Research Metadata Management at the Australian Synchrotron

Richard Farnsworth, Alistair Grant, Andrew Rhyder, Nick Hauser (Ansto)

Abstract

This paper details the approach the Australian Synchrotron is using, in collaboration with the Australian neutron source (OPAL), for the data and metadata management issues of a selection of instruments. The paper explores the data and user policies, describes the quantity and quality of data and demonstrates the way forward based on both existing and future directions in e-research, network communications, user proposal and material databases, portal technologies and integration techniques. This work is funded by an educational infrastructure grant administered by Australian National Data Services.

Note: OPAL is run by the Bragg institute, part of ANSTO

`Where is it?

Delivered by Andrew Rhyder for:

Richard
Farnsworth

Alistair Grant

Nick Hauser

Data Policy

- There is a growing trend from governments around the world that publicly funded research should result in publicly available data

- There are different attitudes in the scientific communities ranging from the astronomical community where data is publicly available immediately through to protein crystallography where even the association with an experiment may be secret

- Concept of an embargo period where the researcher has exclusive access to the data for a period

- ANSTO has a policy of a 3 year embargo. The Australian Synchrotron is developing its user agreement and looking towards something similar

Data Quantity

- The Australian Synchrotron operates nine beamlines producing around two to three TB of experimental data per day across a wide variety of disciplines from protein crystallography, medical through to the conservation and restoration of cultural objects and works of art. Last year (2009) over 500 groups conducted research at the Australian Synchrotron.

- The Australian Synchrotron expects to be producing at least eight terabytes per day as instruments produce more data and the number of beamlines increases.

MeCAT

- The objective of this project is to provide services to researchers to manage their experimental data and to provide data search and access to the broader research community. These services will provide better use and reuse of the data. The ultimate aim is to combine these services into a collaboration environment providing project teams easy access to experimental data, metadata, and analytical tools.

- Collaboration between ANSTO and the Australian Synchrotron

- Step towards using the same software in facilities and institutions around Australia

MeCAT – details

- Deliverables
  - The software: An extensible, searchable, web based meta-data catalogue.
  - Deployment: In use on the SAXS and IR beamlines at the Australian Synchrotron and six of the seven instruments at the Bragg institute
  - Producing a reference spectral database of different materials (e.g. paints, proteins, inks) for the Australian Synchrotron IR beamline

- Requirements
  - Compatible with the ARCS Data Fabric data repository
  - Facilitate harvesting by ARDC of public metadata using RIF-CS

MeCAT – Data sources

Collecting meta-data at each facility from:

- The instruments:
    - SAXS and IR beamlines at the Australian Synchrotron
    - Most of the neutron beam instruments at the Bragg institute

- The proposals database

- The scheduling database

- The control system:
    - EPICS at the Australian Synchrotron
    - SICS at the Bragg institute

- User entered data, e.g. electronic log books

What is TARDIS

Not this one, but

The Australian Repository for Diffraction Images System

TARDIS

- History:
  - The availability of raw data is extremely useful for the development of improved methods of image analysis and data processing.
  - Fostering the archival of raw data at an institutional level is one the best ways of ensuring that this data is not lost (laboratory archives are typically volatile)
  - TARDIS originally facilitated the archiving and sharing of raw X-ray diffraction images from the protein crystallography community.

- The MeCat project is extending TARDIS to create a general purpose store of scientific experimental data
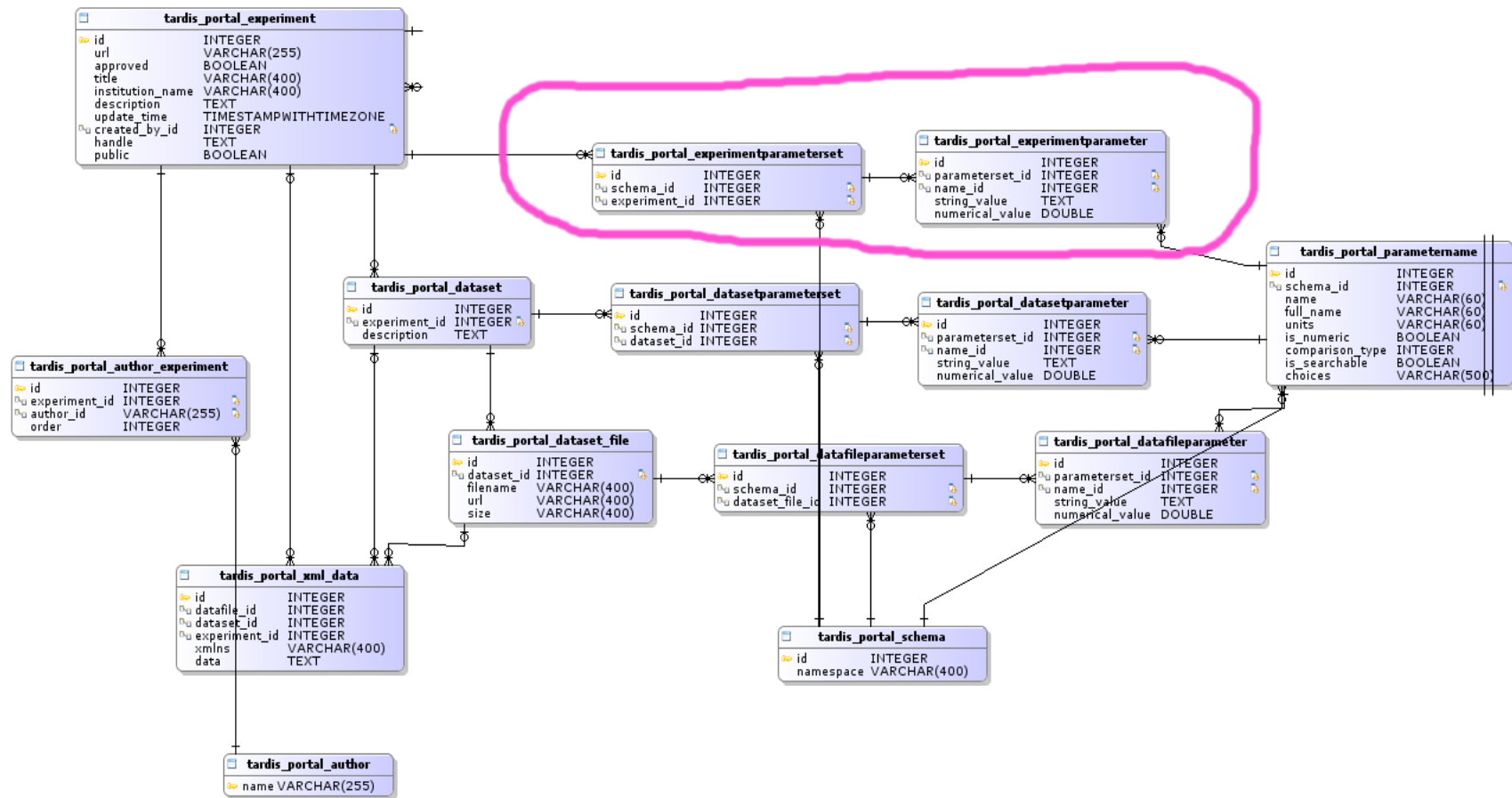
TARDIS – Open source

- The Australian Repository for diffraction imaging Systems is an open source project and can be found at:
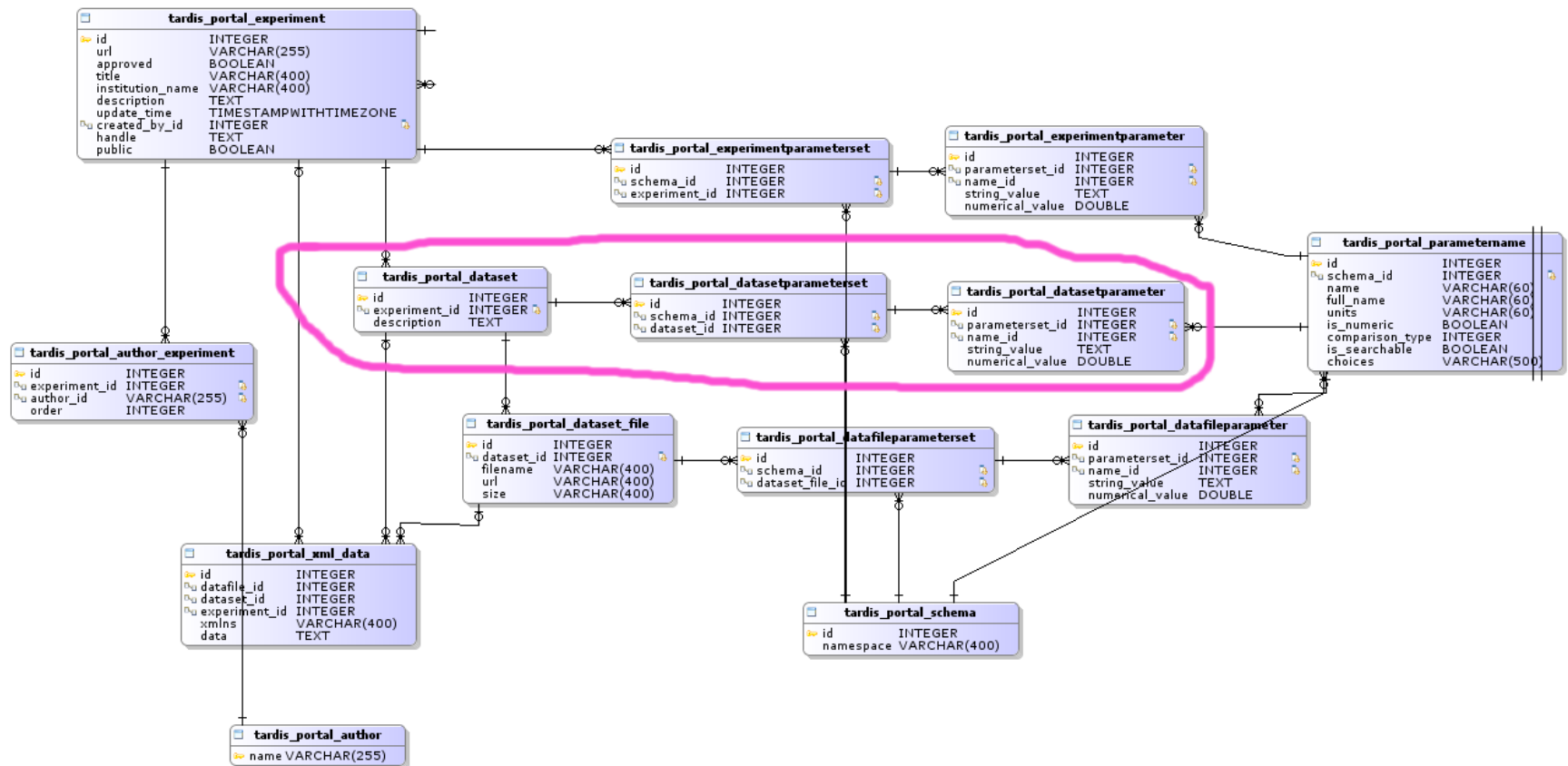
  http://code.google.com/p/mytardis/

TARDIS - Full Portal schema

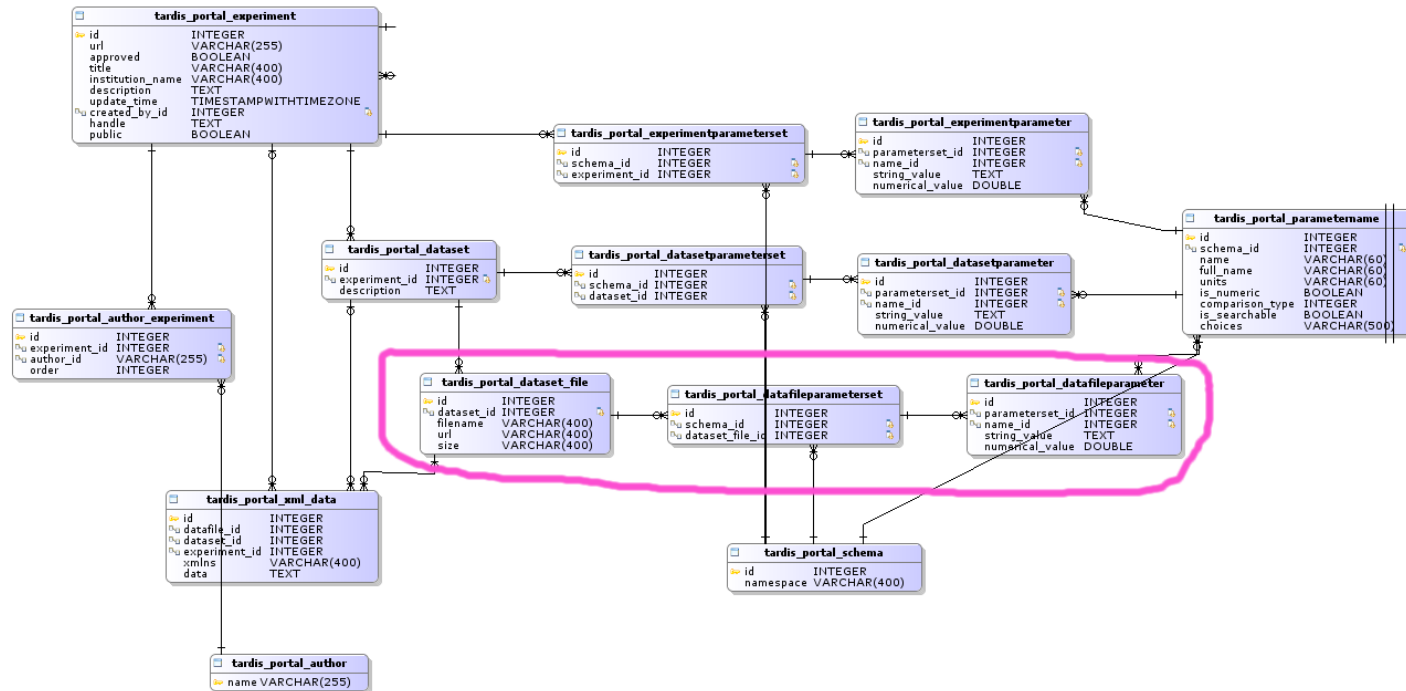# TARDIS - Portal schema – metadata model
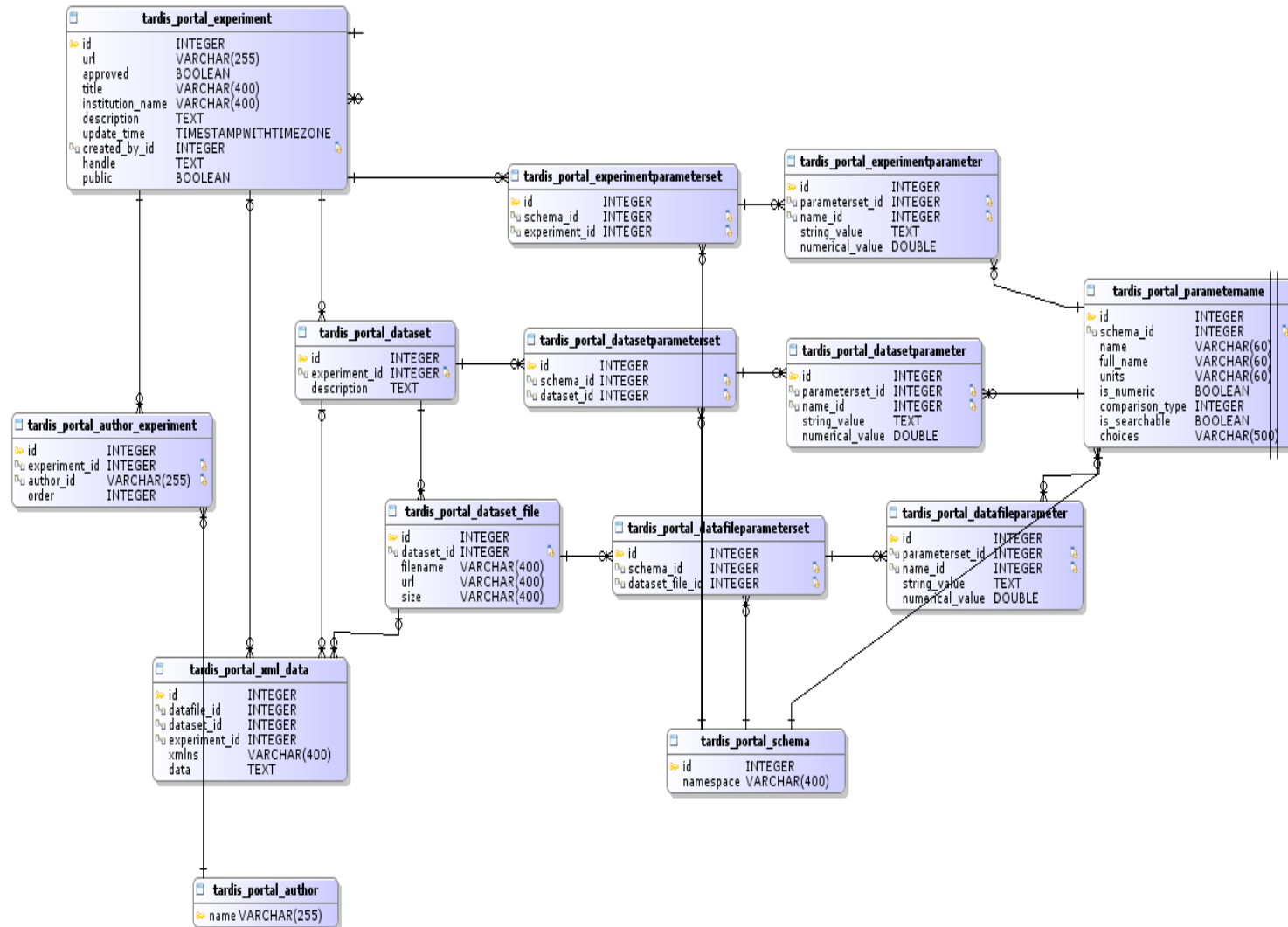
TARDIS - Portal schema - Experiment

TARDIS - Portal schema - Dataset

TARDIS - Portal Schema - Files

Summary

- MeCat is a joint project between the Australian Synchrotron and ANSTO to improve metadata management and publication at the facilities.

- It is using and extending an open source project called TARDIS and will offer greater capabilities to share and reuse data.

Questions?