XML Ingest
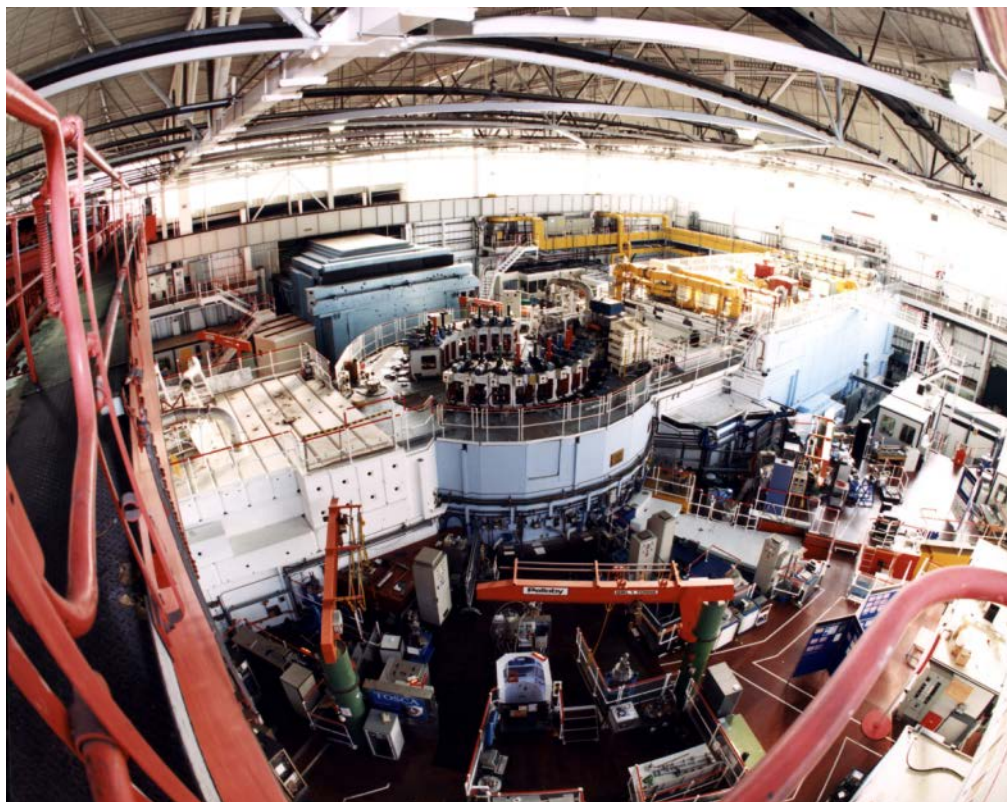

Tom Griffin, STFC ISIS Facility
ICAT Workshop
Lund, Sweden

March 2013

tom.griffin@stfc.ac.uk

Science & Technology Facilities Council
ISIS

Nostalgia

```xml
<?xml version="1.0" encoding="ISO-8859-1"?>
<icat xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="../icatXSD/icatXSD.xsd" version="1.0 RC6">
  <study>
      <name>weld_trans</name>
      <purpose>An investigation</purpose>
      <status>ONGOING</status>
      <study_creation_date>2013-02-24T07:25:45</study_creation_date>
      <investigation trusted="true">
          <inv_number>1210410</inv_number>
          <visit_id>1</visit_id>
          <instrument>ENGINX</instrument>
          <title>weld_trans</title>
          <inv_type>EXPERIMENT</inv_type>
          <bcat_inv_str>Dong,Zhang,Dodge,Zha -</bcat_inv_str>
        <keyword>
              <name>weld_trans</name>
        </keyword>
        <dataset>
            <name>Default</name>
            <dataset_type>EXPERIMENT_RAW</dataset_type>
            <dataset_status>COMPLETE</dataset_status>
            <description>ENG99404 Dong,Zhang,Dodge,Zhaweld_trans 24-FEB-2013 07:25:45 4.87</description>
          <sample>
              <name>weld_trans</name>
              <safety_information>No information available</safety_information>
          </sample>
          <datafile>
              <name>ENGINX00199404.raw</name>
              <location>\\isisdatar80\instruments$\NDXENGINX\Instrument\data\CYCLE_12_5\ENGINX00199404.raw</location>
              <description>weld_trans</description>
              <datafile_format>isis neutron raw</datafile_format>
              <datafile_format_version>2</datafile_format_version>
              <datafile_create_time>2013-02-24T07:25:45</datafile_create_time>
              <datafile_modify_time>2013-02-24T07:30:17</datafile_modify_time>
              <file_size>25879552</file_size>
            <parameter>
                <name>run_number</name>
                <numeric_value>199404</numeric_value>
                <units>N/A</units>
            </parameter>
            <parameter>
                <name>run_header</name>
                <string_value>ENG99404 Dong,Zhang,Dodge,Zhaweld_trans 24-FEB-2013 07:25:45 4.87</string_value>
                <units>N/A</units>
            </parameter>
            <parameter>
                <name>number_of_detectors</name>
                <numeric_value>2668</numeric_value>
                <units>N/A</units>
            </parameter>
            <parameter>
```

Why it is a good idea?

- Anything / anyone can write xml
- Takes a bit more to create a sample, sampleparameter, sampletype etc
- Insulates code
- Easier ingestion
- Supported in Nexus

Why it is a bad idea?

- Controllability
  - Calibration datafiles
  - Facility specific rules
  - What if......?
- Schema changes

Ideas

- Simplified schema
- Best effort / strict flags
- Control file/parameters
  - Calibration numbers
  - Date ranges
- Plugin
  - Could be modified per facility if required

If I could offer you only one tip for the future, xml ingest would be it.