ExPaNDS
European Open Science Cloud Photon and Neutron Data Services

Demonstrate ICAT and SciCat released with APIs compatible with ExPaNDS federated EOSC services

Document Control Information

Settings	Value
Document Identifier:	D3.3
Project Title:	ExPaNDS
Work Package:	WP3
Document Author(s):	Alejandra Gonzalez-Beltran (UKRI STFC), Carlo Minotti (PSI), Louise Davies (UKRI STFC), Marco Leorato (MAXIV), Matthew Richards (UKRI STFC), Rolf Krahl (HZB), Sudha Padmanabhan (MAXIV), Viktor Bozhinov (UKRI STFC)
Document Reviewer(s):	Alun Ashton (PSI), Patrick Fuhrmann (DESY), Paul Millar (DESY), Sophie Servan (DESY)
Doc. Issue:	1.1
Dissemination Level:	Public
Date:	16/03/2022

Abstract

We present the deliverable achieved for a *demonstration release of ICAT and SciCat with APIs compatible with ExPaNDS federated EOSC services* and its connection to milestone *Metadata catalogue as EOSC service* in the domain of photon and neutron (PaN) science. The work represents the achievement of deliverable D3.3 and milestone MS13 of the Horizon 2020 ExPaNDS project.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Executive Summary

We report on the status of the implementation and deployment of ICAT[1] and SciCat[2], by providing real examples of their adoption at different facilities. This deliverable has the primary goal of describing the two data catalogues stack and their integration with the EOSC services. It thus provides examples that other facilities in the process of adopting a data catalogue and willing to integrate it with EOSC can follow. In addition, it will contribute to the ExPaNDS training material collected by *Work Package 5*.

The document starts with an overview of the role of the data catalogues, from the perspective of *Work Package 3*, in the ExPaNDS architecture [1] and concentrates on its compatibility with the PaN search API [2] and its connection to the PaN federated search API service[3] [3] and the EOSC marketplace[4]. It then outlines the existing ICAT open-source tool ecosystem and implementation, providing an environment at the ISIS Neutron and Muon Source facility[5], covering aspects from data storage to data retrieval, from the ICAT low-level APIs to services standardising findability, such as the OAI-PMH[6] and the PaN search API. The same concepts are then described for the SciCat stack, in particular, its adoption at MAXIV[7] and a SciCat integration example with the B2FIND[8] and openAIRE[9] EOSC[10] services is taken from the Paul Scherrer Institut[11]. For both data catalogues software stacks, the dedicated sections focus on explaining in detail their implementation and deployment.

A short mention of the integration with Google Dataset Search[12], even if not strictly related to the EOSC services, is included here, as it represents yet another tool going in the common ExPaNDS direction of making data FAIR[13].

[1] https://icatproject.org
[2] https://scicatproject.github.io
[3] https://federated.scicat.ess.eu
[4] https://marketplace.eosc-portal.eu
[5] https://www.isis.stfc.ac.uk/Pages/home.aspx
[6] https://www.openarchives.org/pmh
[7] https://www.maxiv.lu.se
[8] http://b2find.eudat.eu
[9] https://www.openaire.eu
[10] https://eosc-portal.eu
[11] https://www.psi.ch
[12] https://datasetsearch.research.google.com
[13] https://www.go-fair.org/fair-principles

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Table of Contents

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 857641.

1. ICAT and SciCat demonstrators

1.1 Background

The work outlined in this document is part of the ExPaNDS project[14], carried out in close communication with the PaNOSC project[15], thus representing the majority of European Photon and Neutron sources in a coordinated activity to drive forward Findable, Accessible, Interoperable and Reusable (FAIR) facility data and European Open Science Cloud (EOSC) services.

This deliverable:

> *D3.3: Demonstrate ICAT and SciCat released with APIs compatible with ExPaNDS federated EOSC services.*

aims to demonstrate the flexibility and integrability of the ICAT and SciCat stack with the ExPaNDS federated EOSC services. We present two examples, one for ICAT and one for SciCat, of existing infrastructure at RIs, in particular, we describe the ICAT stack at ISIS and the SciCat instances at MAXIV and PSI.

Milestone related to this deliverable:

> *M13: Metadata catalogue as EOSC service*

We report here on the recent achievement of registering a public PaN facility data repository as an EOSC service[16], under the PSI provider[17].

1.2 Purpose

The main purposes of *D3.3* alluded to in the original proposal and fleshed out by consultation with PaN community representatives, include:

- To provide ICAT and SciCat demonstrations and examples of existing installations at RIs (ISIS Neutron and Muon Source and MAXIV).
- To provide demonstrations and examples of integration with EOSC, both for the PaN federated search-API and other EOSC services, such as B2FIND and openAIRE (ISIS Neutron and Muon Source and PSI).
- To demonstrate the feasibility of finding datasets from Google Dataset Search, showing examples from ISIS Neutron and Muon Source in the UK and PSI in Switzerland.

[14] https://expands.eu/
[15] https://www.panosc.eu/
[16] https://marketplace.eosc-portal.eu/services
[17] https://marketplace.eosc-portal.eu/providers/psi

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

- To provide a status update on the PaN ontologies implementation and the pan-scoring-API and their subsequent adoption by data catalogues.

1.3 Compatibility with ExPaNDS federated EOSC services

1.3.1 Compatibility with the ExPaNDS architecture

To understand the role of data catalogues in the overall ExPaNDS architecture, we quote in figure 1 the architecture overview from D1.6 [1]

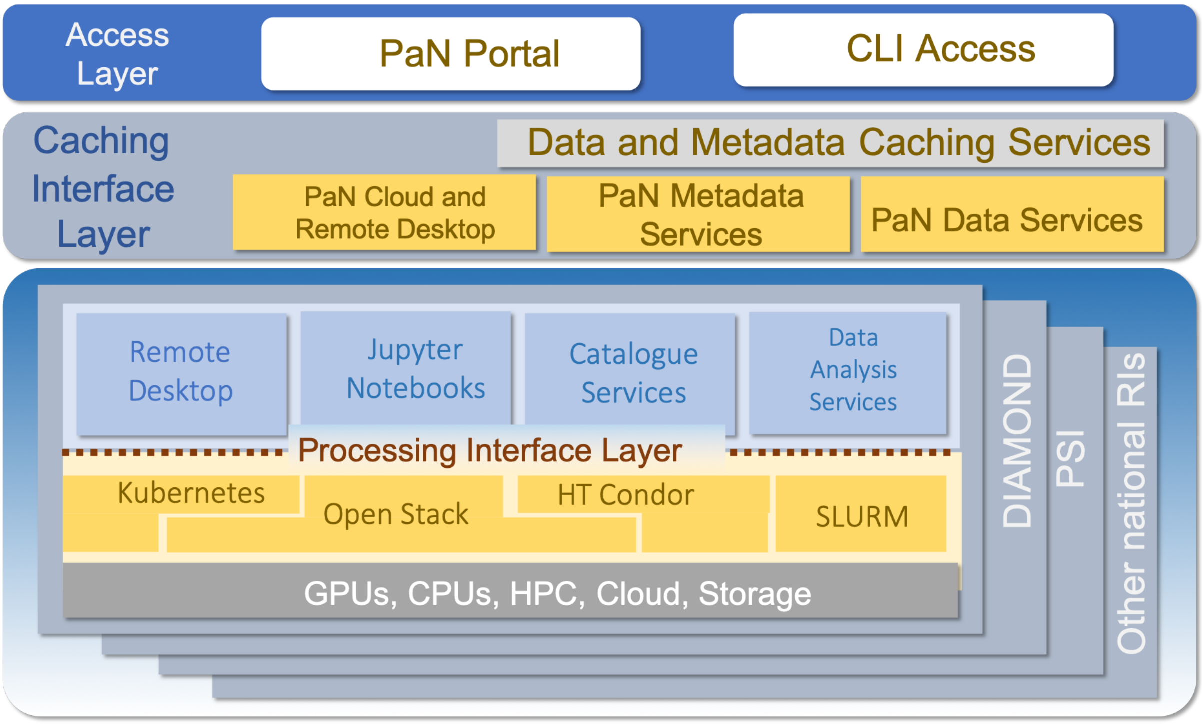


Fig. 1: ExPaNDS architecture - overview

At a high-level, the architecture sketches how data can be accessed and analysed from a federated perspective, in the PaN community and beyond. It consists of a user interface (top layer) pointing to data accessing and analysis tools from different facilities (bottom layer). To enable the communication from the common portal to data at different RIs, the middle layer serves as a hub to standardise the access to facilities and collect the results. This report provides two demonstrations, one for ICAT and one for SciCat, of the *Catalogue Services* component and their connection with the middle layer and with EOSC federated portals (top layer). In particular, we exploit the integrability of data catalogues with the PaN federated search-API and EOSC services such as B2FIND and openAIRE, which, in the diagram of figure 1, sit respectively on the middle and top layer.

Figure 2 depicts a more detailed diagram of the involvement of *Work Package 3* in the ExPaNDS architecture.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

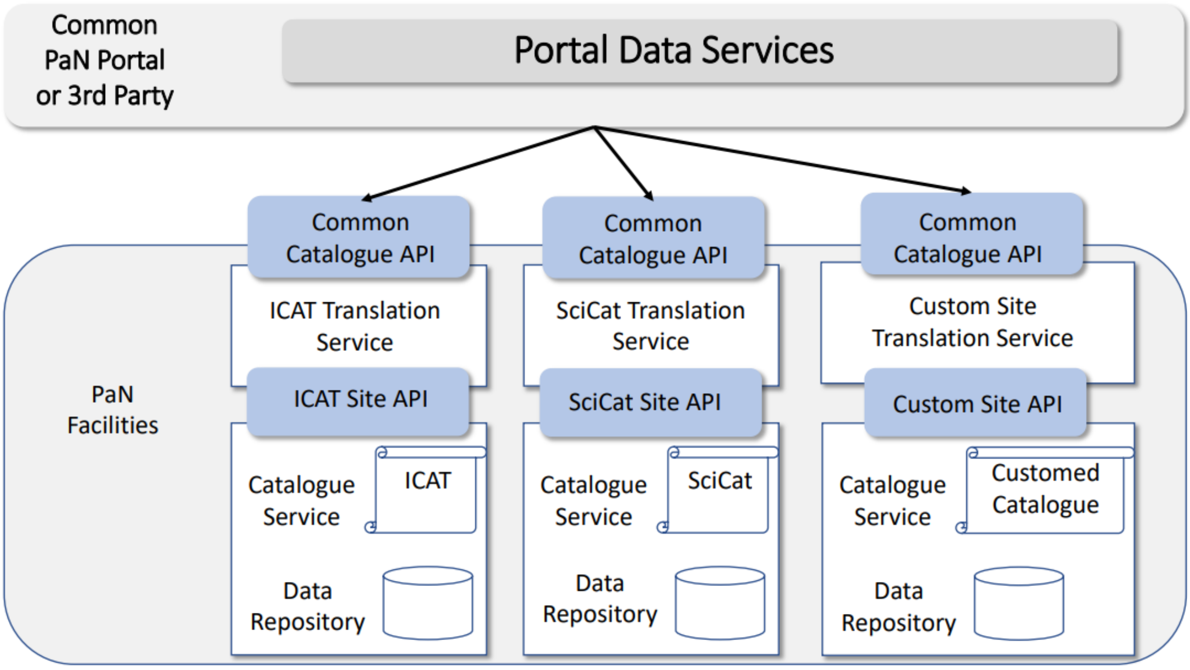


Fig. 2: ExPaNDS architecture - Catalogues services

As mentioned before, the *ICAT stack* will be demonstrated by ISIS Neutron and Muon Source, the *SciCat stack* by MAXIV, the *Common Catalogue API* by the possibility of each to be used in the PaN federated search API and the *Portal Data Services* by the EOSC services. It is worth mentioning that having the *Common Catalogue API* layer available facilitates the integration of additional Portal Data Services, and having chosen the EOSC services was only for the sake of the demonstration. Other examples of *Portal Data Services* include the *PaN Federated Portal*[18] and the *VISA Portal*[19], which are currently under development in the PaNOSC-ExPaNDS collaboration.

The following sections describe concrete examples of the *Catalogue Services* component, covering different use cases in the context of FAIR: PaN federated search-API service, integration with EOSC services and availability of the metadata in web systems such as Google Dataset search.

1.3.2 The PaN federated search-API service

In the PaN community, end-users often need to access and filter PaN data coming from different facilities. Therefore, The PaN federated search-API service is designed as a RESTful service to forward the end-user's query to the data catalogue of each facility and to serve as a hub for collecting PaN data from facilities. There are three components involved:

- The PaN search-API, which sets the APIs for the search interface to the data catalogues,

[18] https://github.com/panosc-portal/frontend
[19] https://visa.ill.eu/login?returnUrl=%2Fhome

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

- the PaN federated search-API service, which is a RESTful service, implementing the PaN search-APIs protocol, forwarding queries and aggregating results from search API services at different facilities.
- The (ICAT, SciCat, Custom Site) search-API RESTful services, implementing the PaN search-APIs protocol, translating the incoming request following the PaN search-API standard to one specific to the Site API and responding accordingly to the PaN search-API standard.

The three of them were designed and developed by collaborators in the PaNOSC and ExPaNDS projects. In the course of the ExPaNDS project, both ICAT and SciCat projects have developed an interface that retrieves PaN data from the data catalogues, compliant with the PaN search-API requirements, and which are then consumed by the PaN federated search-API service. It is noted that the ICAT and SciCat PaN interface, or PaN search-API service, differ from their respective data catalogues service and corresponding API, as they provide functionalities very specific to the PaN community. Figure 3 shows the same diagram as figure 2, but conjugated to the PaN search-API scenario.

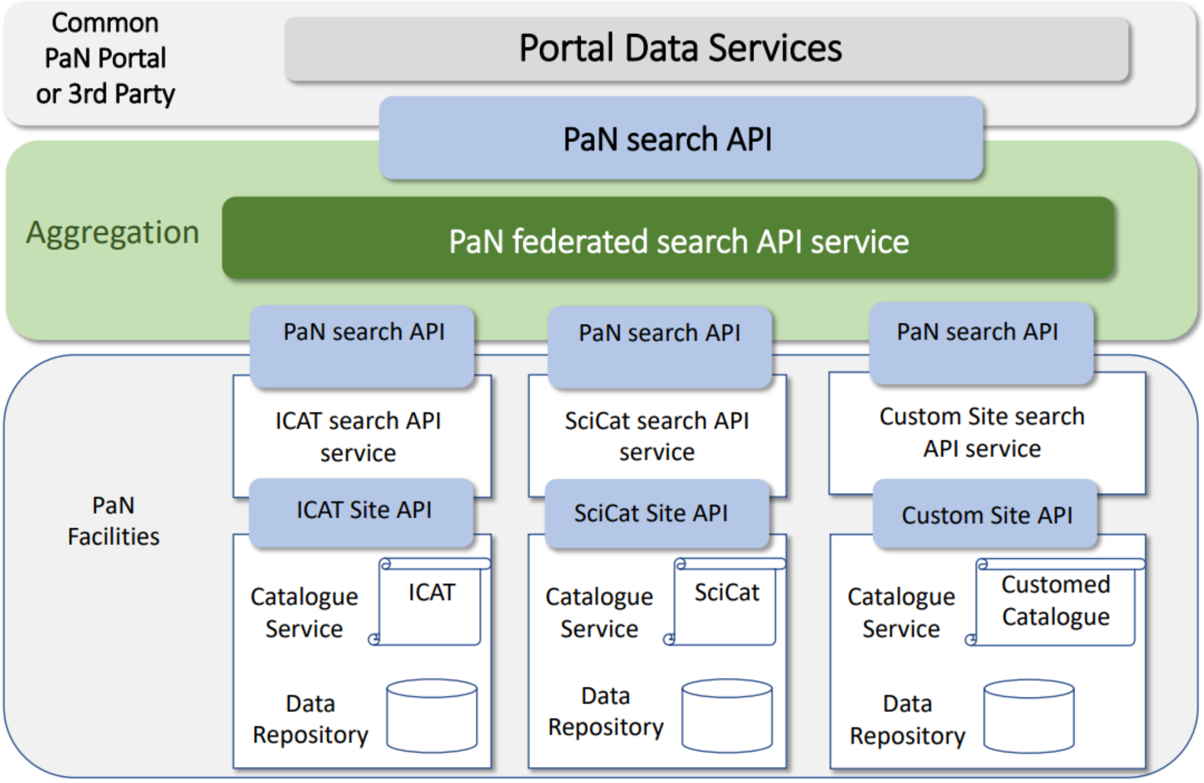


Fig. 3: PaN search-API architecture

As figure 3 shows, each facility will develop and deploy its PaN search-API service and link it to the PaN federated search-API service[20], which is currently hosted by ESS[21], a PaNOSC partner. On this subject, we will provide demonstrations from ISIS Neutron and Muon Source and MAXIV.

[20] https://federated.scicat.ess.eu
[21] https://europeanspallationsource.se

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

1.3.3 The EOSC marketplace: B2FIND and openAIRE

B2FIND and openAIRE cover a use case very similar to the ones aforementioned, differing only from the target community searching for the data. These services provide an interface to data, not limited to the PaN field as their aim is to aggregate as many datasets as possible, relevant to the end-user's query. The data catalogues in ExPaNDS, as a part of their FAIR commitment, have implemented interfaces to be harvested by these EOSC services, named (ICAT, SciCat, Custom Site) OAI-PMH service in figure 4. This interface is usually addressed as specified by the OAI-PMH protocol and follows the Dublin Core metadata guidelines[22].

Figure 4 shows the diagram of figure 2, specialised for the B2FIND and openAIRE services.

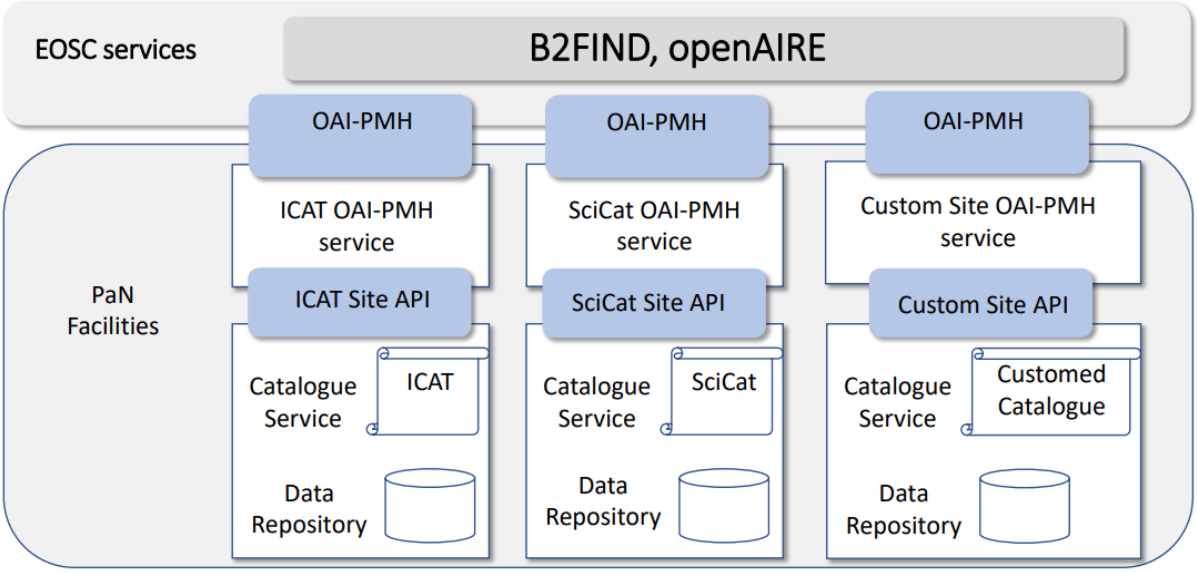


Fig. 4: EOSC services architecture

As figure 4 shows, the flexibility of the *Catalogue Services* architecture allows, as long as a site-specific interface exists, to integrate the data catalogues with different search portals. In this case, B2FIND and openAIRE cover both the aggregation part and the end-user's User Interface (UI).

The ExPaNDS GitHub repository[23] in the ExPaNDS GitHub organisation[24] gives a more detailed explanation of the OAI-PMH and in general delivering data services to EOSC[25].
Further documentation on the harvesting procedure is described by the B2FIND[26] and openAIRE[27] EOSC services.

It is noted that once the OAI-PMH interface is in place, any portal using such a protocol can easily harvest the PaN facilities metadata for integration.

[22] https://dublincore.org
[23] https://github.com/ExPaNDS-eu/ExPaNDS
[24] https://github.com/ExPaNDS-eu
[25] https://github.com/ExPaNDS-eu/ExPaNDS/wiki/Delivering-data-services-to-EOSC
[26] http://b2find.eudat.eu/guidelines/providing.html
[27] https://guidelines.openaire.eu/en/latest

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

1.3.4 A further step towards FAIR: Google Dataset Search

To further demonstrate the wide use of the *Catalogue Service* architecture, we report here on the integration with Google Dataset Search. The underlying idea is similar to the previous ones and ultimately translates into setting up an interface from the data catalogues to the Google Dataset Search engine. This usually means adding the schema.org markup[28] or DCAT markup[29] to a web page[30] owned by the PaN facility which has access to the data. This often takes place in the RI's Landing Page (otherwise called Public Data Repository Dashboard). As before, we present the *Catalogue Service* architecture diagram in the context of Google Dataset Search:

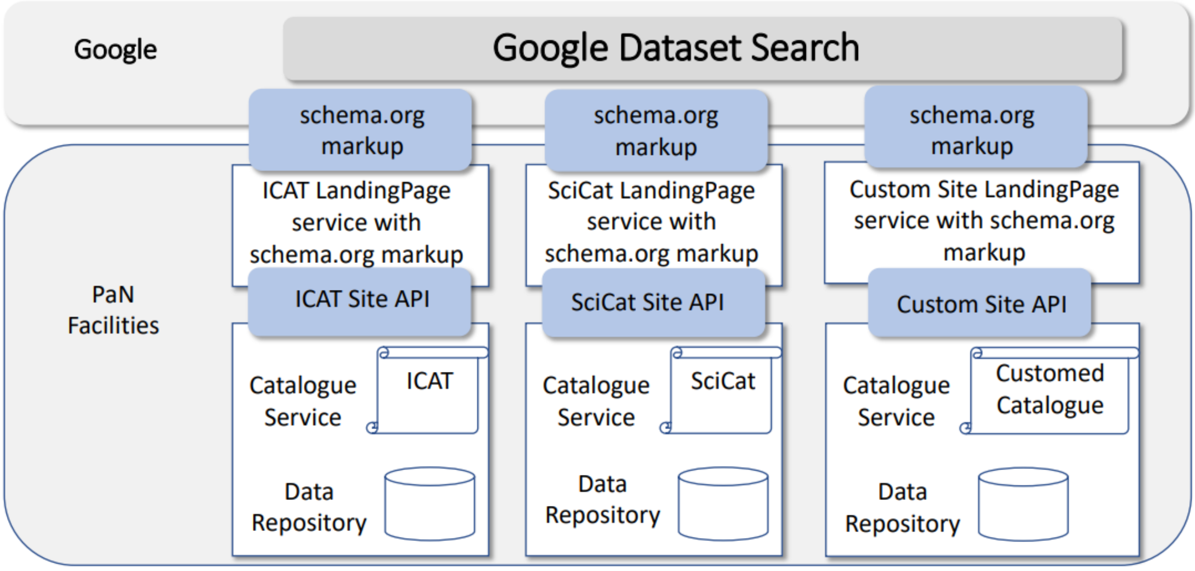


Fig. 5: Landing Page architecture

We will present example implementations of the markup in dataset landing pages for both ISIS Neutron and Muon Source and PSI.

2. ICAT

This section focuses on describing the ICAT data management platform and how it has been extended to support the functionality discussed in this deliverable around making available the metadata in different APIs and/or metadata harvesting platforms.

ICAT is an open-source tool ecosystem[31] to support Photon and Neutron data management that has been used in production systems in large-scale facilities across Europe for over 10 years. The community behind ICAT development is the ICAT collaboration, and it includes partners such as **Diamond Light Source** (ExPaNDS, UK), the **European Synchrotron**

[28] https://schema.org
[29] https://www.w3.org/TR/vocab-dcat-2
[30] https://developers.google.com/search/docs/advanced/structured-data/dataset
[31] https://icatproject.org/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Radiation Facility (PaNOSC, France), **Helmholtz-Zentrum Berlin für Materialien und Energie** (ExPaNDS, Germany), **ISIS Neutron and Muon Source** (ExPaNDS, UK), **ALBA Synchrotron** (ExPaNDS, Spain), **Central Laser Facility** (UK) and **STFC Scientific Computing Department** (ExPaNDS, UK).

ICAT handles all of the data lifecycle, from data creation up to data publication, and has components enabling data discovery, data access and retrieval as well as data preservation, either in one-level or two-level storage (i.e. using tape systems). ICAT successfully manages large-scale volumes of data: for example, Diamond's data archive hosts over 3.4 billion files, which equates to 33.4PB (these figures are from January 2022).

The current installation and deployment process is documented for each component and the main entry point to the documentation is available on the ICAT installation page[32].

2.1 Architecture diagram/data flow

The ICAT stack follows a modular architecture whose control flow can be seen in Figures 6..

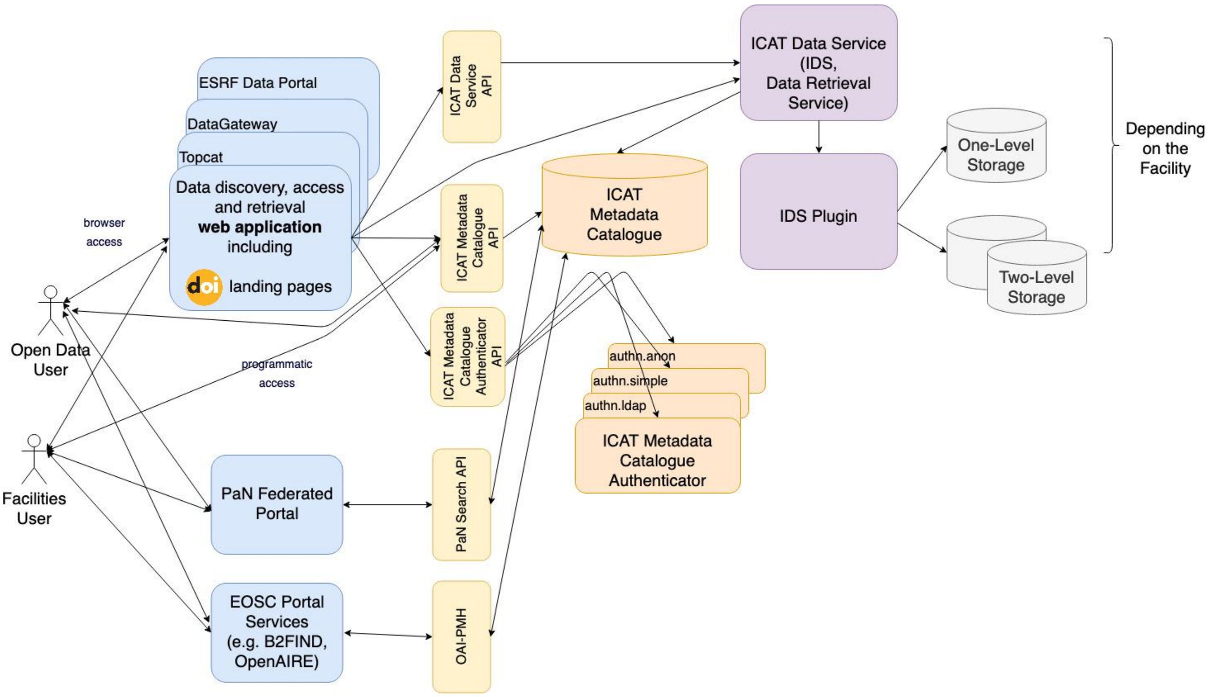

ESRF Data Portal
DataGateway
Topcat
Data discovery, access and retrieval **web application** including
doi landing pages
browser access
Open Data User
programmatic access
Facilities User
PaN Federated Portal
EOSC Portal Services (e.g. B2FIND, OpenAIRE)
ICAT Data Service API
ICAT Metadata Catalogue API
ICAT Metadata Catalogue Authenticator API
PaN Search API
OAI-PMH
ICAT Data Service (IDS, Data Retrieval Service)
ICAT Metadata Catalogue
IDS Plugin
One-Level Storage
Two-Level Storage
Depending on the Facility
authn.anon
authn.simple
authn.ldap
ICAT Metadata Catalogue Authenticator

Fig. 6: High-level ICAT architecture showing the different existing ICAT components - this architecture may be slightly different in each of the implementing facilities

2.2 Services and their implementations

ICAT is designed in a modular fashion with an ICAT instance composed of several components installed into a JavaEE application server.

[32] https://icatproject.org/installation

Component implementations exist for:

- A **metadata catalogue** component (icat.server) that supports large-scale facilities experimental data, dealing with the whole data lifecycle from proposal through to publication.

- A component that manages the **data storage and retrieval** (ids.server) with support for one- or two-level storage. This component handles data access via multiple means, e.g. HTTP downloads for small to medium files, Globus[33] for large files. This component can also handle transferring the data to other target areas such as high-performance computing clusters and/or cloud buckets.

- **Authenticator services** cover different authentication types (for example Lightweight Directory Access Protocol (LDAP)[34] and OpenID Connect[35]).

- A component for **metadata harvesting** implementing the Open Archives Initiative-Protocol for Metadata Harvesting (**OAI-PMH**) interface[36], which is described in more detail in section 2.2.1.

- SOAP[37] & REST[38] and RESTful[39] application programming interfaces.

- Several options for **user interfaces** that allow users to browse, search, share and download their data (eg. Topcat[40], DataGateway[41], ESRF data portal), together with landing pages for published datasets identified with Digital Object Identifiers.

- An extension of the DataGateway-API for ICAT to support the **PaN search API**[42], which can be installed as a separate component and is described in detail in section 2.2.2.

2.2.1 ICAT OAI-PMH component

The ICAT project provides a dedicated component **icat.oaipmh**[43] implementing the Open Archives Initiative Protocol for Metadata Harvesting[44] (OAI-PMH). It listens at a dedicated endpoint for harvesting requests and connects to icat.server as a client, using the standard

[33] https://www.globus.org/
[34] https://github.com/icatproject/authn.ldap
[35] https://github.com/icatproject/authn.oidc
[36] https://github.com/icatproject/icat.oaipmh
[37] https://repo.icatproject.org/site/icat/server/4.9.1/soap.html
[38] https://repo.icatproject.org/site/icat/server/4.9.1/miredot/index.html
[39] https://github.com/ral-facilities/datagateway-api
[40] https://github.com/icatproject/topcat
[41] https://github.com/ral-facilities/datagateway
[42] https://github.com/ral-facilities/datagateway-api
[43] icat.oaipmh – An OAI-PMH interface for ICAT. Version 1.1.1. https://repo.icatproject.org/site/icat/oaipmh/1.1.1/
[44] Open Archives Initiative Protocol for Metadata Harvesting. https://www.openarchives.org/pmh/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

API. The component does not implement any access controls by itself, but rather relies on standard ICAT access rules to restrict what the component is allowed to see. In most use cases, a facility wants to disseminate only public data via OAI-PMH. This can be achieved by configuring icat.oai-pmh to access icat.server as the anonymous user.

One of the major features of **icat.oaipmh** is the simple, yet very powerful run-time configuration. We need a great deal of flexibility, both on input and on output: on input, because different object types from the ICAT schema may need to be exposed: investigation, dataset, study, or (in future ICAT versions) data publication. On output, because different metadata standards may be requested by the harvesting client. We need to support at least Dublin Core and DataCite, but may want to be open to accommodate also community-specific standards to be developed in the future. In order to provide this flexibility, **icat.oaipmh** uses Extensible Stylesheet Language Transformations (XSLT) to translate on the fly from a generic internal XML representation of the ICAT content to the requested metadata standard. Examples of XSLT files for common use cases are provided. In most cases, facilities only need to slightly tweak these examples to their needs.

The **icat.oaipmh** component supports the definition of sets by search conditions in the run-time configuration and the selective harvesting by datestamps and by sets. The component also supports flow control using resumption tokens.

More details on this component can also be found on the ExPaNDS wiki[45].

2.2.2 ICAT PaN search API component

The Search API implementation allows for interfacing with the ICAT metadata catalogue, as described in section 1.3.2. It provides an easy to use RESTful web interface that was built using the lightweight and popular Python framework, Flask RESTful[46]. The Search API uses Python ICAT[47] to query and retrieve data from an ICAT stack. The Search API was implemented as an extension of the code base of the DataGateway API[48], which already was a Flask-based API that exposes the required endpoints for fetching data from an ICAT instance and interfacing with the DataGateway[49] user portal for data browsing, discovery and retrieval. The DataGateway API had the required logic for communicating with ICAT implemented so this was an ideal opportunity for the logic to be reused for the Search API with some tweaking where necessary. By having the DataGateway and Search API as part of one Flask RESTful API, ICAT facilities now have the choice of running both APIs as one instance or separate instances.

[45] https://github.com/ExPaNDS-eu/ExPaNDS/wiki/Delivering-data-services-to-EOSC#the-icat-component-for-oai-pmh

[46] https://flask-restful.readthedocs.io/en/latest/

[47] https://github.com/icatproject/python-icat

[48] https://github.com/ral-facilities/datagateway-api

[49] https://github.com/ral-facilities/datagateway

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

2.3 PaN federated search-API at ISIS Neutron and Muon Source

2.3.1 Deployment

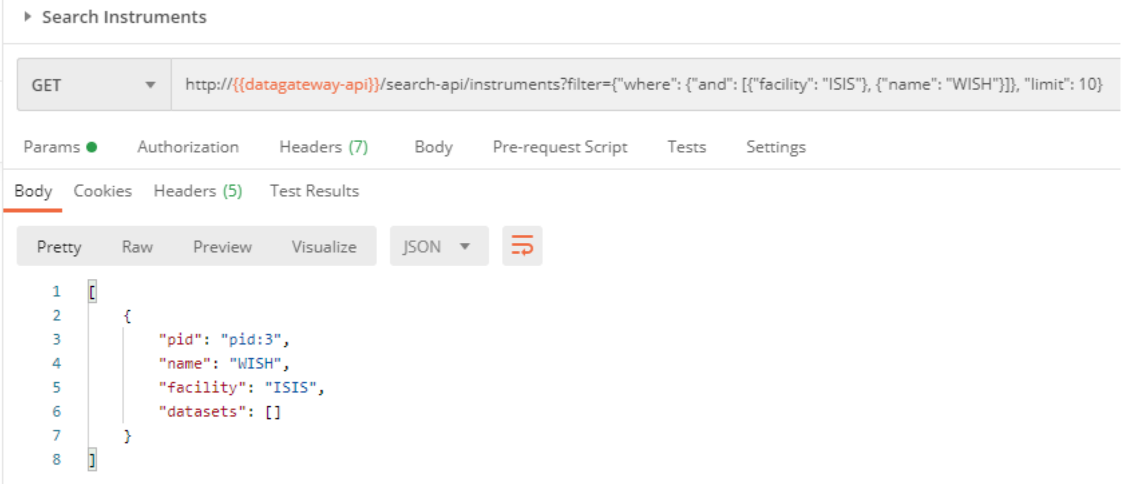
The ISIS data catalogue and archive is deployed in two environments: a production environment made available to the users (ISIS staff and users, as well as available in the open web) and a pre-production environment that is deployed behind a firewall in a private subnet.

The Search API has been deployed on a development ISIS server. This is not externally reachable at the moment, because the API is inside an organisation firewall, but it is being used for testing purposes before the API is deployed on a production instance in the next phase. The Search API at ISIS is running on an Apache web server using mod_wsgi, but any web server that supports WSGI can be used as an alternative. Deployment is automated using an Ansible Playbook[50] so redeploying the API when a new release is made (e.g. to fix a bug) is fast.

In the near future (during the next facility shutdown), the search API will be deployed on the production ISIS instance (using a similar method of deployment). When other ICAT facilities have the capacity and time to do so within their operational cycles, they will also deploy and make use of the ICAT implementation of the Search API.

2.3.2 Usage and examples

As agreed in the specification of the Search API, there are 11 endpoints that can be queried for public ISIS data. Figure 7 shows a request to query a particular instrument called ‘WISH’ that is part of ISIS. No datasets are present in the response because the request doesn’t ask to include them, but a user would be able to view datasets associated with this instrument given the include filter in the request.



```
GET http://{{datagateway-api}}/search-api/instruments?filter={"where": {"and": [{"facility": "ISIS"}, {"name": "WISH"}]}, "limit": 10}
```

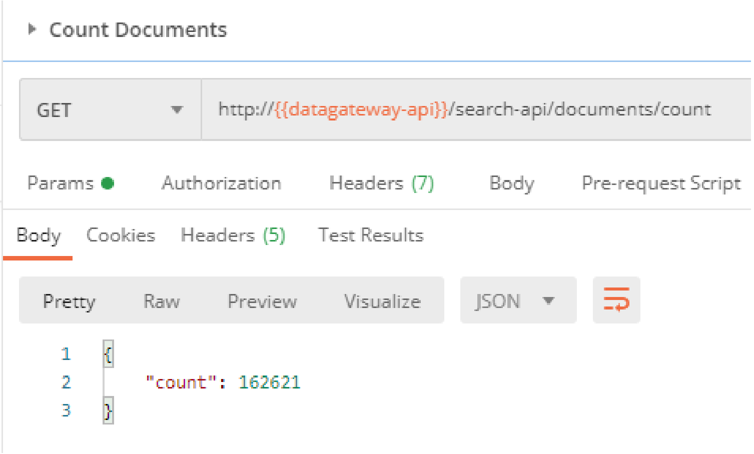
```
[
    {
        "pid": "pid:3",
        "name": "WISH",
        "facility": "ISIS",
        "datasets": []
    }
]
```

Fig. 7: Search instruments using query filters

[50] https://github.com/ral-facilities/scigateway-ansible

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 857641.

A user can also query the API to view the number of documents held in the ICAT stack for ISIS (shown in figure 8).



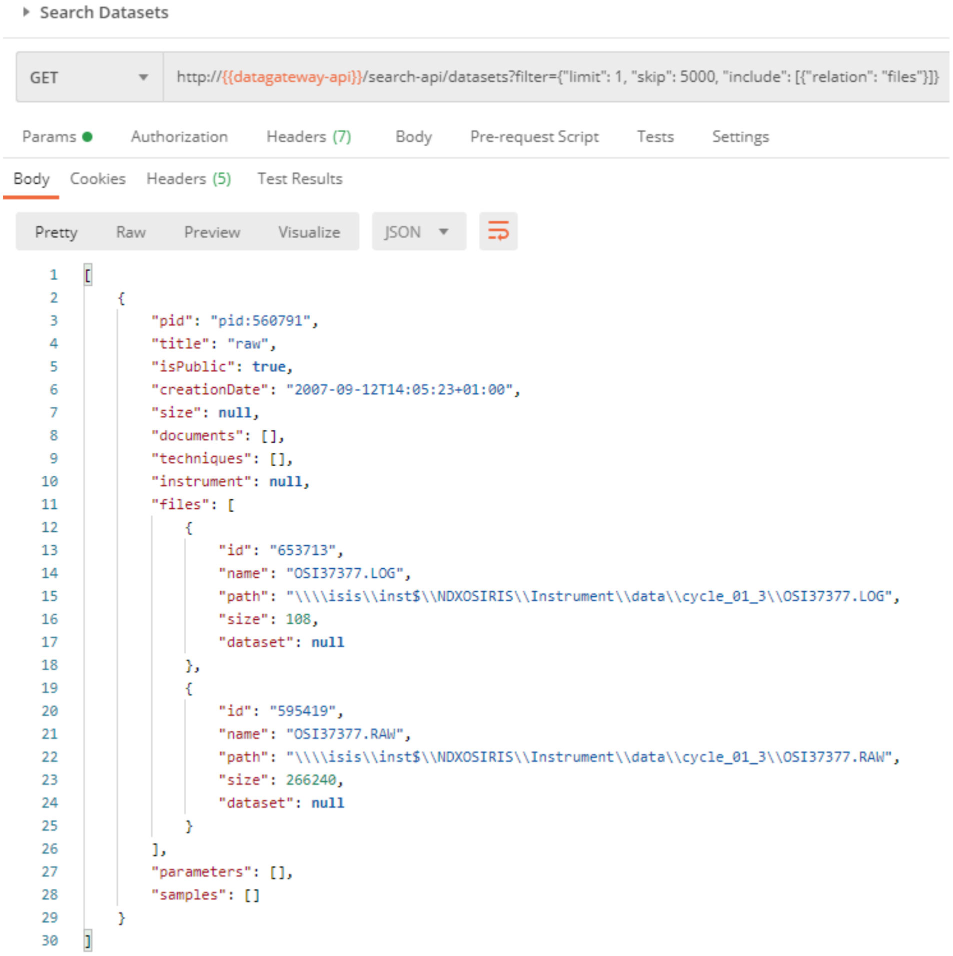
Count Documents

GET http://{{datagateway-api}}/search-api/documents/count

```
{
    "count": 162621
}
```

Fig. 8: Count number of documents for ISIS

If a user requires a deeper understanding of the data, they can send a request to look at the datasets. With an “include” filter, they can also view the dataset’s files (shown in figure 9).



Search Datasets

GET http://{{datagateway-api}}/search-api/datasets?filter={"limit": 1, "skip": 5000, "include": [{"relation": "files"}]}

```
[
    {
        "pid": "pid:560791",
        "title": "raw",
        "isPublic": true,
        "creationDate": "2007-09-12T14:05:23+01:00",
        "size": null,
        "documents": [],
        "techniques": [],
        "instrument": null,
        "files": [
            {
                "id": "653713",
                "name": "OSI37377.LOG",
                "path": "\\\\isis\\inst$\\NDXOSIRIS\\Instrument\\data\\cycle_01_3\\OSI37377.LOG",
                "size": 108,
                "dataset": null
            },
            {
                "id": "595419",
                "name": "OSI37377.RAW",
                "path": "\\\\isis\\inst$\\NDXOSIRIS\\Instrument\\data\\cycle_01_3\\OSI37377.RAW",
                "size": 266240,
                "dataset": null
            }
        ],
        "parameters": [],
        "samples": []
    }
]
```

Fig. 9: Search datasets and include their associated files

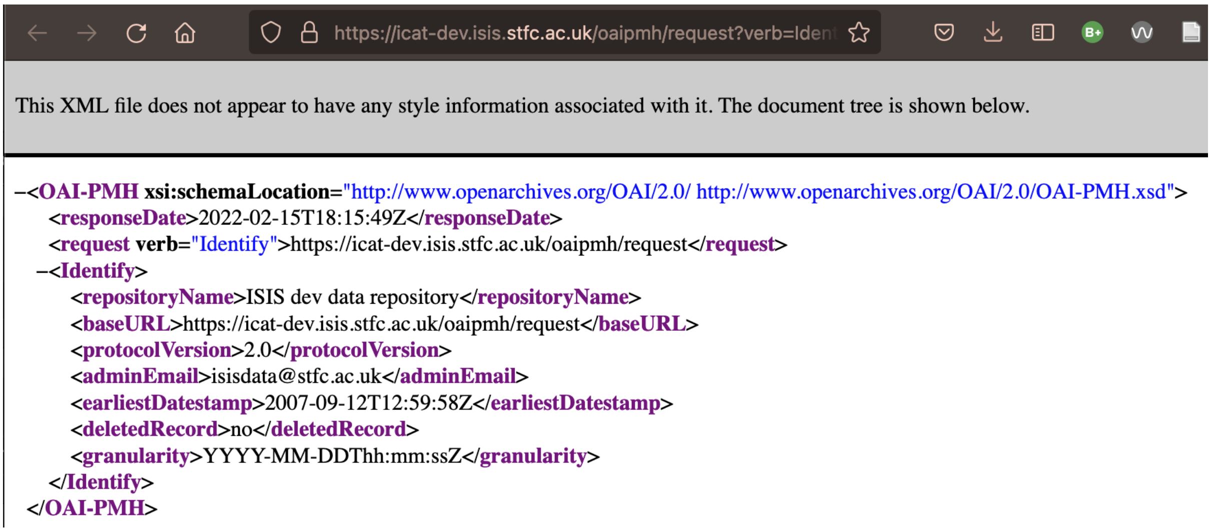
This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 857641.

2.4 Connection to EOSC services at ISIS Neutron and Muon Source

2.4.1 Deployment

In order to integrate the ISIS Neutron and Muon Source open data into the EOSC services such as B2FIND and OpenAire, ISIS has deployed the icat.oaipmh component in their development service. The deployment of this new component was done by ISIS User Programme Software group.

The ISIS OAI-PMH endpoint[51] and the details of the repository can be seen here[52]:



This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
–<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
    <responseDate>2022-02-15T18:15:49Z</responseDate>
    <request verb="Identify">https://icat-dev.isis.stfc.ac.uk/oaipmh/request</request>
  –<Identify>
      <repositoryName>ISIS dev data repository</repositoryName>
      <baseURL>https://icat-dev.isis.stfc.ac.uk/oaipmh/request</baseURL>
      <protocolVersion>2.0</protocolVersion>
      <adminEmail>isisdata@stfc.ac.uk</adminEmail>
      <earliestDatestamp>2007-09-12T12:59:58Z</earliestDatestamp>
      <deletedRecord>no</deletedRecord>
      <granularity>YYYY-MM-DDThh:mm:ssZ</granularity>
    </Identify>
  </OAI-PMH>
```

Fig. 10: description of the ISIS OAI-PMH development endpoint

The ISIS and SCD teams have interacted with B2FIND and are in the process of getting the harvesting pipeline tested. After that, ISIS will provide the OAI-PMH endpoint in the production system so that all the ISIS metadata for the public data can be made available within the B2FIND search engine.

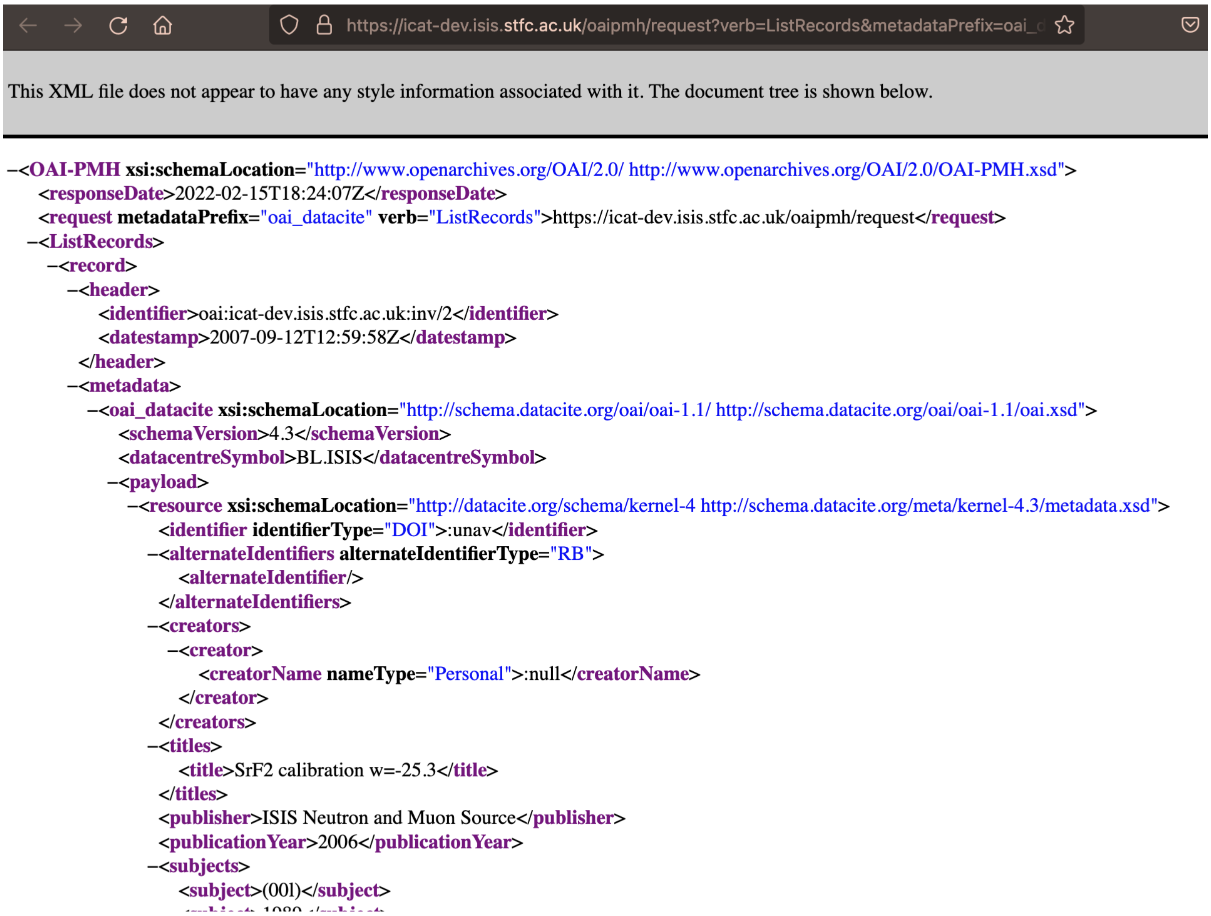
2.4.2 Usage and examples

At the time of writing, the ISIS OAI-PMH development endpoint exposes around 162 thousand entries. Figure 11 shows a screenshot of the request to see the list of records.

[51] https://icatisis.esc.rl.ac.uk/oaipmh/ ; https://icatisis.esc.rl.ac.uk/oaipmh/request?verb=ListRecords&metadataPrefix=oai_datacite

[52] https://icatisis.esc.rl.ac.uk/oaipmh/request?verb=Identify

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.



https://icat-dev.isis.stfc.ac.uk/oaipmh/request?verb=ListRecords&metadataPrefix=oai_d

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
–<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
    <responseDate>2022-02-15T18:24:07Z</responseDate>
    <request metadataPrefix="oai_datacite" verb="ListRecords">https://icat-dev.isis.stfc.ac.uk/oaipmh/request</request>
  –<ListRecords>
    –<record>
      –<header>
          <identifier>oai:icat-dev.isis.stfc.ac.uk:inv/2</identifier>
          <datestamp>2007-09-12T12:59:58Z</datestamp>
        </header>
      –<metadata>
        –<oai_datacite xsi:schemaLocation="http://schema.datacite.org/oai/oai-1.1/ http://schema.datacite.org/oai/oai-1.1/oai.xsd">
            <schemaVersion>4.3</schemaVersion>
            <datacentreSymbol>BL.ISIS</datacentreSymbol>
          –<payload>
            –<resource xsi:schemaLocation="http://datacite.org/schema/kernel-4 http://schema.datacite.org/meta/kernel-4.3/metadata.xsd">
                <identifier identifierType="DOI">:unav</identifier>
              –<alternateIdentifiers alternateIdentifierType="RB">
                  <alternateIdentifier/>
                </alternateIdentifiers>
              –<creators>
                –<creator>
                    <creatorName nameType="Personal">:null</creatorName>
                  </creator>
                </creators>
              –<titles>
                  <title>SrF2 calibration w=-25.3</title>
                </titles>
                <publisher>ISIS Neutron and Muon Source</publisher>
                <publicationYear>2006</publicationYear>
              –<subjects>
                  <subject>(001)</subject>
```

Fig. 11: a list of records from the ISIS OAI-PMH development endpoint[53]

2.5 Connection to Google Dataset Search at ISIS Neutron and Muon Source

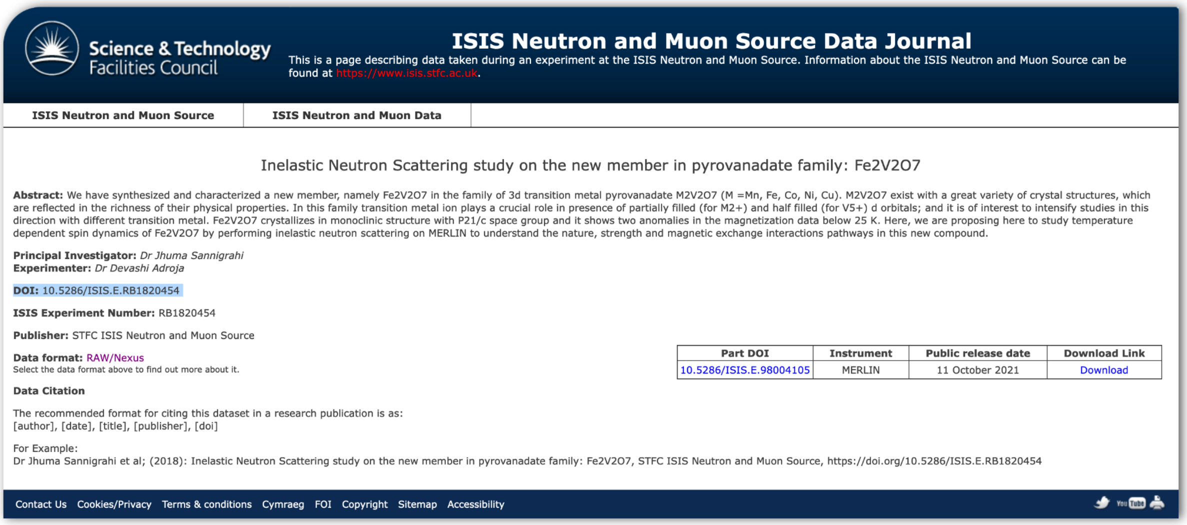
2.5.1 Deployment

The ISIS Neutron and Muon Source data catalogue has adopted an open data policy[54], which states that data is made public after a 3-year long embargo period. Each open dataset is assigned a Digital Object Identifier and landing pages made available. Figure 12 shows an ISIS dataset landing page. The landing pages are made available even for data under the embargo period, and the ICAT system behind the scene determines when the data can be accessible by anyone in the world, or only by those involved in the experiment.

In addition to including all of the DataCite required metadata, the ISIS dataset landing pages include schema.org mark-up as per Google guidelines on structured data[55], so that the

[53] https://icat-dev.isis.stfc.ac.uk/oaipmh/request?verb=ListRecords&metadataPrefix=oai_datacite
[54] https://www.isis.stfc.ac.uk/Pages/Data-Policy.aspx
[55] https://developers.google.com/search/docs/advanced/structured-data/dataset

metadata is made available on the web, and in particular by the Google Dataset Search tool. Figure 14 shows the schema.org markup available on the landing page from Figure 12.



Science & Technology Facilities Council

ISIS Neutron and Muon Source Data Journal

This is a page describing data taken during an experiment at the ISIS Neutron and Muon Source. Information about the ISIS Neutron and Muon Source can be found at https://www.isis.stfc.ac.uk.

ISIS Neutron and Muon Source
ISIS Neutron and Muon Data

Inelastic Neutron Scattering study on the new member in pyrovanadate family: Fe2V2O7

Abstract: We have synthesized and characterized a new member, namely Fe2V2O7 in the family of 3d transition metal pyrovanadate M2V2O7 (M =Mn, Fe, Co, Ni, Cu). M2V2O7 exist with a great variety of crystal structures, which are reflected in the richness of their physical properties. In this family transition metal ion plays a crucial role in presence of partially filled (for M2+) and half filled (for V5+) d orbitals; and it is of interest to intensify studies in this direction with different transition metal. Fe2V2O7 crystallizes in monoclinic structure with P21/c space group and it shows two anomalies in the magnetization data below 25 K. Here, we are proposing here to study temperature dependent spin dynamics of Fe2V2O7 by performing inelastic neutron scattering on MERLIN to understand the nature, strength and magnetic exchange interactions pathways in this new compound.

Principal Investigator: *Dr Jhuma Sannigrahi*
Experimenter: *Dr Devashi Adroja*

DOI: 10.5286/ISIS.E.RB1820454

ISIS Experiment Number: RB1820454

Publisher: STFC ISIS Neutron and Muon Source

Data format: RAW/Nexus
Select the data format above to find out more about it.

Part DOI	Instrument	Public release date	Download Link
10.5286/ISIS.E.98004105	MERLIN	11 October 2021	Download

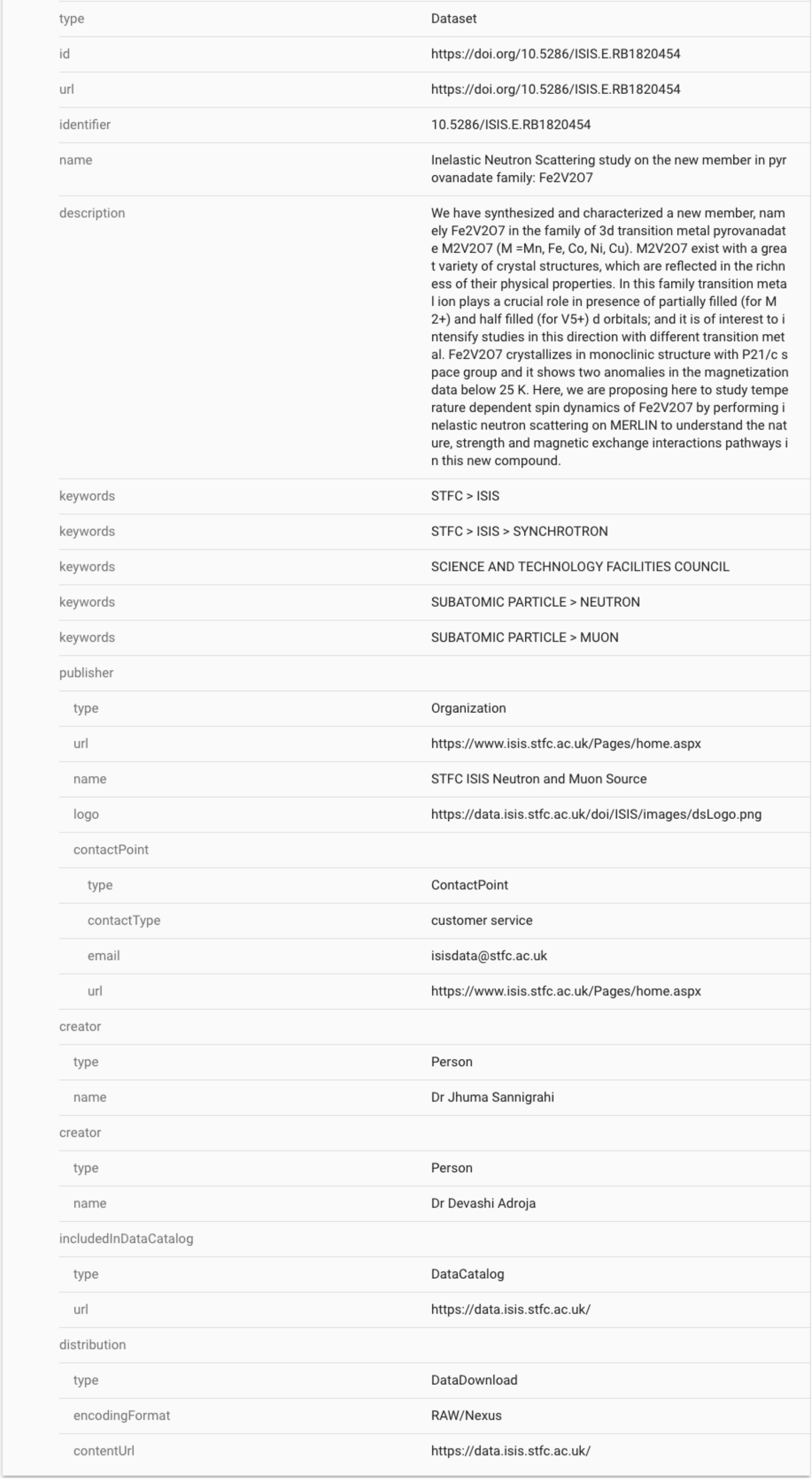
Data Citation

The recommended format for citing this dataset in a research publication is as:
[author], [date], [title], [publisher], [doi]

For Example:
Dr Jhuma Sannigrahi et al; (2018): Inelastic Neutron Scattering study on the new member in pyrovanadate family: Fe2V2O7, STFC ISIS Neutron and Muon Source, https://doi.org/10.5286/ISIS.E.RB1820454

Fig. 12: Landing page for dataset with DOI https://doi.org/10.5286/ISIS.E.98004105, available in the ISIS data catalogue (production instance)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

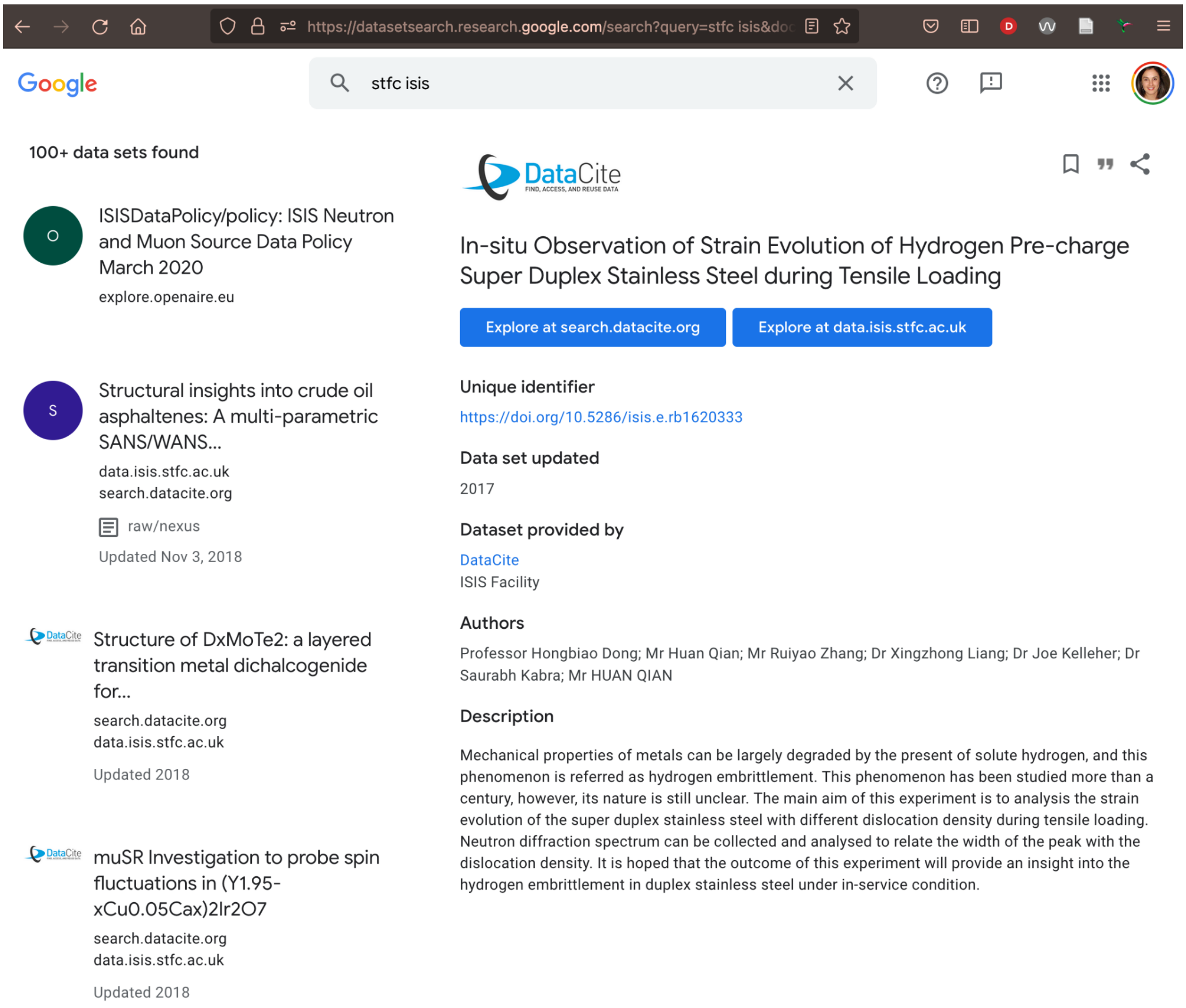


type	Dataset
id	https://doi.org/10.5286/ISIS.E.RB1820454
url	https://doi.org/10.5286/ISIS.E.RB1820454
identifier	10.5286/ISIS.E.RB1820454
name	Inelastic Neutron Scattering study on the new member in pyrovanadate family: Fe2V2O7
description	We have synthesized and characterized a new member, namely Fe2V2O7 in the family of 3d transition metal pyrovanadate M2V2O7 (M =Mn, Fe, Co, Ni, Cu). M2V2O7 exist with a great variety of crystal structures, which are reflected in the richness of their physical properties. In this family transition metal ion plays a crucial role in presence of partially filled (for M 2+) and half filled (for V5+) d orbitals; and it is of interest to intensify studies in this direction with different transition metal. Fe2V2O7 crystallizes in monoclinic structure with P21/c space group and it shows two anomalies in the magnetization data below 25 K. Here, we are proposing here to study temperature dependent spin dynamics of Fe2V2O7 by performing inelastic neutron scattering on MERLIN to understand the nature, strength and magnetic exchange interactions pathways in this new compound.
keywords	STFC > ISIS
keywords	STFC > ISIS > SYNCHROTRON
keywords	SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
keywords	SUBATOMIC PARTICLE > NEUTRON
keywords	SUBATOMIC PARTICLE > MUON
publisher	
type	Organization
url	https://www.isis.stfc.ac.uk/Pages/home.aspx
name	STFC ISIS Neutron and Muon Source
logo	https://data.isis.stfc.ac.uk/doi/ISIS/images/dsLogo.png
contactPoint	
type	ContactPoint
contactType	customer service
email	isisdata@stfc.ac.uk
url	https://www.isis.stfc.ac.uk/Pages/home.aspx
creator	
type	Person
name	Dr Jhuma Sannigrahi
creator	
type	Person
name	Dr Devashi Adroja
includedInDataCatalog	
type	DataCatalog
url	https://data.isis.stfc.ac.uk/
distribution	
type	DataDownload
encodingFormat	RAW/Nexus
contentUrl	https://data.isis.stfc.ac.uk/

Fig. 13: example of the schema.org markup on an ISIS dataset landing page, as seen in the Google Rich Results Test system

2.5.2 Usage and examples

The open datasets from ISIS Neutron and Muon source are available on Google Dataset search and figure 14 shows a screenshot of a search for “stfc isis”[56]. The interface allows the users to use some extra filtering to find specific datasets based on the provided metadata.


https://datasetsearch.research.google.com/search?query=stfc isis&doc
Google
stfc isis

100+ data sets found

ISISDataPolicy/policy: ISIS Neutron and Muon Source Data Policy March 2020
explore.openaire.eu

Structural insights into crude oil asphaltenes: A multi-parametric SANS/WANS...
data.isis.stfc.ac.uk
search.datacite.org
raw/nexus
Updated Nov 3, 2018

Structure of DxMoTe2: a layered transition metal dichalcogenide for...
search.datacite.org
data.isis.stfc.ac.uk
Updated 2018

muSR Investigation to probe spin fluctuations in (Y1.95-xCu0.05Cax)2Ir2O7
search.datacite.org
data.isis.stfc.ac.uk
Updated 2018

In-situ Observation of Strain Evolution of Hydrogen Pre-charge Super Duplex Stainless Steel during Tensile Loading

Explore at search.datacite.org
Explore at data.isis.stfc.ac.uk

Unique identifier
https://doi.org/10.5286/isis.e.rb1620333

Data set updated
2017

Dataset provided by
DataCite
ISIS Facility

Authors
Professor Hongbiao Dong; Mr Huan Qian; Mr Ruiyao Zhang; Dr Xingzhong Liang; Dr Joe Kelleher; Dr Saurabh Kabra; Mr HUAN QIAN

Description
Mechanical properties of metals can be largely degraded by the present of solute hydrogen, and this phenomenon is referred as hydrogen embrittlement. This phenomenon has been studied more than a century, however, its nature is still unclear. The main aim of this experiment is to analysis the strain evolution of the super duplex stainless steel with different dislocation density during tensile loading. Neutron diffraction spectrum can be collected and analysed to relate the width of the peak with the dislocation density. It is hoped that the outcome of this experiment will provide an insight into the hydrogen embrittlement in duplex stainless steel under in-service condition.

Fig. 14: Screenshot of Google Dataset Search showing the STFC ISIS open datasets

[56] https://datasetsearch.research.google.com/search?query=stfc%20isis&docid=L2cvMTFqY2p0NGR2Zw%3D%3D

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 857641.

2.6 Future developments

As part of the work in ExPaNDS, the ICAT collaboration continues to make progress on the required changes to the schema to support FAIRer data, and in particular to support using the PaNET ontology for annotations to improve data harmonisation and search capabilities (e.g. in the keywords from figure 13). These changes involve, for example, including persistent identifiers for the ontology terms that annotate the techniques of instruments as well as of datasets. The discussions of the schema changes being considered can be followed in the icat.server issue tracker[57].

The planned schema extensions also include a new entity needed to store the bibliographic metadata for curated data publications in ICAT (a DataPublication class), which is metadata that can be made open and accessible to any user in the world when a dataset is published. This new class is designed to support rich metadata using the DataCite standard, whose elements are otherwise available in other ICAT entities but are usually only accessible by the data owners. As a result, it will be easier to generate the data publication landing pages from a single class in ICAT, the DataPublication class, and to enable the harvesting of the publication metadata using the OAI-PMH component.

The ICAT collaboration has been working on a continuous deployment based on containerising each of the required ICAT components. A Docker image including a full-stack of ICAT components has been contributed by a member of the ICAT collaboration and is available[58]. This image is used at HZB in production. A second approach to containerise each of the ICAT components individually and orchestrate them for continuous deployment is ongoing[59].

3. SciCat

This section will focus on the SciCat data catalogue, which refers to the SciCat catalogue service, SciCat search API service and data repository box in figure 3. An overview of the architecture is provided first, then some references to the implementation of its components and finally some examples coming from facilities using the SciCat stack, that demonstrate the progress in the project.

3.1 Architecture diagram/data flow

Figure 15 shows the general SciCat architecture. The non-site specific part roughly applies to all facilities using SciCat, even though there might be some minor differences (for example, RIs can connect the Landing Page Server directly to the Catamel API server).

[57] https://github.com/icatproject/icat.server/issues?q=is%3Aissue+is%3Aopen+label%3Aschema

[58] https://hub.docker.com/r/rkrahl/icat

[59] https://github.com/icatproject-contrib/icat-cloud-native-migration

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

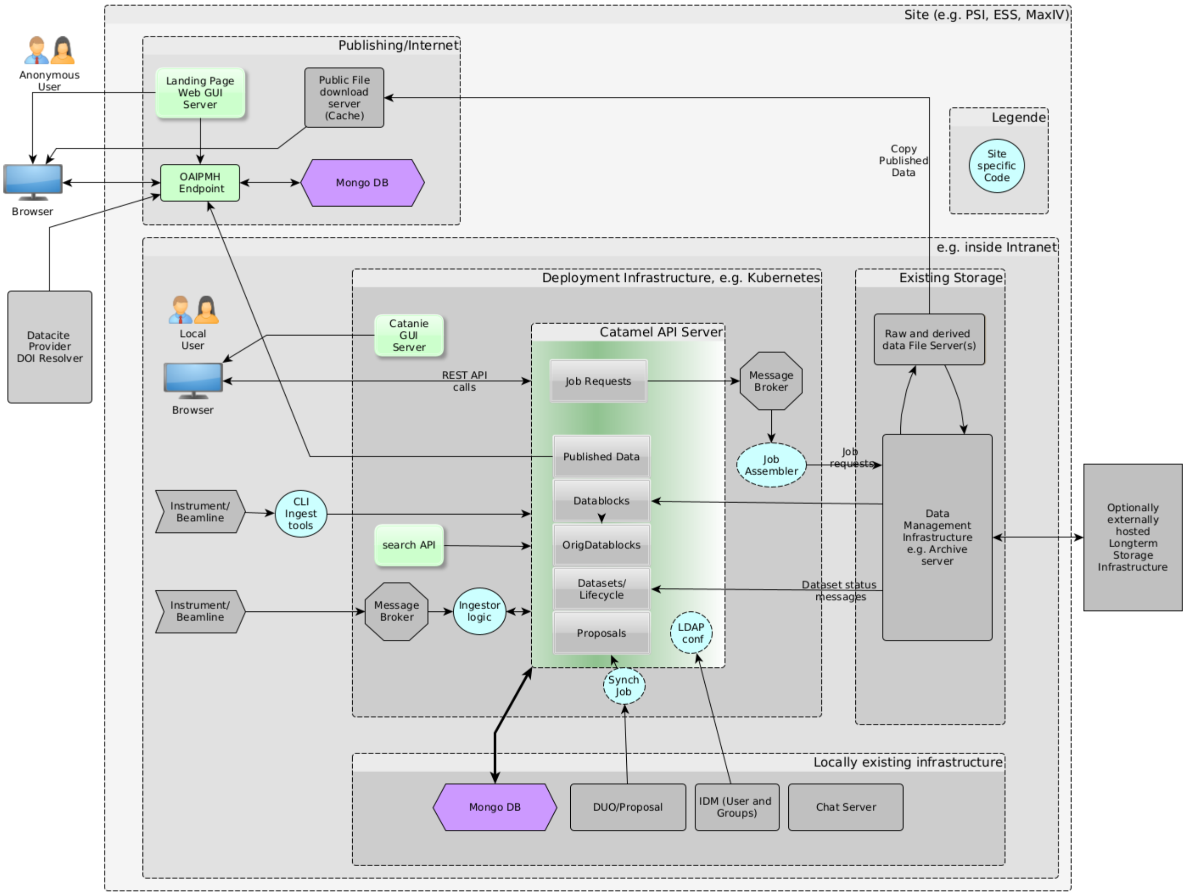


Fig. 15: SciCat architecture

All the green boxes in figure 15 are services provided by SciCat and its community. We see, in particular, the presence of all the interfaces mentioned in section 1 and, in addition, that all are provided as part of the SciCat stack.

The facility still needs to implement some components, because there are very different needs in some aspects of the flow, for example when ingesting data.

3.2 Services and their implementation

On a high level, SciCat provides a set of functionalities to interact with data and metadata. Most of them come from the backend (named Catamel API Server), which represents the layer of communication with the MongoDB instance. Other services usually call the backend, which, depending on the endpoints, processes the data and metadata from the database and returns what is needed.

Clockwise from the top left corner, figure 15 shows (link to the implementation in each bullet point):

- The landing Page Web GUI Server[60]: a single page application enabling anybody with internet access to browse Published data. Once on the publication of interest, the user can download the data.

[60] https://github.com/SciCatProject/LandingPageServer

- The Catamel API Server[61]: the core of SciCat. It is a set of APIs which implement functionalities on data management, from ingestion to publishing to retrieval. It is the central point of the whole architecture and all other services interact with it when accessing data.
- The Search API service[62]: the SciCat implementation of the PaN search-API to the data catalogue, as described in section 1.3.2.
- The Catanie GUI Server[63]: a single page application that enables users to manage data, depending on their permissions. Unlike the Landing Page Web GUI Server, it manages authentication and is not limited to open data.
- The OAI-PMH Endpoint[64]: the SciCat implementation of the OAI-PMH interface to the data catalogue, as described in section 1.3.3.

The SciCat documentation provides a full overview[65]. Other services interacting with SciCat, but not included in figure 15, can be found in the SciCat GitHub collaboration[66].

3.3 The SciCat stack at MAXIV

3.3.1 Deployment

At MAXIV, there are multiple internal networks, each dedicated to different categories of services. A beamline's data acquisition is carried out in the Blue network while web applications are deployed in the White network.

At each beamline in the Blue network, when an experiment is run, a script collects the metadata and posts it to the beamline's dedicated topic in the Kafka cluster.
The Kafka Cluster, a distributed event streaming platform, which can be used as a message broker, with topics dedicated to each beamline, acts as the intermediary to retrieve metadata collected at the beamlines. The cluster is deployed in a Kubernetes platform and is configured to have access to both the White and Blue networks.
A python script in the White network reads the messages from the Kafka topic and posts them to SciCat API to add them to the SciCat database.

The SciCat application, both frontend and backend, is installed in the White network[67], together with the Landing page[68], the search-API[69] and the OAI-PMH Service[70]. All the services are containerised and are deployed using GitlabCI and Ansible.

Datafiles generated during experiments are accessible only to the experimental group through access control applied at the storage level. A separate storage segment has been

[61] https://github.com/SciCatProject/backend
[62] https://github.com/SciCatProject/panosc-search-api
[63] https://github.com/SciCatProject/frontend
[64] https://github.com/SciCatProject/oai-provider-service
[65] https://scicatproject.github.io/documentation
[66] https://github.com/SciCatProject
[67] https://scicat.maxiv.lu.se
[68] https://doi.maxiv.lu.se/
[69] https://searchapi.maxiv.lu.se/
[70] https://scicat.maxiv.lu.se/openaire/oai

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

set up as a pilot in order to provide access to ‘open’ data files. For the purpose of the pilot, any data that is published and is openly available is copied to the public data segment. Data files in this segment can be downloaded with a MAXIV Globus account.
The download of the data files through the data catalogue from the public storage segment is still under development.

MAXIV has a repository with DataCite and can mint DOIs. Currently, the production environment is connected to a test repository provided by DataCite.

A representation of this structure:

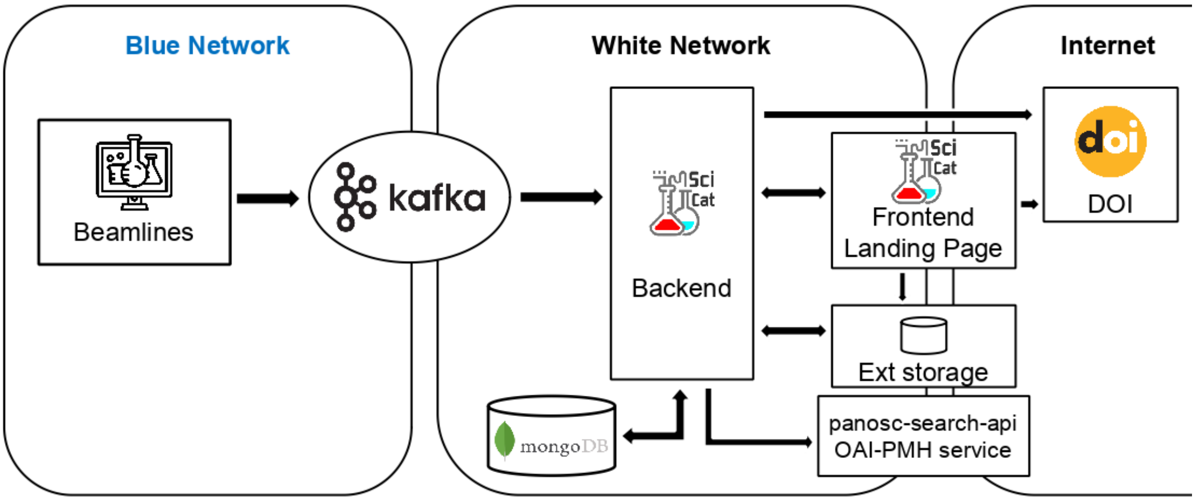


Fig. 16: MAXIV SciCat architecture

3.3.2 Usage and examples

Eight beamlines out of fourteen as of January 2022 have been integrated into the SciCat environment, i.e. scripts to post metadata have been deployed at eight beamlines and post metadata to Kafka topics.

MAXIV has started testing the environment for publishing open datasets. A few commissioning datasets from the Beamline BioMAX have been published to the test DataCite repository and DOIs have been minted. These are available through the search and OAI-PMH APIs.

Search-API

The search-API service[71] allows a client to search the available publications. Multiple endpoints are available, for example, the datasets route[72] shows all the public datasets in the MAXIV SciCat and the documents route [73] shows all the datasets underpinning publications.

[71] https://searchapi.maxiv.lu.se/
[72] https://searchapi.maxiv.lu.se/api/Datasets
[73] https://searchapi.maxiv.lu.se/api/Documents

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 857641.

The search functionality supports filtering. This is better documented in the search-API docs[74]. An example[75] of usage is showing the documents in which the title contains the word "biomax" without considering uppercase/lowercase.
A more specific example[76] returns the results for sample 57531.

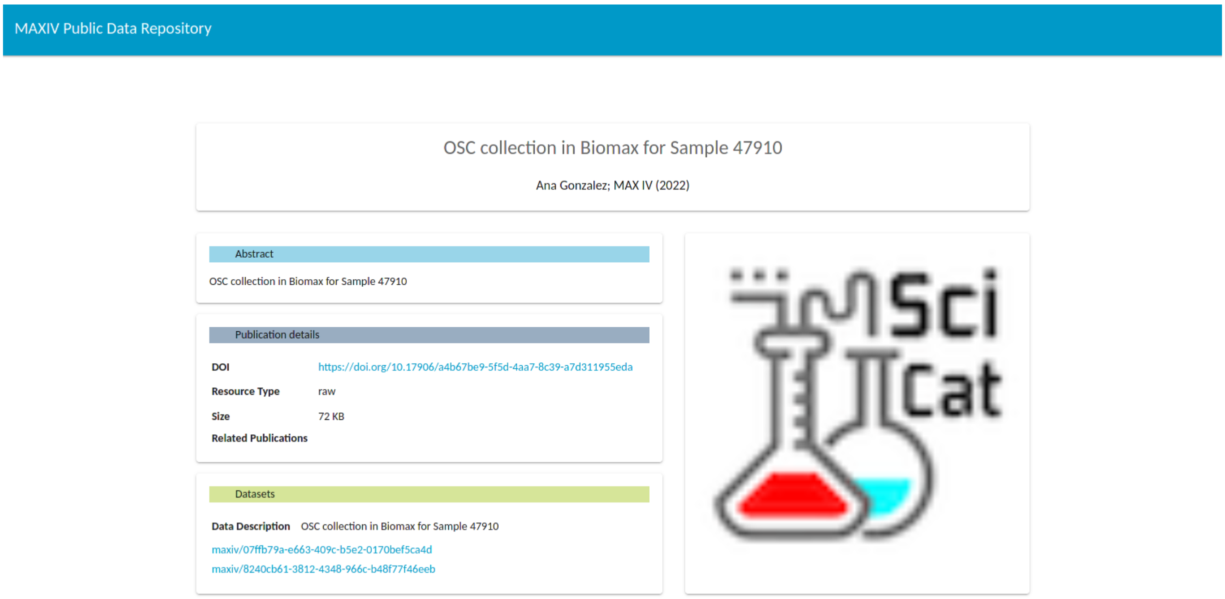
OAI-PMH

The OAI-PMH service[77] allows a client to search publications similar to the search-API but following the openarchives structure. The result of the request is presented in XML format which is not easily human-readable but, being strictly structured, can be used by applications easily.
This service offers endpoints to search through the published dataset; for example:

- to list all the public records[78].
- To return the same record we received from the search-API selecting it by its ID[79].

Landing Page Server

The Landing page[80] is a simple static webpage. Opening it shows all the available publications and selecting a publication shows the information about the corresponding datasets. Figure 17 shows an example of the view of a publication:



MAXIV Public Data Repository

OSC collection in Biomax for Sample 47910

Ana Gonzalez; MAX IV (2022)

Abstract

OSC collection in Biomax for Sample 47910

Publication details

DOI	https://doi.org/10.17906/a4b67be9-5f5d-4aa7-8c39-a7d311955eda
Resource Type	raw
Size	72 KB
Related Publications	

Datasets

Data Description OSC collection in Biomax for Sample 47910

maxiv/07ffb79a-e663-409c-b5e2-0170bef5ca4d
maxiv/8240cb61-3812-4348-966c-b48f77f46eeb

Fig. 17: MAXIV SciCat Landing Page

[74] https://github.com/panosc-eu/search-api/tree/master/doc
[75] https://searchapi.maxiv.lu.se/api/Documents?filter={"where":{"title":{"like":"biomax","options":"i"}}}
[76] https://searchapi.maxiv.lu.se/api/Documents?filter={"where":{"title":{"like":"57531"}}}
[77] https://scicat.maxiv.lu.se/openaire/oai
[78] https://scicat.maxiv.lu.se/openaire/oai?verb=ListRecords&metadataPrefix=oai_dc
[79] https://scicat.maxiv.lu.se/openaire/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=10.17906/a4b67be9-5f5d-4aa7-8c39-a7d311955eda
[80] https://doi.maxiv.lu.se/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Publishing datasets

At MAXIV, currently there is no automatic publication of data. However, a user can at any time publish data manually using the SciCat interface following the steps below.

The usual steps in which a dataset is published:

1. The user conducts an experiment at the beamline and the metadata is recorded automatically
2. The user logs in on SciCat[81] where she can see her datasets
3. The User adds the datasets she wishes to the cart and uses the Publish action
4. At this point, the user is required to add the proper title, abstract of the publication, and then press the ‘Publish’ button
5. Now the datasets are public although the metadata publications process is not finished
6. The user can review the information in the publication one last time before selecting the ‘Register’ button.
7. Once the user presses the "Register" button, the publication is finalised. The datasets are public, and the publication gets assigned a DOI and is registered.

7. After this process, the publication is included on the MAXIV Landing page[82], it is possible to find it through the search-API and it is visible through the OAI-PMH service. This information is accessible from anywhere on the web. From the publication landing page, links to access the datasets that are part of the publication are available.

3.3.3 Connection with the PaN federated search-API and EOSC

MAXIV has requested that their search-API endpoint be added to the PaN federated search API currently hosted by ESS.
Adding the publications to the B2FIND and openAIRE EOSC services, now that the OAI-PMH interface is developed and accessible, only requires following the procedure described in section 1.3.3.

To increase the visibility of Photon and Neutron data in the EOSC community, we are additionally planning to make the MAXIV Landing page directly available as an EOSC service. In the context of milestone 13: *Metadata catalogue as EOSC service*, this has already been done for the PSI’s published data repository which is accessible in the EOSC marketplace[83]. Several other PaN facilities will also register their public data repositories as EOSC services before the end of the project.

[81] https://scicat.maxiv.lu.se
[82] https://doi.maxiv.lu.se
[83] https://marketplace.eosc-portal.eu/services/psi-public-data-repository

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 857641.

3.4 B2FIND, openAIRE and Google Dataset Search at PSI

This section demonstrates the connection to B2FIND and openAIRE (EOSC services) and Google Dataset Search, made possible using the SciCat interfaces. This gives working examples of how any EOSC user can find data and in particular, we report step by step instructions on the use of EOSC and Google Dataset Search to find PSI datasets.

The step by step instructions will show some commonalities in the use of EOSC services and Google Dataset Search and highlight that by enabling findability on many platforms, we increase the exposure of datasets to a bigger researchers community, accustomed to the same flow to find data, but using different entry points.

3.4.1 EOSC services: B2FIND and openAIRE

The component to enable findability through EOSC services is, as mentioned already, the OAI-PMH service, which follows the architecture outlined in 1.3.3 which, since PSI uses SciCat, can be seen in more detail in section 3.1.

The section in the ExPaNDS GitHub wiki page[84] provides a more detailed explanation of its high-level implementation and here we only mention the two major responsibilities of the SciCat OAI-PMH: it exposes a well-defined set of APIs, as required by OAI-PMH, and it translates the content of the SciCat MongoDB database to the XML schema format[85] specified in the API call and returns it.

B2FIND

1. Go to the B2FIND web portal[86].
2. Input the query of interest, for example: “Lesions in Human Pulmonary Hypertension”
3. After the query is submitted, B2FIND collects the list of results[87].
   Selecting the PSI dataset we land on the detail page[88]
   where we find metadata about the publication and in particular a link to doi.org
4. Selecting the doi.org link the user lands on the PSI landing page for that publication[89].
5. Data can now be downloaded by selecting the "Access data" button.

openAIRE

1. Go to the openAIRE explore web portal[90].

[84] https://github.com/ExPaNDS-eu/ExPaNDS/wiki/Delivering-data-services-to-EOSC#scicat-implementation

[85] https://www.w3.org/XML/Schema

[86] http://b2find.eudat.eu/

[87] http://b2find.eudat.eu/dataset?q=Lesions+in+Human+Pulmonary+Hypertension&sort=score+desc%2C+metadata_modified+desc&ext_timeline_start=&ext_timeline_end=&ext_startdate=&ext_enddate=&ext_bbox=&ext_prev_extent=-154.68749999999997%2C-80.17871349622823%2C154.68749999999997%2C80.17871349622823

[88] http://b2find.eudat.eu/dataset/27649182-9ee3-592b-9c95-1b8e666b2752

[89] https://doi.psi.ch/detail/10.16907/d699e1f7-e822-4396-8c64-34ed405f07b7

[90] https://explore.openaire.eu/

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 857641.

2. Input the query of interest, for example: “Lesions in Human Pulmonary Hypertension”
3. After the query is submitted, openAIRE collects the list of results[91].
4. Clicking on the PSI dataset we land on the detail page[92] where we find metadata about the publication and in particular a link to doi.org
5. Selecting the doi.org link the user lands on the PSI landing page for that publication[93].
6. Data can now be downloaded by selecting the "Access data" button.

3.4.2 Google Dataset Search

To enable Google Dataset Search, PSI added the schema.org markup on their SciCat Landing Page Server and a detailed explanation on how to do that is available in the google guide[94].

1. Go to the Google Dataset Search web portal[95].
2. Input the query of interest, for example: “Lesions in Human Pulmonary Hypertension”
3. After the query is submitted, Google Dataset Search collects the list of results[96].
4. The result titled: “Synchrotron Imaging of Complex Vascular Lesions in Human Pulmonary Hypertension: Pathology Distribution in 3D Space” is from PSI
5. Selecting the doi.org link or the “explore at doi.psi.ch” button, the user lands on the PSI landing page for that publication[97].
6. Data can now be downloaded by selecting the "Access data" button.

3.5 Future developments

The facilities using SciCat will deploy and enable the pan-ontologies-API and the scoring service - see section 4 for a more detailed description. The two services interact with the SciCat search-API and the SciCat community has implemented the code to enable their connection, which is respectively covered by the panet-service[98] and the pss-service[99] in the SciCat search-API code base.

[91] https://explore.openaire.eu/search/find?resultbestaccessright=%22Open%2520Access%22&fv0=Lesions%20in%20Human%20Pulmonary%20Hypertension&f0=q&active=result

[92] https://explore.openaire.eu/search/other?orpId=r31e55174ff7::0bb7372c8f548336cf2d0bba6482912c

[93] https://doi.psi.ch/detail/10.16907/d699e1f7-e822-4396-8c64-34ed405f07b7

[94] https://developers.google.com/search/docs/advanced/structured-data/dataset

[95] https://datasetsearch.research.google.com/

[96] https://datasetsearch.research.google.com/search?query=Lesions%20in%20Human%20Pulmonary%20Hypertension&docid=L2cvMTFybms5Z2hjbg%3D%3D

[97] https://doi.psi.ch/detail/10.16907/d699e1f7-e822-4396-8c64-34ed405f07b7

[98] https://github.com/minottic/panosc-search-api/blob/pans/common/panet-service.js

[99] https://github.com/minottic/panosc-search-api/blob/pans/common/pss-service.js

4. Future developments

4.1 The scoring service [5]

The role of this service, developed by the PaNOSC partner ESS, is to attach a score based on the user's query to each dataset returned by the PaN federated search-API, which can be used as a ranking criterion. Figure 18 shows the scoring service data flow in the *Catalogue Service* ecosystem.

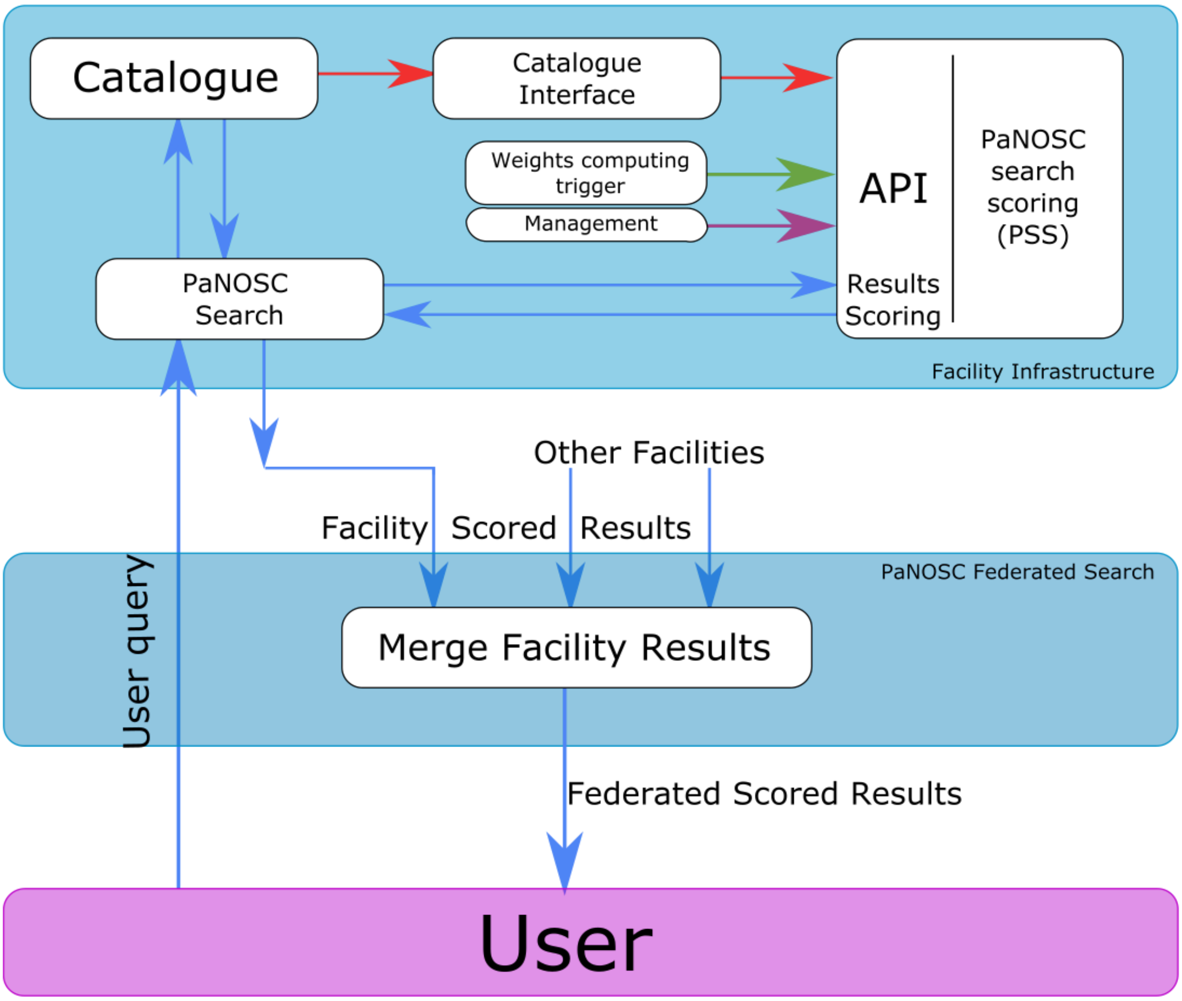


Fig.18: PaNOSC search scoring data flow

To map figure 18 to figure 3, please note that the top layer of figure 18 represents the PaN facilities layer in figure 3, the PaNOSC Federated Search the middle layer and the bottom layer the PaN portal.

Once the PaNOSC search scoring is deployed, the facility is responsible for the following three steps:

1. Populate the scoring information
2. Trigger the weight computing
3. Configure the PaNOSC search API to integrate with the scoring

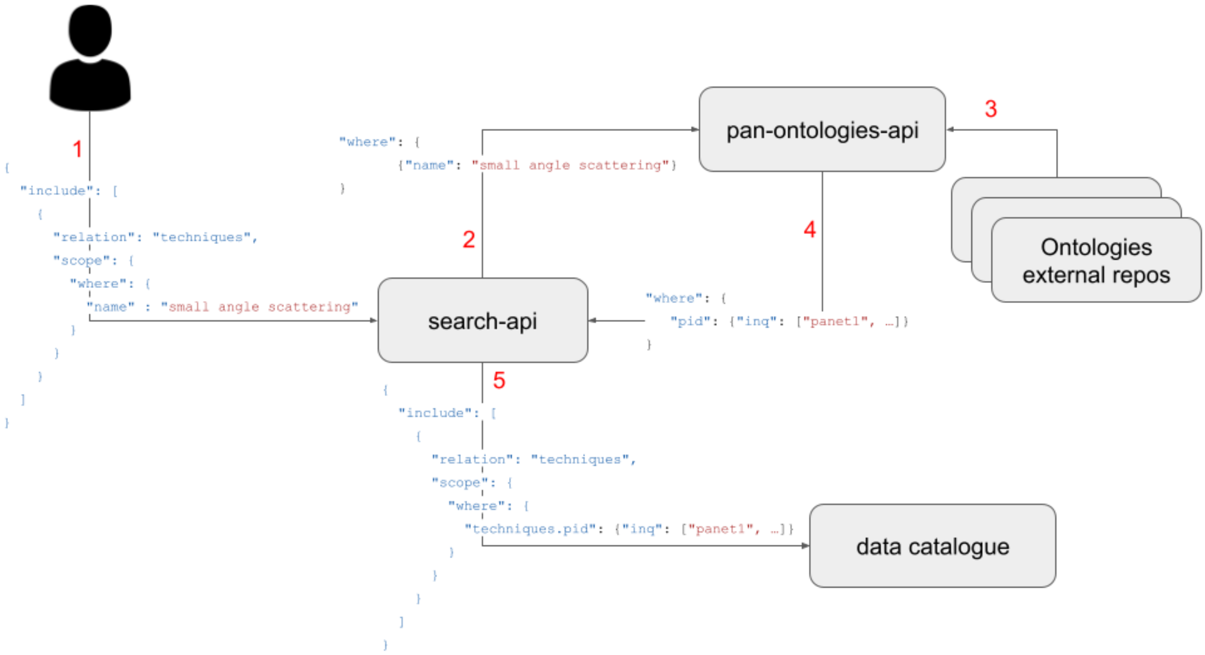
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Populating the scoring information can be achieved by writing an ad-hoc integration service and deploying it on the facility's IT infrastructure. This service periodically updates the scoring information in the scoring service from the relevant fields from the catalogue system. Such service can be based on the Jupyter notebook provided as an example under the notebook folder of the code repository[100].
Triggering the weight computation is achieved by a POST request to the dedicated endpoint of the scoring service. Each facility has to decide if the computation should happen periodically or on-demand and make sure to configure their IT infrastructure accordingly.
Integration with and management of the scoring service is done through the API, which is highlighted in the document PaN Federated Search results scoring API which can be found in the docs folder of the code repository[101].

As a future development in ExPaNDS *Work Package 3*, the ExPaNDS facilities will deploy the needed service to score the datasets and enable their use in the PaN search-API.

4.2 The pan-ontologies-API

The pan-ontologies-API is a RESTful service that links the ontologies - currently only the PaNET ontology as defined in [4] - to the search-API. At a high level, the service works by pulling and caching information from external ontology repositories, and in particular, the GET /techniques/pan-ontology endpoint translates the query from the user input (e.g. in the PaN portal) to a new query sent to the search-API, having added the logic from the PaNET ontology. In PaNET, we want to expand the query from the user, when looking for a particular technique, to include all the techniques which descend from it, namely the techniques which are different specifications of the same input class. We reference here the pan-ontologies-API data flow for the PaNET ontology.



```
{
  "include": [
    {
      "relation": "techniques",
      "scope": {
        "where": {
          "name" : "small angle scattering"
        }
      }
    }
  ]
}
```

```
"where": {
    {"name": "small angle scattering"}
}
```

```
"where": {
   "pid": {"inq": ["panet1", …]}
}
```

```
{
  "include": [
    {
      "relation": "techniques",
      "scope": {
        "where": {
          "techniques.pid": {"inq": ["panet1", …]}
        }
      }
    }
  ]
}
```

Fig. 19: pan-ontologies-API data flow

[100] https://github.com/panosc-eu/panosc-search-scoring/tree/master/notebooks
[101] https://github.com/panosc-eu/panosc-search-scoring/tree/master/docs/pdf

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Following the order depicted in the data flow:

1. The search-API takes a Loopback filter[102] per the search-API specification.
2. The "where" part of the filter, related to, e.g., techniques, is forwarded to the pan-ontologies-API.
3. The pan-ontologies-API fetches the ontology from an external source (if not already cached).
4. The fetched ontology is processed depending on the ontology logic (in the diagram the PaNET one) and a processed Loopback "where"[103] condition is returned.
5. The search-API takes the "where" condition from step two and sends it to the data catalogue, replacing the generic term "pid" with the one available to the *Site API* of figure 3 (in the figure the SciCat search-API output is shown, where the term "pid" is replaced with "techniques.pid").

Given its role in the data model, namely the fact that it does not depend on the data catalogue specific implementation (e.g. ICAT and SciCat can use the same implementation), the pan-ontologies-API could work either at the level of the PaN federated search-API service (the middle layer of figure 3), thus becoming a federated service or at the level of the data catalogue of each facility (each box named "search-API" in the bottom layer of figure 3).
If the pan-ontologies-API is deployed at the level of each facility, steps 2 and 5 from the data flow aforementioned will need to be implemented individually, while deploying it at the federated level will enable a transparent integration at every facility, since the response from the pan-ontologies-API can be made compliant with the PaN search-API requirements in the PaN federated search-API service.
PSI has initiated an internal discussion to understand the feasibility of federating the pan-ontologies-API and will include PaNOSC members, especially from *Work Package 3*, to finalise and take action on the decision.

As a future development in ExPaNDS *Work Package 3*, the ExPaNDS facilities will deploy the required service to use the PaNET ontology in the PaN search-API.

Conclusions

In this document we have shown the ICAT and SciCat releases at ISIS, MAXIV and PSI and their ability to integrate with the PaN search-API and multiple EOSC services. To highlight the FAIR commitment, we have also shown the interoperability with Google Dataset Search, which is used both by the ICAT and the SciCat stacks.
Referencing the ExPaNDS architecture at the beginning of the document, we have enabled the reader to contextualise the facilities' implementation and deployment efforts into the project architecture and direction.
We have presented the connection to *Milestone 13*, in particular mentioning the achievement of having added a Published Data Repository to the EOSC services.
In the last section, we have commented on the future plans to deploy the newly developed services and described their high-level behaviour.

[102] https://loopback.io/doc/en/lb3/Querying-data.html
[103] https://loopback.io/doc/en/lb3/Where-filter.html

References

[1] Scardaci, Diego, Salvat, Daniel, Fuhrmann, Patrick, Barty, Anton, Ashton, Alun, & Servan, Sophie. (2020). ExPaNDS General Architecture description in relation to the EOSC services. Zenodo. https://doi.org/10.5281/zenodo.3697704

[2] Richter, Tobias et al. Common search API definition. (2020). PaNOSC https://www.panosc.eu/wp-content/uploads/2020/12/D3.1_API-definition.pdf

[3] Schrettner Lajos et al. Federation of search APIs Demonstrator Implementation. (2021). PaNOSC https://www.panosc.eu/wp-content/uploads/2021/11/PaNOSC_D3.2_DemonstratorImplementation_20210324.pdf

[4] Collins, Steve P., da Graça Ramos, Silvia, Iyayi, Daniel, Görzig, Heike, González Beltrán, Alejandra, Ashton, Alun, Egli, Stefan, & Minotti, Carlo. (2021). ExPaNDS ontologies v1.0. Zenodo. https://doi.org/10.5281/zenodo.4806026

[5] Tobias Richter, Fredrik Bolmsten, Massimiliano Novelli, Luis Maia, Axel Bocciarelli, Alex de Maria, Marjolaine Bodin, Carlo Minotti. (2022). PaNOSC Photon and Neutron Open Science Cloud H2020-INFRAEOSC-04-2018 Grant Agreement Number: 823852 Deliverable: D3.3 Catalogue Service (federation of search APIs)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.