PSI Center for Photon Science

SciCat at TOMCAT

Use cases and requirements for handling X-ray tomographic data

Dr. Goran Lovric
SciCatCon Meeting 2024 (PSI), 3 July 2024

Agenda

1 Introduction
2 Use cases
3 Feature requirements
4 Conclusion and Outlook

Introduction

TOMCAT at a glance

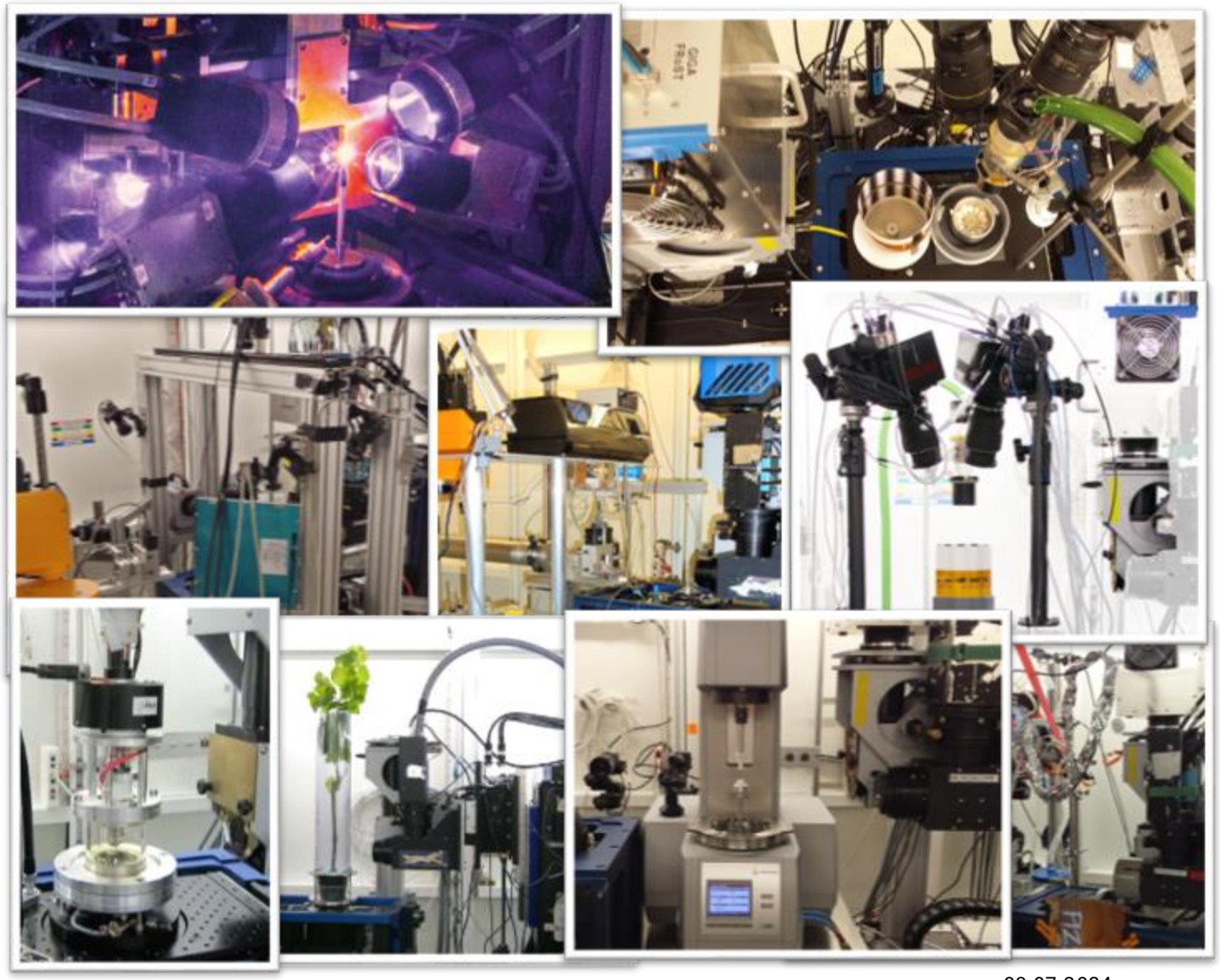
- Superbend source: 2.9 T
- Critical energy: 11.1 keV
- X-ray energies: 9- 45 keV
- Spatial resolutions: 250 nm – 20 µm
- Enhanced density resolution (inline phase contrast, TXM etc.)
- Fast Tomo @ 20 Hz

Geoscience | Biomedical research

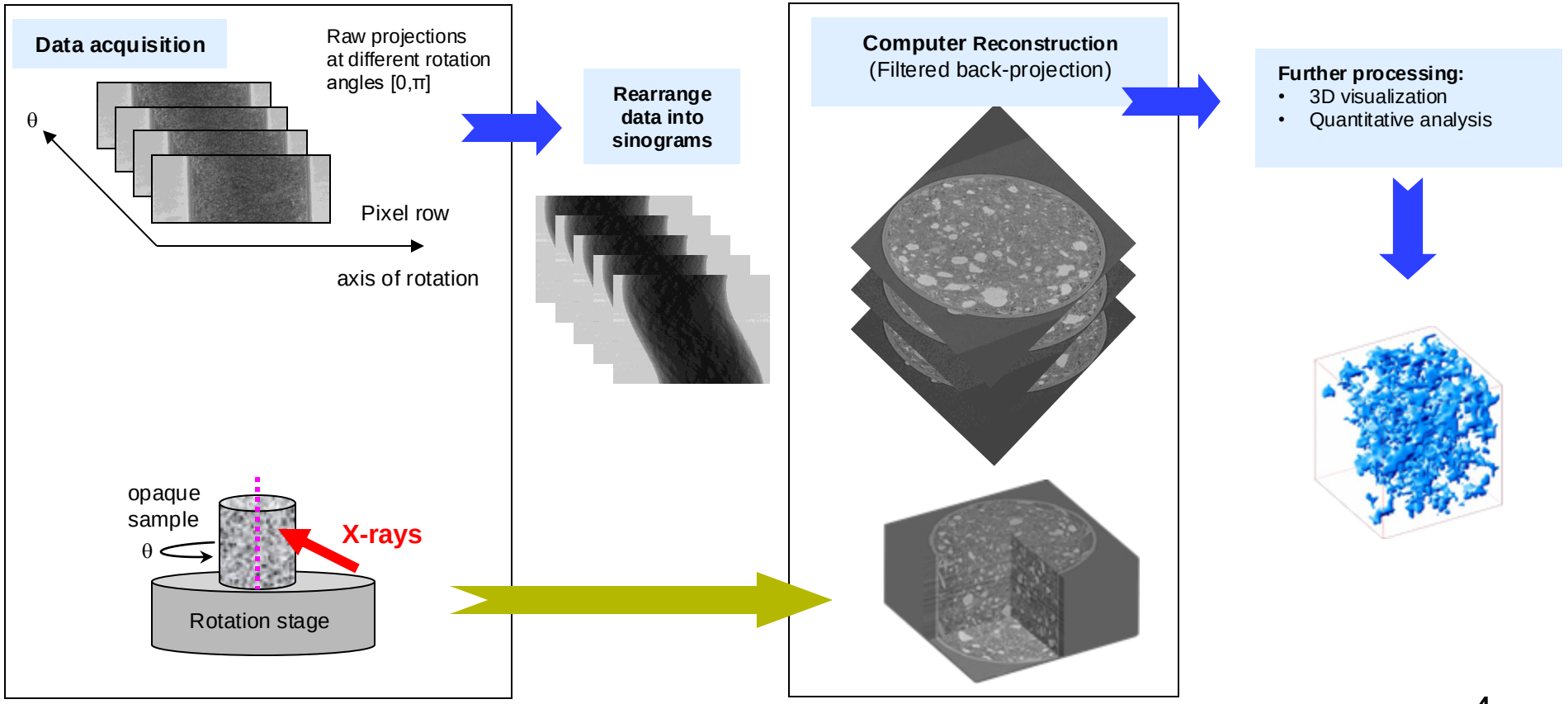
Energy & Battery research

Paleontology | Materials science

Soft condensed matter | ...



Introduction to CT data



4

SLS 2.0 and TOMCAT 2.0 upgrade

Construction of new I-TOMCAT hutch (August 2023)

Construction of new S-TOMCAT hutch (May 2024)

S-TOMCAT

- will outperform TOMCAT @ E > 20 keV
- Will give access to up to E = 80 keV

I-TOMCAT

- Will outperform TOMCAT by 3 orders of magnitude @ E = 10-25 keV
- Will outperform TOMCAT by 2 orders of magnitude @ E = 30-60 keV

Implications

- Doubling of data («standard» experiments)
- Stitched scans acquisitions: from 24h to ~1h
- Radiation dose
- Multi-scale & multi-modal

Data size

Acquisition

Dose-efficient imaging

Reconstruction

Image enhancement

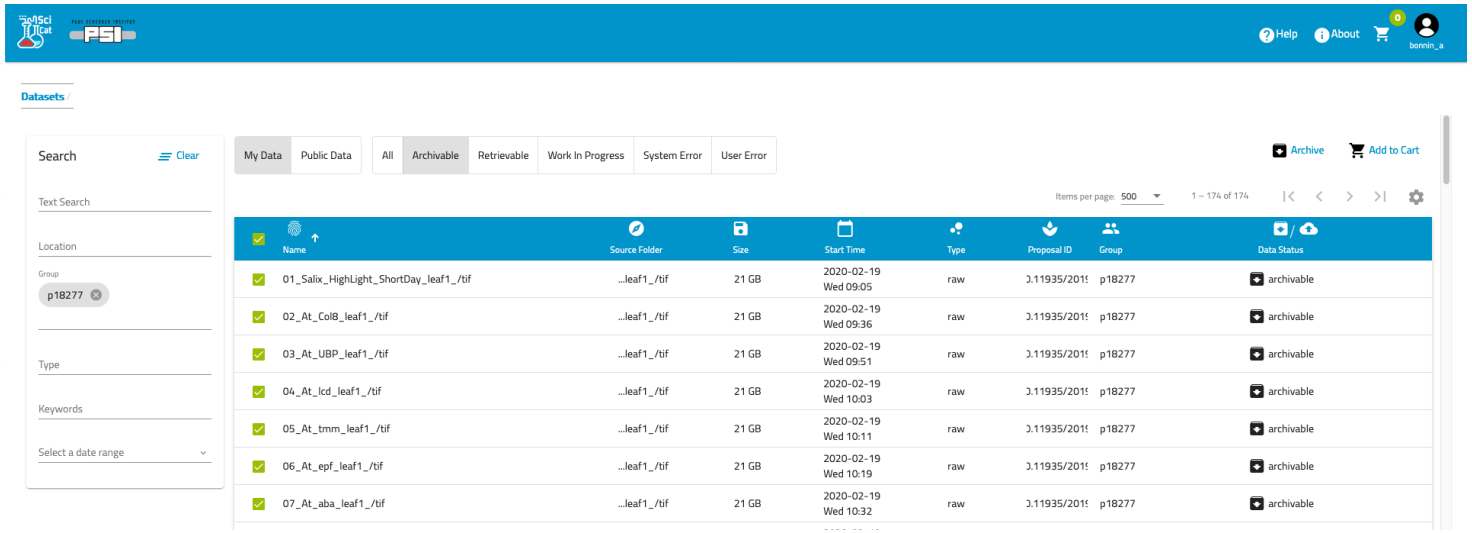
Visualization

Fusion

Use Cases

Type #1: Archive and restore raw tomographic data (Beamline)

- Ingest data after or during a standard tomographic experiment at the beamline
- Manual step after „obsolete” data removal
- Raw data is single source of truth
- Data typically restored during “offline” post-processing
- Standard restore point: RA Cluster
- ORD guidelines:
  - Publish datasets
  - Metadata management



Use Cases

Type #2: Archive custom experimental data (X-ray labs, prototype detectors etc.)

- Customized setups with varying equipment
- Metadata is amended manually
- Non-standard settings (student projects etc.)

Manual ingestion

This short manual explains how to do **ingest** data manually. This might be useful if the data has been acquired at TOMCAT I

1. Define your data which needs to be archived, i.e. by defining the raw datasets as well as appropriate metadata. Beware c
   - a single dataset should currently not have more than 200k files
   - a single dataset should not be larger than 50 TB
   - recommended size of a single dataset: between 1GB and 1TB

For the example below, we assume it's one folder.

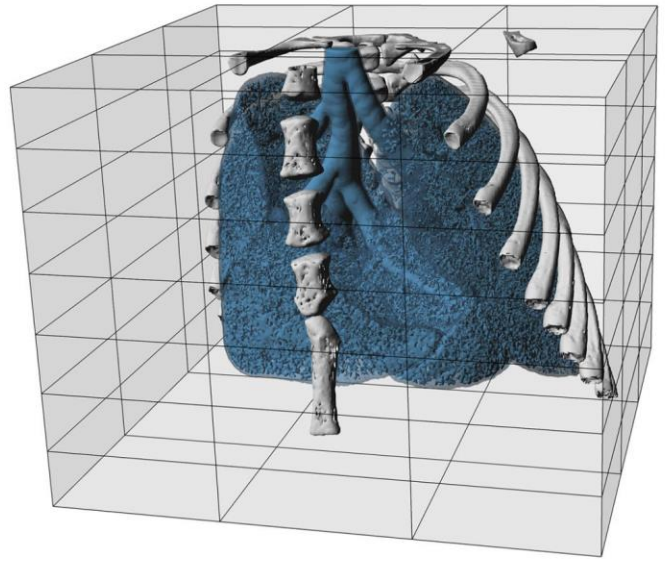
2. Create a `JSON` file in the folder and refer to the example below. For the definition and explanationof the necessary and

```
{
  "principalInvestigator": "marco.stampanoni@psi.ch",
  "creationLocation": "/PSI/SLS/TOMCAT",
  "sourceFolder": "/sls/X02DA/Data20/e15889/BunkerSetup",
  "owner": "Polikarpov Maxim and Joan Vila",
  "type": "raw",
  "ownerGroup": "p15889",
  "datasetName": "Joan_Vila_2016_2019_Bunker3_setup_BunkerSetup",
  "keywords": [
    "polikarpov.maxim@mail.ru",
    "Eiger Data",
    "DPC data",
    "2016-2019",
    "assymetric setup Joan",
    "Sigray source\n",
    "MOENCH Detector",
    "other detectors (?)"
  ]
}
```

Use Cases

Type #3: Archive of reconstructed datasets (derived data)

- Datasets which require complex reconstruction & post-processing:
  - ”Mosaic” scanning (stitching datasets)
  - Time-resolved datasets
- Lack of knowledge/resources from user community to do reconstructions
- General availability to other potentially interested scientific communities



Requirements – Part 1

- Improved search capabilities:
  - Metadata full-text search
  - File name search with instantaneous results
  - Part of sample name is known (e.g. regular expressions?)
- Allow for metadata adjustments (submit a change request, form?)
- Automatic notifications:
  - Embargo period approaching
  - Storage deadline is approaching – data will be deleted from Tape archive
- Easy download capabilities:
  - Browser
  - Resume downloads
  - Only selected datasets

Requirements – Part 2

- Metadata versioning:
  - Metadata information might also be changing in the future
- Improved debugging information when archiving (or restoring) is failing:
  - For instance: ingestion seems OK, but data does not appear on the Web interface
  - Data expected to be archived in XX days
- Potential future requirements:
  - Different restore points (Cloud?)

Conclusion and Outlook

- Overall very satisfied with the service
- A few improvements would be required to further simplify general usability
- Combine service with experimental logbook, metadata and reconstruction algorithms to guarantee full reproducibility
- Perhaps a publicly available roadmap of future developments:
  - current/future developments
  - release notes