

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

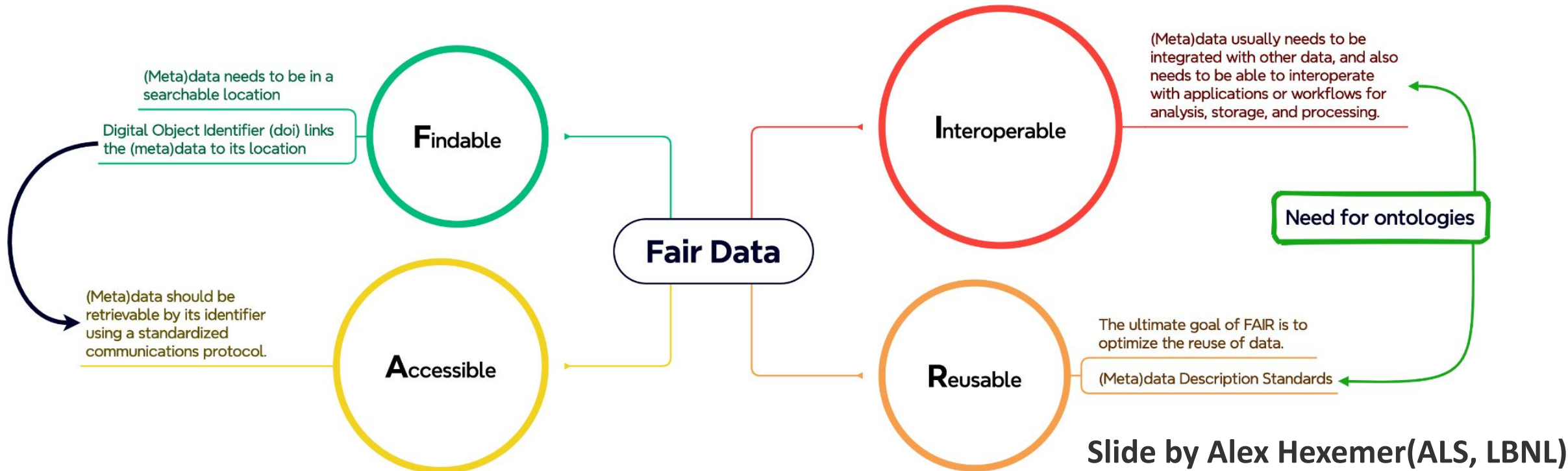
FAIR Data

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.



FAIR Data

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

Check

FAIR assessment of web resources

Resource identifier (URL/DOI)

 FAIR resource URL or DOI

 All metrics

 Clean results

[Dataset Dataserve](#)

[Workflow](#)

[Publication Datasite](#)

[Dataset](#)

[Tool](#)

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



F-UJI is a web service to programmatically assess FAIRness of research data objects at the dataset level based on the FAIRsFAIR Data Object Assessment Metrics [↗](#)

[Click here to assess a dataset](#)

STANDARDS

DATABASES

POLICIES

COLLECTIONS

ORGANISATIONS

ADD CONTENT

STATS

Clear All

Query string: small angle scattering ✕

SASBDB

Small angle scattering biological data bank



Curated repository for small angle scattering data and models. SASBDB contains X-ray (SAXS) and neutron (SANS) scattering data from biological macromolecules in solution.

Molecular ... Life Science Protein Str... All

- ▶ Linked Collections 1
- ▶ Linked Databases 4
- ▶ Linked Policies 1
- ▶ Linked Standards 2

PED

Protein Ensemble Database



The Protein Ensemble Database (PED) is an open access database for the deposition of structural ensembles, including intrinsically disordered proteins (IDPs). Manually curated data of structural ensembles measured with nuclear magnetic resonance spectroscopy, small-angle X-ray scattering, fluorescence resonance energy transfer are annotated in PED.

Structural ... Proteomics Protein Str... Curated Inf... Intrinsicall... All

- ▶ Linked Collections 3
- ▶ Linked Databases 4
- ▶ Linked Policies 0
- ▶ Linked Standards 7

<https://www.sasbdb.org/>



<https://simplescattering.com/>



DARA

Rapid search of structural neighbours using solution SAXS data

GNOM file (*.out) or
experimental data (*.dat) or
simulated data (*.int) or
model (*.pdb)

No file chosen

Angular units $s = 4\pi\sin(\theta)/\lambda$

Macromolecule type

Show neighbours

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



<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

234,611 result(s) found

Sort by Best match

Versions

 View all versions

Access status

 Open Restricted Embargoed

Resource types

 Publication Dataset Image Software

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

226,447

8,072

92

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 respectively) 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 affect...

Uploaded on June 25, 2015

174 14

199,906

14,368

13,530

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 comparable

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

The screenshot shows the SciCat web interface. At the top, there's a navigation bar with 'Help', 'About', and 'Sign in' options. Below that, a search bar and navigation tabs for 'My Data', 'Public Data', 'All', 'Archivable', 'Retrievable', 'Work in Progress', 'System Error', and 'User Error'. The main content area displays details for a dataset with the following information:

- About the data:**
 - Name: 20210608_OBS143_test_BF
 - Description: 20210608 OBS143 test bf
 - Owner: None
 - PID: PutfourPIDPefukHere/ID17f65-af66-4819-df6a-412557693e1
 - Source Folder: /data_mover/8.3.2/raw/mhme
- Structural information:**
 - Type: raw
 - Version: 3.1.0
 - Sample: 20210608_OBS143_test_BF
 - Size: 734 MB
- Administrative information:**
 - Creation Time: 2021-06-08 08:51
 - Principal Investigator: None
 - Creation Location: b032
 - Owner Group: ingestor
 - Access Groups: None, B032
 - Contact Email: dmcreynolds@slf.gov
- Scientific Metadata:**

```

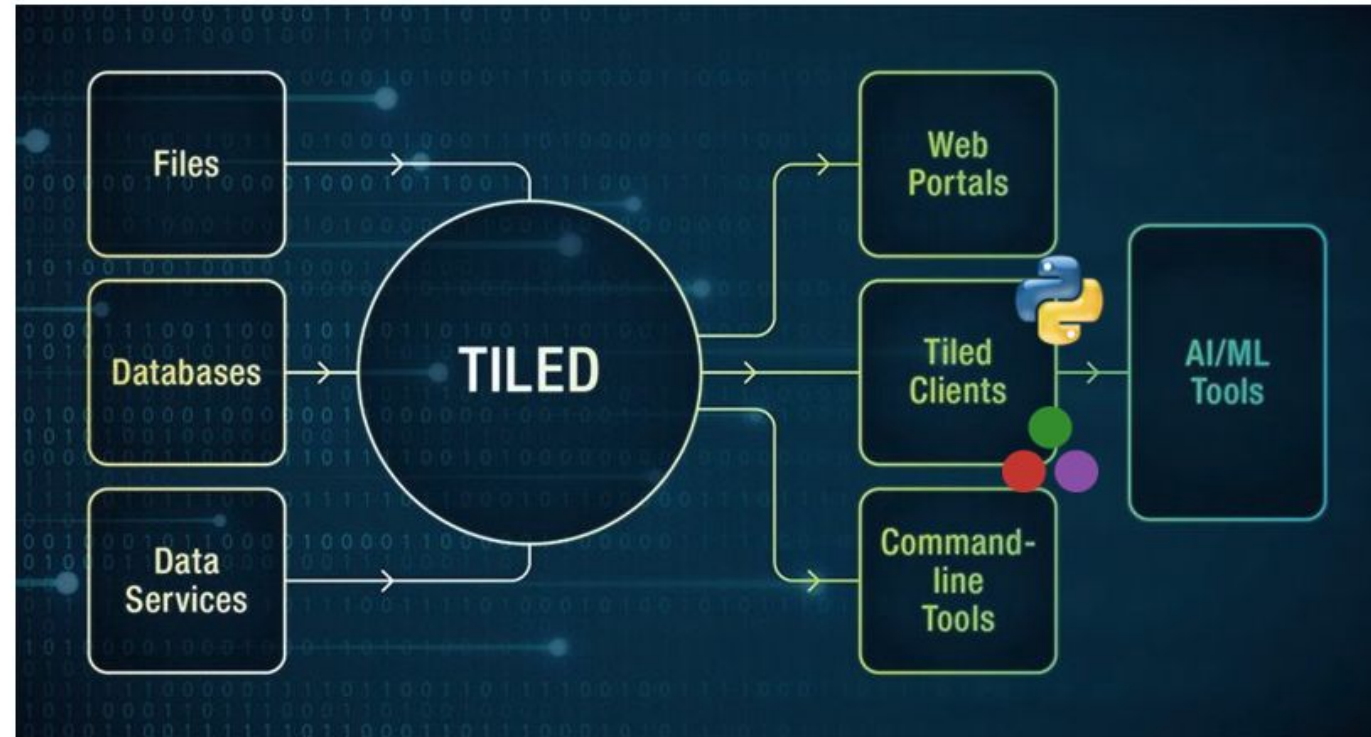
- /exchange/data: Object {"@": "3f53c198-becc-4ad3-9353-ff72a50b2a37/0", "65": "3f53c198-becc-4ad3-9353-ff72a50b2a37/65", "130": "3f53c198-becc-4ad3-9353-ff72a50b2a37/65", "195": "3f53c198-becc-4ad3-9353-ff72a50b2a37/65"}
- /measurement/instrument/attenuator/setup/filter_y: Array [1] [0]
- /measurement/instrument/camera_motor_stack/setup/camera.distance: Object {"@": "192-234492", "65": "192-234492", "130": "0", "195": "0", "260": "0", "325": "0", "390": "0"}
- /measurement/instrument/camera_motor_stack/setup/camera.elevation: Object {"@": "0", "65": "0", "130": "0", "195": "0", "260": "0", "325": "0", "390": "0"}
- /measurement/instrument/camera_motor_stack/setup/tilt_motor: Array [1] [0.1565]
- /measurement/instrument/detection_system/objective/camera_objective: "Optique 10X"
- /measurement/instrument/detection_system/scintillator/scintillator_type: "50 um LuAG"
- /measurement/instrument/detector/binning_x: Array [1] [1]
- /measurement/instrument/detector/binning_y: Array [1] [1]
- /measurement/instrument/detector/dark_field_value: Array [1] [100]
- /measurement/instrument/detector/delay_time: Array [1] [6]
- /measurement/instrument/detector/dimension_x: Array [1] [200]

```



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



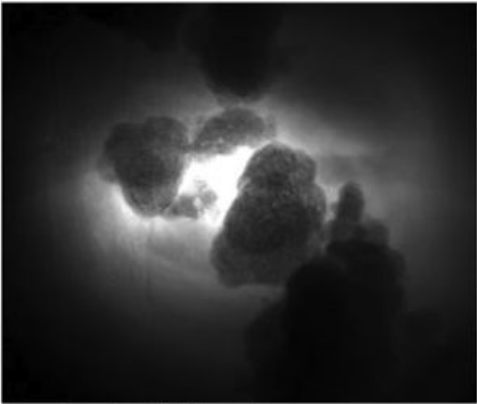
Data in a web browser, on a phone

TILED BROWSE

Top / fxi / raw / 1b0b4d73-6d87-43ab-8d62-ed035c51b9b4 / primary / data / Andor_image

VIEW DOWNLOAD METADATA DETAIL

Andor_image



Choose a planar cut through this 4-dimensional array.

i This large array has been downsampled by a factor of 2. Use the "Download" tab to access a full-resolution image.

0 22 44

0 10 19

TILED BROWSE

Top / fxi / raw / 1b0b4d73-6d87-43ab-8d62-ed035c51b9b4 / primary / data / Andor_image

VIEW DOWNLOAD METADATA DETAIL

Andor_image



Choose a planar cut through this 4-dimensional array.

i This large array has been downsampled by a factor of 3. Use the "Download" tab to access a full-resolution image.

0 22 44

0 10 19

TILED BROWSE

Top / fxi / raw / 1b0b4d73-6d87-43ab-8d62-ed035c51b9b4 / primary / data / Andor_image

VIEW **DOWNLOAD** METADATA DETAIL

Dimensions: 45 × 20 × 1080 × 1280


Slice (Optional) **EXAMPLES**

if blank, access entire array

Format *
CSV

DOWNLOAD **LINK** **OPEN**

Link
https://tiled-demo.blueskyproject.io/a



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?