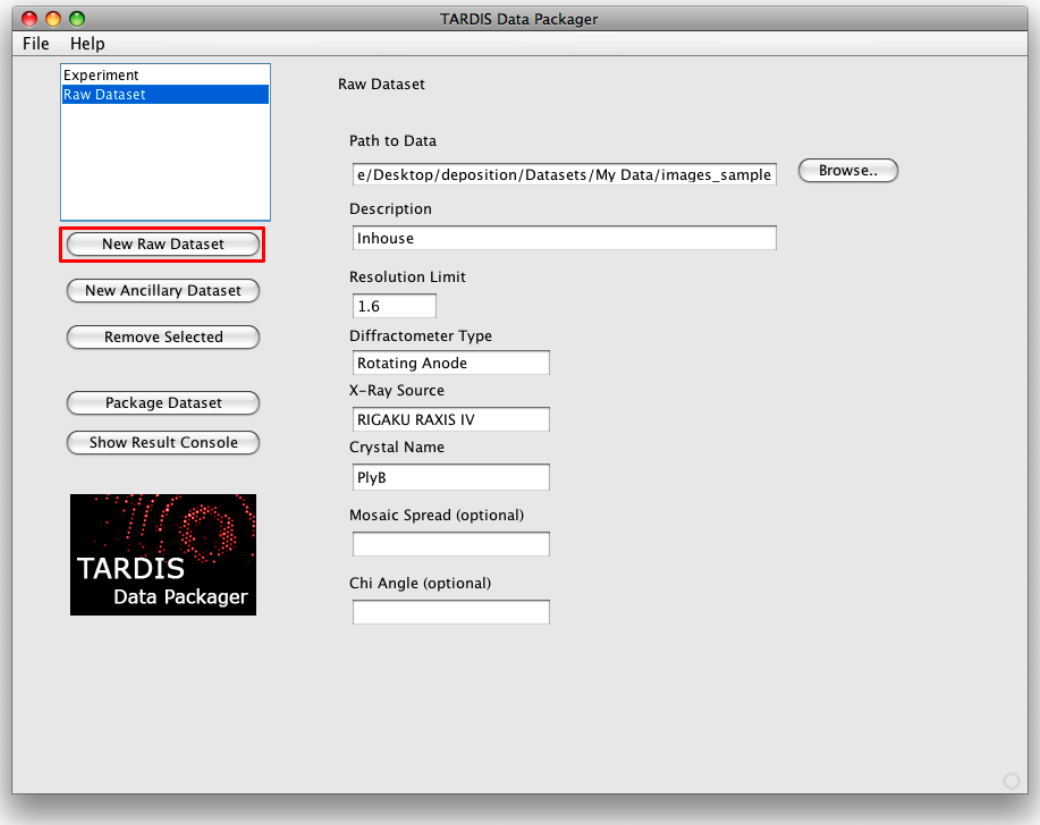
'ancillary' – a directory containing all other types of files to be included

Annotating Data

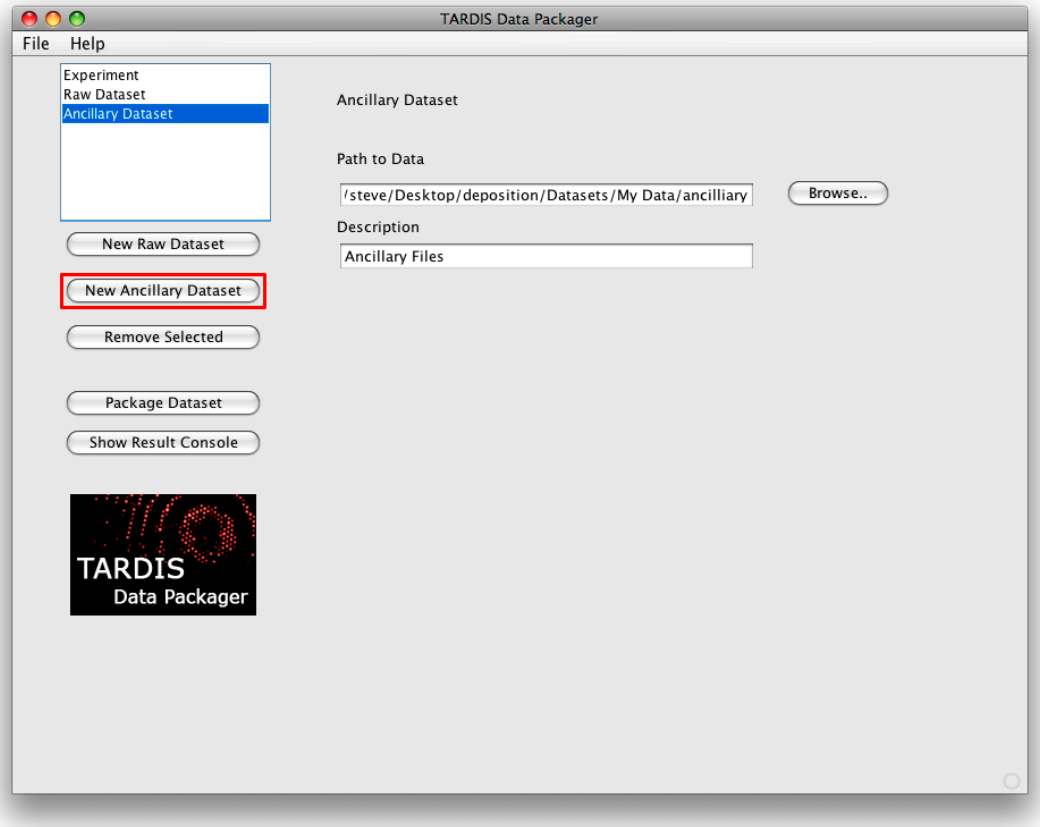
Download the TARDIS Data Packager program and run it. The initial screen gives users the ability to describe the overall experiment (a collection of datasets). Also important is the "Destination for Data" field. Choose an empty directory for the annotated data to be output to.



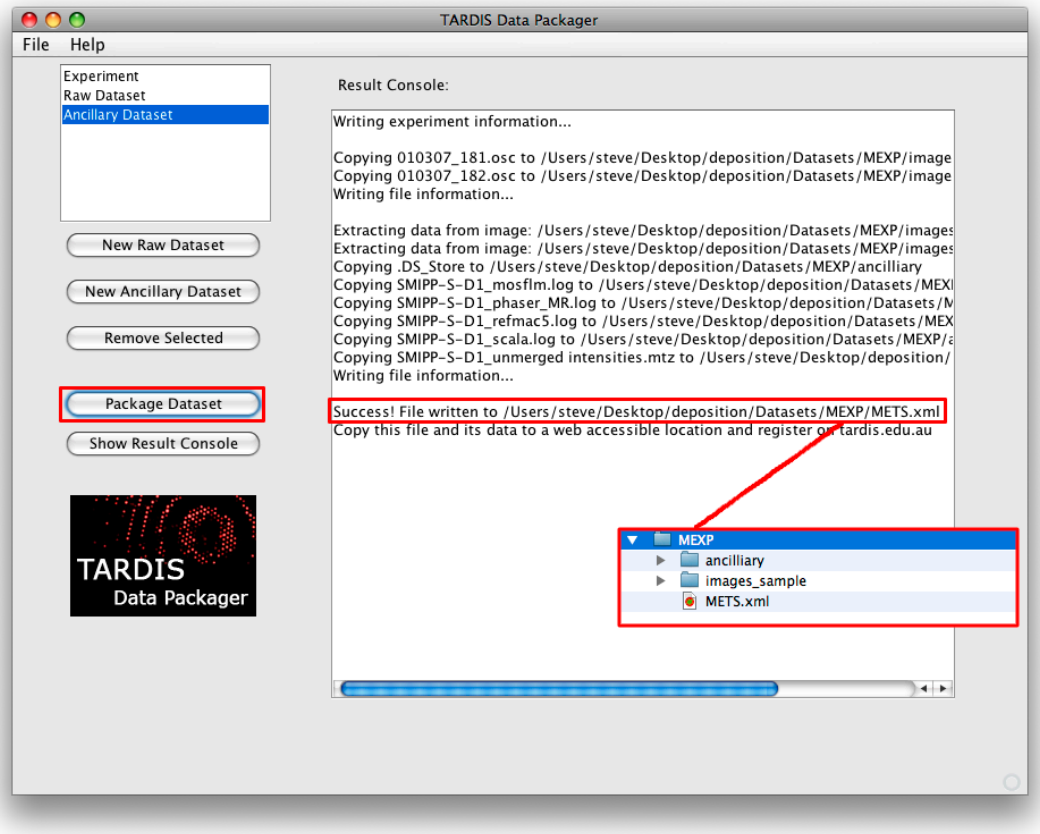
Click 'New Raw Dataset' to locate and annotate a set of diffraction images.



Click ‘New Ancillary Dataset’ to choose the path to extra files, and describe the set.



Click ‘Package Dataset’ to begin the annotation and organisation process. Metadata is automatically extracted from raw diffraction images, files are copied to their destination and an annotation file ‘METS.XML’ is created.

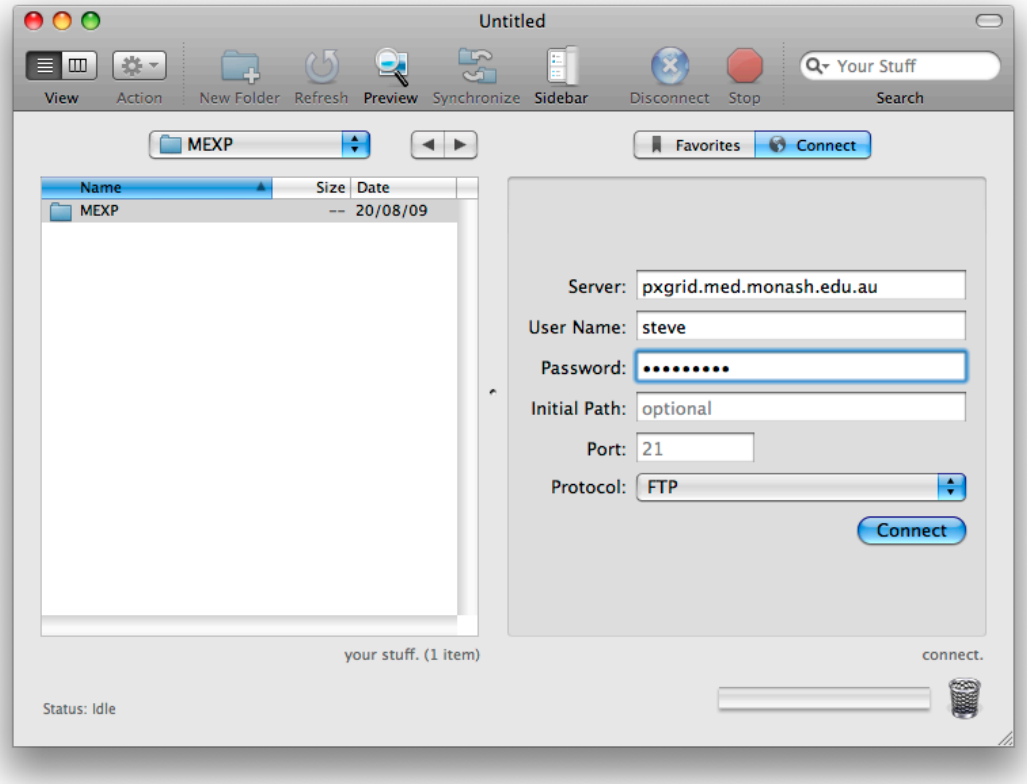


Files are copied along with the annotation file (METS.XML) to the destination path chosen in the experiment section. All of these files should be copied to a publically web-accessible location, so a URL can be created and submitted to TARDIS.

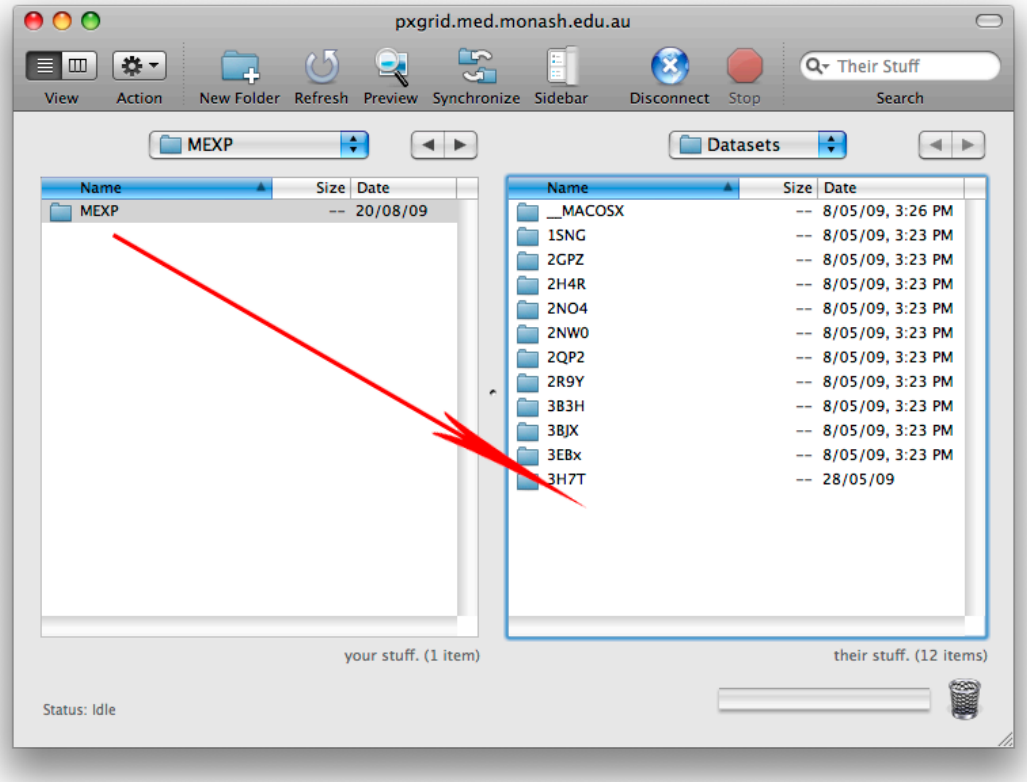
Copying Data To A Public Location

The copying of files to a web server can vary widely and a system administrator should be contacted for assistance. For example purposes, this section outlines how one might copy files via FTP to their web server.

Open an FTP client (Transmit for the Mac is pictured here) and connect to the web server that will store the files.

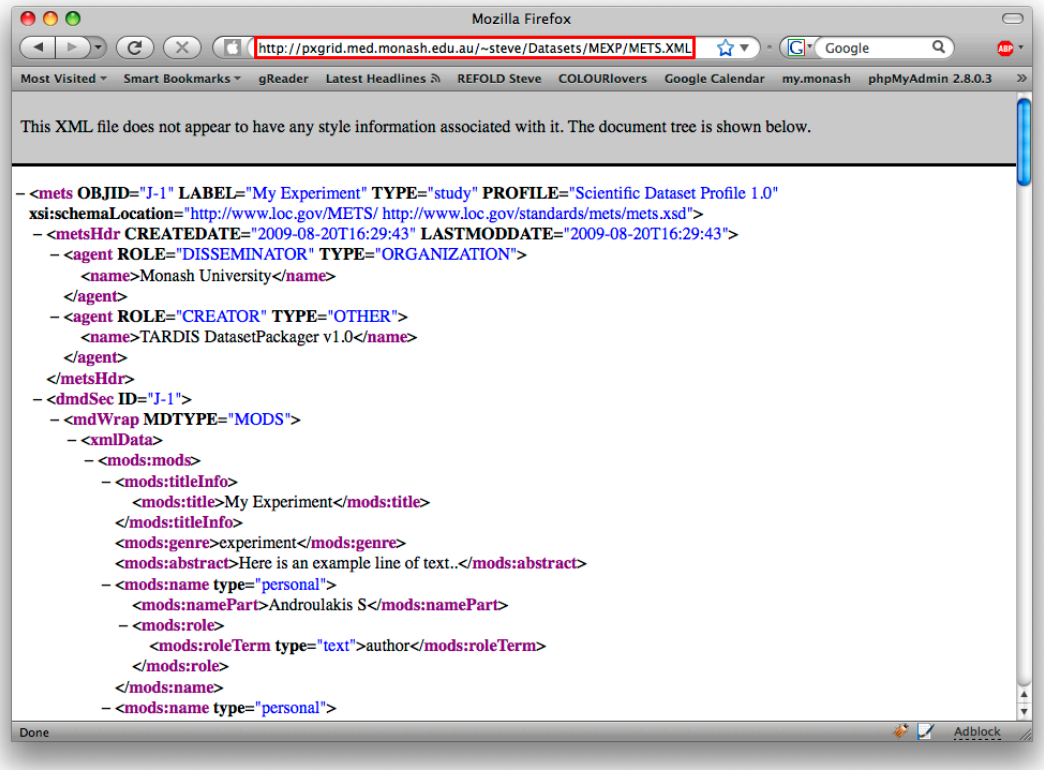


Copy the entire directory holding the datasets and the METS.XML file across to the server.



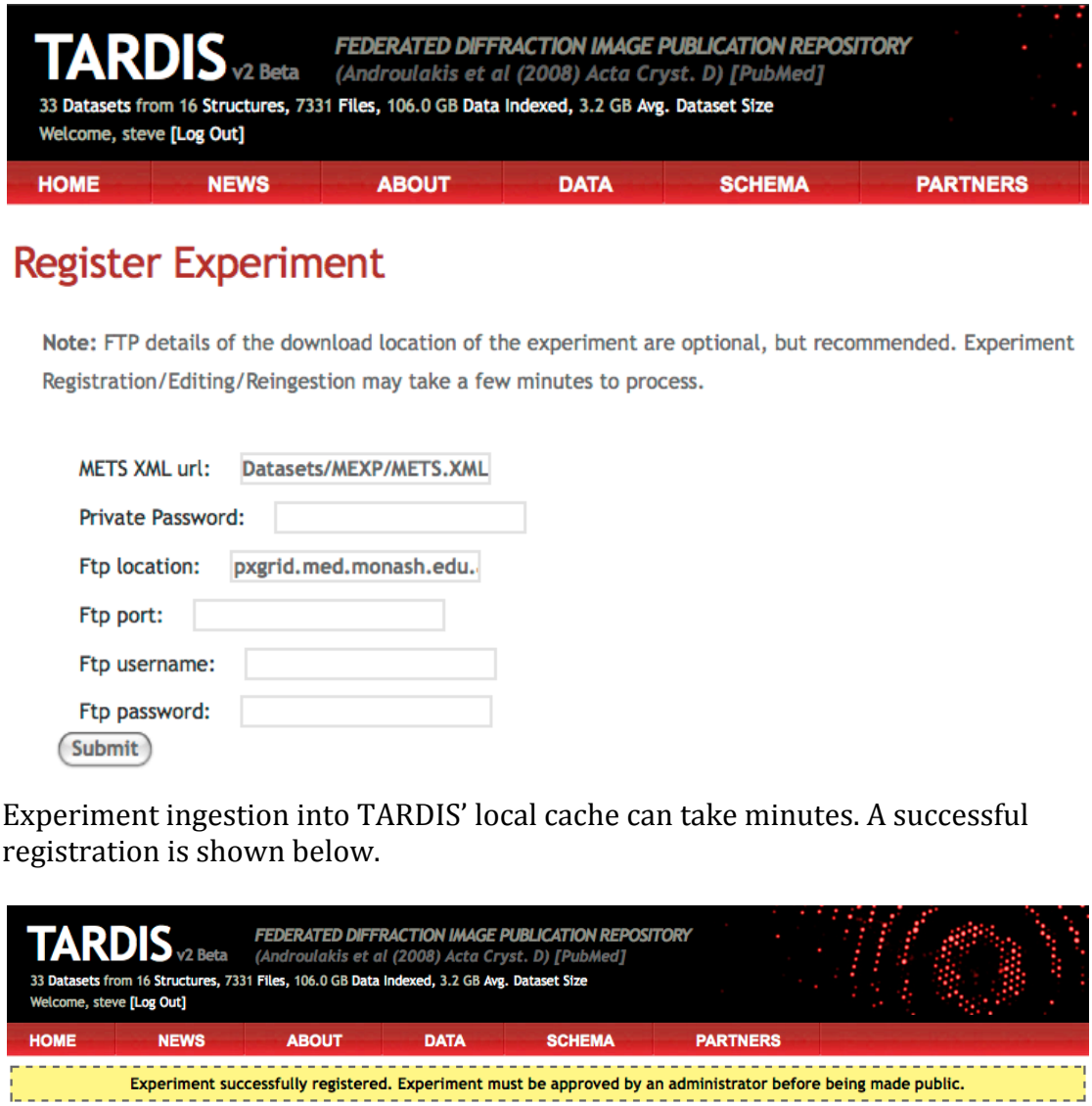
Register Data on TARDIS

Once on a web server, the data should be accessible via URL. To verify this, a good test would be to try and access the METS.XML file in a browser. A successful display of METS.XML in Firefox is shown below.




```
– <mets OBJID="J-1" LABEL="My Experiment" TYPE="study" PROFILE="Scientific Dataset Profile 1.0"
  xsi:schemaLocation="http://www.loc.gov/METS/ http://www.loc.gov/standards/mets/mets.xsd">
  – <metsHdr CREATEDATE="2009-08-20T16:29:43" LASTMODDATE="2009-08-20T16:29:43">
    – <agent ROLE="DISSEMINATOR" TYPE="ORGANIZATION">
        <name>Monash University</name>
      </agent>
    – <agent ROLE="CREATOR" TYPE="OTHER">
        <name>TARDIS DatasetPackager v1.0</name>
      </agent>
    </metsHdr>
  – <dmdSec ID="J-1">
    – <mdWrap MDTYPE="MODS">
      – <xmlData>
        – <mods:mods>
          – <mods:titleInfo>
              <mods:title>My Experiment</mods:title>
            </mods:titleInfo>
            <mods:genre>experiment</mods:genre>
            <mods:abstract>Here is an example line of text..</mods:abstract>
          – <mods:name type="personal">
              <mods:namePart>Androulakis S</mods:namePart>
            – <mods:role>
                <mods:roleTerm type="text">author</mods:roleTerm>
              </mods:role>
            </mods:name>
          – <mods:name type="personal">
```

This URL will provide as the basis for experiment registration. Log in to TARDIS and click ‘Data’ then ‘New Experiment’. Paste the full URL to METS.XML in the ‘METS XML url’ field. A password can be entered to protect the experiment from the public view. Optionally, FTP details can be entered for easy downloading of full datasets.



Experiment ingestion into TARDIS' local cache can take minutes. A successful registration is shown below.

Note: Data needs to be approved by an administrator before attaining a persistent handle or showing in the public experiment index.