Topical Discussion: Databases

Wiebke Köpp

Computing Group, Advanced Light Source, Lawrence Berkeley National Lab

November 3rd, 2024

Topical Discussion: Databases

- Publicly accessible databases
  - FAIR
  - ML-ready data
  - Raw and/or reduced data?
  - What metadata to include?
  - General or specialized?
- Facility- or instrument-specific databases / data catalogues
  - Access controls through connections to proposal systems
  - Pathway to publicly shared data

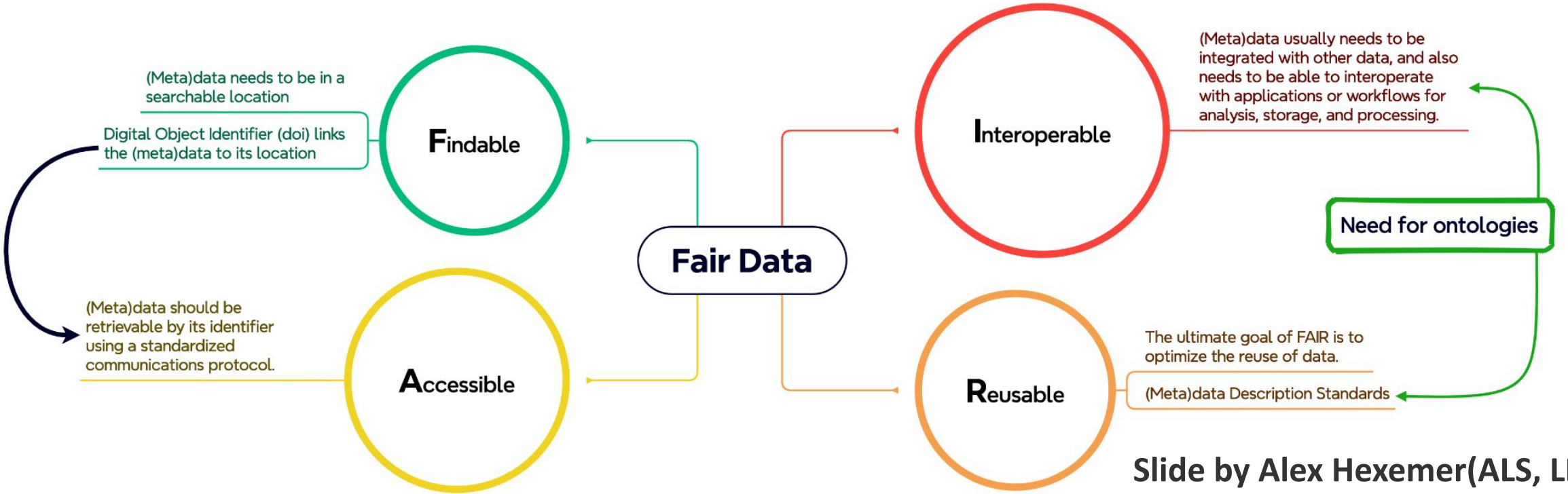
FAIR Data

Findable: Data and supplementary materials have sufficiently rich metadata and a unique and persistent identifier.

Accessible: (Meta)data are understandable to humans and machines. Data is deposited in a trusted repository.

Interoperable: (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

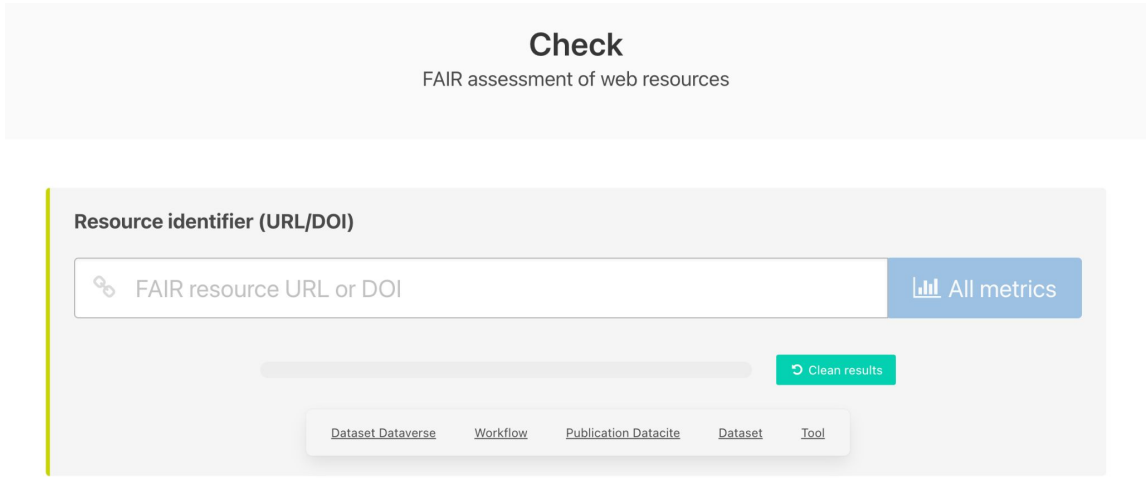
Reusable: Data and collections have a clear usage licenses and provide accurate information on provenance.



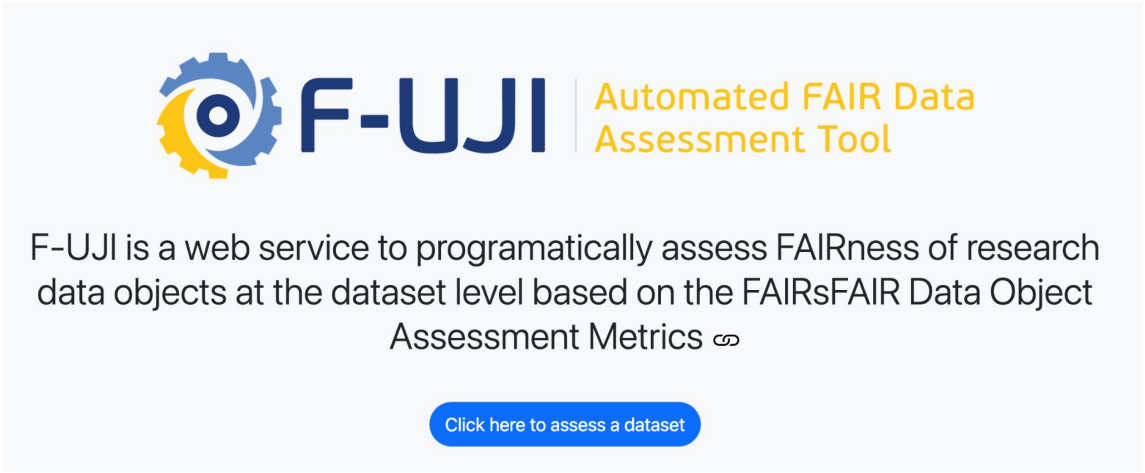
Slide by Alex Hexemer(ALS, LBNL)

FAIR Data

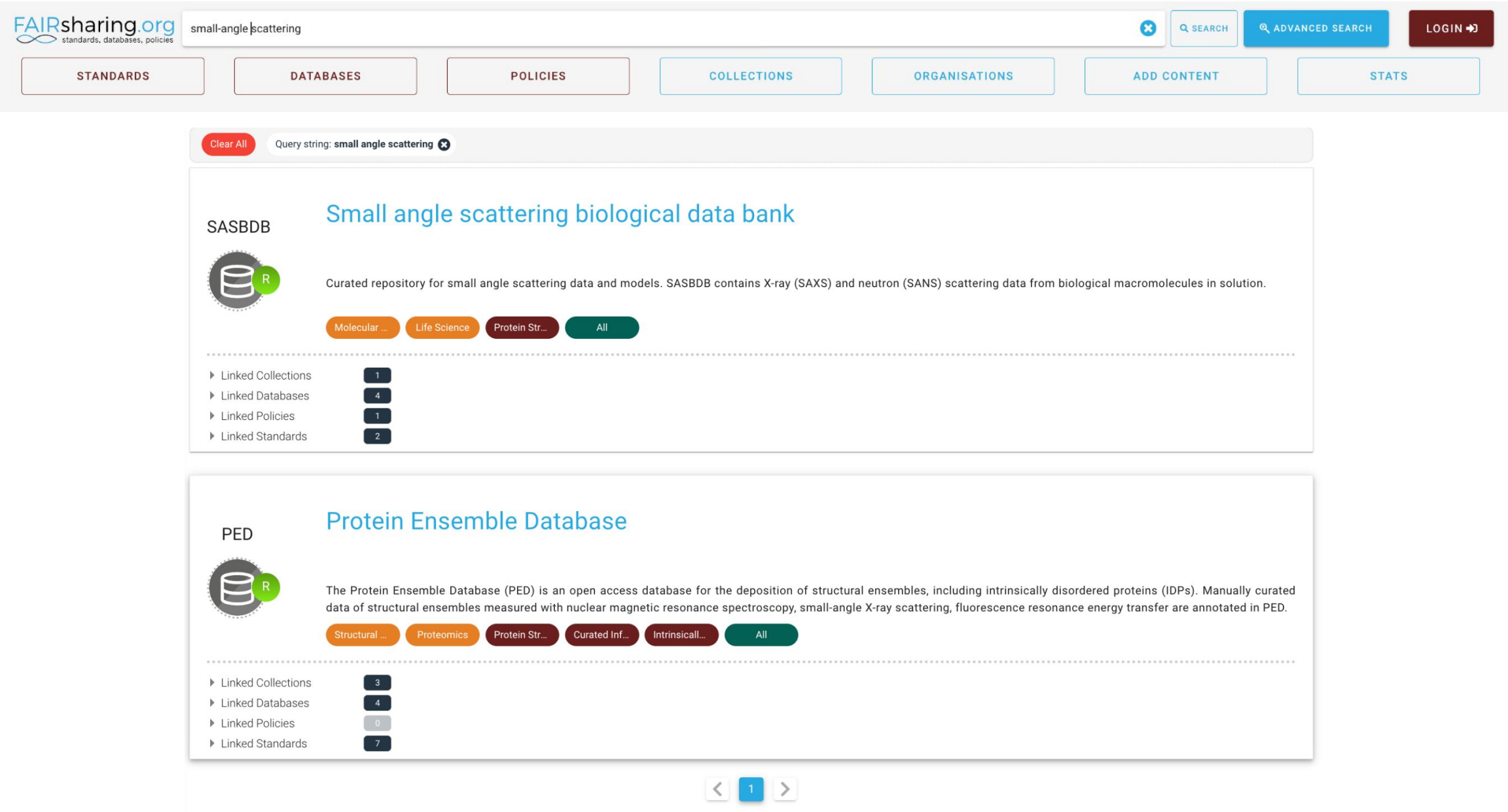
https://www.f-uji.net/



https://fair-checker.france-bioinformatique.fr/check



Public Databases / Data catalogues



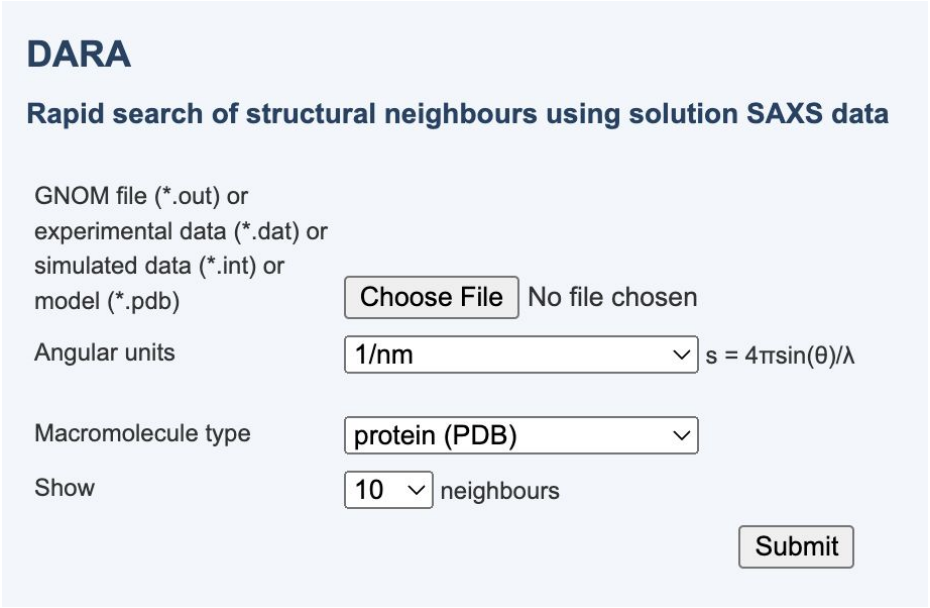
Public Databases / Data catalogues

https://www.sasbdb.org/

SASBDB

Small Angle Scattering Biological Data Bank

https://simplescattering.com/



https://dara.embl-hamburg.de/

BIOISIS™ .net

https://www.bioisis.net/

Databases at SAS 2024

November 4, 2024 (Monday)

1F 101CD	
MOD-1	**Methods, SAS Database-application & standards** (Chair: Guan-Rong Huang/Brian Richard Pauw)
15:45-16:10	**The Small Angle Scattering Biological Data Bank - SASBDB** Cy Jeffries (Invited Speaker)

16:50-17:20	(Chair: An-Chung Su) **Flash Talk B** (15 talks)

Chair: An-Chung Su

Order	Poster Number	Speaker	Title
5	PJ-10	Glen Smales	DACHS: Database for Automation, Characterisation, and Holistic Synthesis

Public Databases / Data catalogues

zenodo | small angle scattering | Communities | My dashboard | Log in | Sign up

234,611 result(s) found

Sort by: Best match

Versions

View all versions

Access status

- Open 226,447
- Restricted 8,072
- Embargoed 92

Resource types

- Publication 199,906
- Dataset 14,368
- Image 13,530
- Software 2,705

September 18, 2013 (v1) | Proposal | Open

ESS Construction Proposal : LoKI - A broad-band SANS Instrument

Jackson, Andrew; Kanaki, Kalliopi

Proposal for a SANS instrument at the European Spallation Source (ESS). Proposal submitted to ESS in October 2012. Recommended for construction by ESS SAC in April 2013. Approved by ESS Steering Committee for preliminary design in October 2013.

Uploaded on January 15, 2015 — 222 — 136

March 16, 2015 (v1.0.0) | Software | Open

SCT: Analysis and Modelling of Small Angle Scattering Data

David W Wright; Stephen J Perkins

Small angle X-ray and neutron-scattering techniques (known as SAXS and SANS nrespectively) are low resolution techniques that can characterize protein structure in solution. They are of particular utility when large proteins cannot be crystallized and in systems where solution conditions affe…

Uploaded on June 25, 2015 — 174 — 14

March 26, 2024 (v1) | Other | Open

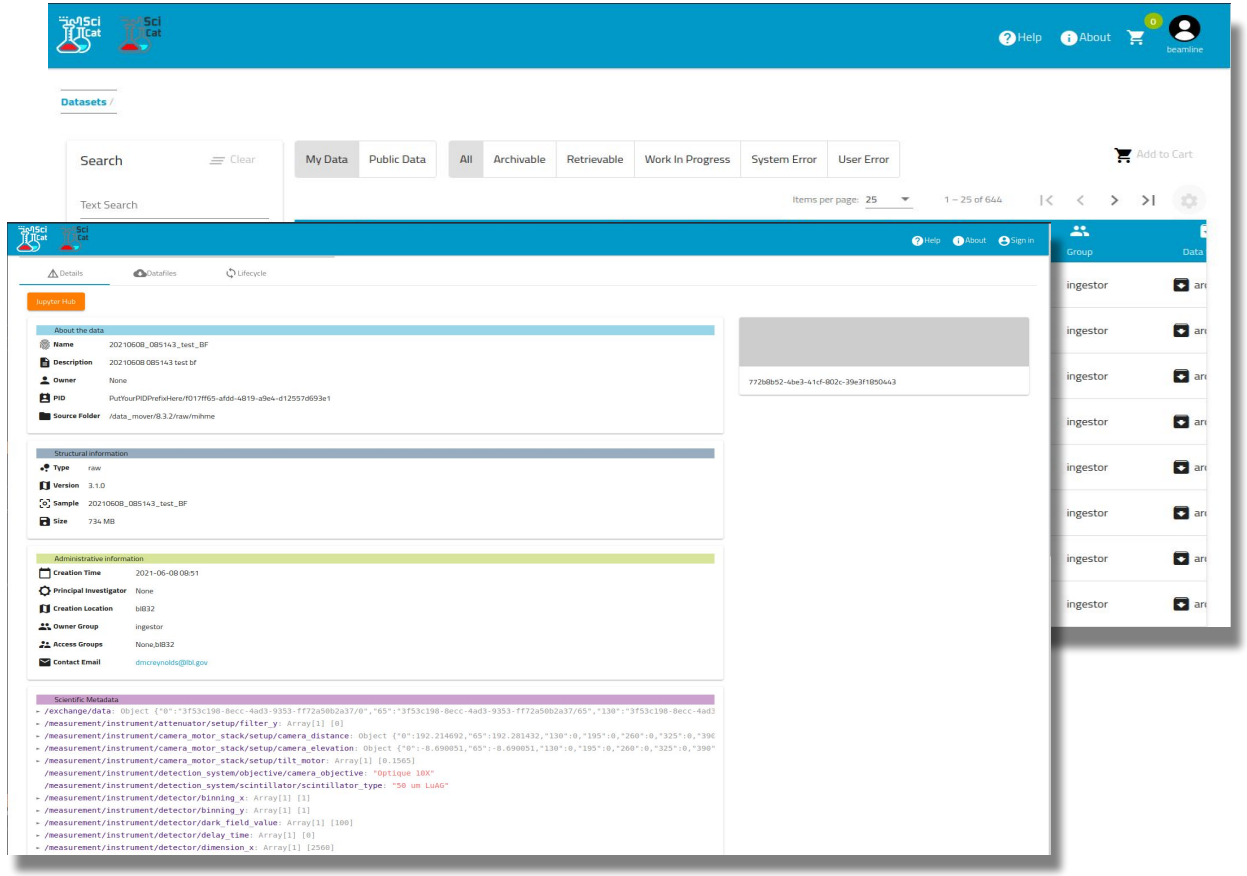
Small-Angle X-ray and Neutron Scattering

Creative Biolabs

Small-angle X-ray and neutron scattering (SAXS and SANS) measurements are theoretically simple. The essential principles of them are samplable

SciCat

- Loading into datasets
  - extract metadata from raw data set and insert into SciCat's database
  - Add extra data (thumbnails, etc)
- Search datasets
- Display dataset details
- Access controls
- Downloading Datasets
- Work to be done with metadata



EUROPEAN SPALLATION SOURCE

PAUL SCHERRER INSTITUT PSI

MAX IV

The Rosalind Franklin Institute

MLZ Heinz Maier-Leibnitz Zentrum

DESY.

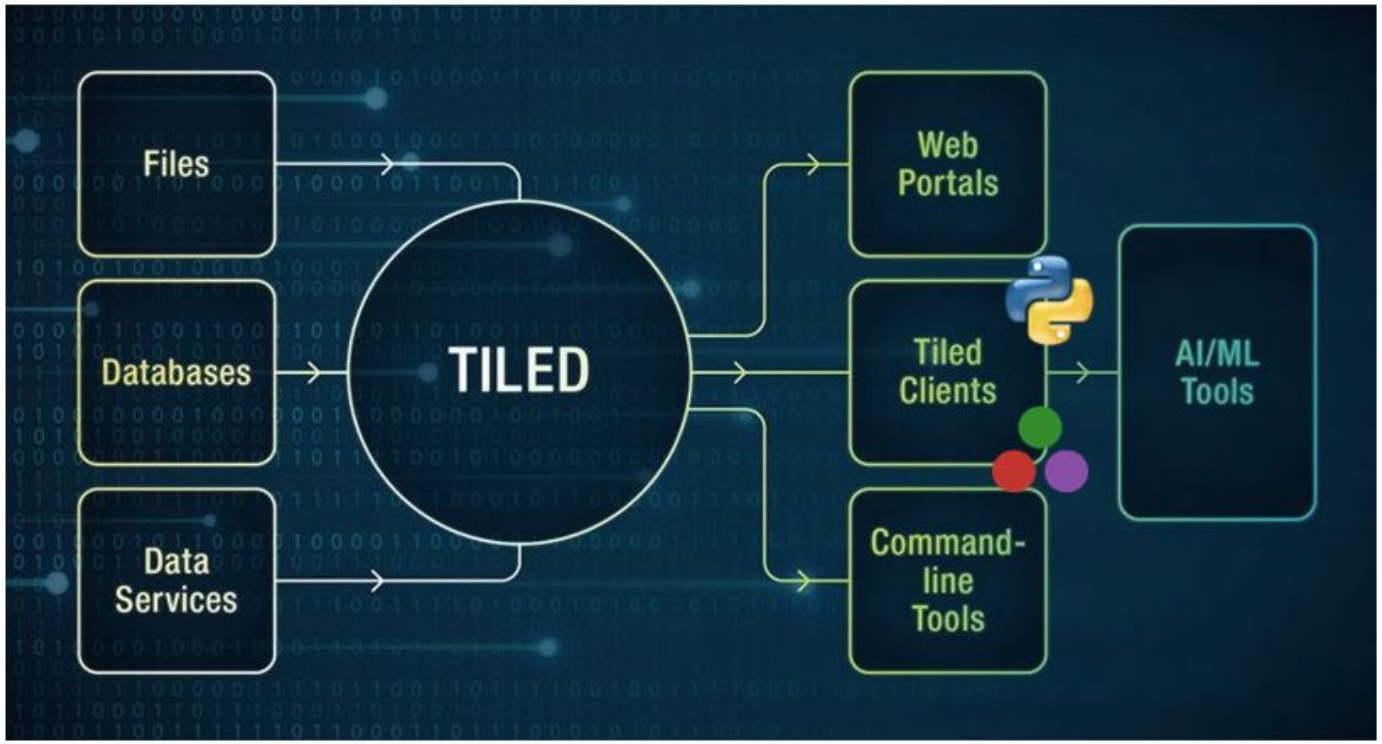
SOLEIL SYNCHROTRON

DAPHNE 4NFDI

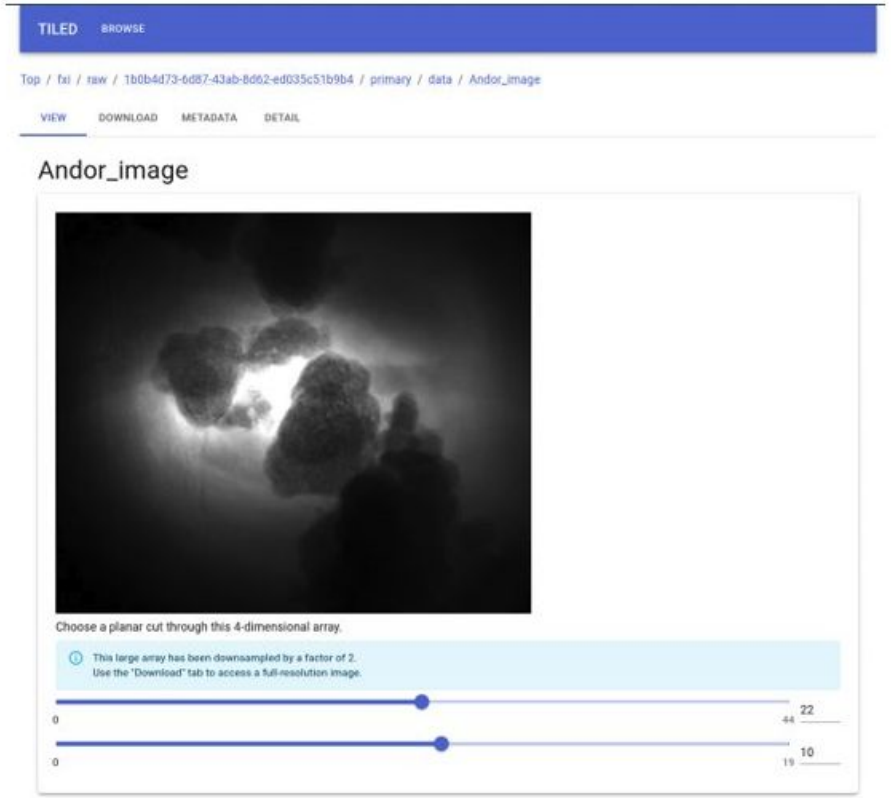
Slide by Dylan McReynolds (ALS, LBNL)

Tiled: A Structured API to Data

- **Search** on metadata
- **Slice** into remote datasets
- **Transcode** between formats
- **Download** partial or whole datasets
- Or **find** data storage location(s) for direct access
- Implement web **security** standards and authorization



Data in a web browser, on a phone



Topical Discussion: Databases

- Publicly accessible databases
  - FAIR
  - ML-ready data
  - Raw and/or reduced data?
  - What metadata to include?
  - General or specialized?
- Facility- or instrument-specific databases / data catalogues
  - Access controls through connections to proposal systems
  - Pathway to publicly shared data?