



| The European Synchrotron

Status of metadata at ESRF



ICAT deployment
Metadata ingestion
Current status
Next steps

Current software stack

- ✓ Glassfish 4.0
- ✓ ICAT 4.3.1 + authn_db 1.1.1
- ✓ TopCAT 1.11 + IDS 1.0.0 (+ custom plugin)

Databases

- ✓ Oracle (test and production)
- ✓ Derby (development)

Machines

- ✓ Oracle VM
- ✓ Sun JDK 1.7.0

Security

- ✓ Full SSL with signed certificates (TERENA)
- ✓ Apache reverse proxy (mod proxy http, address translation)

IDS not fully suitable yet for ESRF setup

- We need more operations delegated to the plugin, including logic for available/not available, direct/archived, path reconstruction, ...

Problem with TopCAT behind a reverse proxy (GWT issue)

Address translation ignored:

- All .gwt.rpc files need to be copied

or

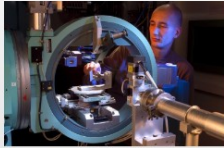
- Archive needs to be modified to change the context root

Other issues with proxy or firewall

- Problems connecting to external ICAT on port 8181 (solved)
- Problems connecting to ILL using ill.eu URL (ill.fr works)

Click to edit Master text styles

The beamline



The ICAT ingester

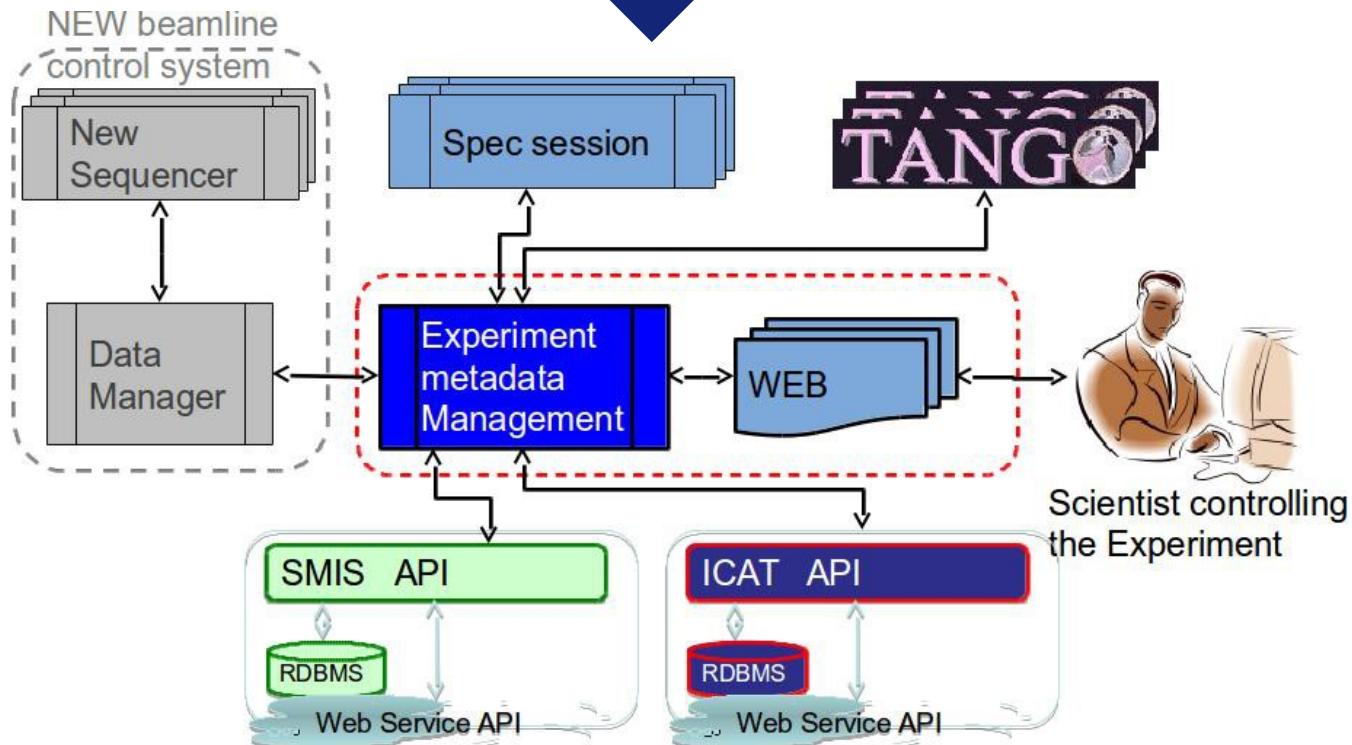
Second level

Third level

Fourth level

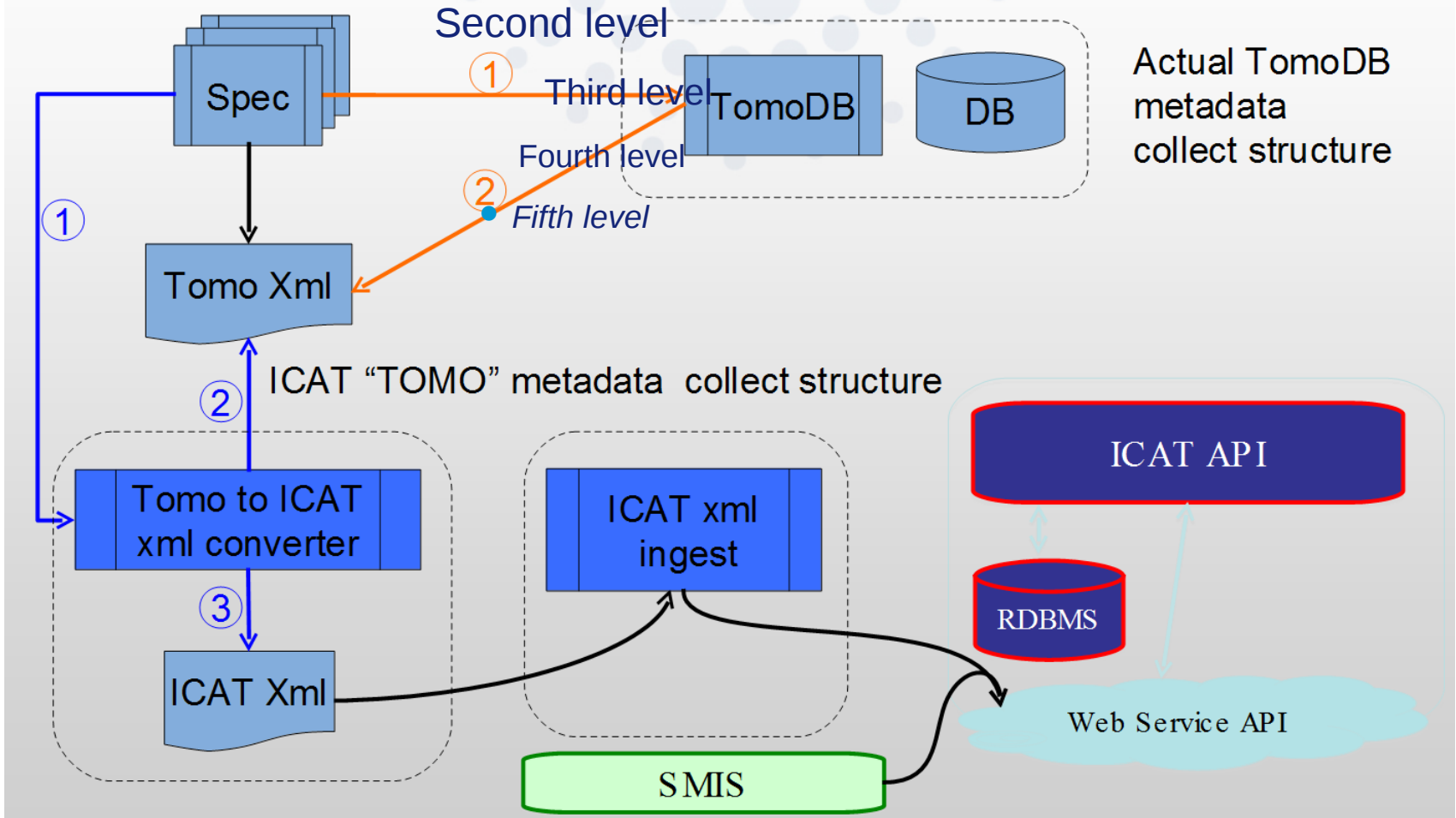
- Fifth level

The ICAT installation



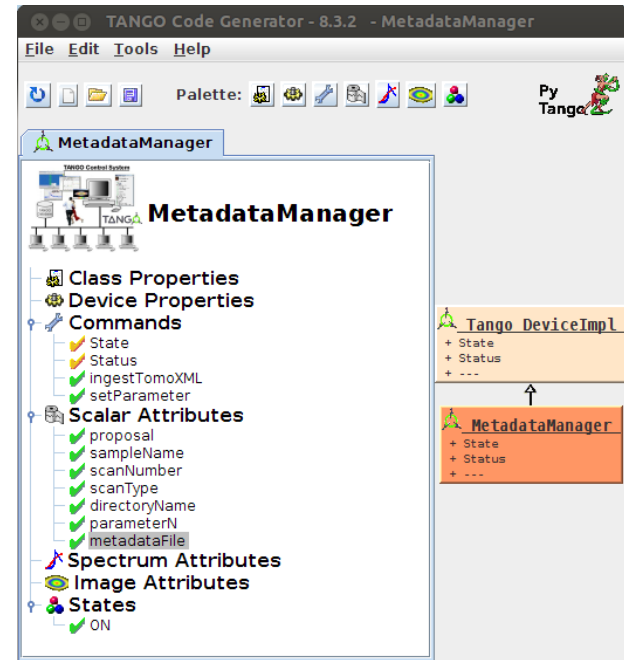
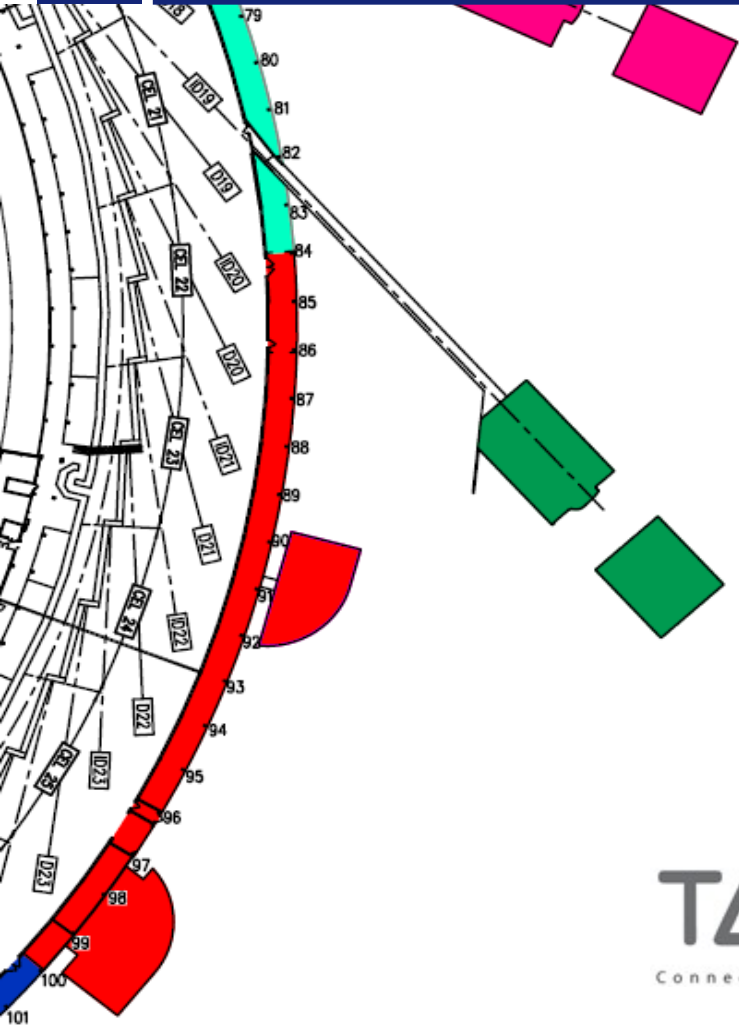
First Pilot project for ID 19

Click for ID 19 Master text styles



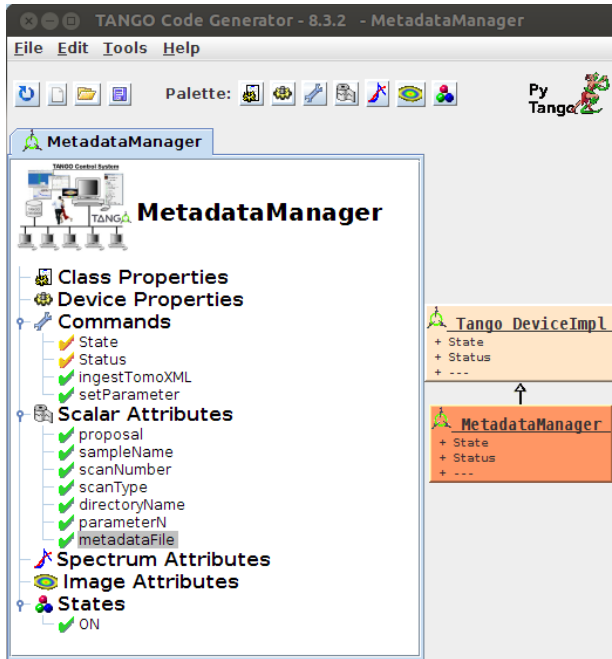
Data used for service verification imported using the Pilot

General case

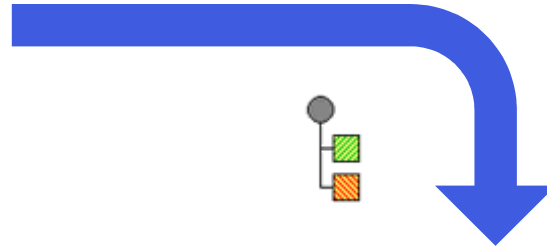


HALL EXPERIMENTAL

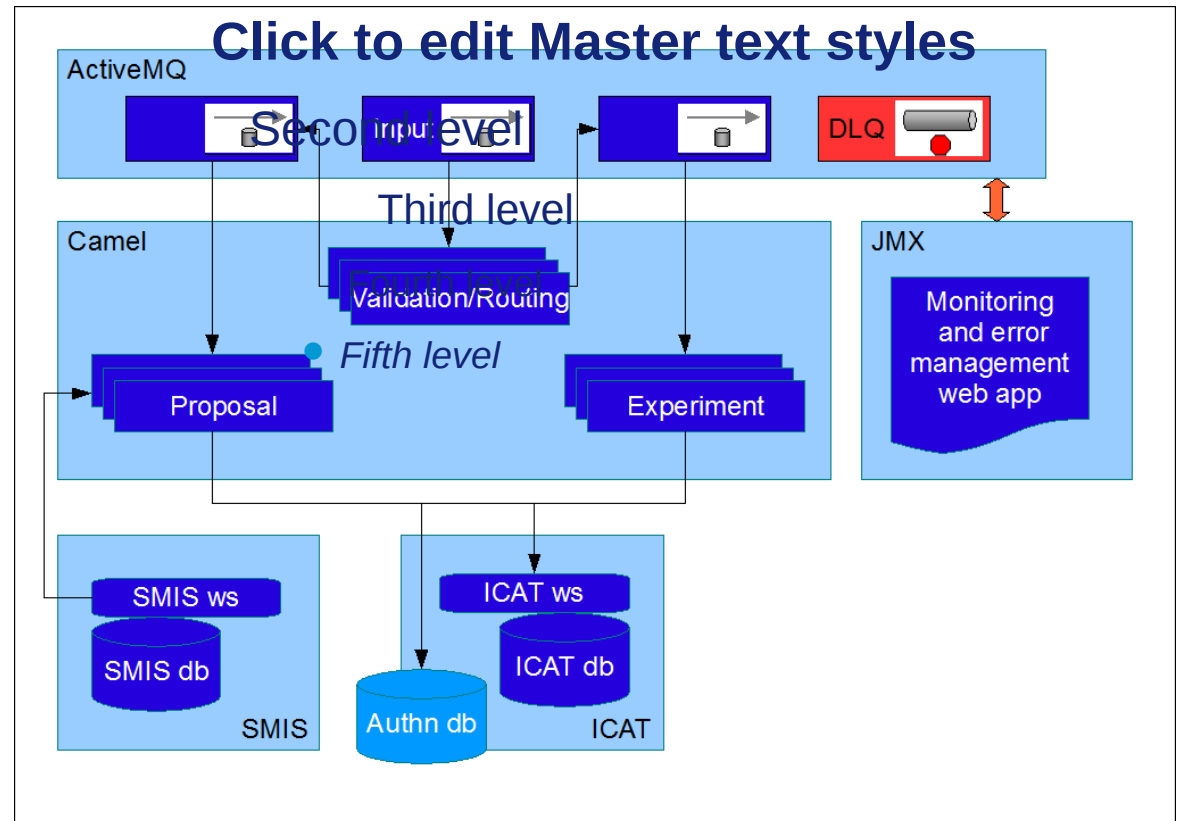
Proposed implementation



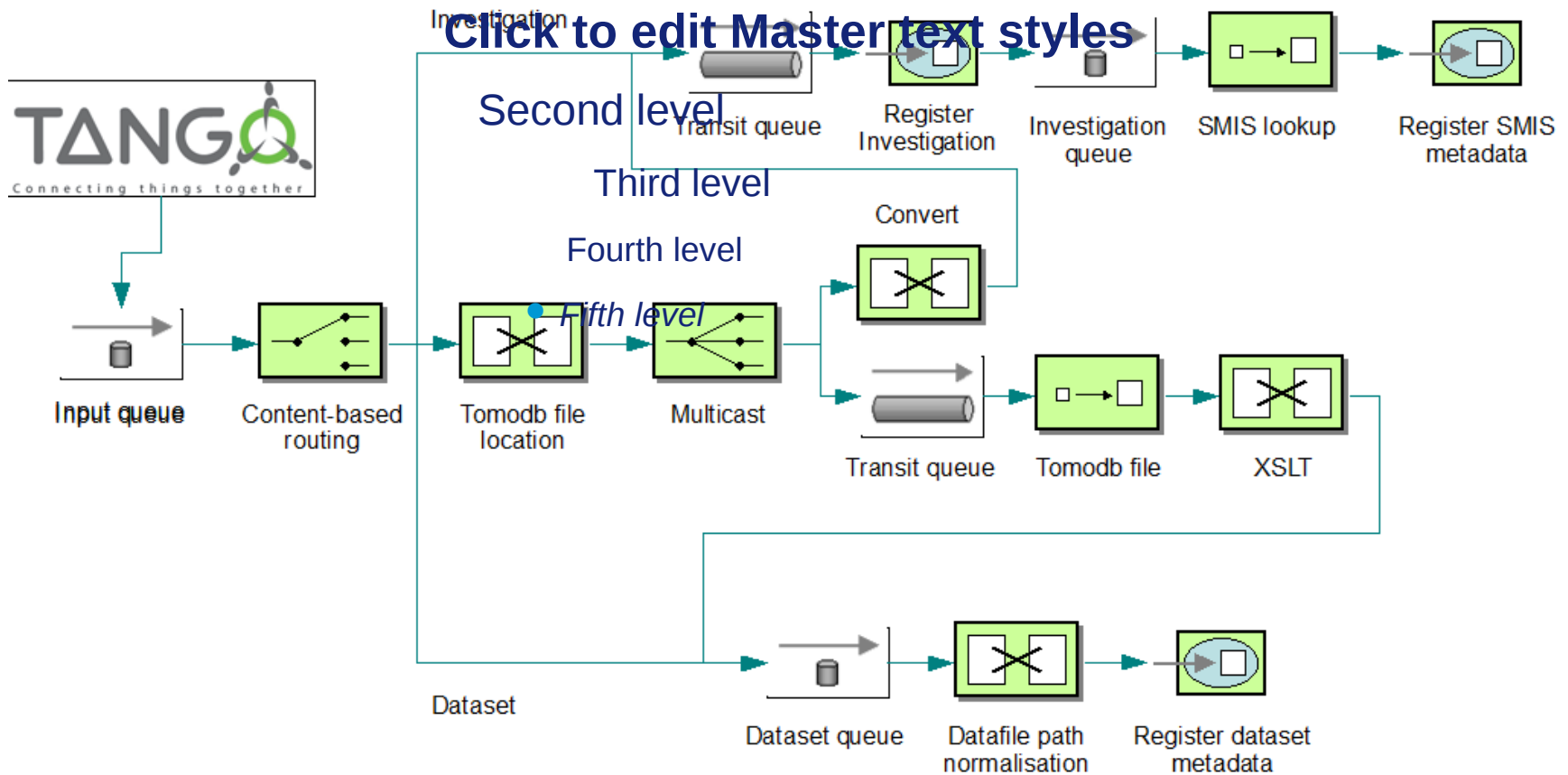
1 stateful TANGO device
on each beamline



1 stateless queuing and
processing system to
bind them all



Processing details



Asynchronous processing

- Non-blocking TANGO devices (fire and forget)
- Queues serve as buffer in case of peak activity or slow/dead process

Data integrity

- Message persistence
- Transacted processing
- Error management, redelivery policy

Scalability

- Concurrent processing with configurable thread pools
- Several instances of queues and/or routes can be deployed

Input queue is a single point of failure

- Failover configuration / cluster deployment
- Local hdf5 file for offline processing

Most features already implemented

- ✓ Concurrent processing
- ✓ Transactions
- ✓ Automatic message translation (marshall/unmarshall, conversion)
- ✓ Logging / message tracing
- ✓ Error management

Technology agnostic

- ✓ Implements well-defined Enterprise Integration Patterns
- ✓ Independent of queuing system, transport technology, ...

Very well integrated

- ✓ DSL for Java, Spring, Scala, Groovy, Annotations, Blueprint, ...
- ✓ Options for deployment (standalone, EJB container, OSGi container)
- ✓ Pure Maven (archetypes, plugins), JUnit support, Spring integration

Prototype development

- Embedded Apache ActiveMQ
- Standalone Java application (using Spring)

Project metrics

- c.a. 1 month in the making
- c.a. 600 LoC (250 tests, 200 ICAT client, 150 data objects, beans)
- c.a. 800 lines of XML (incl. blank lines) (600 test, 100 configuration, 100 XSLT)
- Stored on ESRF gitlab

Not done yet

- SMIS integration
- Monitoring and error management interface

Authentication with Umbrella

Reusing SMIS mechanism already in place

Prototype almost ready for testing with TANGO device

Validation of the design

Testing and performance analysis

Stress testing, failure recovery

Dimensioning of the different components

Data volume for queues, number of concurrent processors for each steps

1 year = 1200 experimental session, 4800 visits, 1800 publications

on 30 public beamlines (+10 CRG)

Live testing on one volunteer beamline

We still need to define

- ESRF data policy
- Infrastructure needed for deployment
- Maintenance and support policy



Thank you for your attention

Solved issues

- ❑ No migration script for Derby db (no python driver), no migration of rules
 - ❑ Python code (partially) ported to Java, rules recreated
- ❑ Some modification of the package required for TopCAT
 - ❑ To change the JPA log level
- ❑ Problem with TopCAT behind a reverse proxy (GWT issue)
 - ❑ All .gwt.rpc files need to be copied (address translation ignored)

Current issues

- ❑ ICE 1.0.0 not compatible with ICAT 4.3.x
 - ❑ Need to be updated at the same time than ICAT if considered as an admin module
- ❑ IDS not fully suitable for ESRF setup
 - ❑ We need more operations delegated to the plugin, including logic for available/not available, direct/archived, path reconstruction, ...
 - ❑ The plugin currently needs a special packaging

Issues with clients

- DAWN ICAT plugin
 - Only supports LDAP as authentication method
 - Cannot configure new ICAT repositories
- Most Python tools need at least Python 2.6
 - Version 2.4 stated in documentation
- XML ingestion tool does not work with ICAT 4.3.x
 - Would need a complete rebuild as it uses a modified 4.2 client

Issues with other ICATs

- Problems connecting to external ICAT on port 8181 (solved)
- Problems connecting to ILL using ill.eu URL (ill.fr works)
- Failed to connect to ISIS during SV 7 (investigation ongoing)
 - Likely linked to our proxy and firewall configuration too**