

Introduction to FOXDEN



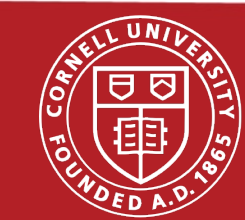
CHES Scientific Computing Team

Valentin Kuznetsov, Keara Soloway, Werner Sun, Rolf Verberg

Cornell University

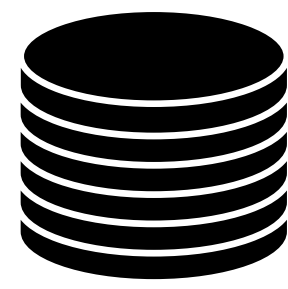
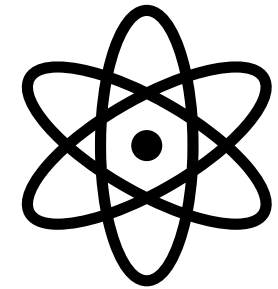
X-Cite workshop

June 17, 2026

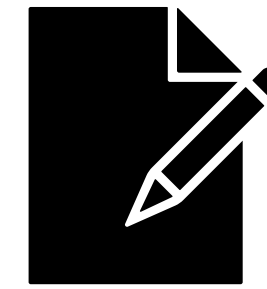
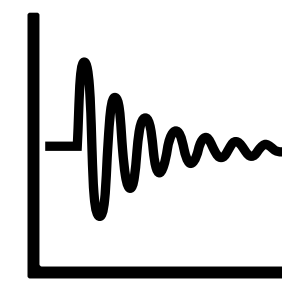
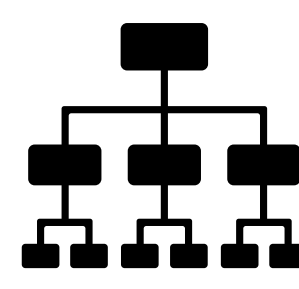


Scientific workflow(s)

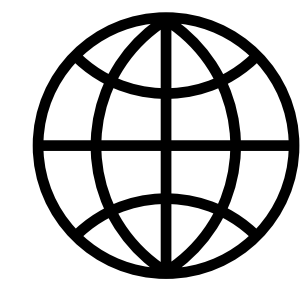
Data
Collection



Data
Processing

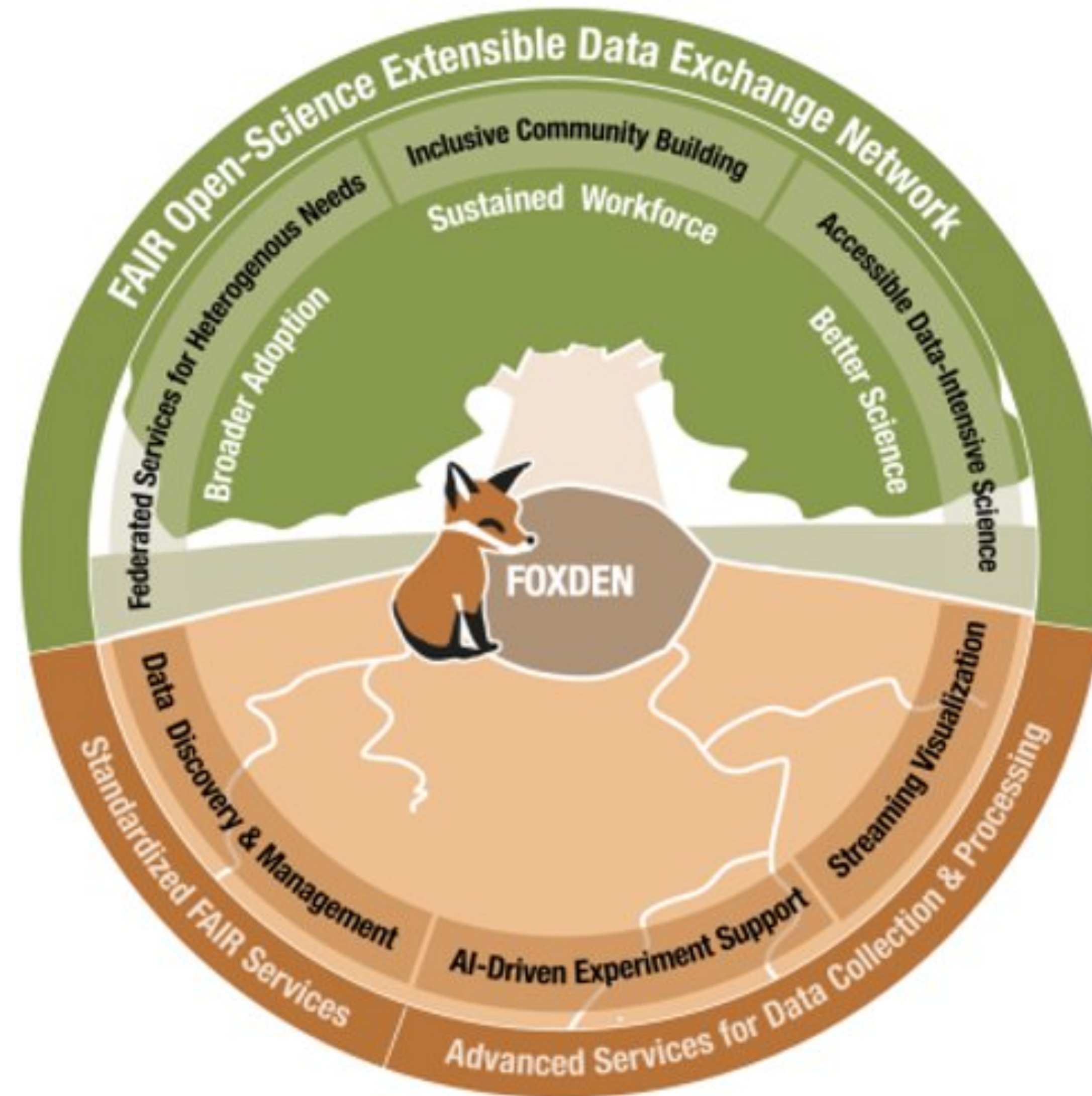


Publication



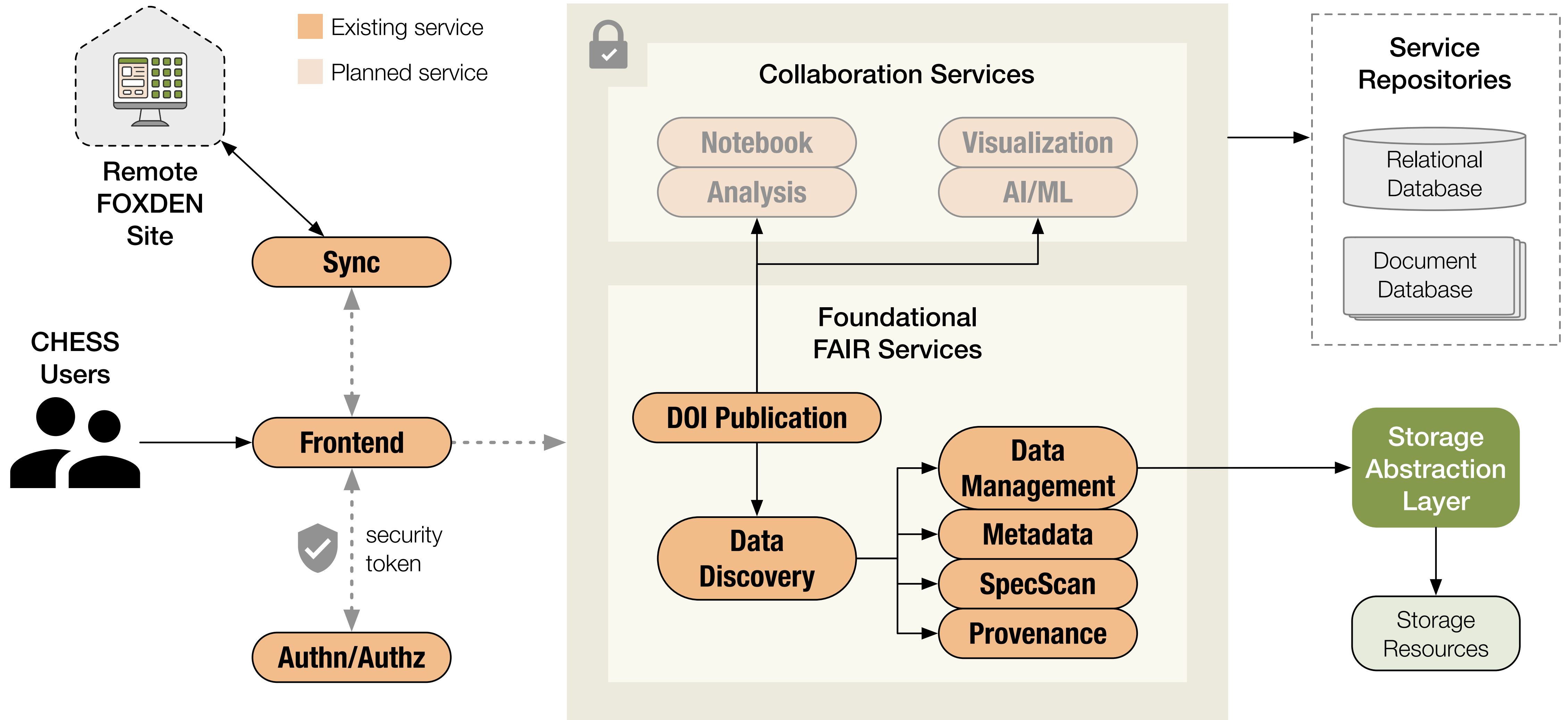


FAIR
Open-Science
eXtensible
Data
Exchange
Network

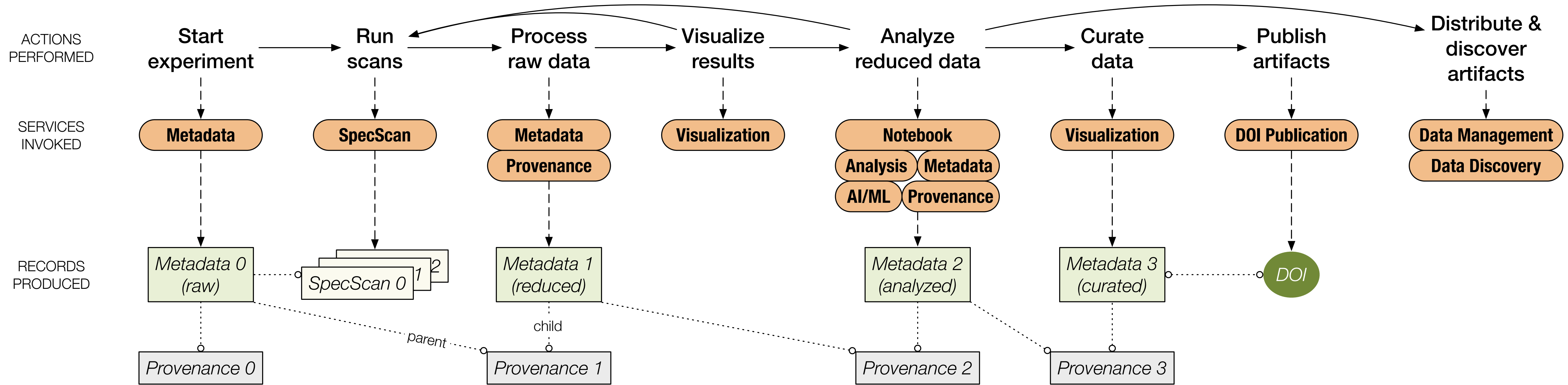


- A set of data services to:
 - Augments *existing* systems and workflows with capabilities for managing research artifacts
 - Service = web application with backend repository and command line API
 - Research artifacts = raw/reduced/analyzed data, metadata, provenance, code, visualizations, AI/ML models
- An encapsulation of data science expertise and best practices for domain scientists

FOXDEN Services



FOXDEN & Typical Research Workflow



FOXDEN services shadow the research workflow and help organize its metadata, capture provenance, and easy publication process





FOXDEN: datasets

User vek3, BTRs: [test-demo test-987-b test-123-a]



Show entries

ⓘ CASE INSENSITIVE SEARCH:

Filter:

Date	Beamline	Btr	Cycle	SampleName	Action
2025-10-24T15:20:02Z	3A	test-123-a	2025-3	meta-test-6	<input type="button" value="Record"/> <input type="button" value="Add note"/>
2025-09-19T13:21:18Z	3A	test-123-a	2024-1	test_sample	<input type="button" value="Record"/> <input type="button" value="Add note"/>
2025-09-19T13:21:12Z	3A	test-123-a	2024-1	test_sample	<input type="button" value="Record"/> <input type="button" value="Add note"/>

Page 1 of entries related to your BTR

Prev 2 3 4 5 ... 1,146 Next

TOTAL RECORDS IN FOXDEN: 11451

TOTAL RECORDS IN FOXDEN: 11451



FOXDEN: Record



CHES Data Management

Search Tools Docs New issue Logout

Showing 1—1 records out of 1 Sort by: Date ▾ first prev next last

User vek3, BTRs: [test-demo test-987-b test-123-a]

DatasetID (DID): /beamline=3a/btr=test-123-a/cycle=2025-3/sample_name=meta-test-6

Schema: ID3A, [SpecScan record\(s\)](#), [User metadata](#), [Raw data](#), [Reduced data](#), Date: Fri, 24 Oct 2025 15:20:02 UTC

Record Description Provenance User Amend Notes JSON Save **Create DOI**

```
affiliation      : [Basic]
alignment       : false
beam_energy     : 61.332
beamline        : [3A]
beamline_funding_partner : [CHESS_internal]
btr             : test-123-a
calibration     : false
cesr_conditions : [9x5_bunch_mode]
cycle           : 2025-3
data_location_beamtime_notes : /nfs/chess/aux/cycles/2025-3/id3a/test-123-a/metadata/meta-test-4/meta-te
data_location_meta : /nfs/chess/aux/metadata/cycles/2025-3/id3a/test-123-a/meta-test-6
data_location_raw : /nfs/chess/id3a/2025-3/test-123-a/meta-test-6
data_location_reduced : /nfs/chess/aux/reduced_data/cycles/2025-3/id3a/test-123-a/meta-test-6
data_location_scratch : /nfs/chess/scratch/cycles/2025-3/id3a/test-123-a/meta-test-6
q9t9_j0c9ftou_2c19fcp : \u1f2\cpe22\2c19fcp\cyc\2025-3\1q39\162f-153-9\w6f9-162f-0
q9t9_j0c9ftou_16qnc6q : \u1f2\cpe22\9nx\16qnc6q_q9t9\cyc\2025-3\1q39\162f-153-9\w6f9-162f-0
q9t9_j0c9ftou_19M : \u1f2\cpe22\1q39\2025-3\162f-153-9\w6f9-162f-0
q9t9_j0c9ftou_w0rg : \u1f2\cpe22\9nx\w0rg\cyc\2025-3\1q39\162f-153-9\w6f9-162f-0
```

FOXDEN: Command line tool



```
foxden <command> <action> [options]
```

```
foxden meta add metadata.json
```

```
foxden prov add provenance.json
```

```
foxden search pi:name
```

```
foxden view <did>
```

```
foxden prov ls datasets
```

```
foxden ml predict ...
```

```
# every command has its own help
```

```
foxden <command> -help
```

```
# get token with relevant scope
```

```
foxden token create write
```

```
# add new meta-data record
```

```
foxden meta add metadata.json
```

```
# add new provenance record
```

```
foxden prov add provenance.json
```

```
# find meta-data
```

```
foxden search {
```

```
---
```

```
DID      : abc
```

```
Schema   : ID3A
```

```
Cycle    : 2023-3
```

```
Beamline : [3A]
```

```
BTR      : 3731-b
```

```
Sample   : <nil>
```

```
# view meta-data
```

```
foxden view abc
```

```
---
```

```
### MetaData records:
```

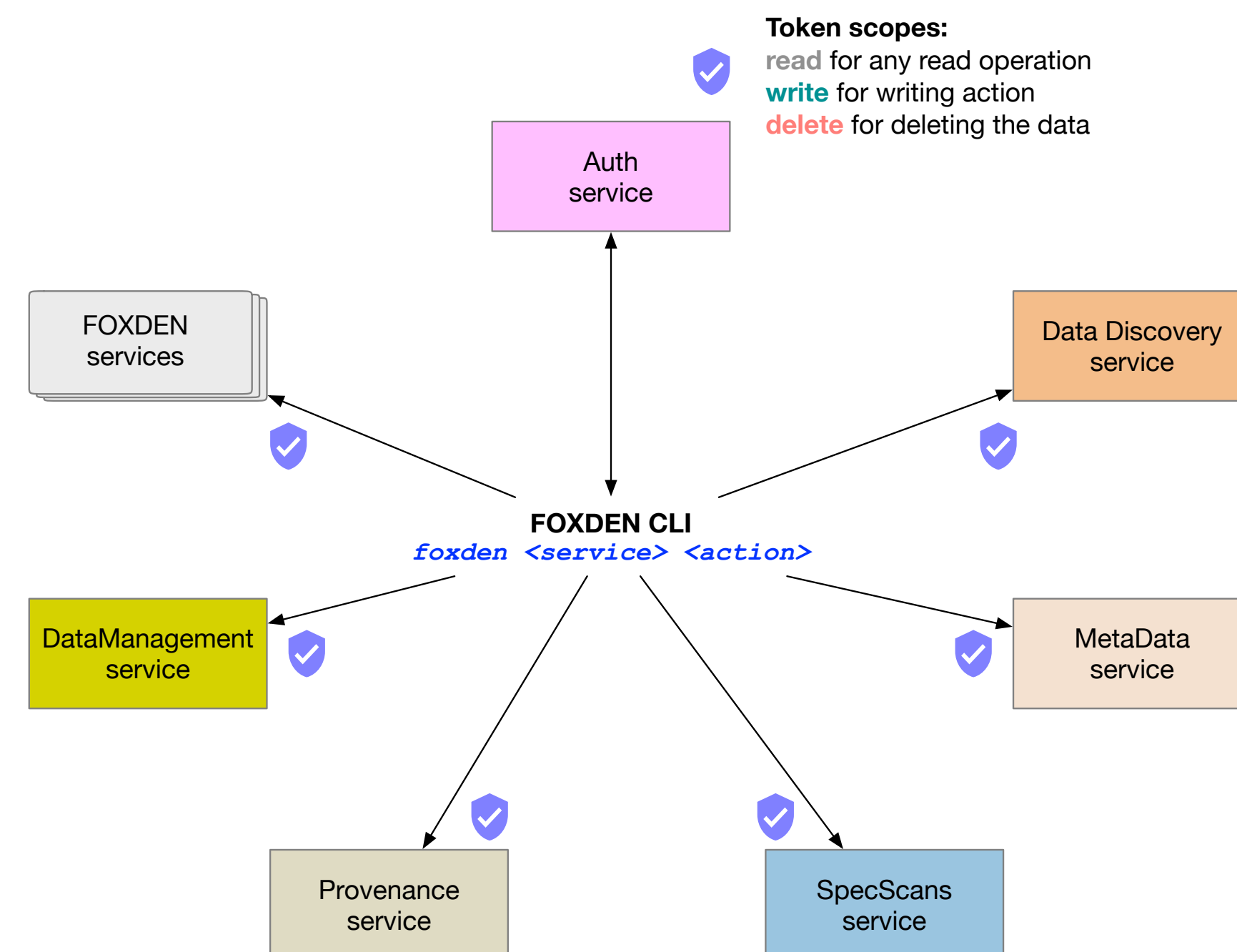
```
{"BTR": "3731-b", ...}
```

```
### Provenance dataset records:
```

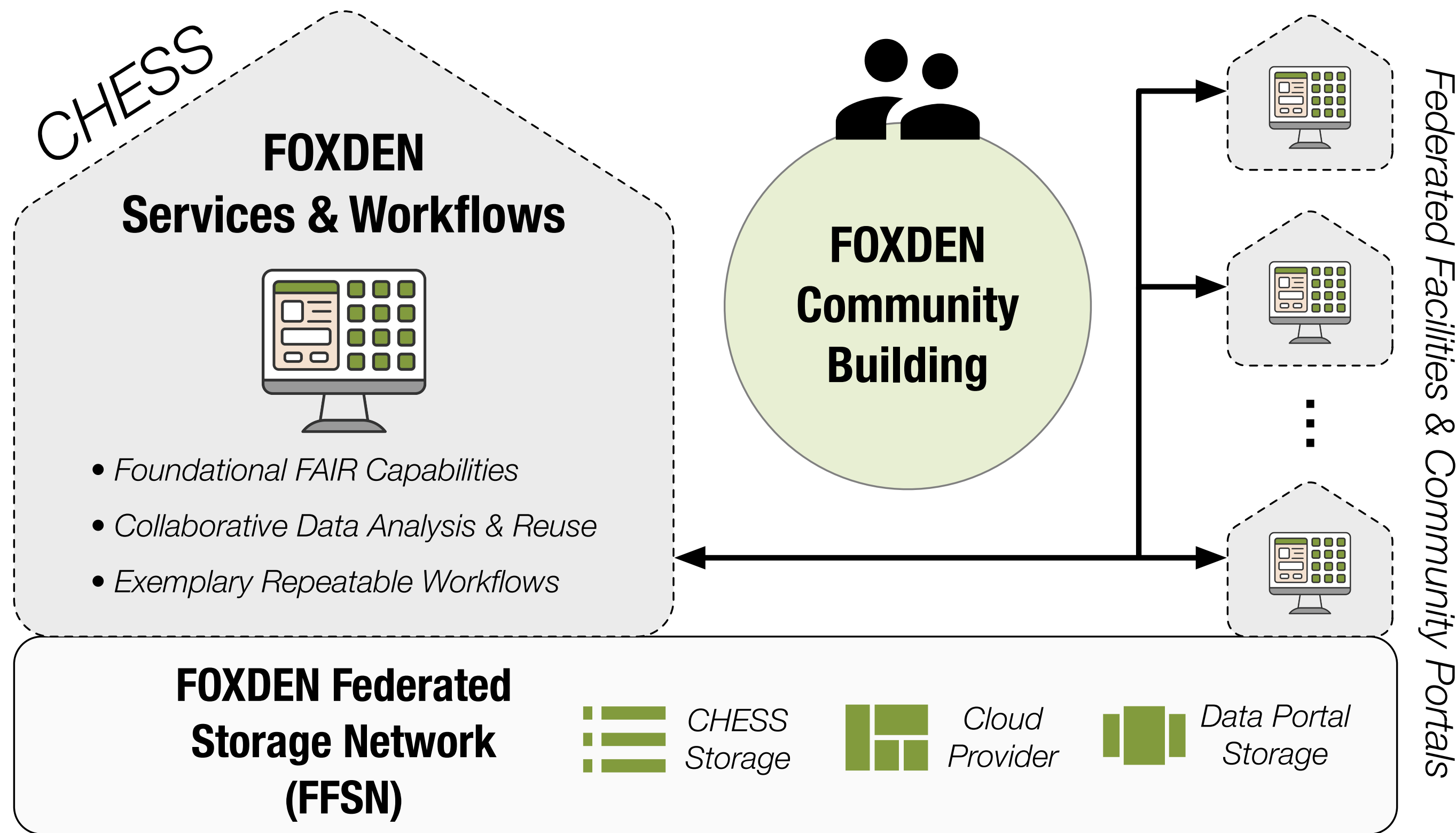
```
[{... "dataset": "/a/b/c", "meta_id": "abc" ...}]
```

```
### Provenance files records:
```

```
[{... "file": "/a/b/c/file1.png", ...}]
```





Future Vision: Federated Model




- **FOXDEN is domain-agnostic**
 - Services can be deployed at multiple facilities
 - Repositories periodically synchronized
 - Improves findability of facility data
- **Federated data storage**
 - FFSN: network of interoperable resources
 - Members contribute storage to a common pool
 - Facilitates dissemination, distributed computing
- **Streamline data sharing**
 - Hub for community building & data reuse by other users beyond original researchers
 - Use case: cross-facility datasets
 - Uniform interfaces to heterogeneous datasets, standardizes multi-modal analysis

FOXDEN: documentation

chesscomputing.github.io/FOXDEN/  

[Home](#) | [Git](#)



FOXDEN

FAIR Open-Science Extensible Data Exchange Network

FOXDEN documentation

- [Introduction](#)
- [Quick start guide](#)

FOXDEN internals

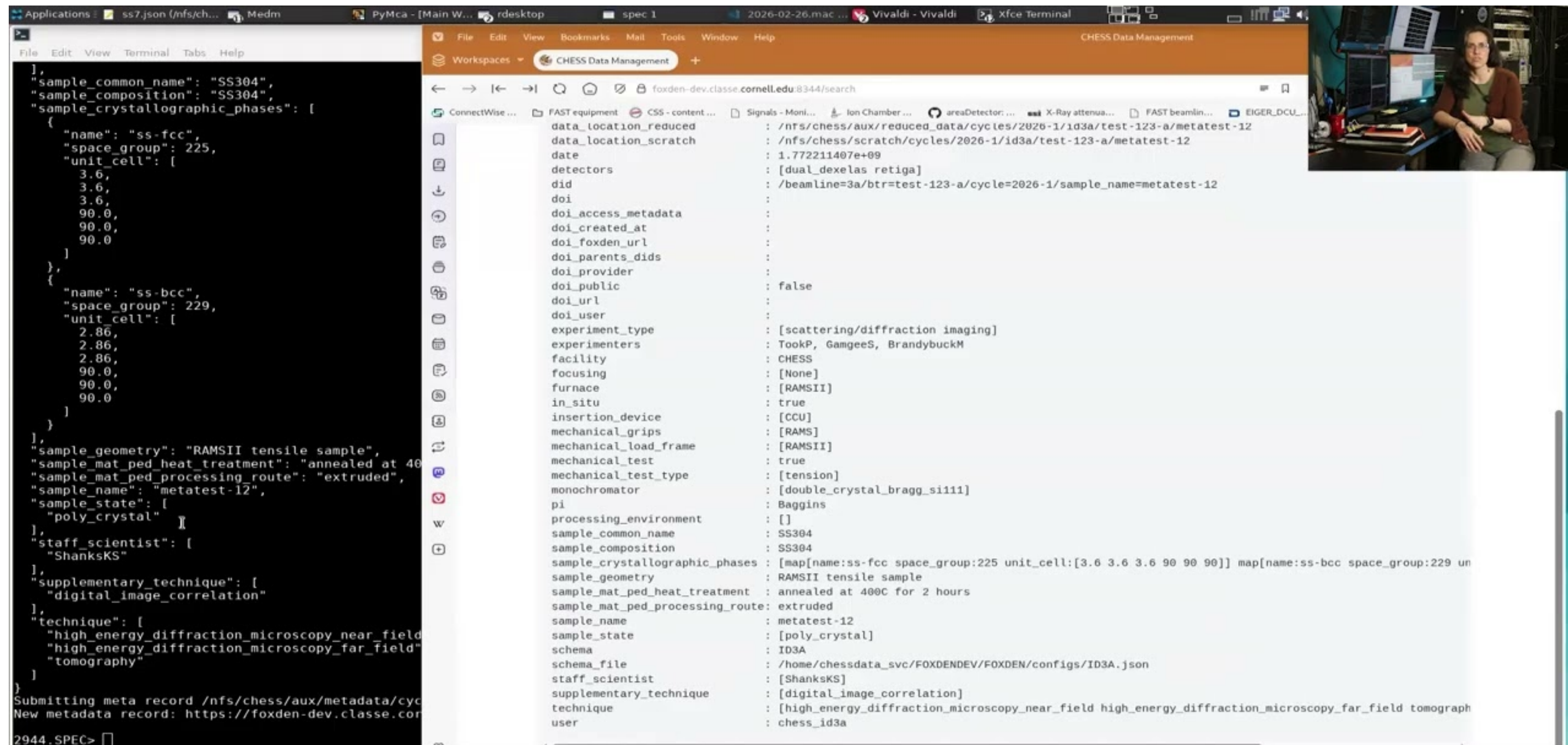
- [Architecture](#)
- [Configuration](#)
- [Deployment](#)
- [Schemas](#)
- [Dataset Identifier \(DID\)](#)
- [DOI support](#)

FOXDEN services

- [Frontend service \(web UI\)](#)
- [FOXDEN tools](#)



FOXDEN: overview



The screenshot shows a terminal window on the left and a web browser on the right. The terminal displays JSON metadata for a sample, including sample name, composition, crystallographic phases, geometry, and processing details. The web browser shows the same metadata rendered in a structured, key-value format.

```
},
"sample_common_name": "SS304",
"sample_composition": "SS304",
"sample_crystallographic_phases": [
  {
    "name": "ss-fcc",
    "space_group": 225,
    "unit_cell": [
      3.6,
      3.6,
      3.6,
      90.0,
      90.0,
      90.0
    ]
  },
  {
    "name": "ss-bcc",
    "space_group": 229,
    "unit_cell": [
      2.86,
      2.86,
      2.86,
      90.0,
      90.0,
      90.0
    ]
  }
],
"sample_geometry": "RAMSII tensile sample",
"sample_mat_ped_heat_treatment": "annealed at 400C",
"sample_mat_ped_processing_route": "extruded",
"sample_name": "metatest-12",
"sample_state": [
  "poly_crystal"
],
"staff_scientist": [
  "ShanksKS"
],
"supplementary_technique": [
  "digital_image_correlation"
],
"technique": [
  "high_energy_diffraction_microscopy_near_field",
  "high_energy_diffraction_microscopy_far_field",
  "tomography"
]
}
Submitting meta record /nfs/chess/aux/metadata/cyc
New metadata record: https://foxden-dev.classe.cornell.edu/3344/search/2944.SPEC>
```

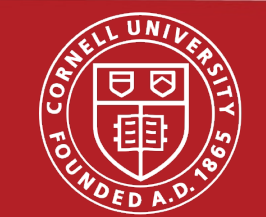
data_location_reduced	: /nfs/chess/aux/reduced_data/cycles/2026-1/1d3a/test-123-a/metatest-12
data_location_scratch	: /nfs/chess/scratch/cycles/2026-1/1d3a/test-123-a/metatest-12
date	: 1.772211407e+09
detectors	: [dual_dexelas retiga]
did	: /beamline=3a/btr=test-123-a/cycle=2026-1/sample_name=metatest-12
doi	:
doi_access_metadata	:
doi_created_at	:
doi_foxden_url	:
doi_parents_dids	:
doi_provider	:
doi_public	: false
doi_url	:
doi_user	:
experiment_type	: [scattering/diffraction imaging]
experimenters	: TookP, GamgeeS, BrandybuckM
facility	: CHESS
focusing	: [None]
furnace	: [RAMSII]
in_situ	: true
insertion_device	: [CCU]
mechanical_grips	: [RAMS]
mechanical_load_frame	: [RAMSII]
mechanical_test	: true
mechanical_test_type	: [tension]
monochromator	: [double_crystal_bragg_si111]
pi	: Baggins
processing_environment	: []
sample_common_name	: SS304
sample_composition	: SS304
sample_crystallographic_phases	: [map[name:ss-fcc space_group:225 unit_cell:[3.6 3.6 3.6 90 90 90]] map[name:ss-bcc space_group:229 un
sample_geometry	: RAMSII tensile sample
sample_mat_ped_heat_treatment	: annealed at 400C for 2 hours
sample_mat_ped_processing_route	: extruded
sample_name	: metatest-12
sample_state	: [poly_crystal]
schema	: ID3A
schema_file	: /home/chessdata_svc/FOXDENDEV/FOXDEN/configs/ID3A.json
staff_scientist	: [ShanksKS]
supplementary_technique	: [digital_image_correlation]
technique	: [high_energy_diffraction_microscopy_near_field high_energy_diffraction_microscopy_far_field tomograph
user	: chess_id3a

Now I said I would also show an alternative to this interactive approach.

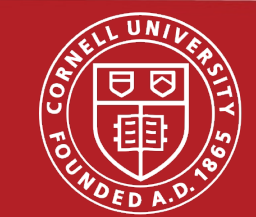


FOXDEN DEMO

<https://foxden-demo.classe.cornell.edu:18344/>



BACKUP SLIDE(S)



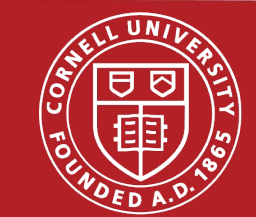
FOXDEN internals

- FOXDEN is designed to work with arbitrary schemas
 - but each schema file should define data-type, optionality of attribute, etc.
 - CHESS beamlines provide their own schemas
 - schema can be nested, DOI schema can be embedded into all CHESS beamlines schema, etc.
- The Dataset Identifier (DID) is defined as a path of key-value pairs, e.g. /key1=value1/key2=value2/...
- DID are used to connect all FOXDEN services, like Provenance, Spec Scans, Data Management, Publication and refer to identify our datasets and define parent-child relationships



FOXDEN Services

- **Frontend:**
 - represent data in different views; provides uniform web UI and access to all FOXDEN services
- **Authz:**
 - handles LDAP, OAuth, OIDC providers and manage user's tokens
- **Metadata:**
 - experiment metadata are stored according to beamline schemas; user's metadata can be in free (unstructured) form
- **Provenance:**
 - provenance information is stored in RDBMS to allow easily scaling with data growth
 - we capture provenance information at different levels: OS, Environment, Scripts, Files, etc.
- **DataDiscovery**
 - discover and aggregate results across multiple FOXDEN services using common Query Language (QL)
- **DOI publication**
 - full integration with DataCite, MaterialsCommons and Zenodo DOI providers
- **AI/ML, Data Management, Notebook, Sync services**
 - supplementary services for specific tasks (still evolving)



FOXDEN Architecture

• Services

- FOXDEN is composed by loosely coupled services via HTTP protocol/APIs
- Each service is designed to take care of single task
 - currently we operate with 10 individual services
 - choice of service backend is driven by its use-case and scalability, e.g. Metadata are stored into document-oriented database (e.g. MongoDB) while Provenance data are backed by RDBMS (multiple backends are supported)
- Data are interconnected via common Dataset Identifier (DID)

• Web frontend

- Users mostly interact with FOXDEN via web frontend
- Data exploration, publication and management
- Describes raw and reduced datasets

• FOXDEN CLI tool

- One tool to answer all your questions about FOXDEN
- Authentication, Metadata, SpecScans, Provenance management, DOI publications, synchronization, etc.

• Technical details

- Code is written in Go language and (mostly) compiled into static executables
- Zero setup/dependency requirements, OS agnostic, available on all hardware architectures (x86, ARM, Power8, RISC)
- Build-in concurrency, minimal hardware requirements, JSON for data exchange and Yaml for configuration
- Easily integrated with any software stack (via HTTP protocol)

