

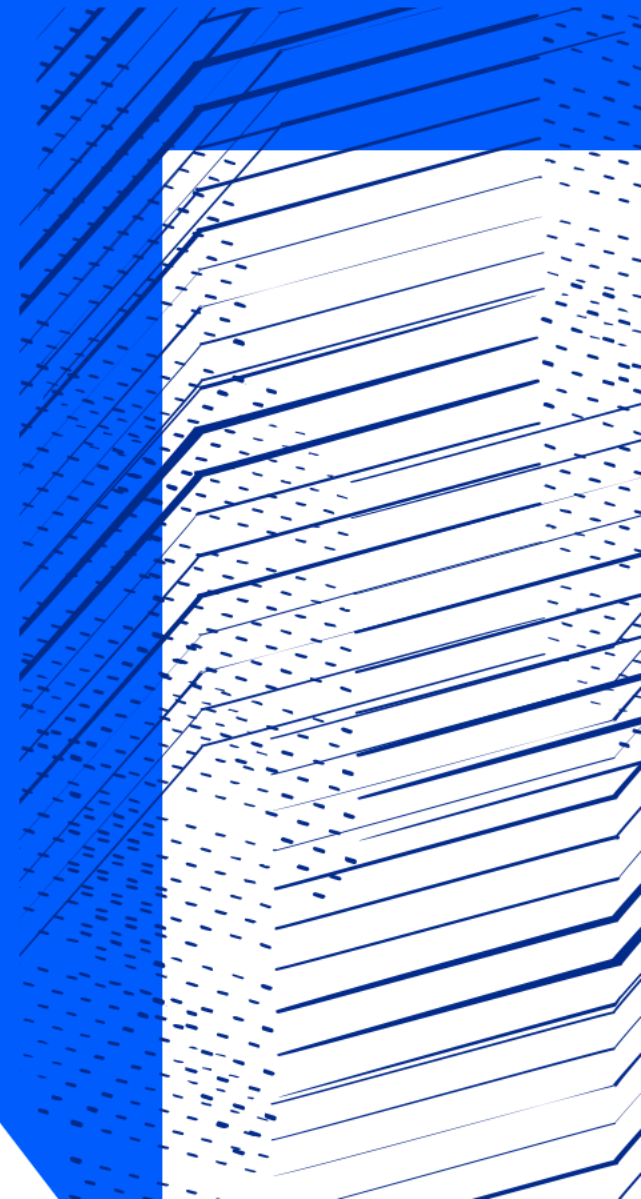


Science and
Technology
Facilities Council

ICAT Metadata Catalogue

SciCatCon Meeting 2024

Louise Davies



About me

- Worked at STFC for ~8 years
- Worked on ICAT related software for 6 years
- Lead frontend engineer in our group
- Developed STFC's ICAT frontend

Agenda

1 ICAT overview

2 ICAT data model

3 Handling scientific metadata

4 Data Retrieval





Science and
Technology
Facilities Council

ICAT Overview



ICAT Collaboration

STFC
• ISIS
• CLF
Diamond

HZB

ALBA

ESRF

SESAME

Sirius

CERIC

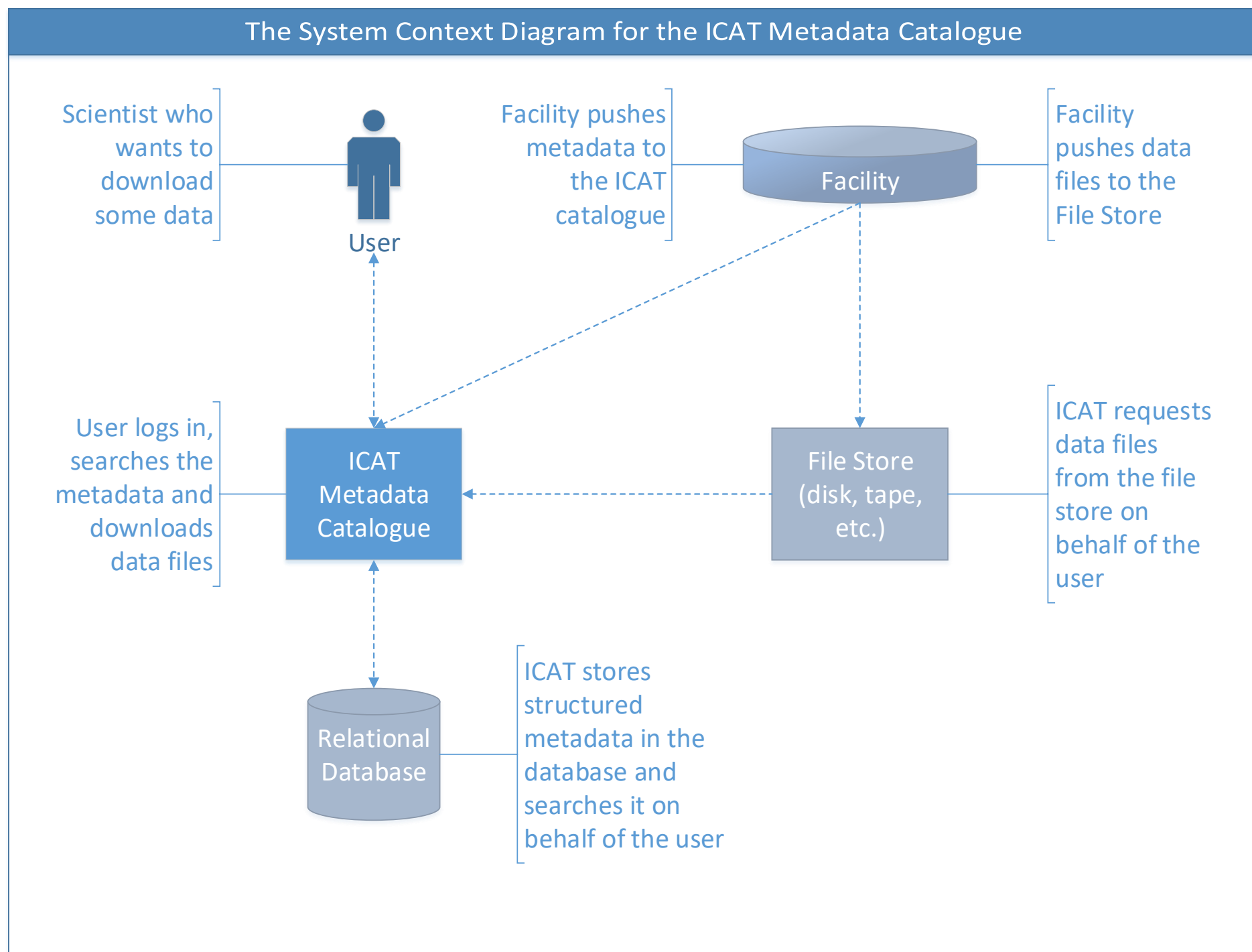
- **ICAT** has been developed for at least 17 years
- **ICAT** is used at several facilities worldwide
- Handles many years of data (e.g. ISIS - 40 years) and large volumes (e.g. DLS - ~60PB data, $\sim 5 \times 10^9$ files stored in **ICAT**)



Science and
Technology
Facilities Council

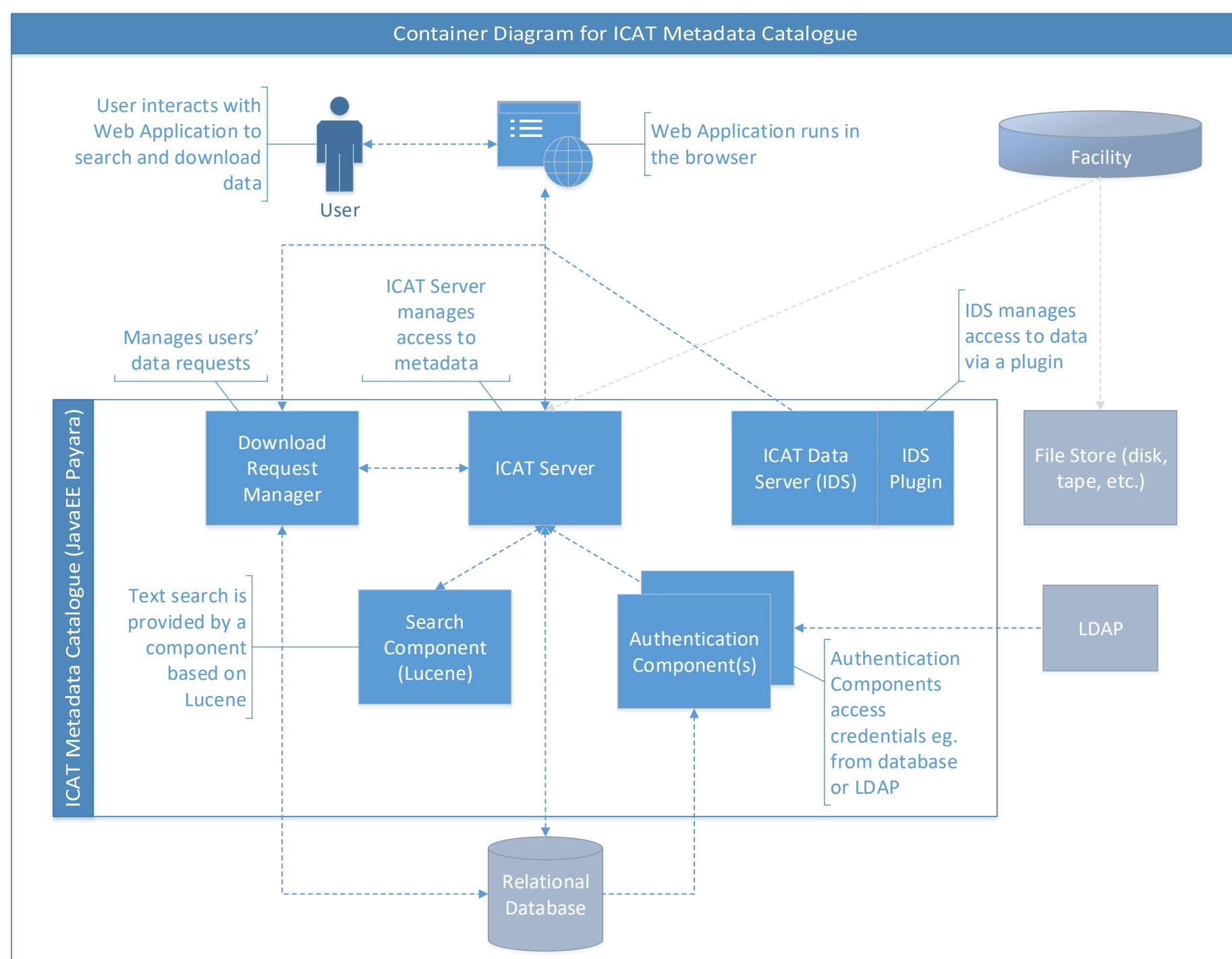
ICAT

Overview



ICAT

Detailed View



Integrations

Technology Overview

- Written in **Java EE**
- Runs on **Payara**
 - open-source successor to Sun/Oracle Glassfish Application Server

Databases

- **MySQL/MariaDB & Oracle** supported
- Any JPA-compatible database ought to be possible

Storage

- Simple **filesystem access** supported
- Other storage integrations (eg. tape archive) achieved via **plugin** system

Frontends

- Two “main” frontends – DataGateway (STFC developed) and DataPortal (ESRF developed)
 - Both written in React
- Also some more domain specific frontends
 - eCAT, Human Organ Atlas, Paleontology Database etc.
- Older frontend, TopCAT, written in Angular.js v1 (deprecated)

Frontends – DataGateway

📄 Display as cards ✕ Clear filters

<input type="checkbox"/>	Title	RB Num	DOI	Size	Principa	Start Da	End Dat	Actions
	<u>Include</u>	<u>Inclt</u>	<u>Include</u>		<u>Inclt</u>	<u>From.</u>	<u>From.</u>	
						<u>To...</u>	<u>To...</u>	
<input type="checkbox"/>	Hydrogen Bond Dynamics in Organi...	1910554	10.5286/ISIS.E.RB1910554	109.59 MB	Dr Sangha...	2019-03-2...	2019-05-2...	
<input type="checkbox"/>	Investigation of segmental dynami...	1868044	10.5286/ISIS.E.RB1868044	136.48 MB	Professor ...	2019-03-1...	2019-03-2...	
<input type="checkbox"/>	Water diffusion in small peptide ba...	1820530	10.5286/ISIS.E.RB1820530	38.38 MB	Dr Kathari...	2019-03-1...	2019-03-1...	
<input type="checkbox"/>	Inelastic neutron scattering on a go...	1990058		0 B	Mr Daniel ...	2019-03-1...		
<input type="checkbox"/>	Probing the dynamics of silica sup...	1910021	10.5286/ISIS.E.RB1910021	238.62 MB	Professor ...	2019-03-0...	2019-07-1...	
<input type="checkbox"/>	Probing Ligand Flexibility using Qu...	1910587	10.5286/ISIS.E.RB1910587	87.29 MB	Dr David N...	2019-03-0...	2019-03-0...	
<input type="checkbox"/>	Inelastic Scattering on a Gold(I) Hy...	1890389		0 B	Mr Daniel ...	2019-03-0...	2019-03-0...	
<input type="checkbox"/>	A QENS Study of the Effect of Anti...	1890251	10.5286/ISIS.E.RB1890251	31.8 MB	Professor ...	2019-03-0...	2019-03-0...	
<input type="checkbox"/>	Phase transition in HAP polymorphs	1890388	10.5286/ISIS.E.RB1890388	6 MB	Professor ...	2019-03-0...	2019-03-0...	
<input type="checkbox"/>	A QENS study of polyethylene furan...	1890274		0 B	Dr Mariela ...	2019-03-0...	2019-03-0...	
<input type="checkbox"/>	QENS test B1 DC002	QENS		78.06 MB		2019-03-0...	2019-03-0...	

Frontends – DataPortal

Data Portal My Data [Open Data](#) Closed Data

 Feedback  Log out **Anonymous**

[Open Data](#)

Investigations

[Data collections](#)

[Documents](#)







Filter between dates



Filter between dates



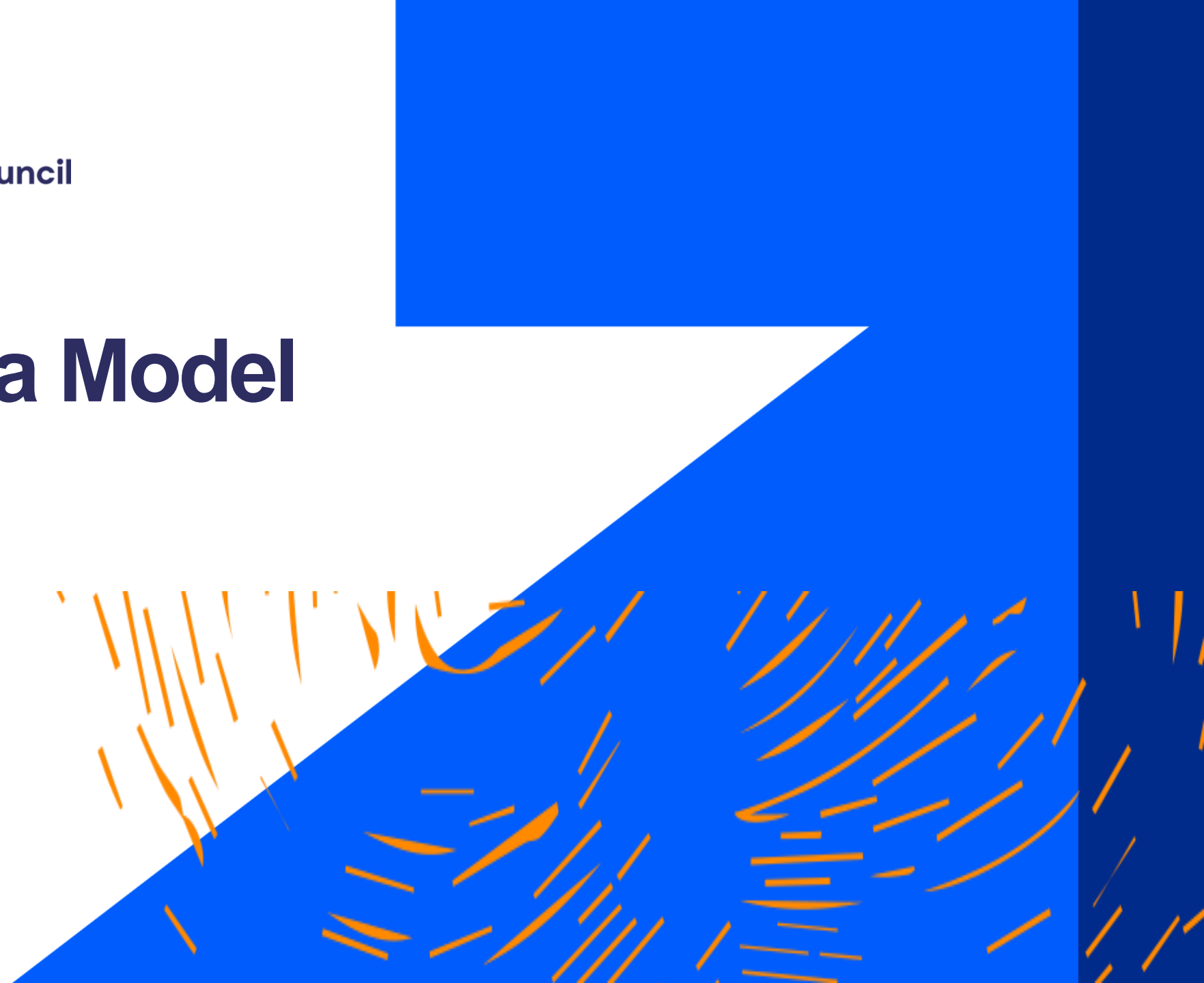
Search

Proposal	Beamline	Start	Title	Datasets	Files	Release	DOI
 HC-5496	BM02	13/02/2024	Investigation of the nematic phase of FeSe by resonant X-ray diffraction	3 27.1 GB	1815	19/02/2024	DOI 10.15151/ESRF-ES-1437864712
 ID23-1-0000	ID23-1	11/11/2023	Data collected for the ISPyB Collaboration	173 148.8 GB	23711	14/11/2023	DOI ESRF-DOI/ISPyB-12345567894
 ES-1294	ID15B	11/07/2023	Redox reactions between FeCO ₃ and H ₂ O in the deep mantle: Significance for the Earth's carbon-hydrogen-oxygen cycles	2675 301.4 GB	26538	16/11/2023	DOI 10.15151/ESRF-ES-1220324761
 IH-ES-74	ID19	09/05/2022	Phase contrast imaging of saline spray ice	74 2.4 TB	621	12/01/2023	DOI 10.15151/ESRF-ES-810702957
 MI-1355	ID13	29/06/2021	Scanning and In-situ MLL Development for X-ray Nanodiffraction on Thin Film, Space and Additively Manufactured Applications	8 13.3 GB	135	01/07/2024	DOI 10.15151/ESRF-ES-452641547
 HC-4681	ID02	29/06/2021	Temperature evolution of the compressibility and correlation length in the vicinity of the lambda transition in sulfur	1 130.1 GB	23668	01/07/2024	DOI 10.15151/ESRF-ES-465961344



Science and
Technology
Facilities Council

ICAT Data Model



Data Model

Implementation

- Each facility maps relevant entities from data model onto locally relevant concepts
- **Different at each ICAT instance**

Metadata stored in Parameters

- Attached to Investigation, Dataset, Datafile, Sample, DataCollection
- Name
- Value – Numeric/String/DateTime
- No JSON, no hierarchy – except with ICAT+ (ESRF)

Data Model Comparison

- Dataset fairly equivalent between ICAT and SciCAT
 - However, Investigation is more “central” to ICAT*, whereas Dataset is more “central” to SciCAT
- ICAT Investigation = SciCAT proposal
 - Investigation is non-optional and is more “core”
- Datafile equivalent
 - No DataBlocks in ICAT
 - Datafiles are queryable individually
- ICAT Job != SciCAT Job

Authorization

- Rules stored in ICAT database **not** in an external system
- CRUD flags paired with simple table permissions or more complex SQL 'what' clauses
 - crudFlags: R, what: Instrument
 - crudFlags: RU, what:

```
SELECT df FROM Datafile df JOIN df.dataset d JOIN d.investigation i JOIN i.investigationInstruments ii JOIN ii.instrument inst JOIN inst.instrumentScientists instSci JOIN instSci.user u WHERE d.name!='raw' AND u.name = :user
```
- Rules can apply to user groups or to all users
- Can change authorization rules at any time

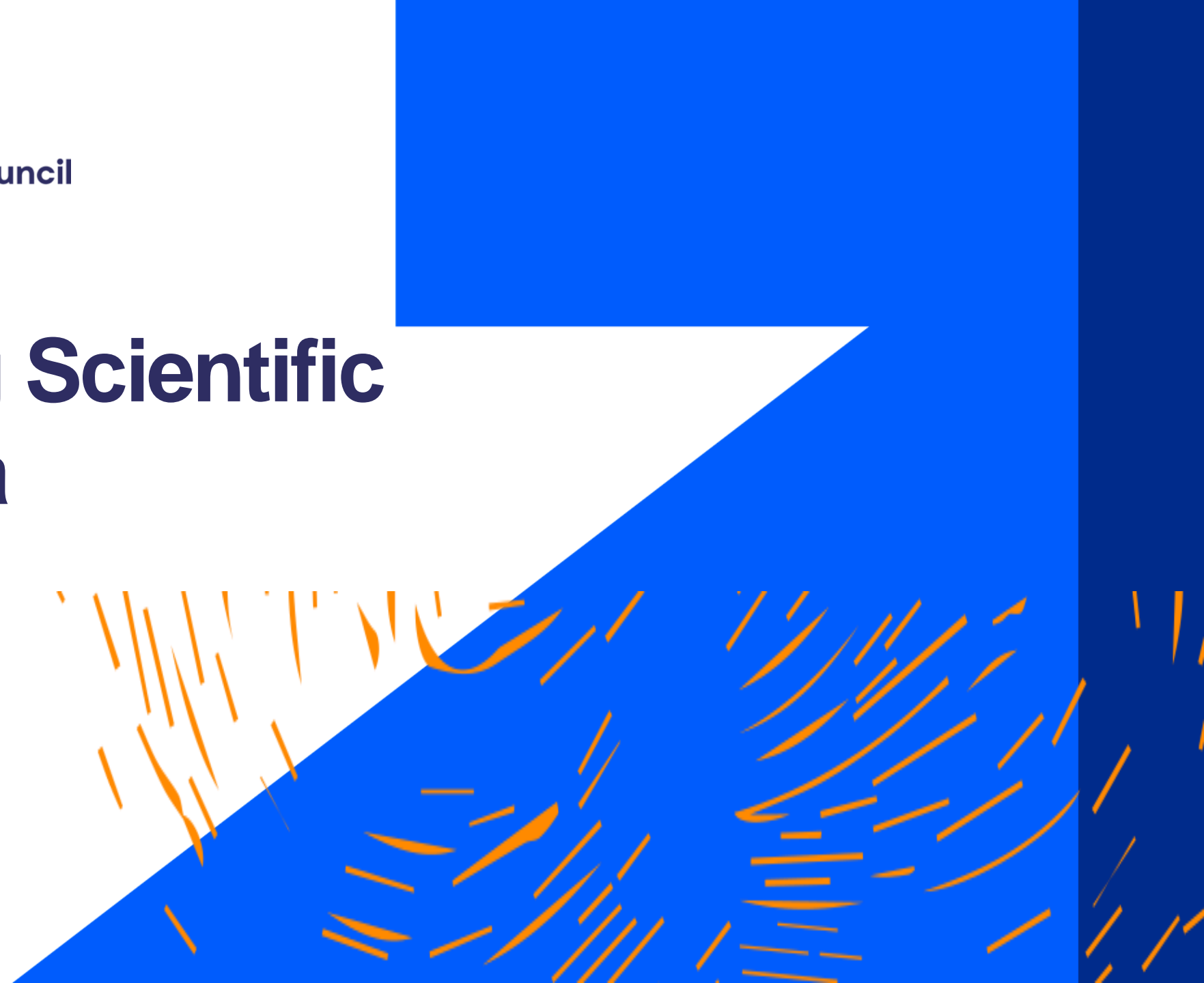
Authorization comparison

- ICAT rules are more flexible but also more complex
- SciCAT “rules” are baked in, ICAT each new facility needs to create their own rules – even “standard” ones
 - E.g. isPublished makes data open, ICAT needs an explicit rule (e.g. ISIS has one based on releaseDate)
- Facilities can change their ICAT rules at any time
- Authorization can be a performance bottleneck in ICAT



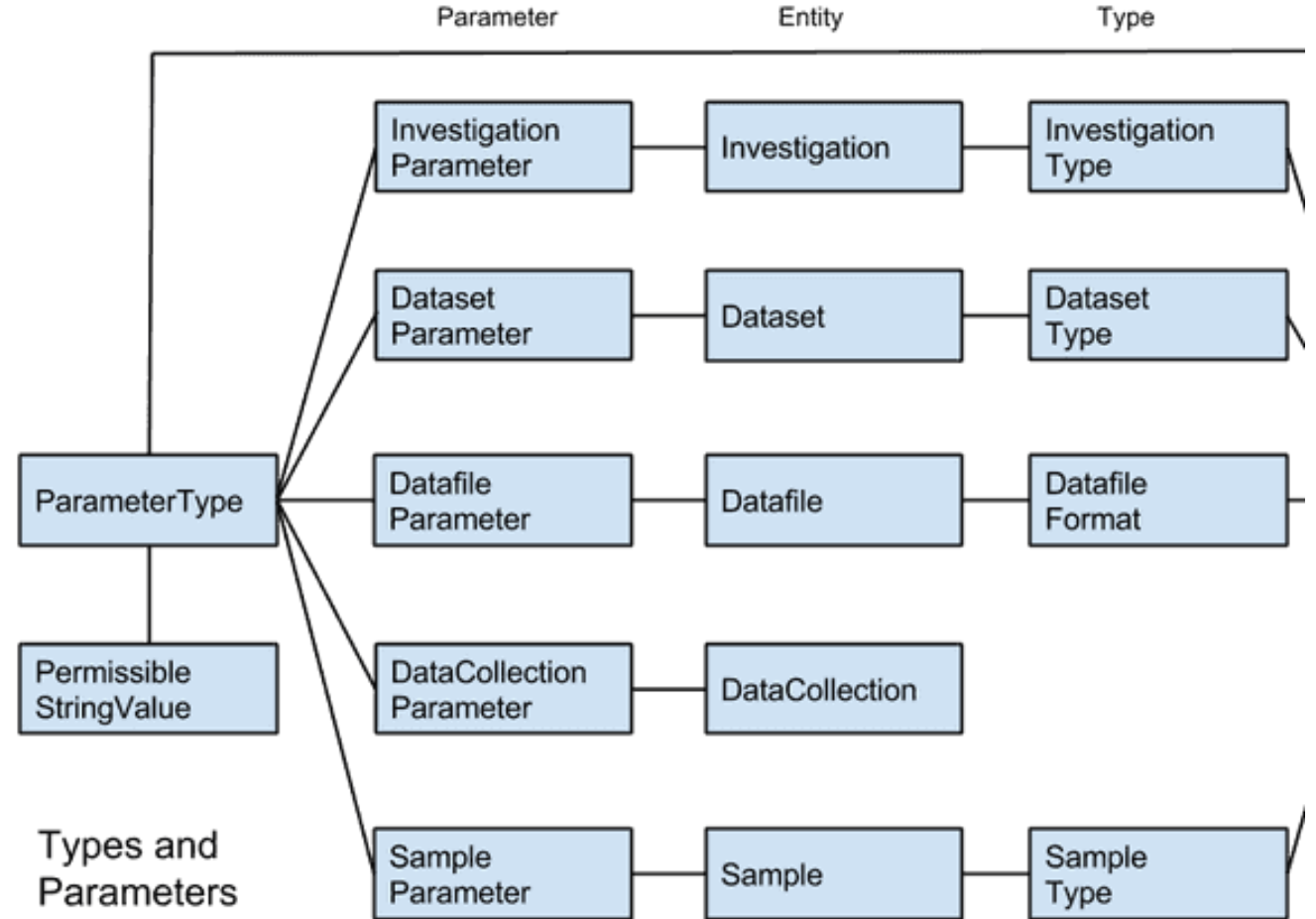
Science and
Technology
Facilities Council

Handling Scientific Metadata



Parameters

- _____Parameter entities contain value
- ParameterType gives metadata about that parameter (e.g. name, units, constraints etc.)
- Need to create ParameterType before creating parameters that link to it
- ESRF use parameters extensively, and use name prefixes to create “hierarchies” e.g. CRYOEM_xparameter



Searching Parameters

- ParameterType is its own table, and so can be queried across to find e.g. Investigations which have a certain parameter type with certain value
- At STFC, we index parameters in Lucene search component and can filter using them

Search for investigations, datasets and datafiles

tomato

Types (3) Start d

For example "instrument calibration" or neutron AND scattering. See all search options.

Display as cards Clear filters

Investigation 2 Dataset 2 Datafile 8

Filters

Type

Parameter name

bcat_inv_str 2

run_number_r... 2

Parameter filters +

No parameter filters

Sample

<input type="checkbox"/>	Title	Experiment Part
<input type="checkbox"/>	Waxed tomato cuticle	25 - LOQ
<input type="checkbox"/>	De-wax tomato cuticle	24 - LOQ

New parameter filter

Parameter name

run_number_range

Parameter type

String

Parameter equals

9921, 9924 1

9922, 9925 1

Searching Parameters

Human Organ Atlas EXPLORE SEARCH 3D RECONSTRUCTIONS TUTORIALS HELP 2 RESULTS

PATIENT

Sex

male

female

Age

Height (cm)

Weight (kg)

SAMPLE

Organ

lung

kidney

heart

spleen

brain

Pathology

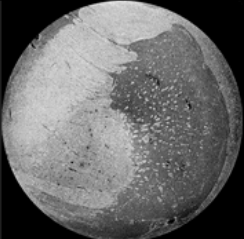
SCAN PARAMETERS

Pixel size (um)

6.36um_ROI-13

LADAF-2021-17 brain 19/06/2024

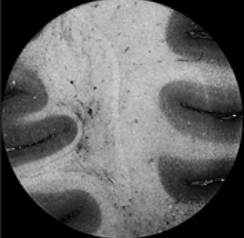
Vertical column in local tomography at 6.36 um voxel size performed by HiP-CT on the beamline BM18 of the brain from the body donor LADAF-2021-17 using half-acquisition protocol.



6.54um_ROI-02

LADAF-2021-17 brain 19/06/2024

Vertical column in local tomography at 6.54 um voxel size performed by HiP-CT on the beamline BM18 of the brain from the body donor LADAF-2021-17 using half-acquisition protocol.





Science and
Technology
Facilities Council

Data Retrieval



Data Retrieval

- User selects data to add to their “cart”
- User chooses download type
- Download will be submitted for restoring
- User can opt-in to an email notification
- User visits web service to check download status & download HTTPS downloads
- Other download types require interaction with 3rd party (e.g. Globus server, facility file system)

Confirm Your Download

Download Name (optional)
my download

Enter a custom file name or leave as the default format (facility_date_time).

Access method
Globus

- HTTPS
- Globus
- Restore to Diamond online

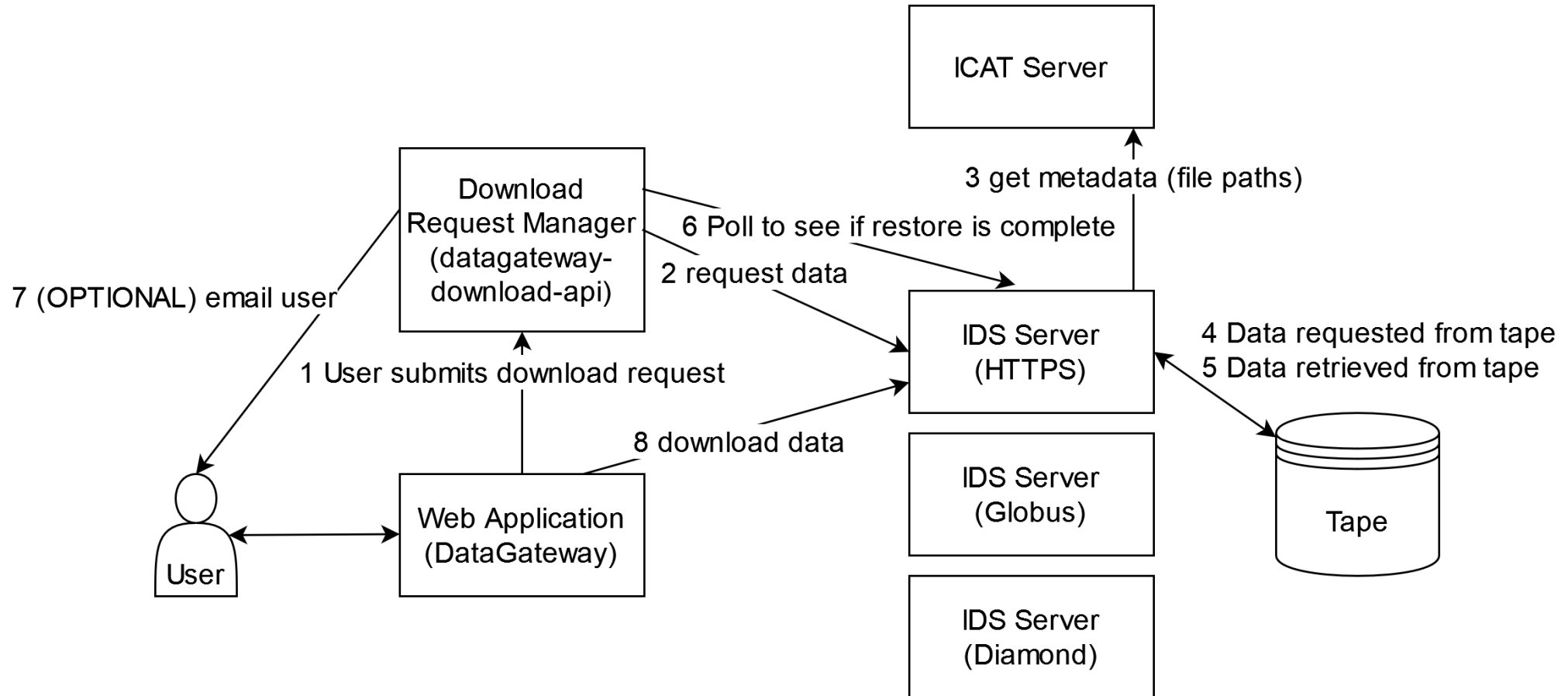
Download Size: 139.53 KB

Email Address (optional)

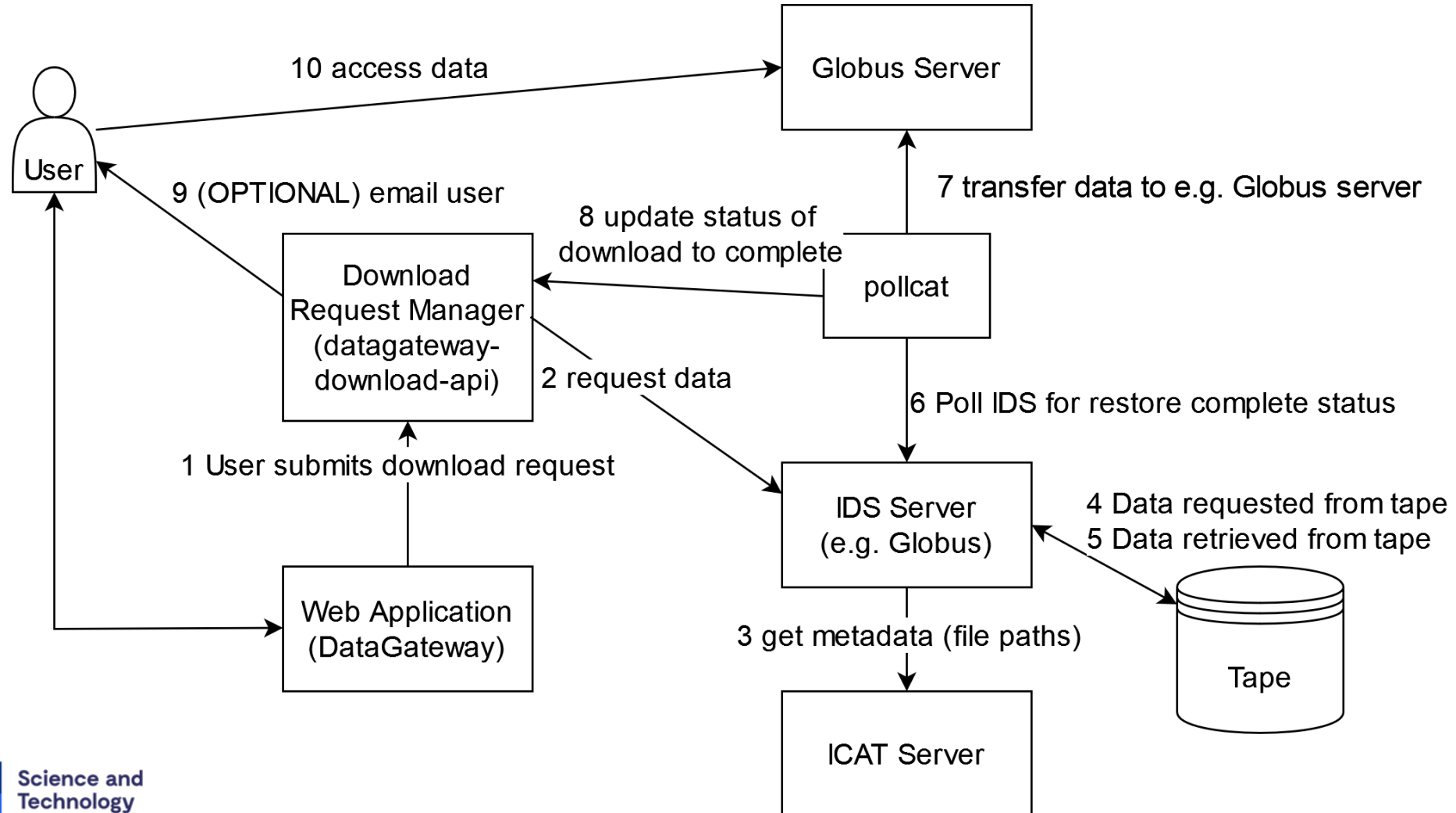
Send me download status messages via email.

Download

Data Retrieval - Technical



Data Retrieval - Technical





Science and
Technology
Facilities Council

Questions?



Science and
Technology
Facilities Council

Thank you



Science and Technology Facilities Council



@STFC_matters



Science and Technology Facilities Council