Experimental data transfer system BENTEN at SPring-8

Takahiro Matsumoto, Shigeru Yokota, Tomohiro Matsushita, Kengo Nakada,
Yuji Hiraoka, Masahiko Kodera, Yukito Furukawa, Akihito Yamashita
Japan Synchrotron Radiation Research Institute (JASRI)

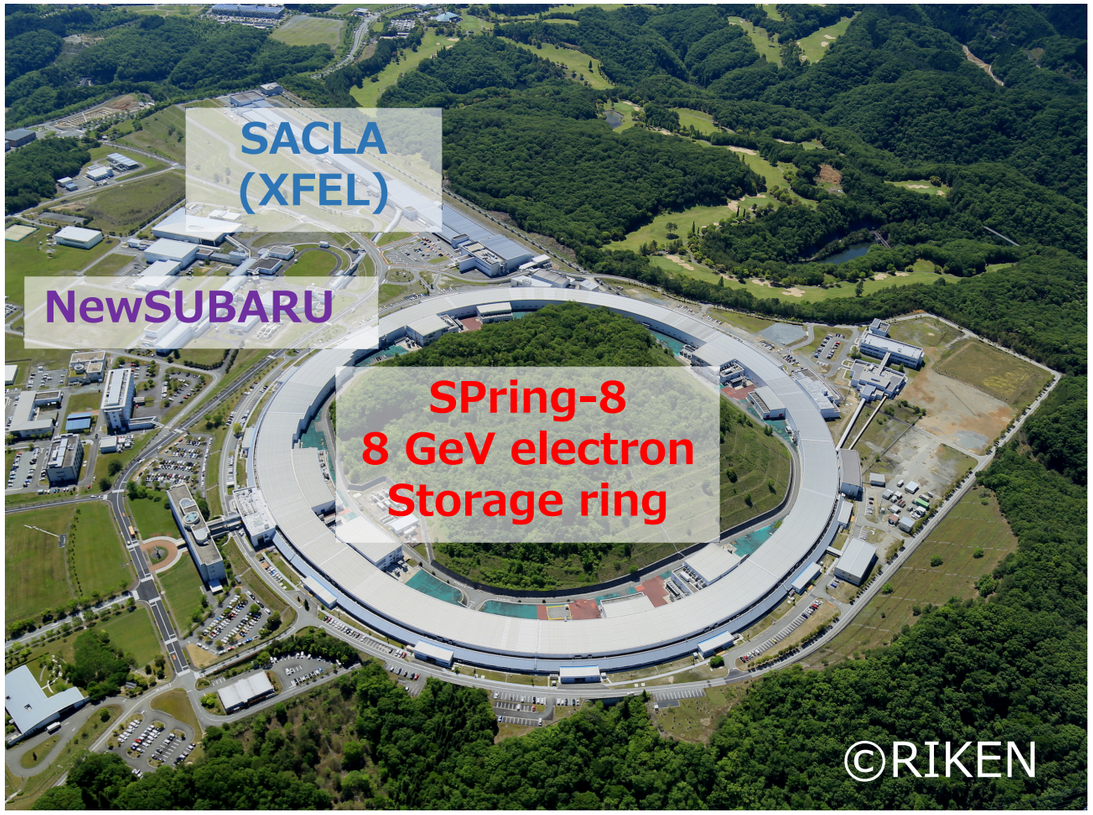
8th October, 2019

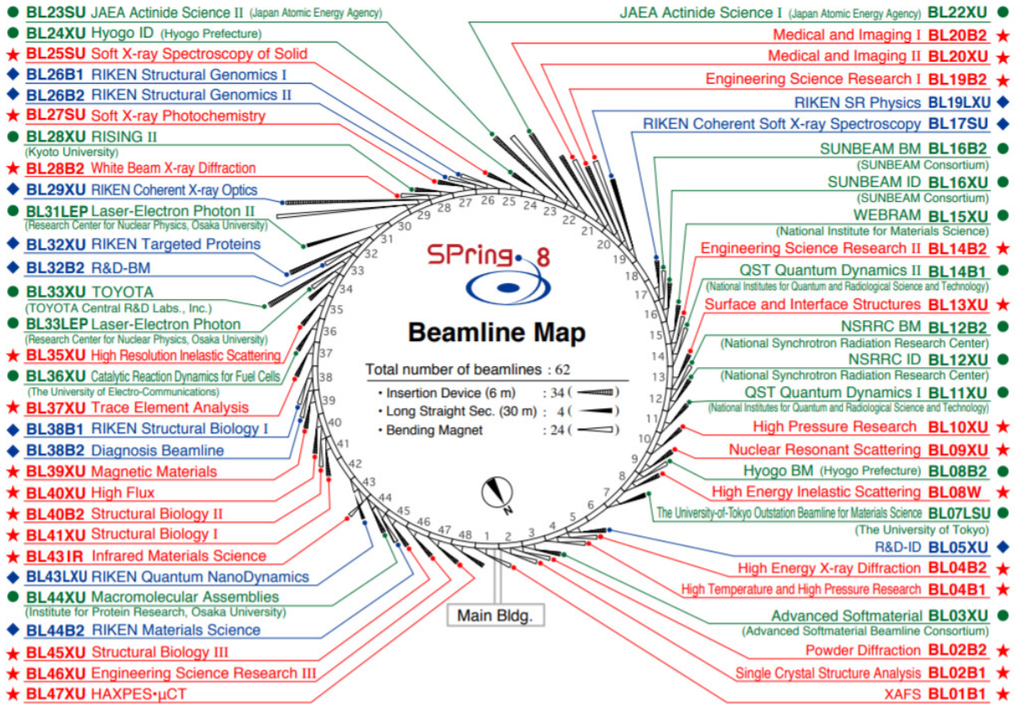
Outline

- Introduction
- Experimental data transfer system BENTEN
- Operation of BENTEN at SPring-8
- Summary and Future plan

Introduction

- SPring-8 is opened for scientific researchers and industrial users
- In total, ~17,000 users/year come to SPring-8 for experiments



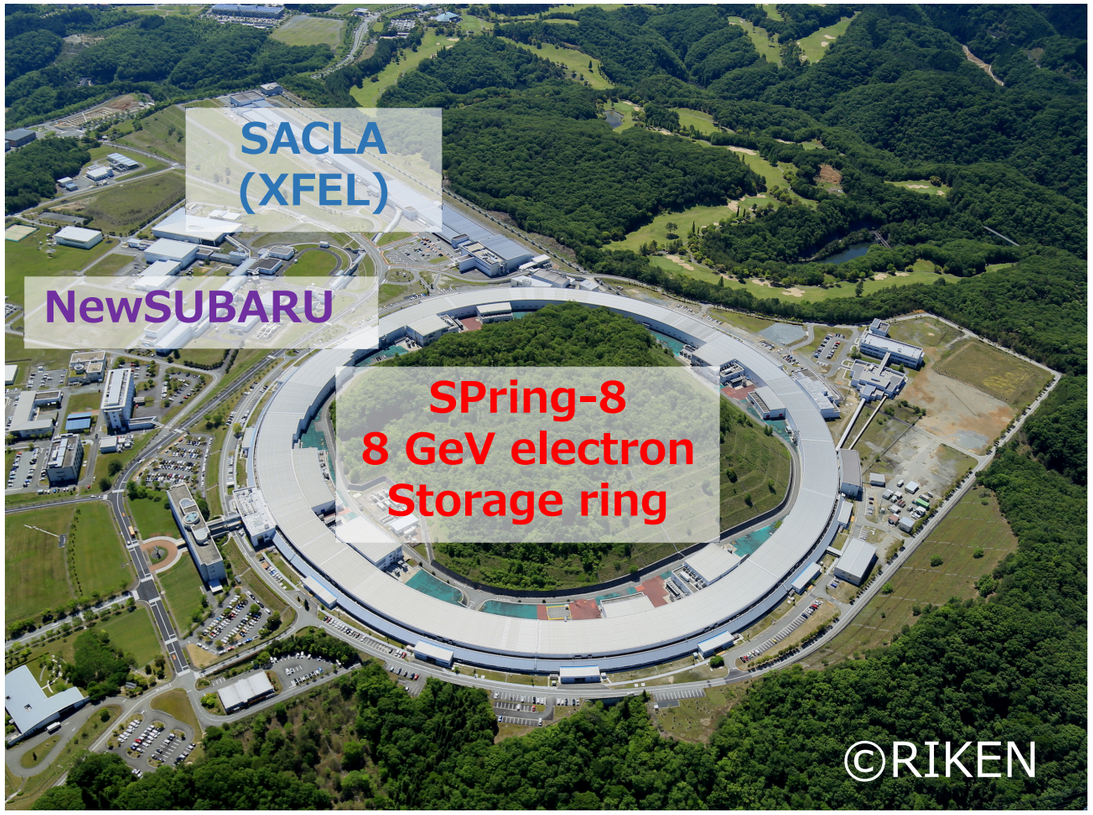


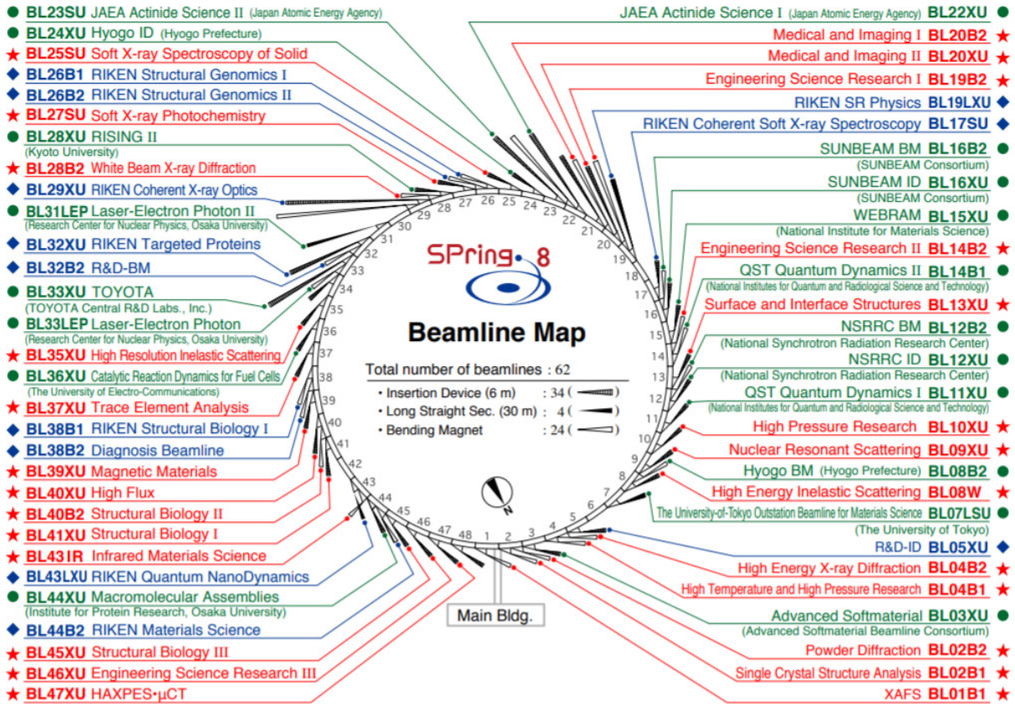
Experimental stations in 57 beamlines

Introduction

- SPring-8 is opened for scientific researchers and industrial users
- In total, ~17,000 users/year come to SPring-8 for experiments

➔Required Experimental data transfer for Data access through the Internet





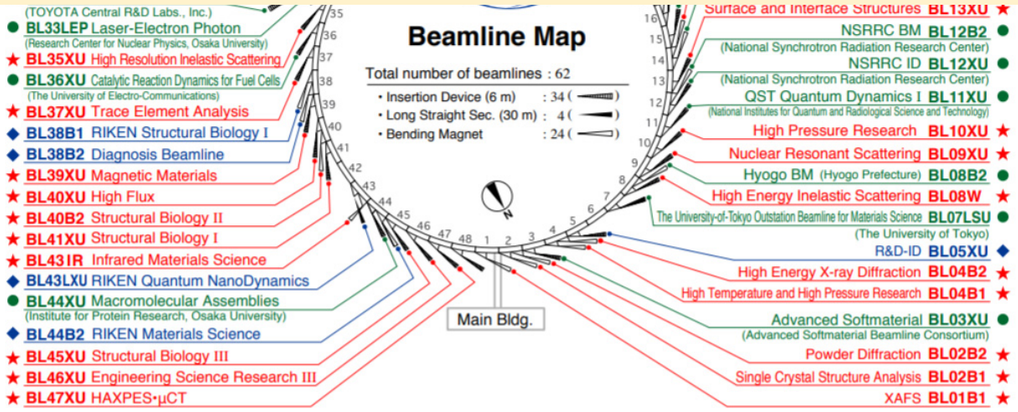
Experimental stations in 57 beamlines

Introduction

- SPring-8 is opened for scientific researchers and industrial users
- In total, ~16,000 users/year come to SPring-8 for experiments

➔Required Experimental data transfer for Data access through the Internet

- Open data access
  - Data science such as material informatics
  → Recently attracting public attentions



Experimental stations in 57 beamlines

Introduction

- SPring-8 is opened for scientific researchers and industrial users
- In total, ~16,000 users/year come to SPring-8 for experiments

➔Required Experimental data transfer for Data access through the Internet

- Open data access
  - Data science such as material informatics
  → Recently attracting public attentions
- Restricted data access
  - Remote experiment
  - Proxy measurements

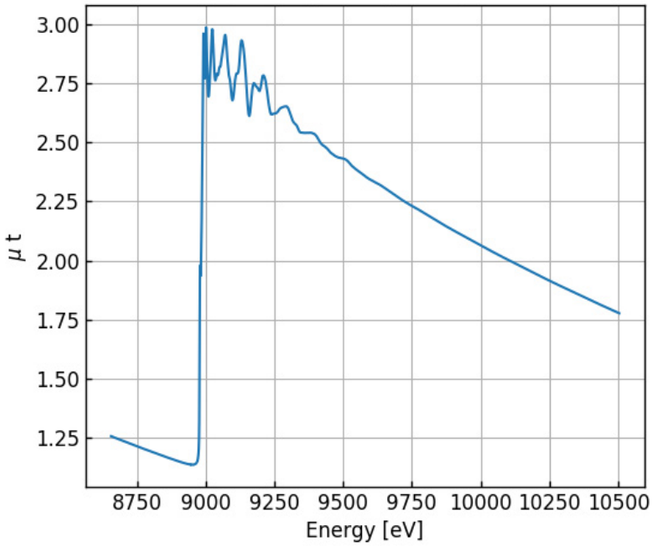
Experimental stations in 57 beamlines

An example of open data at SPring-8

- XAFS spectrum data for Standard sample
  - **XAFS**: **X**-ray **A**bsorption **F**ine **S**tructure
  - ~ 800 measured data at SPring-8: 2nd in the world

  → Utilized as reference for the measurement

Example of
XAFS spectrum



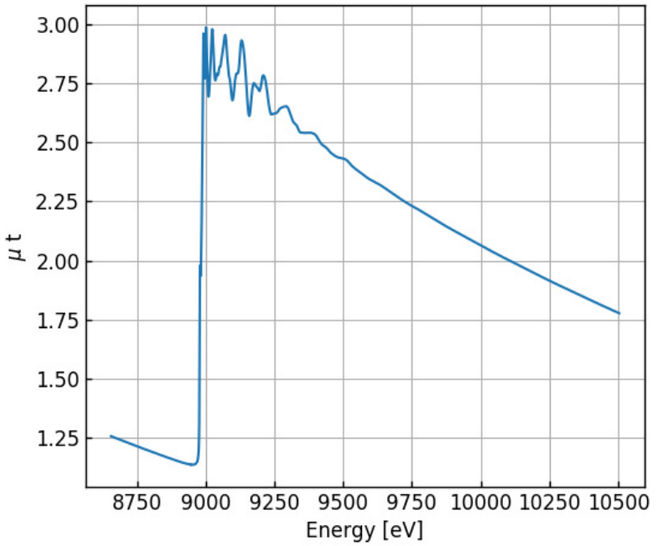
An example of open data at SPring-8

- XAFS spectrum data for Standard sample
  - **XAFS**: **X**-ray **A**bsorption **F**ine **S**tructure
  - ~ 800 measured data at SPring-8: 2nd in the world

  → Utilized as reference for the measurement

- Operated Web portal
  for data access through the Internet (2013~)

Example of
XAFS spectrum



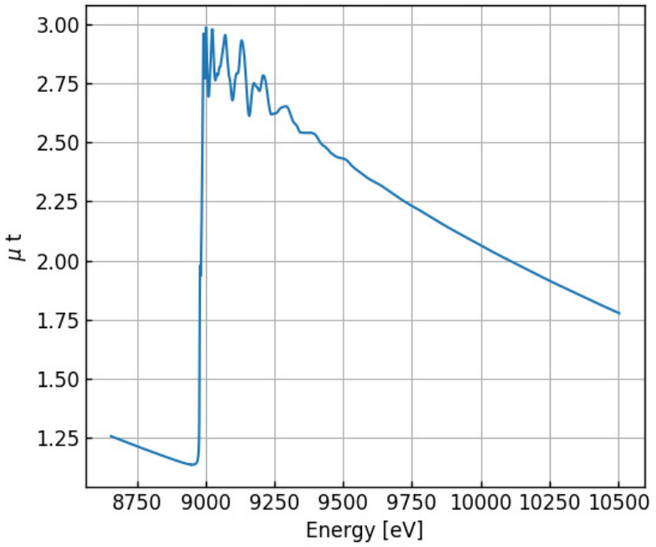
An example of open data at SPring-8

- XAFS spectrum data for Standard sample
  - **XAFS**: **X**-ray **A**bsorption **F**ine **S**tructure
  - ~ 800 measured data at SPring-8: 2nd in the world

  → Utilized as reference for the measurement

- Operated Web portal
  for data access through the Internet (2013~)

Example of
XAFS spectrum



We newly developed

Experimental data transfer system BENTEN
to promote data utilization for
public experimental stations at SPring-8

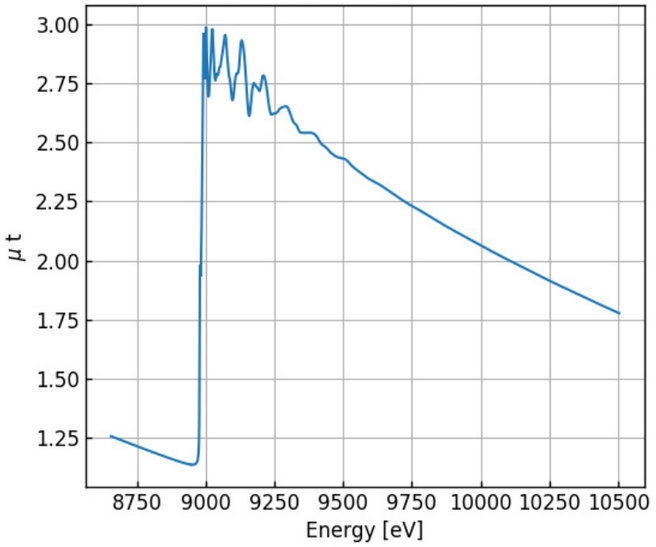
An example of open data at SPring-8

- XAFS spectrum data for Standard sample
  - **XAFS**: **X**-ray **A**bsorption **F**ine **S**tructure
  - ~ 800 measured data at SPring-8: 2nd in the world

  → Utilized as reference for the measurement

- Operated Web portal
  for data access through the Internet (2013~)

Example of
XAFS spectrum



We newly developed

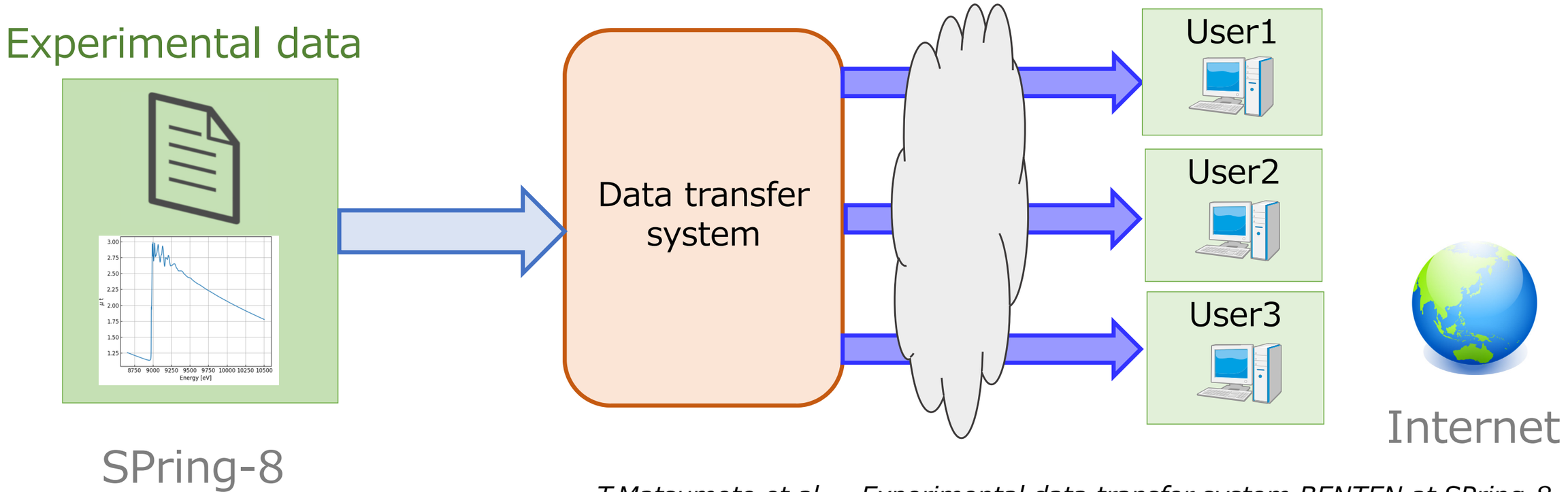
Experimental data transfer system BENTEN
to promote data utilization for
public experimental stations at SPring-8

Requirements in Experimental data transfer (1)

- Easy-to-use user interface

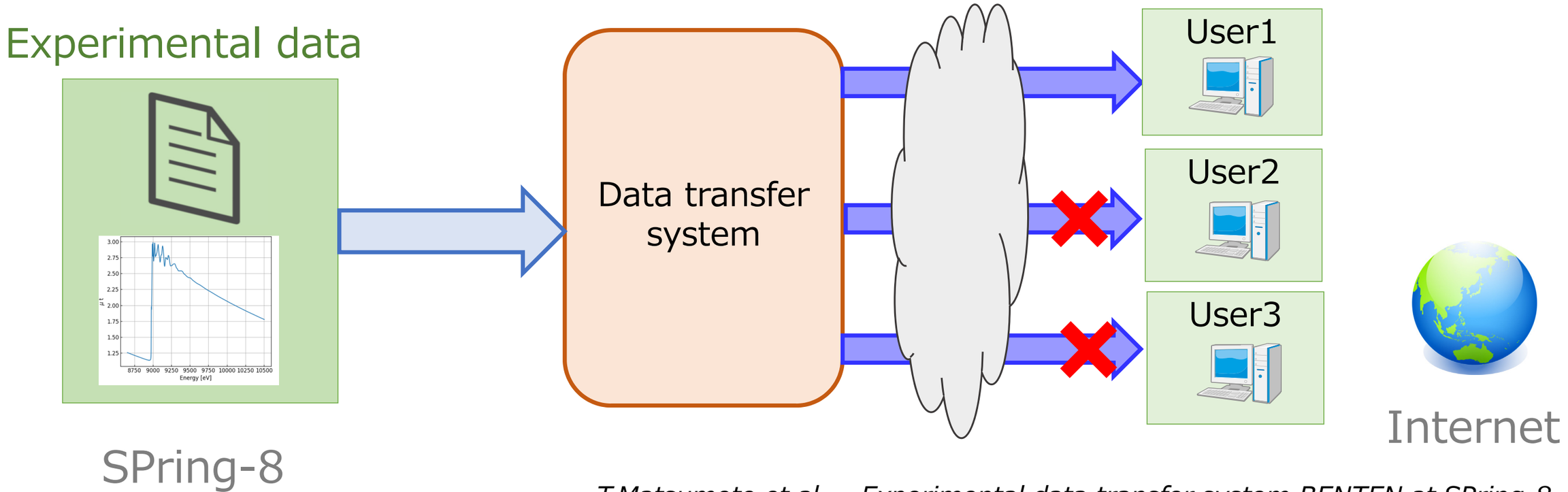
Requirements in Experimental data transfer (1)

- Easy-to-use user interface
- Control range of data sharing
  - **Case1: Open data access**



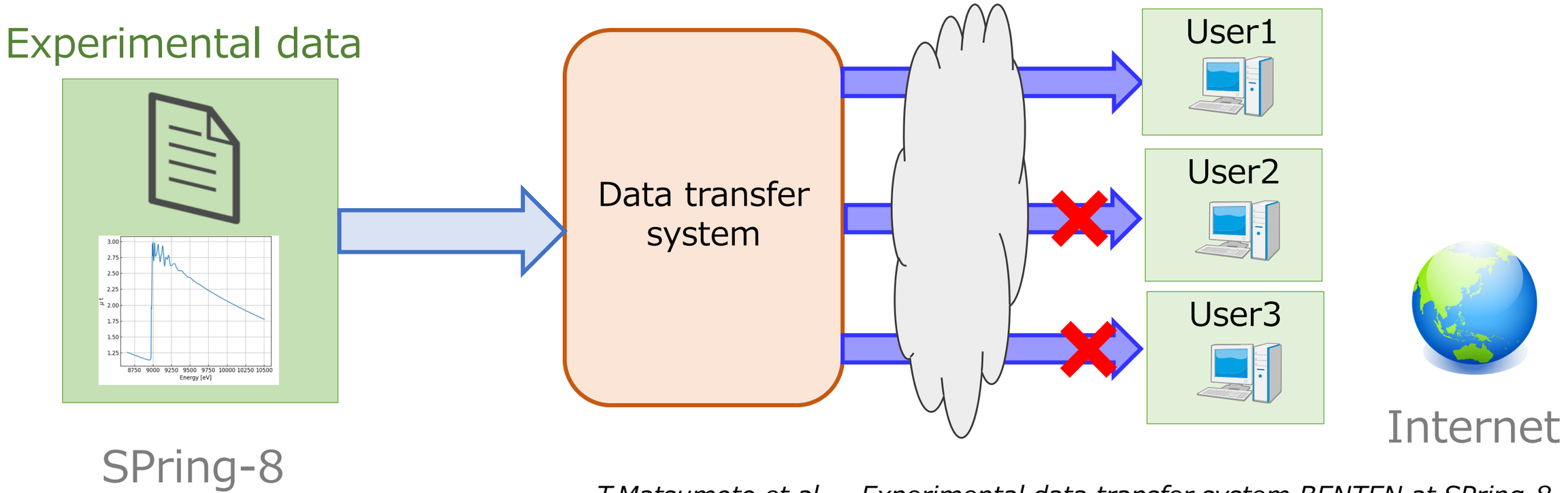
Requirements in Experimental data transfer (1)

- Easy-to-use user interface
- Control range of data sharing
  - **Case2: Restricted data access**



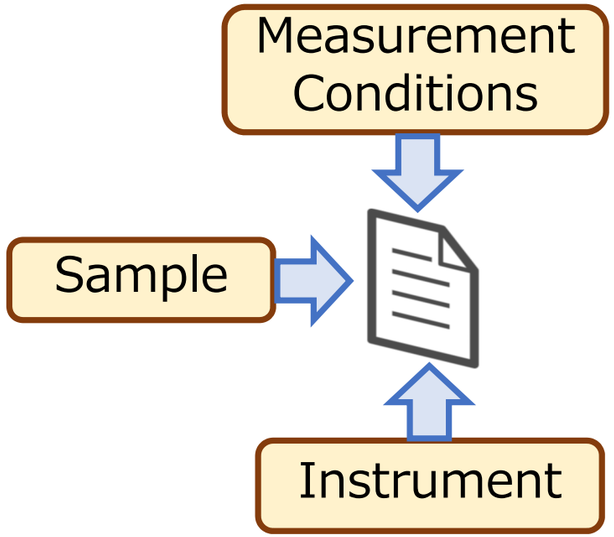
Requirements in Experimental data transfer (1)

- Easy-to-use user interface
- Control range of data sharing ➔ **Authentication**
  - **Case2: Restricted data access**



Requirements in Experimental data transfer (2)

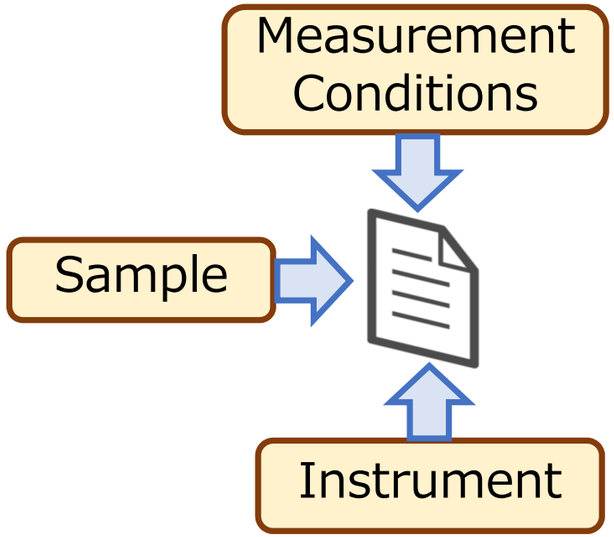
- Metadata management



- Easy to describe metadata in the experiments
  → To provide sufficient metadata to utilize data
  - Automated metadata extraction is desired

Requirements in Experimental data transfer (2)

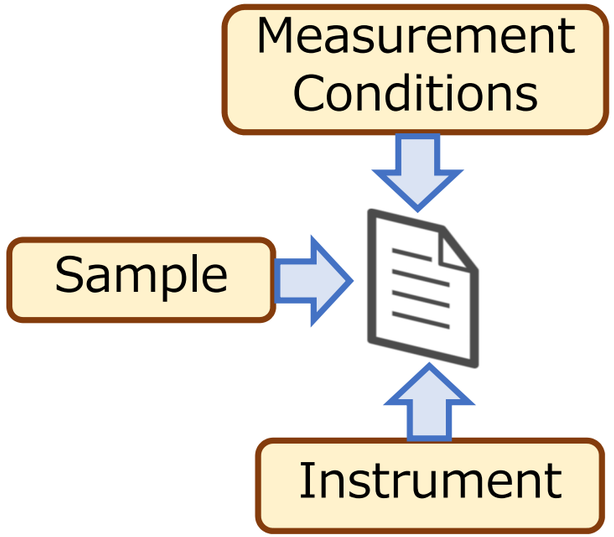
- Metadata management



- Easy to describe metadata in the experiments
  → To provide sufficient metadata to utilize data
  - Automated metadata extraction is desired
- Machine readability to utilize data in AI

Requirements in Experimental data transfer (2)

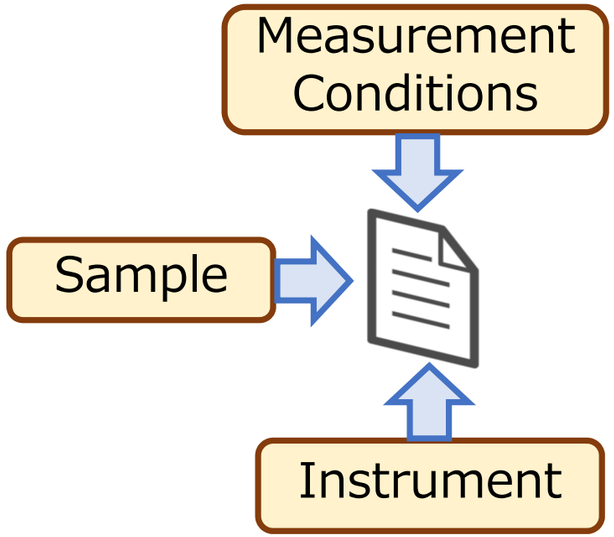
- Metadata management



- Easy to describe metadata in the experiments
  → To provide sufficient metadata to utilize data
  - Automated metadata extraction is desired
- Machine readability to utilize data in AI
- Flexible data search using metadata items

Requirements in Experimental data transfer (2)

- Metadata management



- Easy to describe metadata in the experiments
  → To provide sufficient metadata to utilize data
  - Automated metadata extraction is desired
- Machine readability to utilize data in AI
- Flexible data search using metadata items

- Data management
  - Data life cycle
    - Create, Update, Open/Closed access, Delete
  - Define items for Open Data
    - PID (Persistent ID), Contact name and its affiliation

BENTEN system Overview

Experimental station

BENTEN system

User

Data registration terminal

Data access terminal

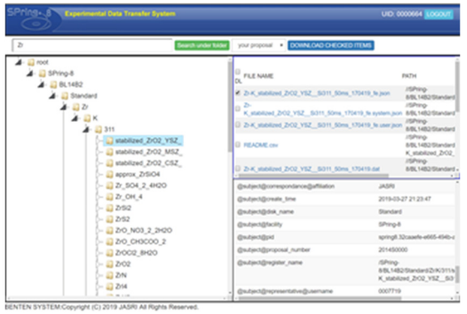
Local storage

Data register command

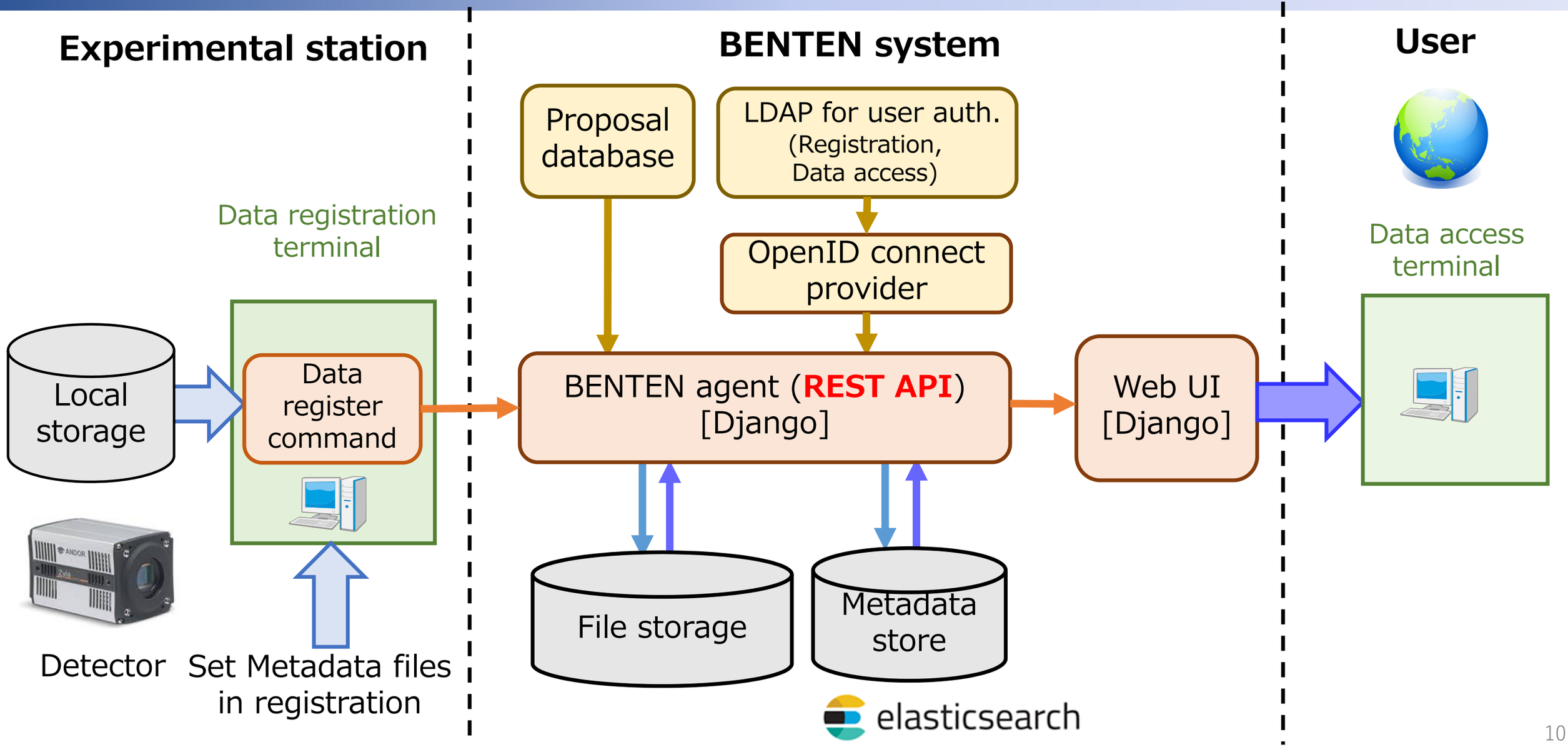
Data transfer system

Detector

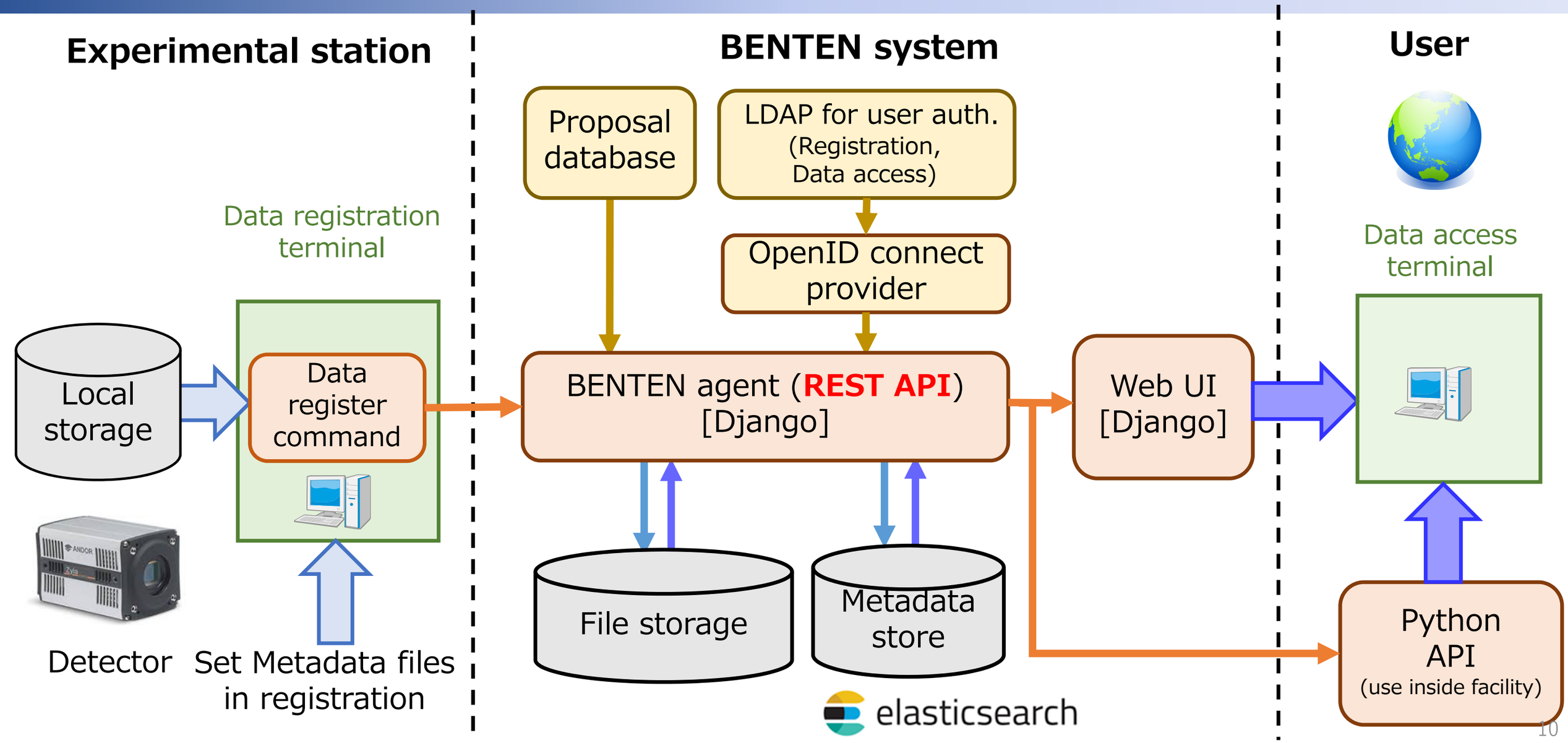
Set Metadata files in registration



BENTEN system Overview



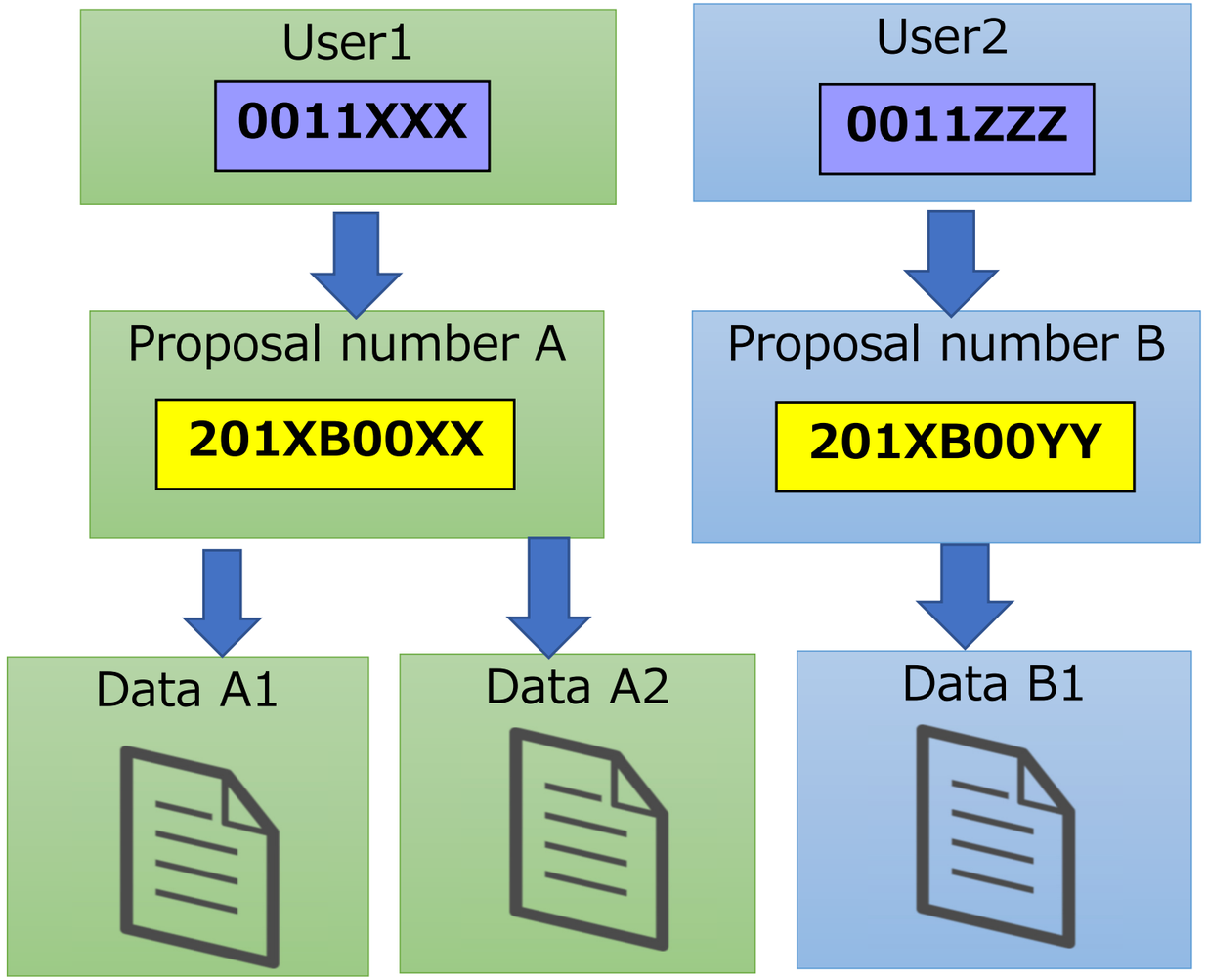
BENTEN system Overview



Authorization for data access

OpenID connect 1.0
for authentication

How to control
range for data sharing?



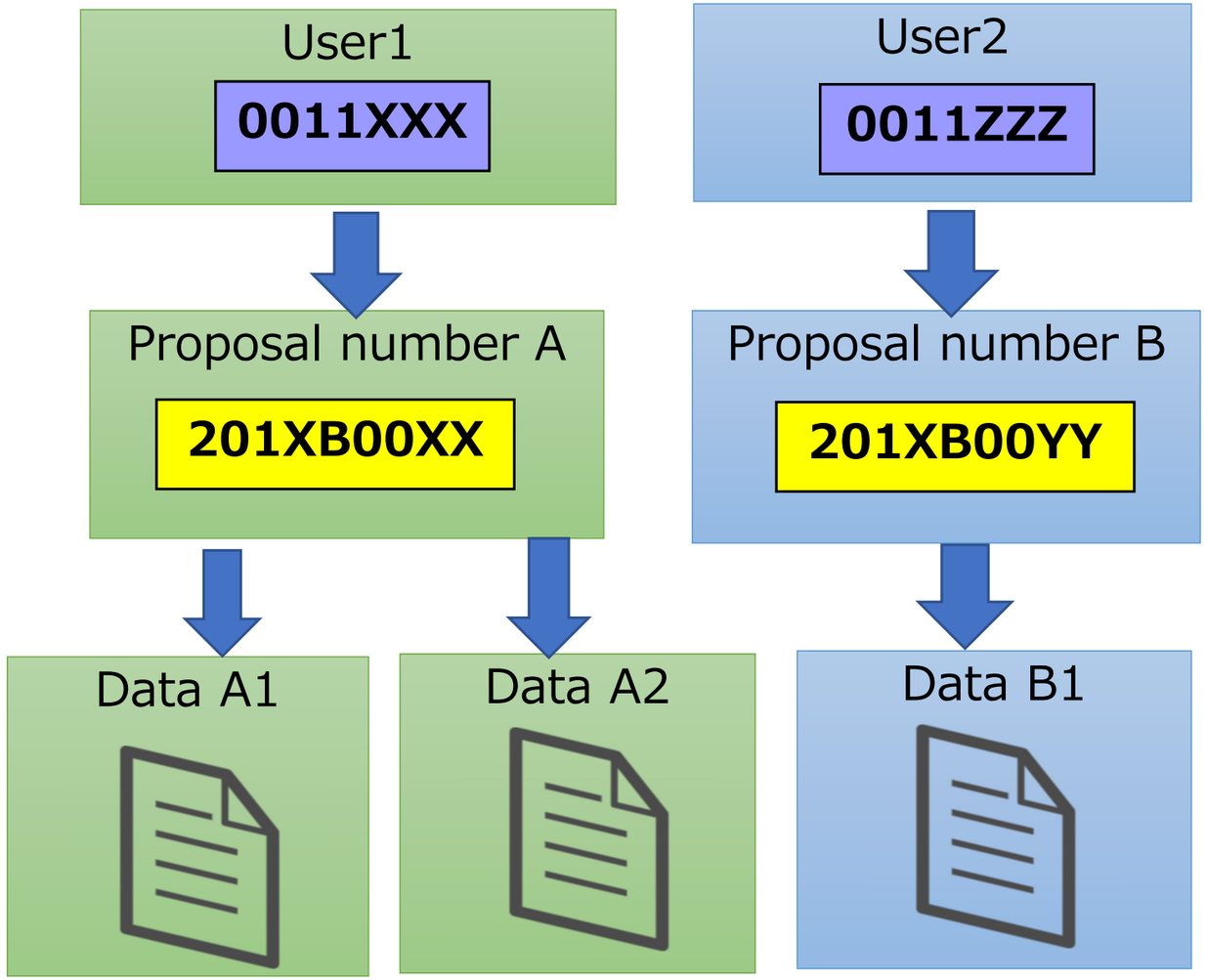
Authorization for data access

OpenID connect 1.0
for authentication

How to control range for data sharing?

- **Proposal number** is assigned for each experiment at SPring-8

→ Use the proposal number for authorization



Authorization for data access

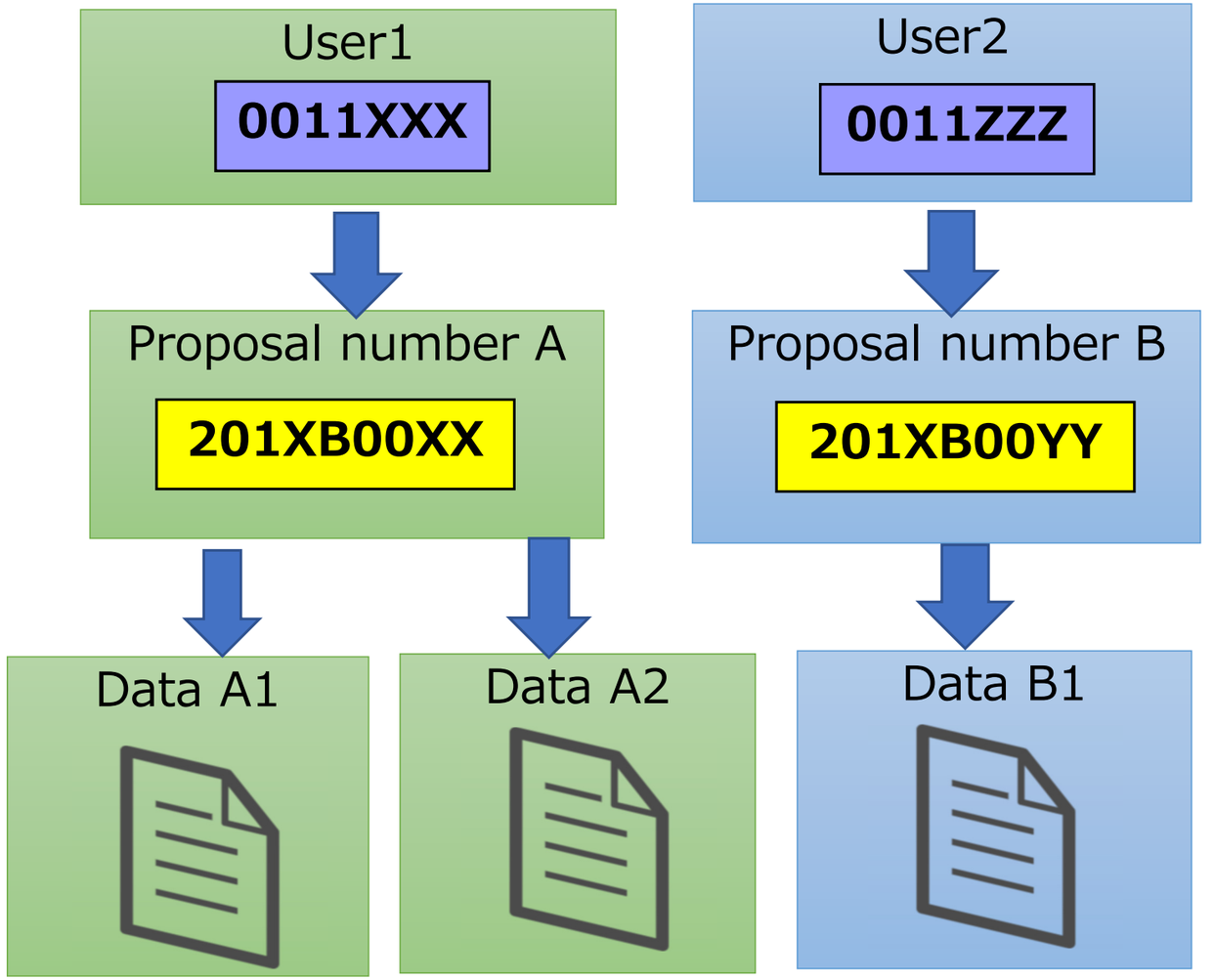
OpenID connect 1.0
for authentication

How to control range for data sharing?

- **Proposal number** is assigned for each experiment at SPring-8

→ Use the proposal number for authorization

- Procedure
  - Request user to login



Authorization for data access

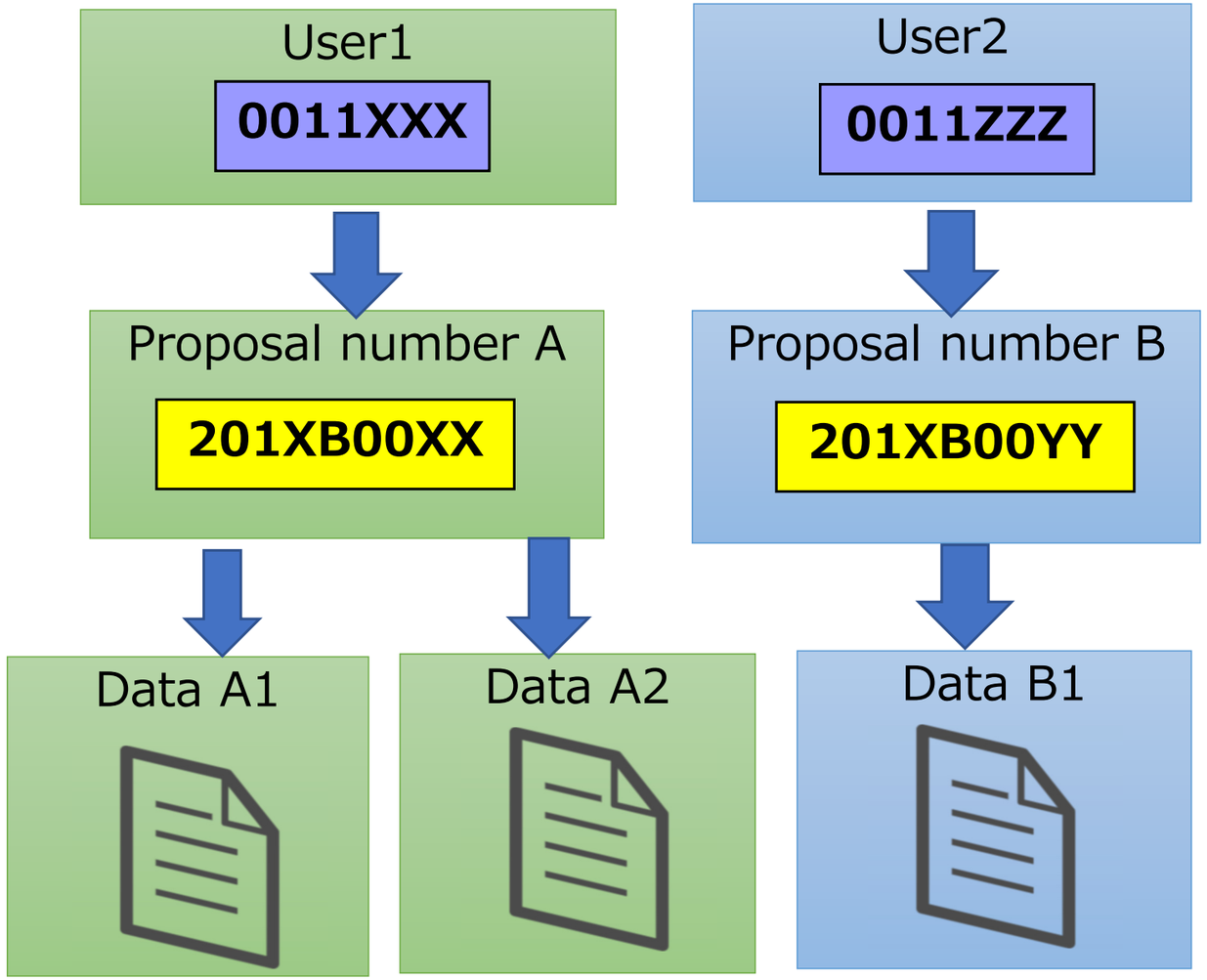
OpenID connect 1.0
for authentication

How to control range for data sharing?

- **Proposal number** is assigned for each experiment at SPring-8

→ Use the proposal number for authorization

- Procedure
  - Request user to login
  - List up proposal number list assigned to the user



Authorization for data access

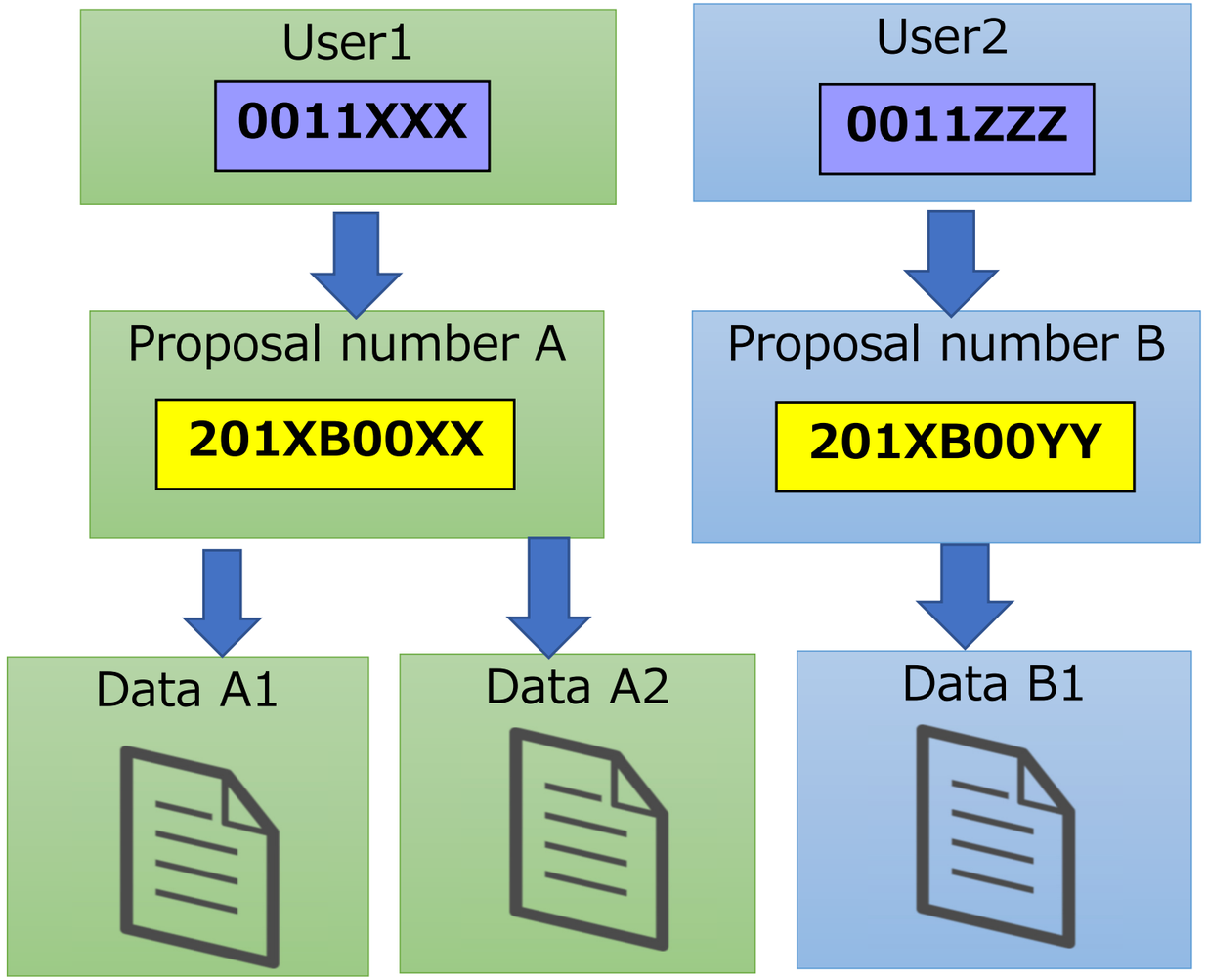
OpenID connect 1.0
for authentication

How to control range for data sharing?

- **Proposal number** is assigned for each experiment at SPring-8

→ Use the proposal number for authorization

- Procedure
  - Request user to login
  - List up proposal number list assigned to the user
  - Allow access only for data associated with the proposal number



Authorization for data access

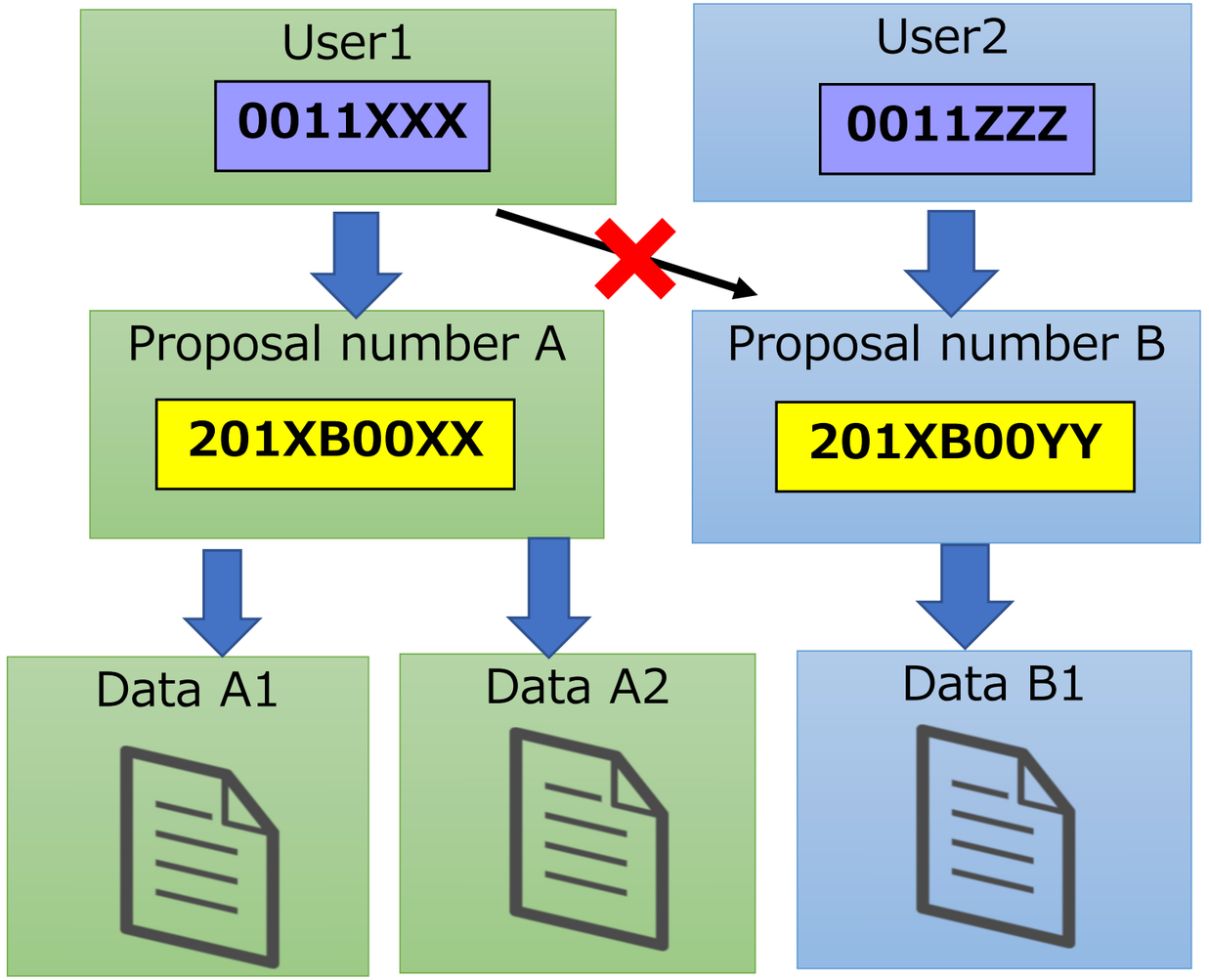
OpenID connect 1.0
for authentication

How to control range for data sharing?

- **Proposal number** is assigned for each experiment at SPring-8

→ Use the proposal number for authorization

- Procedure
  - Request user to login
  - List up proposal number list assigned to the user
  - Allow access only for data associated with the proposal number



Authorization for data access

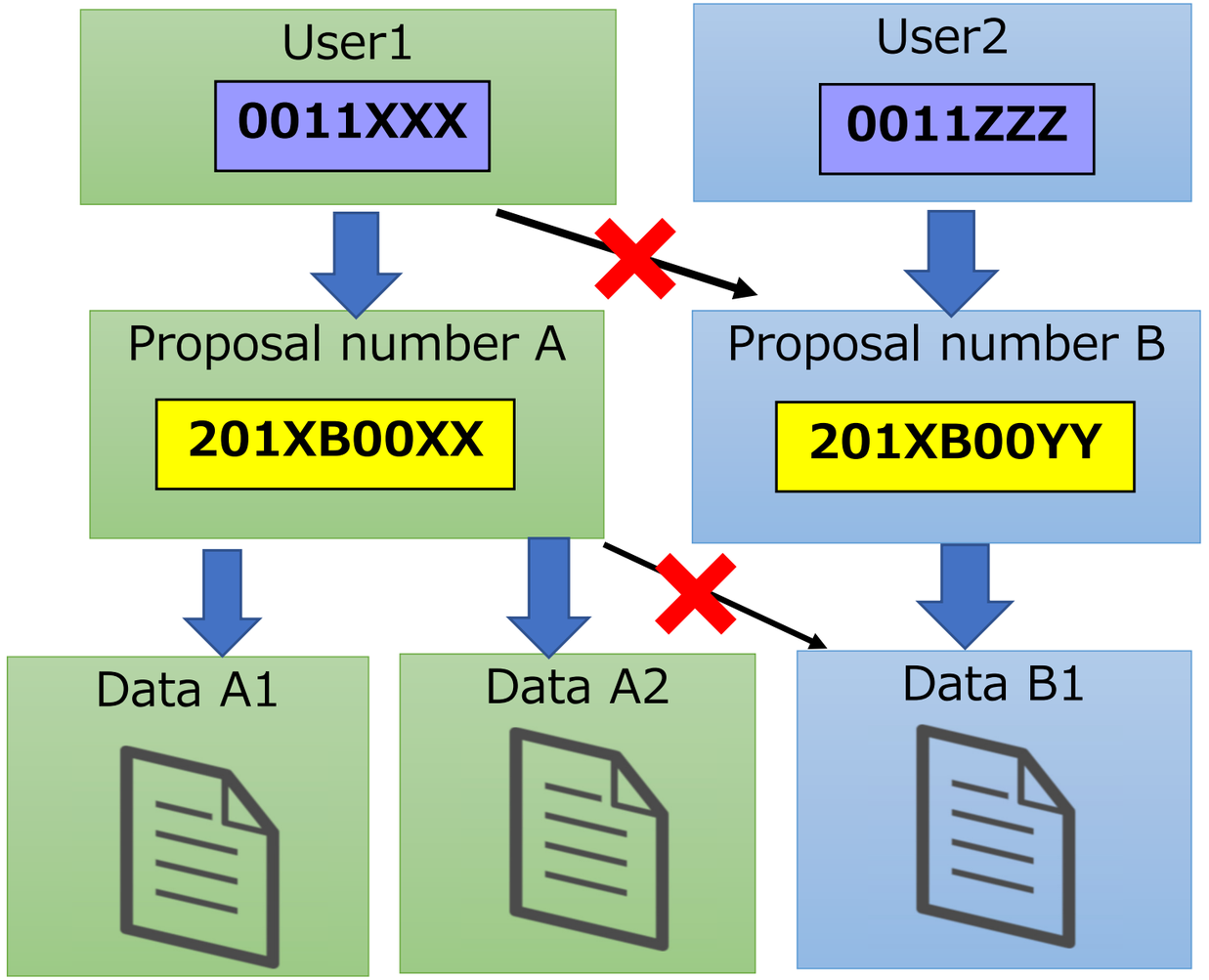
OpenID connect 1.0
for authentication

How to control range for data sharing?

- **Proposal number** is assigned for each experiment at SPring-8

→ Use the proposal number for authorization

- Procedure
  - Request user to login
  - List up proposal number list assigned to the user
  - Allow access only for data associated with the proposal number



Our approach to Metadata

- To describe metadata, common data format is desired
  - ex.) Nexus format based on HDF5 data container

Experimental Data file

Data register command

Data transfer system

Our approach to Metadata

- To describe metadata, common data format is desired
  - ex.) Nexus format based on HDF5 data container
- However, various data formats are used in Japan
  - Text, Tiff etc.

Experimental Data file

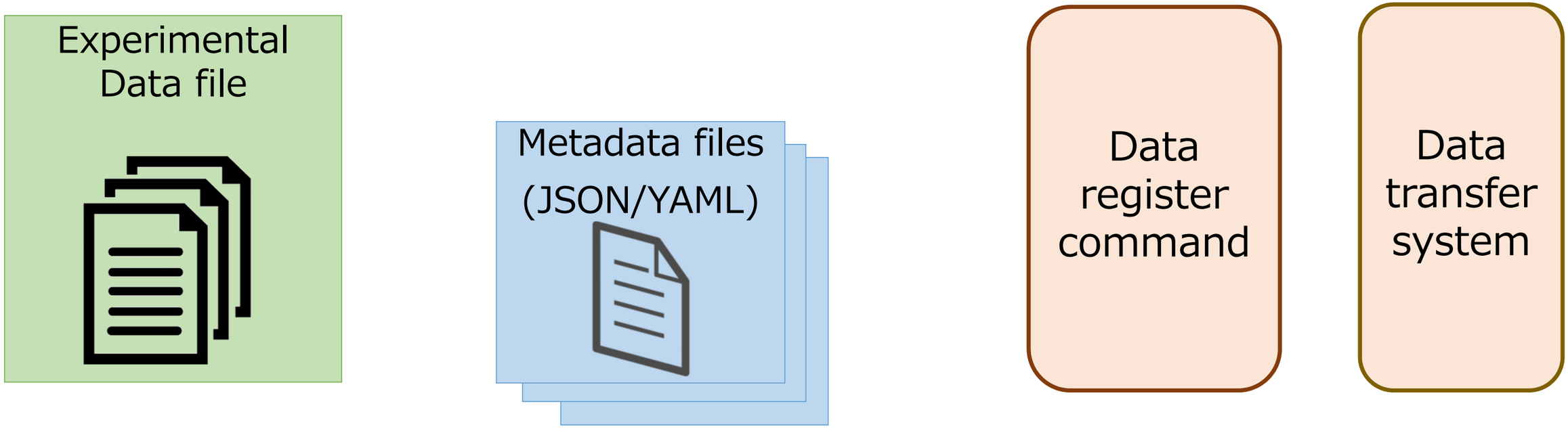
Data register command

Data transfer system

Our approach to Metadata

- To describe metadata, common data format is desired
  - ex.) Nexus format based on HDF5 data container
- However, various data formats are used in Japan
  - Text, Tiff etc.

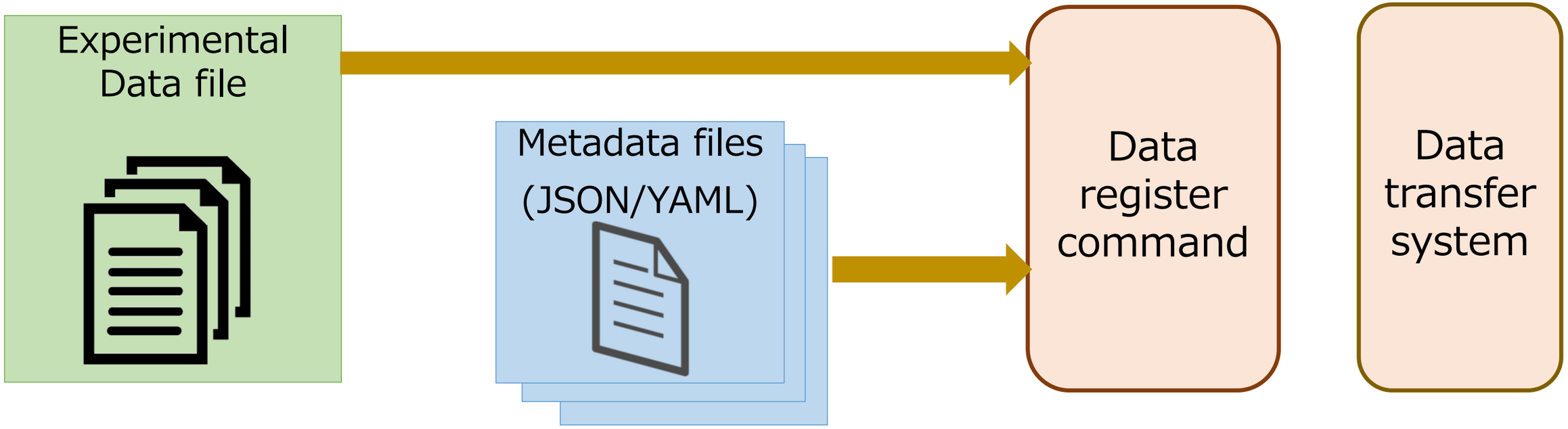
→Attach metadata files in JSON or YAML format separately



Our approach to Metadata

- To describe metadata, common data format is desired
  - ex.) Nexus format based on HDF5 data container
- However, various data formats are used in Japan
  - Text, Tiff etc.

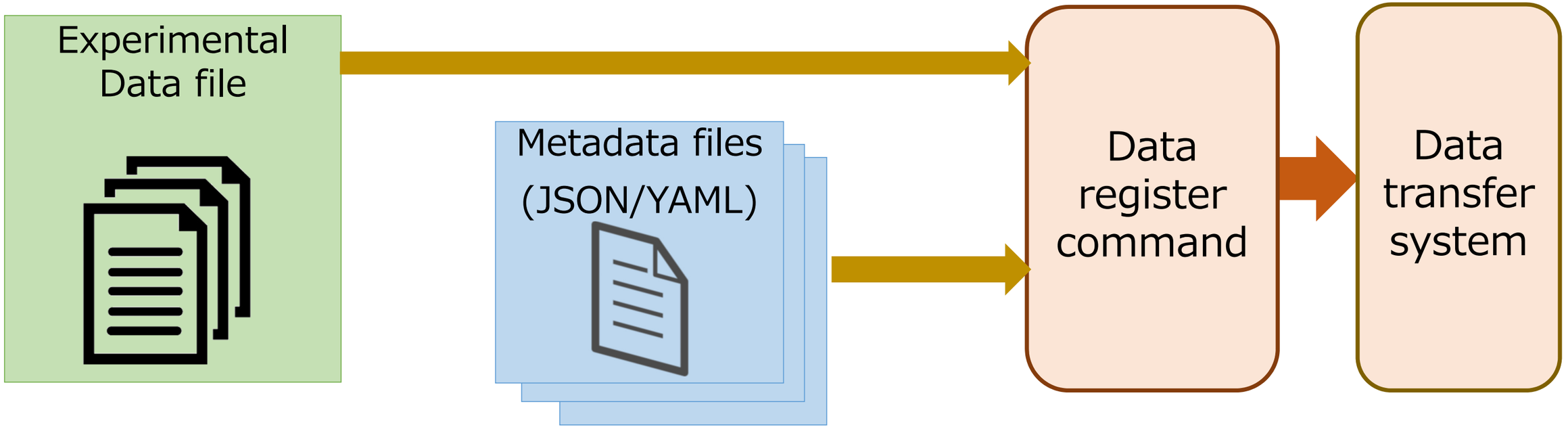
→Attach metadata files in JSON or YAML format separately



Our approach to Metadata

- To describe metadata, common data format is desired
  - ex.) Nexus format based on HDF5 data container
- However, various data formats are used in Japan
  - Text, Tiff etc.

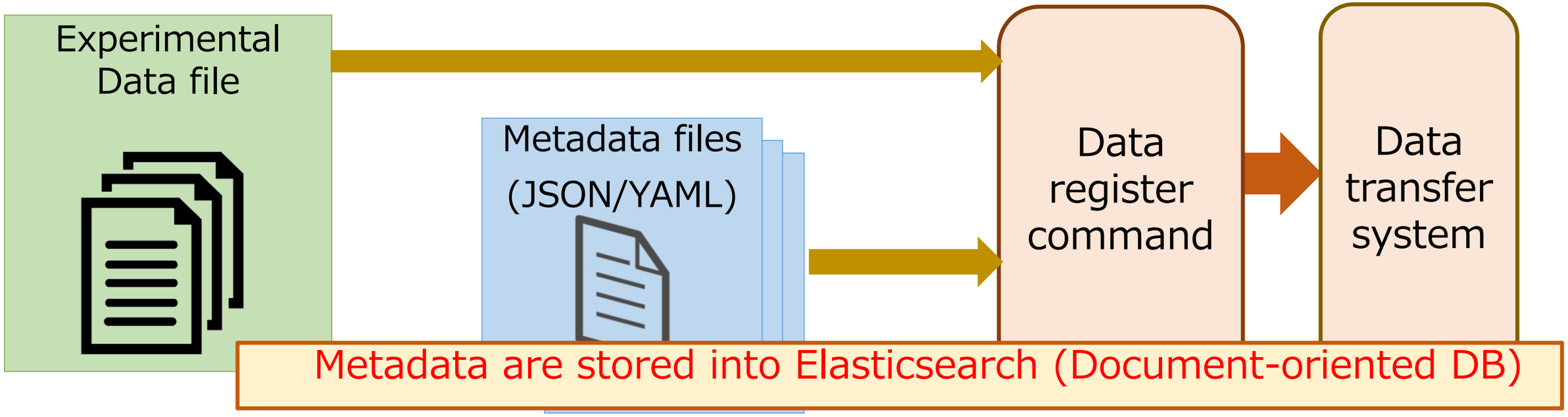
→Attach metadata files in JSON or YAML format separately



Our approach to Metadata

- To describe metadata, common data format is desired
  - ex.) Nexus format based on HDF5 data container
- However, various data formats are used in Japan
  - Text, Tiff etc.

→Attach metadata files in JSON or YAML format separately



Metadata description by User

```yaml
sample:
  name: Cupper (1) oxide
  chemical_formula: Cu2O
  element:
  - name: Cu
    suffix: 2
  - name: O
    suffix: 1
  vendor: KOJUNDO CHEMICAL LABORATORY
  model_number: CU01PB
measurement:
  method:
    name: XAFS
    detection: Transmission
    absorption_edge: Cu K-edge
  date:
    start_time: 2012-06-13 02:28:00
    end_time: 2012-06-13 02:32:00
```

Ex.) Metadata with YAML

- User prepares metadata in files with JSON or YAML
  - Human/Machine readable

Metadata description by User

```yaml
sample:
  name: Cupper (1) oxide
  chemical_formula: Cu2O
  element:
  - name: Cu
    suffix: 2
  - name: O
    suffix: 1
  vendor: KOJUNDO CHEMICAL LABORATORY
  model_number: CU01PB
measurement:
  method:
    name: XAFS
    detection: Transmission
    absorption_edge: Cu K-edge
  date:
    start_time: 2012-06-13 02:28:00
    end_time: 2012-06-13 02:32:00
```

Ex.) Metadata with YAML

- User prepares metadata in files with JSON or YAML
  - Human/Machine readable
  - Defined for each category
    - Sample
    - Measurement
    - Instrument
    - Dataset
    - Etc.

Metadata description by User

Sample category

```
sample:
  name: Cupper (1) oxide
  chemical_formula: Cu2O
  element:
  - name: Cu
    suffix: 2
  - name: O
    suffix: 1
  vendor: KOJUNDO CHEMICAL LABORATORY
  model_number: CU01PB
measurement:
  method:
    name: XAFS
    detection: Transmission
    absorption_edge: Cu K-edge
  date:
    start_time: 2012-06-13 02:28:00
    end_time: 2012-06-13 02:32:00
```

Ex.) Metadata with YAML

- User prepares metadata in files with JSON or YAML
  - Human/Machine readable
  - Defined for each category
    - Sample
    - Measurement
    - Instrument
    - Dataset
    - Etc.

Metadata description by User

Sample category

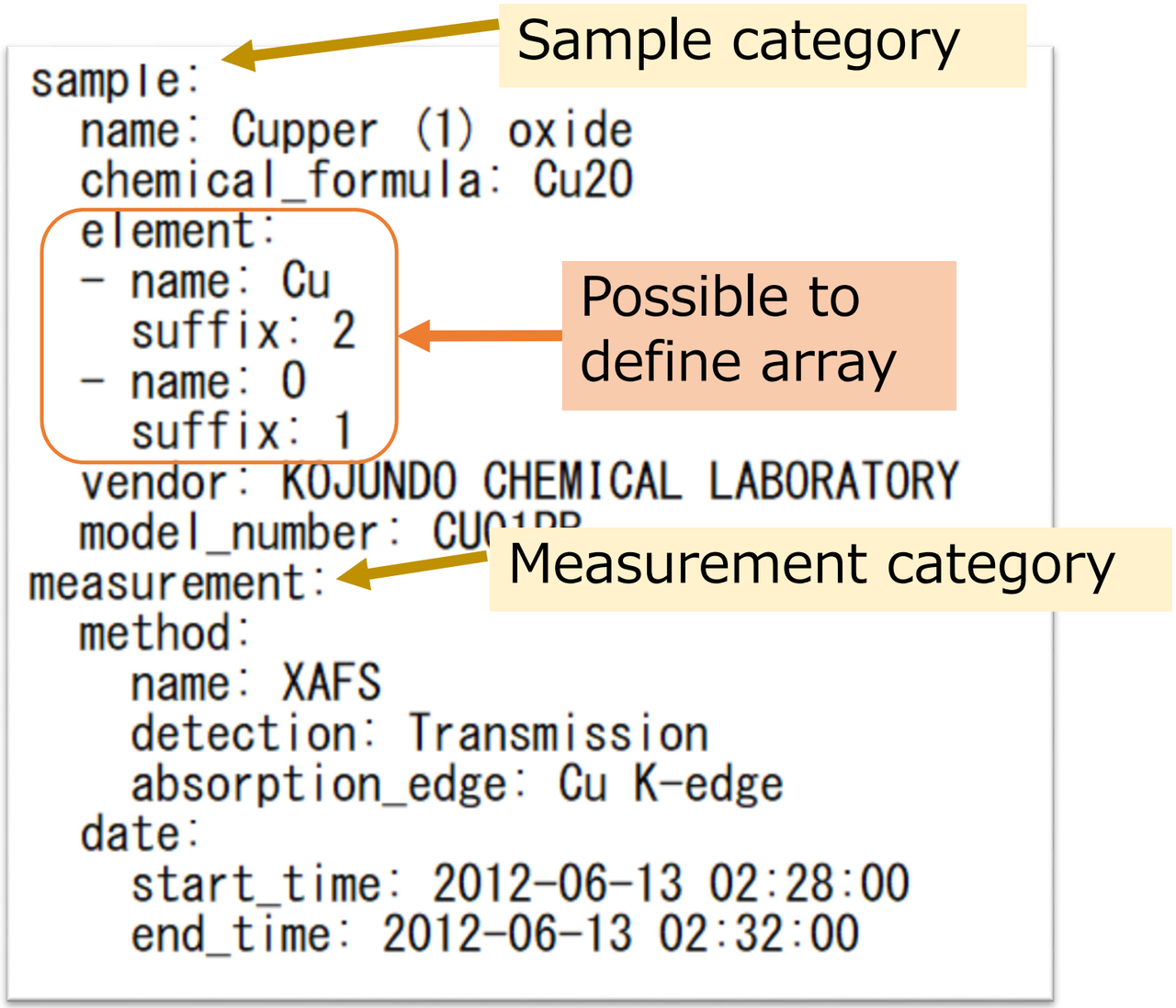
```
sample:
  name: Cupper (1) oxide
  chemical_formula: Cu2O
  element:
  - name: Cu
    suffix: 2
  - name: O
    suffix: 1
  vendor: KOJUNDO CHEMICAL LABORATORY
  model_number: CUO1PB
measurement:
  method:
    name: XAFS
    detection: Transmission
    absorption_edge: Cu K-edge
  date:
    start_time: 2012-06-13 02:28:00
    end_time: 2012-06-13 02:32:00
```

Measurement category

Ex.) Metadata with YAML

- User prepares metadata in files with JSON or YAML
  - Human/Machine readable
  - Defined for each category
    - Sample
    - Measurement
    - Instrument
    - Dataset
    - Etc.

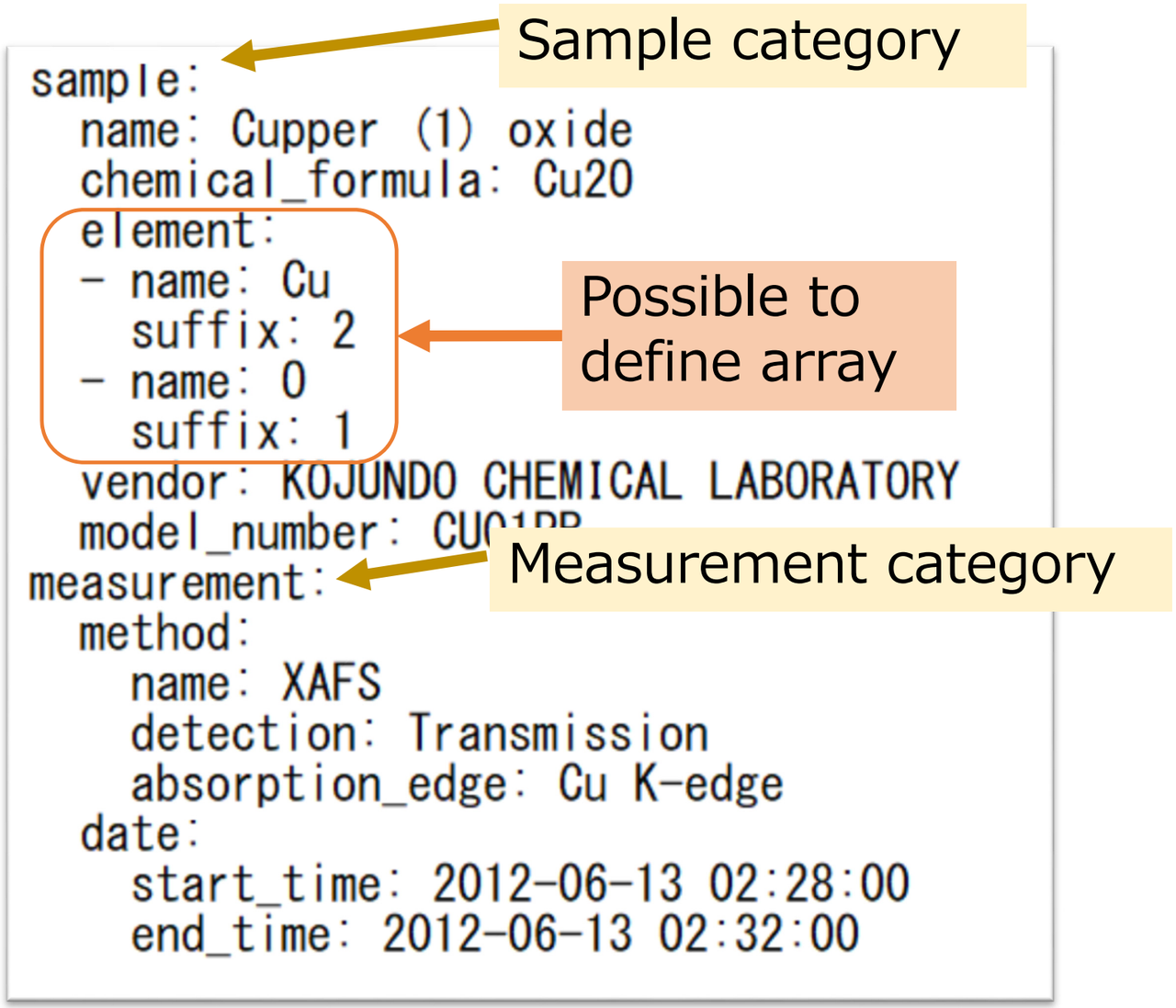
Metadata description by User



Ex.) Metadata with YAML

- User prepares metadata in files with JSON or YAML
  - Human/Machine readable
  - Defined for each category
    - Sample
    - Measurement
    - Instrument
    - Dataset
    - Etc.

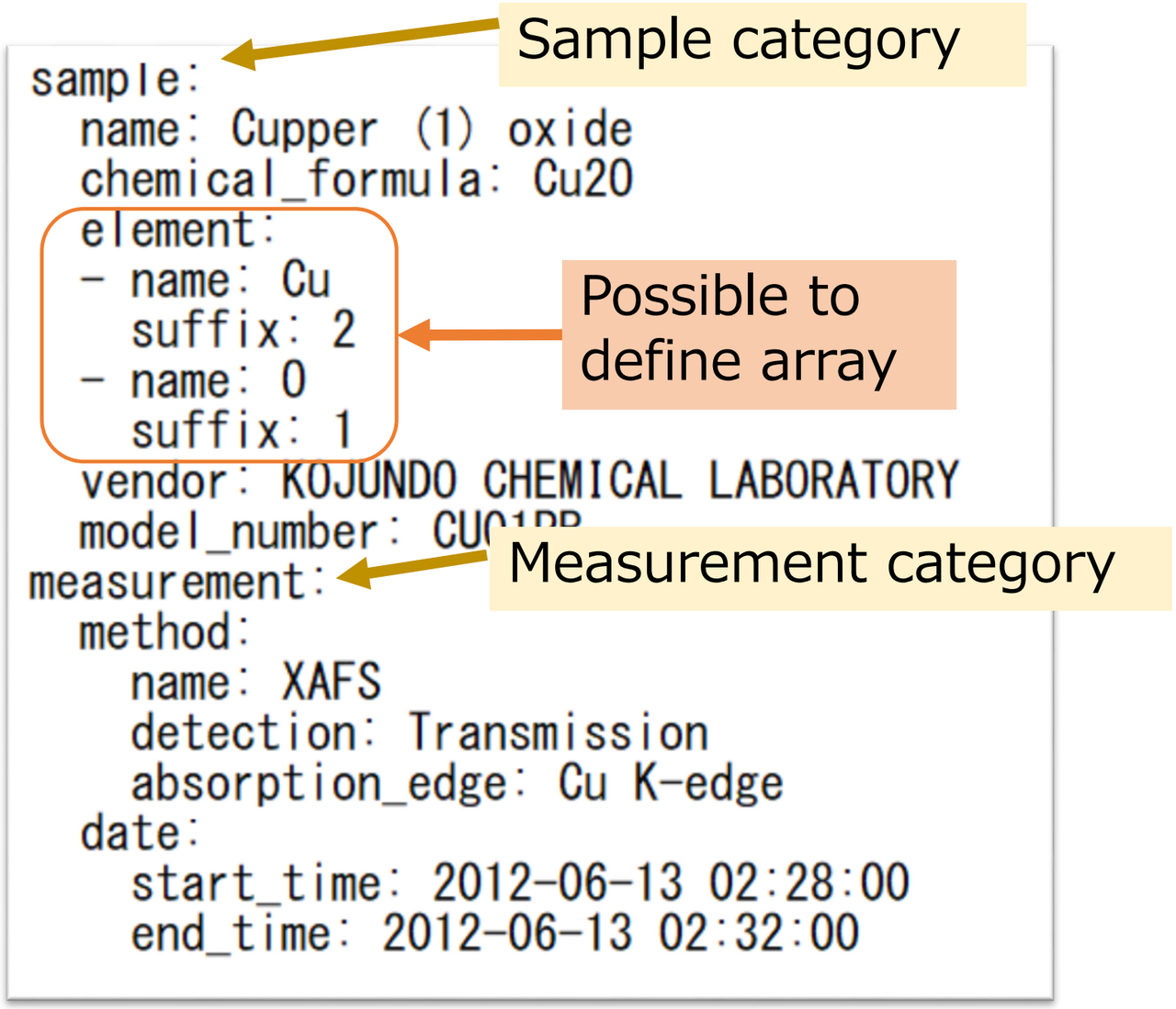
Metadata description by User



Ex.) Metadata with YAML

- User prepares metadata in files with JSON or YAML
  - Human/Machine readable
  - Defined for each category
    - Sample
    - Measurement
    - Instrument
    - Dataset
    - Etc.
  - Possible to define hierarchical structure

Metadata description by User



Ex.) Metadata with YAML

- User prepares metadata in files with JSON or YAML
  - Human/Machine readable
  - Defined for each category
    - Sample
    - Measurement
    - Instrument
    - Dataset
    - Etc.
  - Possible to define hierarchical structure
- User can add items required for measurements

Operation of BENTEN at SPring-8

- Started Operation at SPring-8 since March, 2019
  → Applicable to data transfer in public experimental stations

Operation of BENTEN at SPring-8

- Started Operation at SPring-8 since March, 2019
  → Applicable to data transfer in public experimental stations
- Web portal : https://benten.spring8.or.jp
  - 2 factor authentication was imposed for secured data access

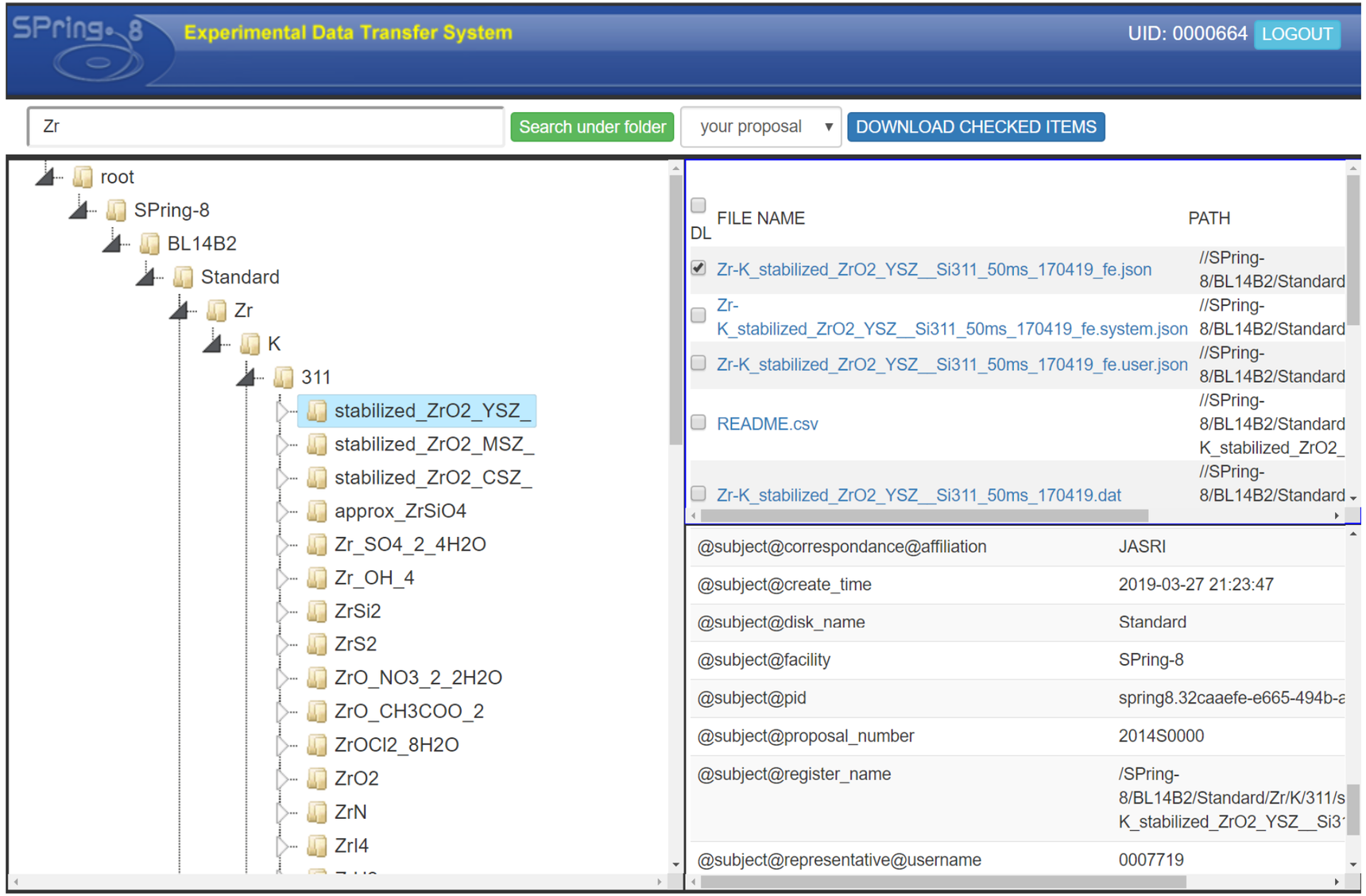
Operation of BENTEN at SPring-8

- Started Operation at SPring-8 since March, 2019
  → Applicable to data transfer in public experimental stations
- Web portal : https://benten.spring8.or.jp
  - 2 factor authentication was imposed for secured data access
- First attempt:
  - Restricted data access for user experiments at BL14B2
    - Automated data transfer with BENTEN was implemented

Operation of BENTEN at SPring-8

- Started Operation at SPring-8 since March, 2019
  → Applicable to data transfer in public experimental stations
- Web portal : https://benten.spring8.or.jp
  - 2 factor authentication was imposed for secured data access
- First attempt:
  - Restricted data access for user experiments at BL14B2
    - Automated data transfer with BENTEN was implemented
  - Open data access for XAFS standard sample
    - Registered offline with sufficient metadata
    - Major items were automatically extracted
    - However, manual input was still required (sample, instrument etc.)

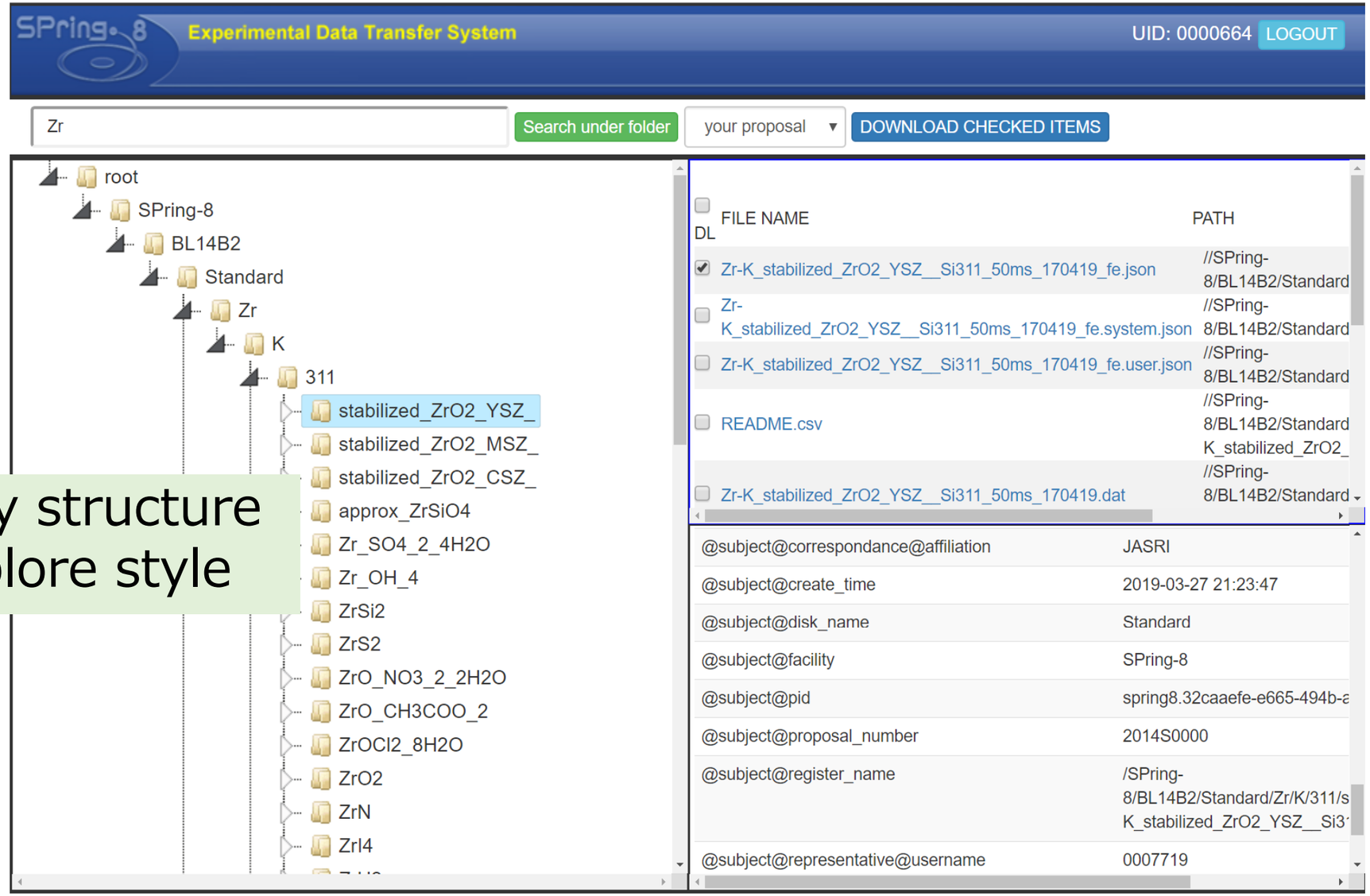
Web portal (Accessible data for user)



BENTEN SYSTEM:Copyright (C) 2019 JASRI All Rights Reserved.

Web portal (Accessible data for user)

Directory structure with explore style

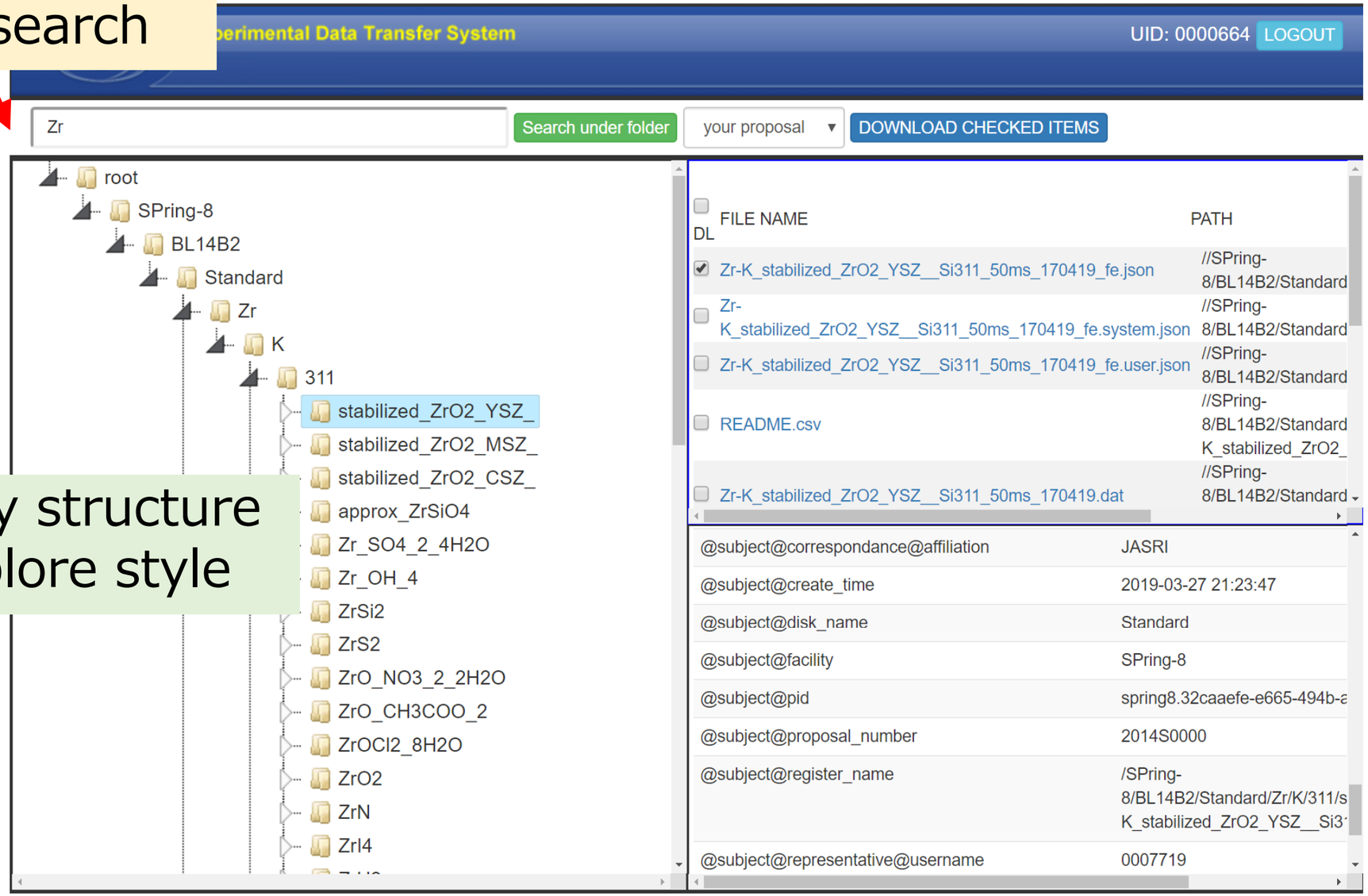


BENTEN SYSTEM:Copyright (C) 2019 JASRI All Rights Reserved.

Web portal (Accessible data for user)

Full-text search

Directory structure with explore style



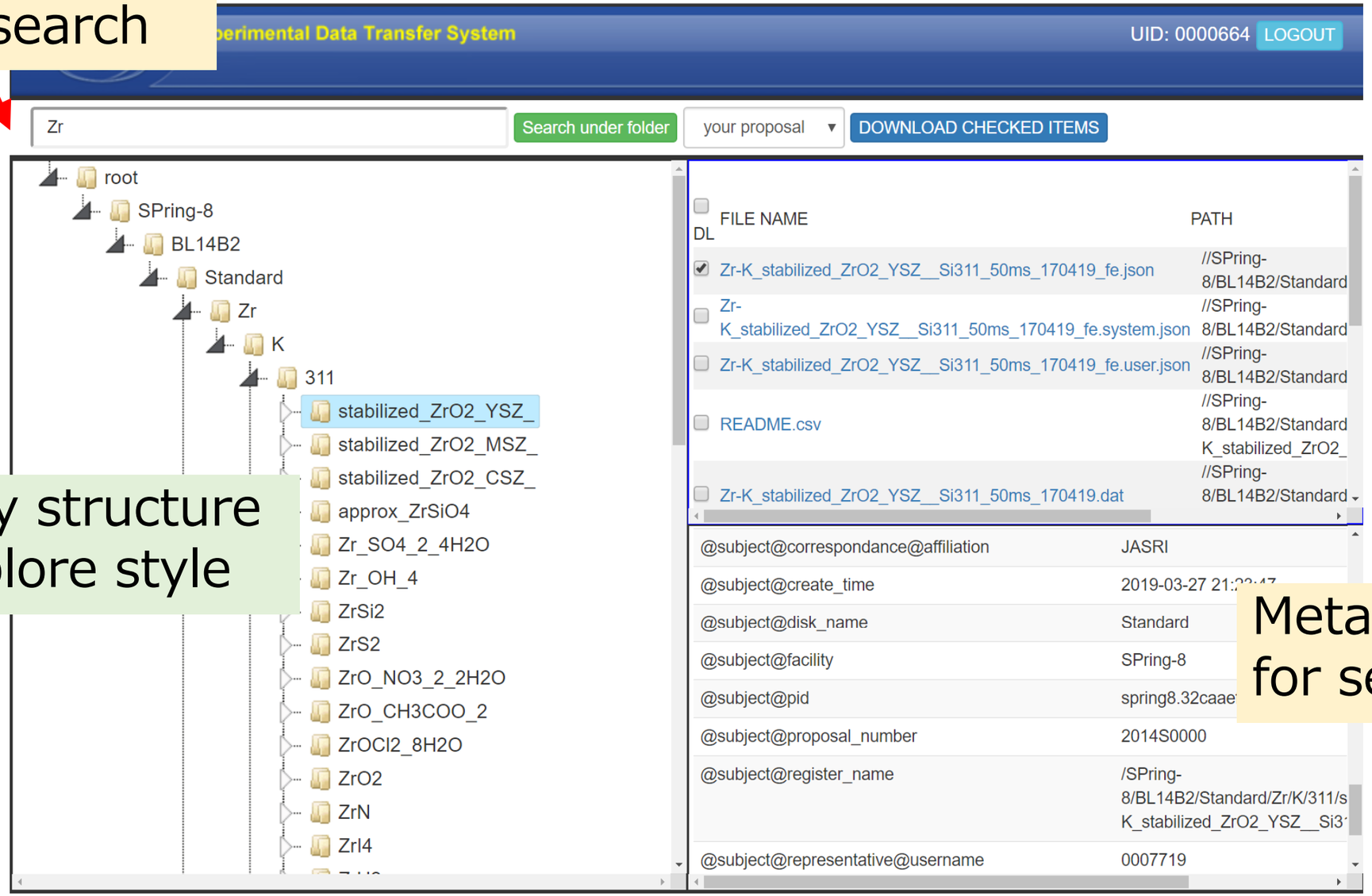
BENTEN SYSTEM:Copyright (C) 2019 JASRI All Rights Reserved.

Web portal (Accessible data for user)

Full-text search

Directory structure with explore style

Metadata view for selected file



BENTEN SYSTEM:Copyright (C) 2019 JASRI All Rights Reserved.

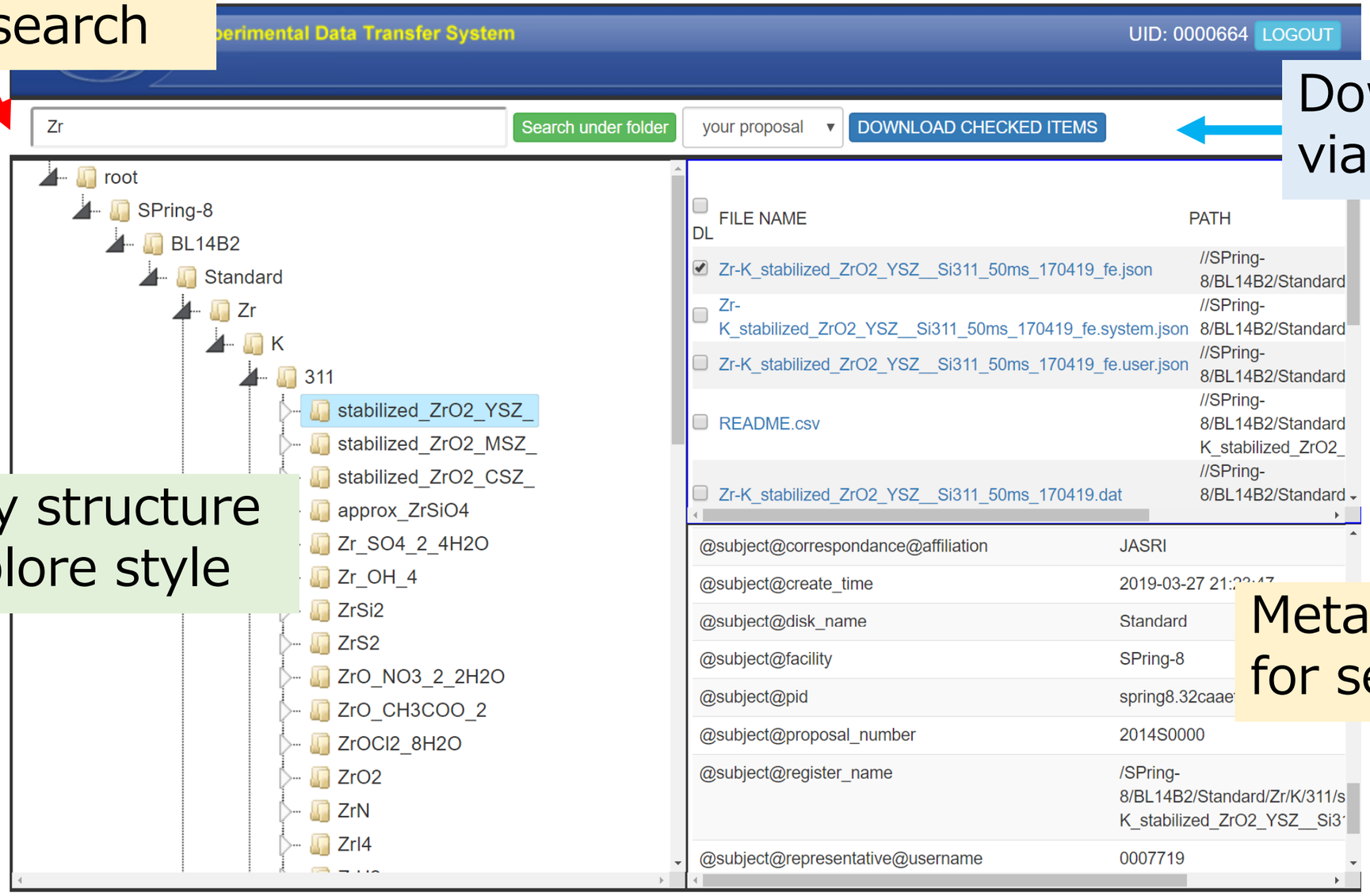
Web portal (Accessible data for user)

Full-text search

Download file via zip

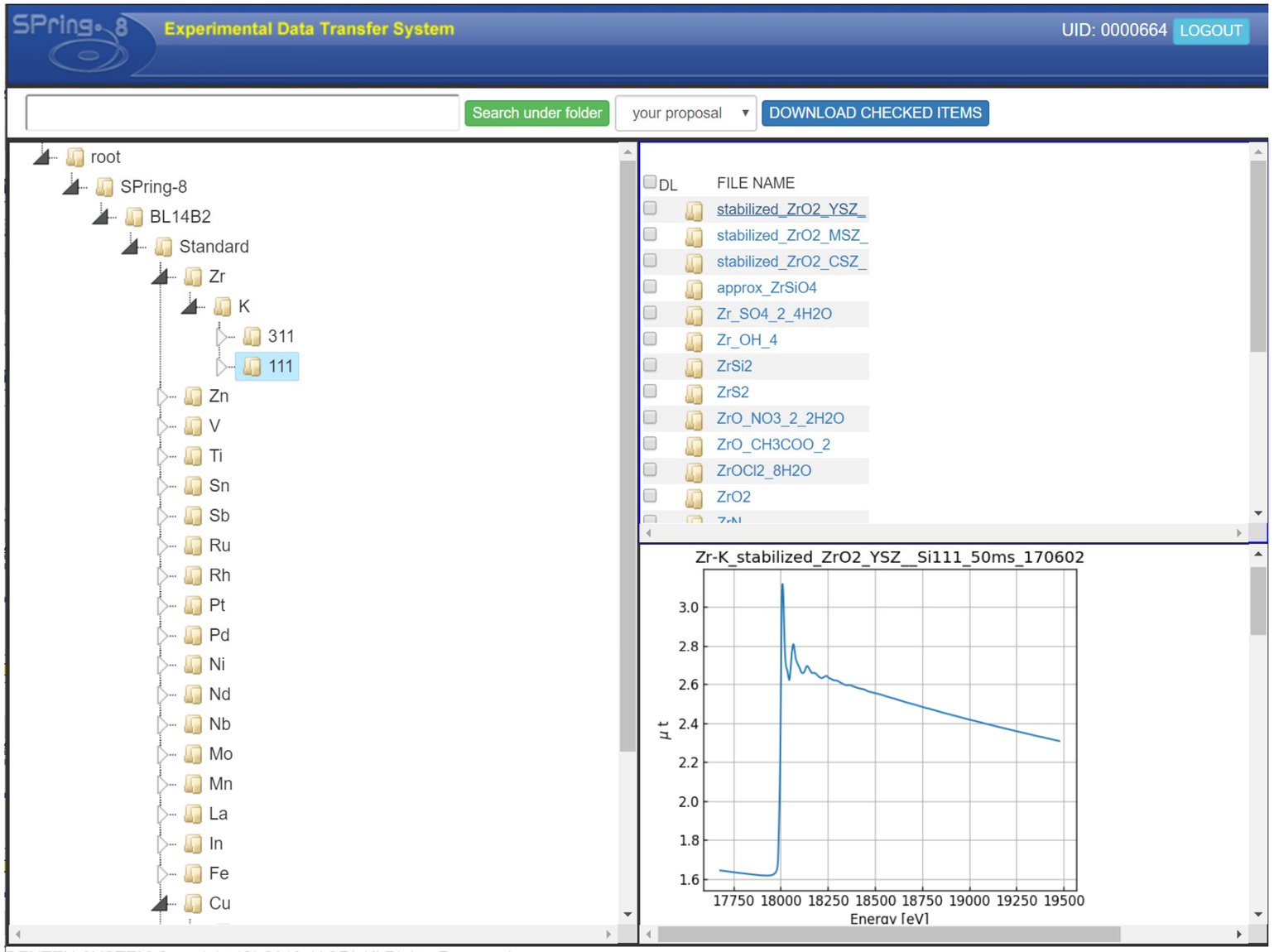
Directory structure with explore style

Metadata view for selected file



BENTEN SYSTEM:Copyright (C) 2019 JASRI All Rights Reserved.

Thumbnail for dataset

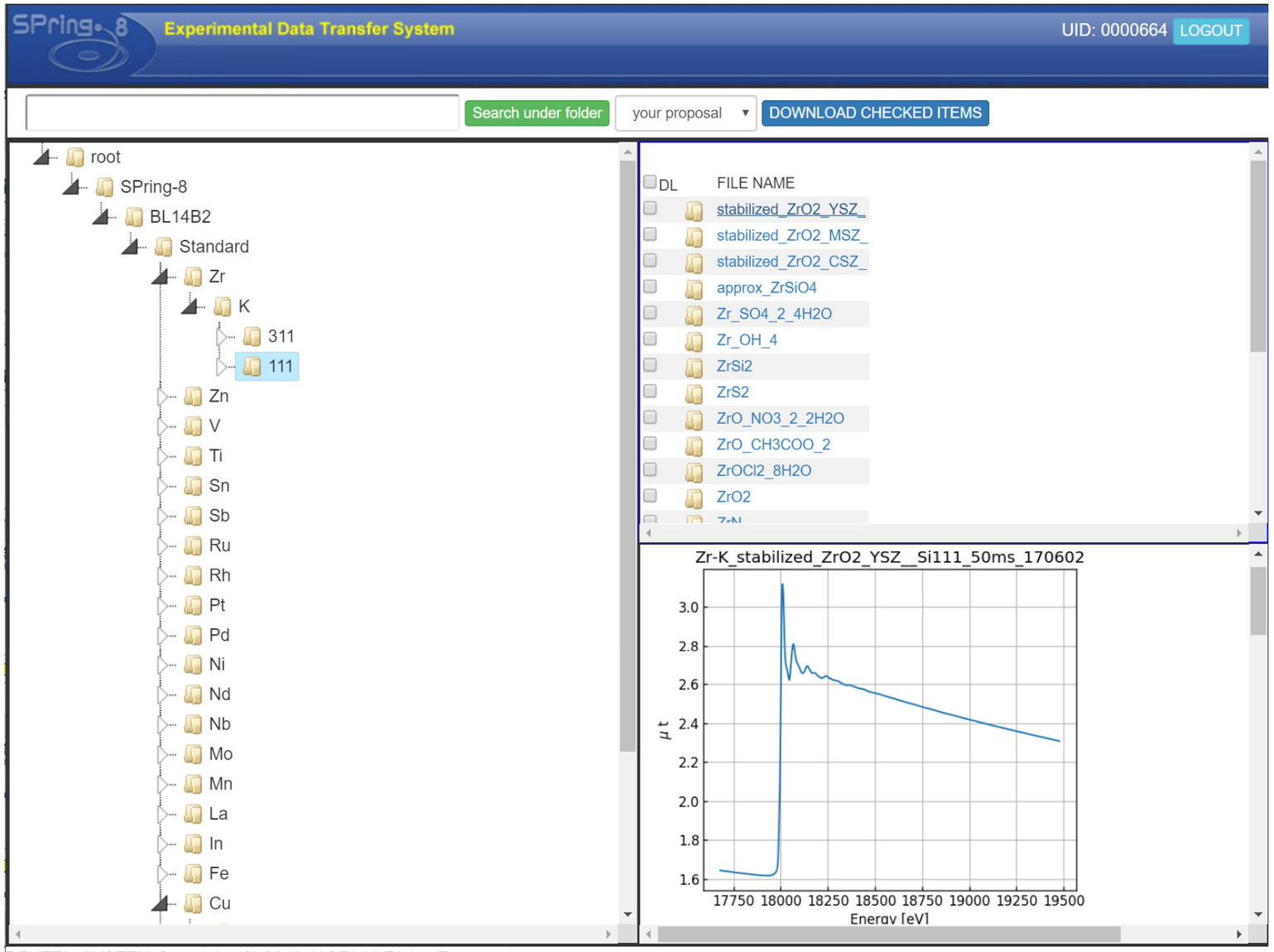


BENTEN SYSTEM:Copyright (C) 2019 JASRI All Rights Reserved.

- Thumbnail can be attached to dataset

→monitored in Web with metadata items

Thumbnail for dataset



BENTEN SYSTEM:Copyright (C) 2019 JASRI All Rights Reserved.

- Thumbnail can be attached to dataset

→monitored in Web with metadata items

Example:

XAFS spectrum plot

→ Useful for users

Summary: BENTEN features list

- Easy-to-use and generic experimental data transfer at SPring-8
  - Unified REST API for all the functions

Summary: BENTEN features list

- Easy-to-use and generic experimental data transfer at SPring-8
  - Unified REST API for all the functions
- Implement authentication
  - For restricted data access / Open data access
  - Use proposal number to derive range of data sharing

Summary: BENTEN features list

- Easy-to-use and generic experimental data transfer at SPring-8
  - Unified REST API for all the functions
- Implement authentication
  - For restricted data access / Open data access
  - Use proposal number to derive range of data sharing
- Flexible metadata management
  - Metadata can be easily described in separate files (JSON or YAML format)
    → Existing data format in the measurements can be used as they are
  - Full-text search for flexible data access using Elasticsearch

Summary: BENTEN features list

- Easy-to-use and generic experimental data transfer at SPring-8
  - Unified REST API for all the functions
- Implement authentication
  - For restricted data access / Open data access
  - Use proposal number to derive range of data sharing
- Flexible metadata management
  - Metadata can be easily described in separate files (JSON or YAML format)
    → Existing data format in the measurements can be used as they are
  - Full-text search for flexible data access using Elasticsearch
- Data management
  - Data life cycle
    - Creation, Update, Open/Closed access, Delete
  - Clarify Contact name for Open data

Current status and Future plan

- Operation of BENTEN at SPring-8 since March, 2019
  - Provided data access thorough the Internet
    - Open data for XAFS standard sample
    - Restricted data access in user experiments at BL14B2

Current status and Future plan

- Operation of BENTEN at SPring-8 since March, 2019
  - Provided data access thorough the Internet
    - Open data for XAFS standard sample
    - Restricted data access in user experiments at BL14B2

- Future plan
  - Promote BENTEN to other experimental stations
    - Open data for HAXPES (hard X-ray photoemission spectroscopy) standard sample
    - Remote data access for CT (Computed Tomography) image

Current status and Future plan

- Operation of BENTEN at SPring-8 since March, 2019
  - Provided data access thorough the Internet
    - Open data for XAFS standard sample
    - Restricted data access in user experiments at BL14B2

- Future plan
  - Promote BENTEN to other experimental stations
    - Open data for HAXPES (hard X-ray photoemission spectroscopy) standard sample
    - Remote data access for CT (Computed Tomography) image

→ Will start within a year