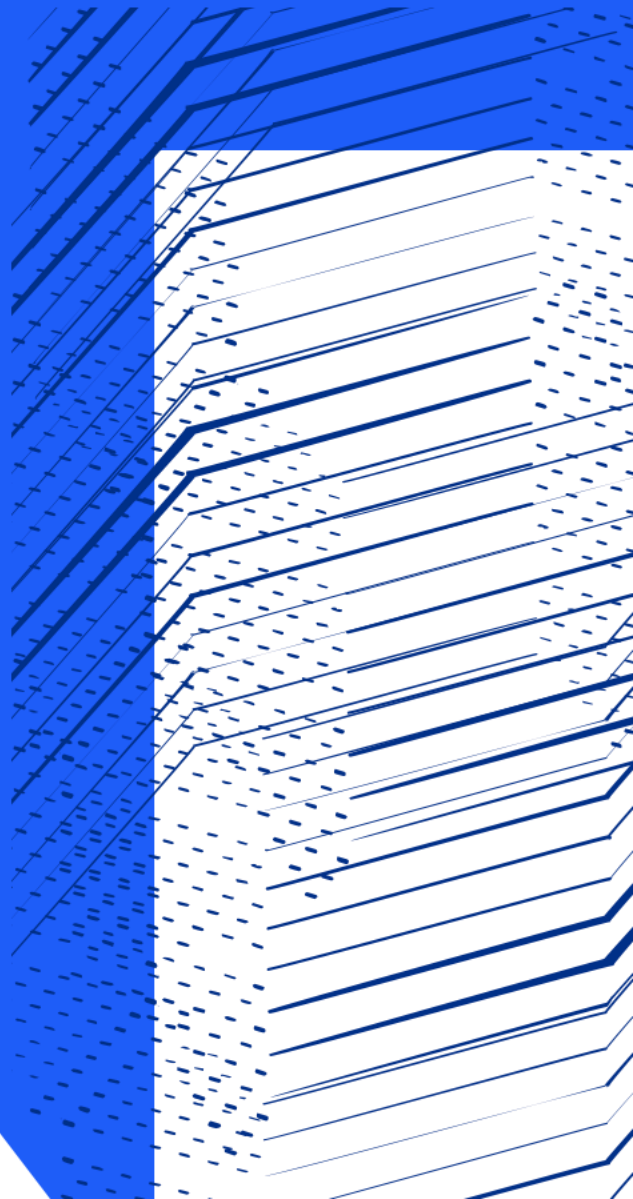
Science and Technology Facilities Council

DataGateway

Louise Davies

ICAT Collaboration F2F Meeting
Wednesday 3rd May 2023



Agenda

1 What is DataGateway?

2 New features

3 Upcoming features

Image © STFC Alan Ford

Science and Technology Facilities Council

Science and Technology Facilities Council

Introduction

- What is DataGateway?

Science and Technology Facilities Council

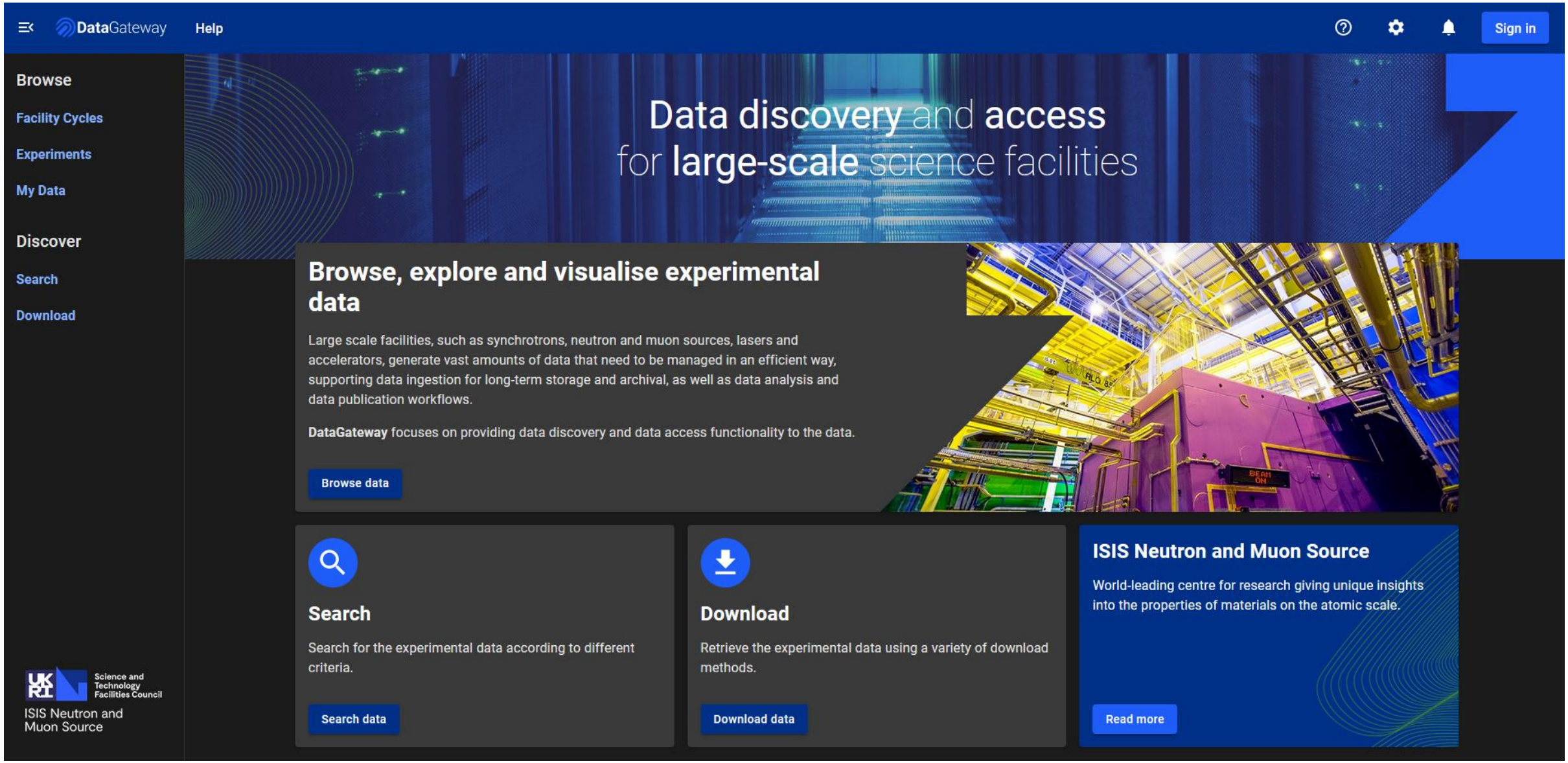
DataGateway

- An interface reflecting the users **data journeys**
  - Considering both facility users and open data users
  - From data creation to data publication
    - Data provenance: Associate instrument setup with raw & processed data
    - DOI creation & workflows for data publication
  - Data discovery & data access
  - Rich metadata (moving to FAIR data)
  - Specialised data catalogue - information about the data hierarchies in each facility

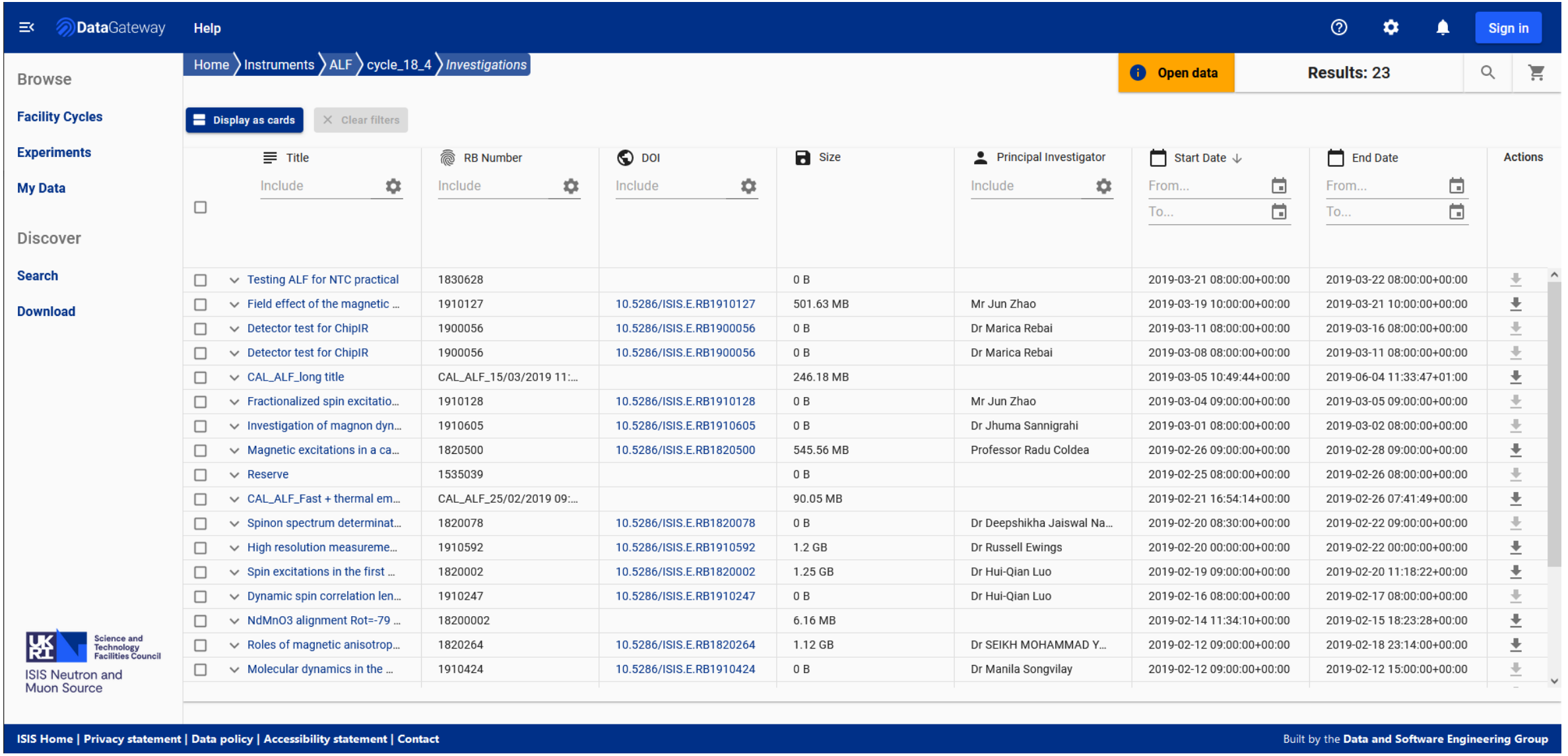
DataGateway

- DataGateway is a plugin to SciGateway
  - SciGateway provides common features to other front-ends such as:
    - Authentication, cookies management, dark mode, notifications etc
- React JS web application written in TypeScript using the MUI (Material UI) component library
- Replaced TopCAT for both ISIS and Diamond

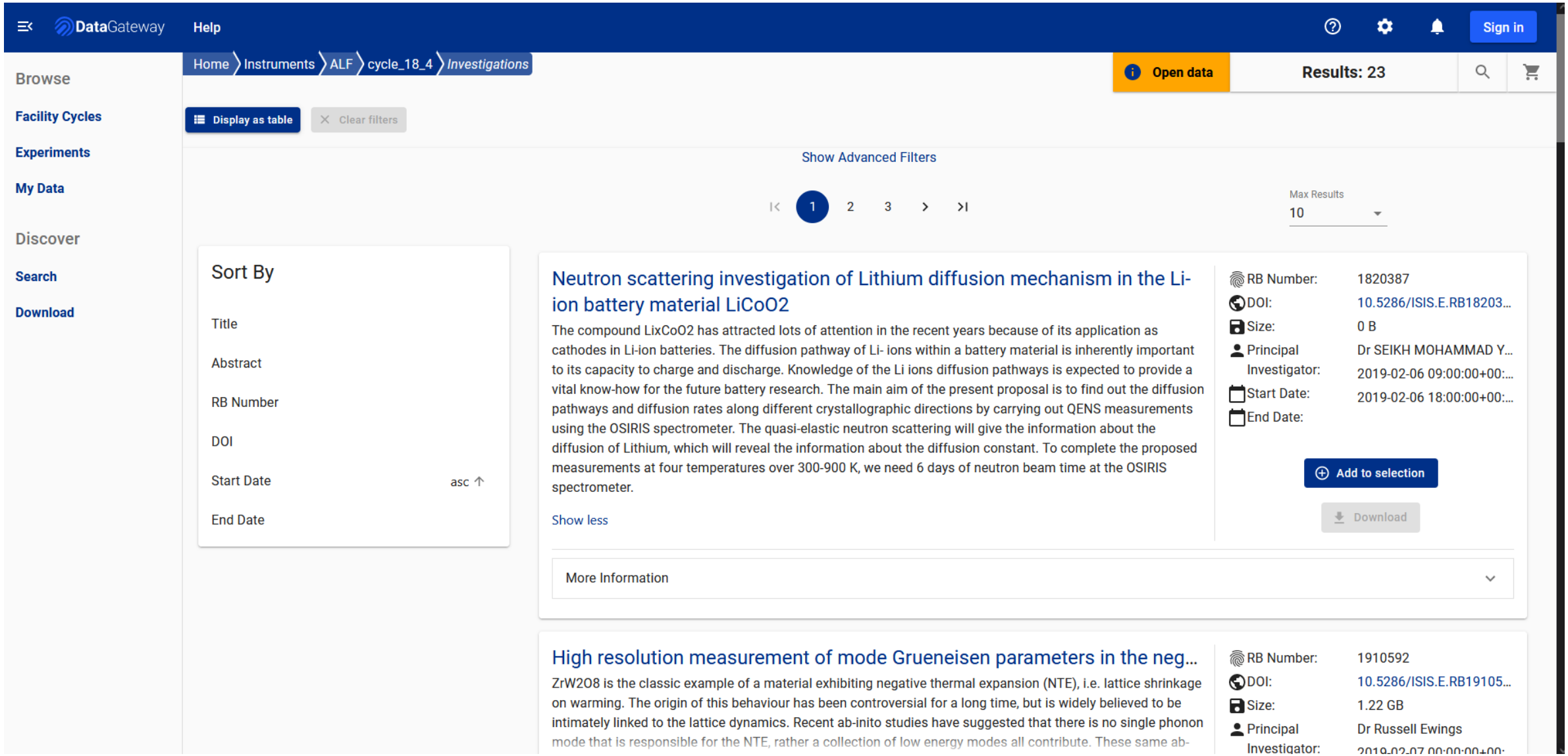
Screenshots!



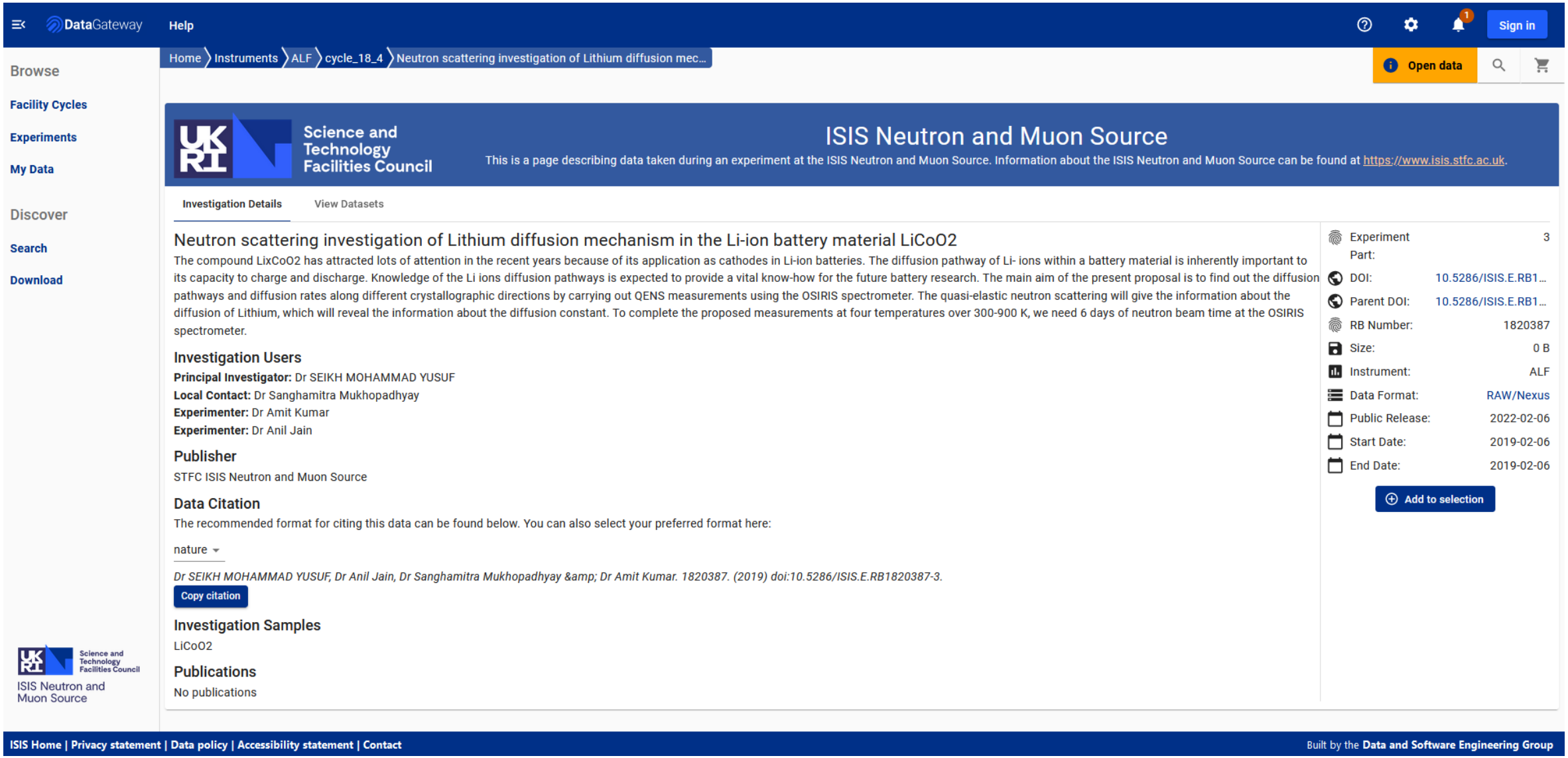
Screenshots!



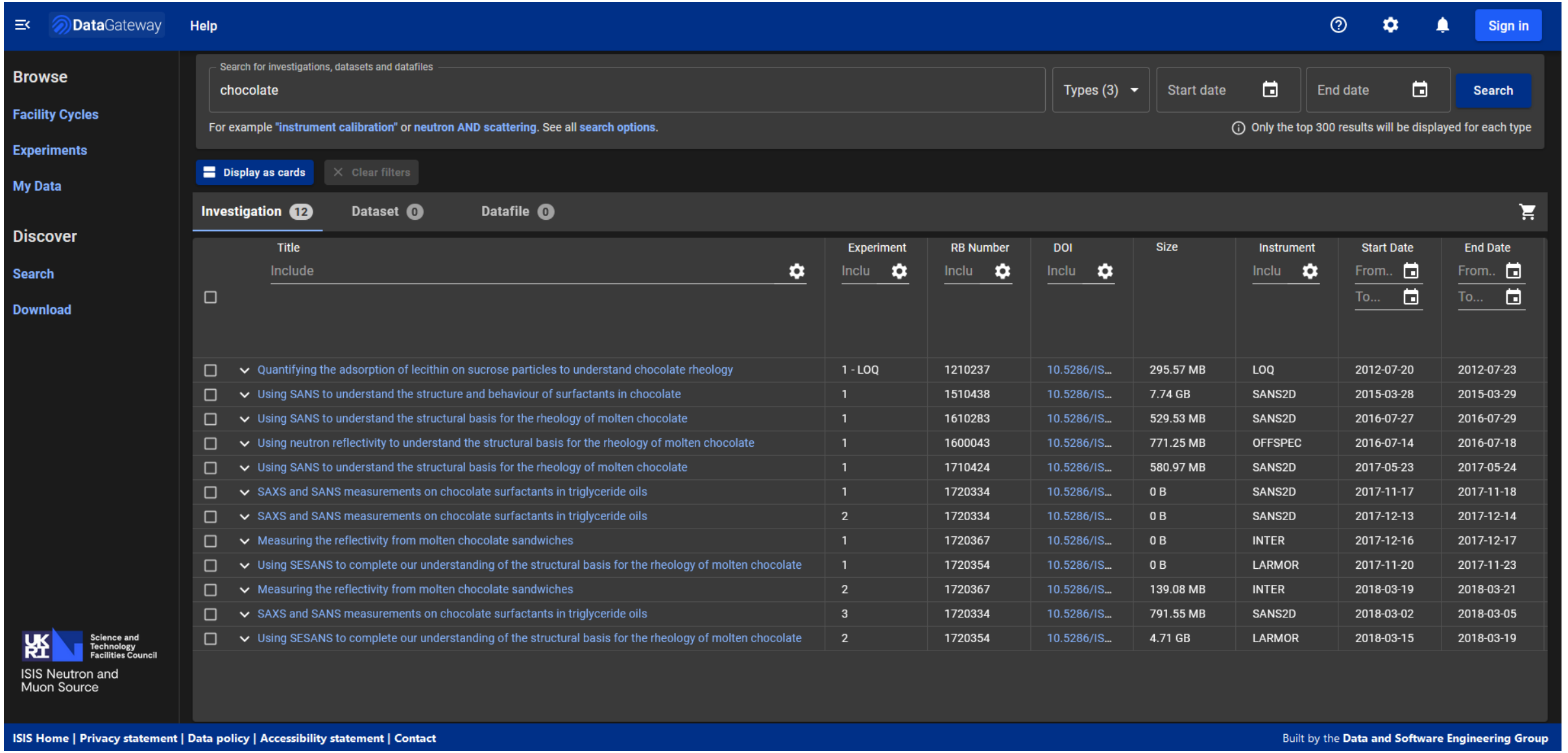
Screenshots!



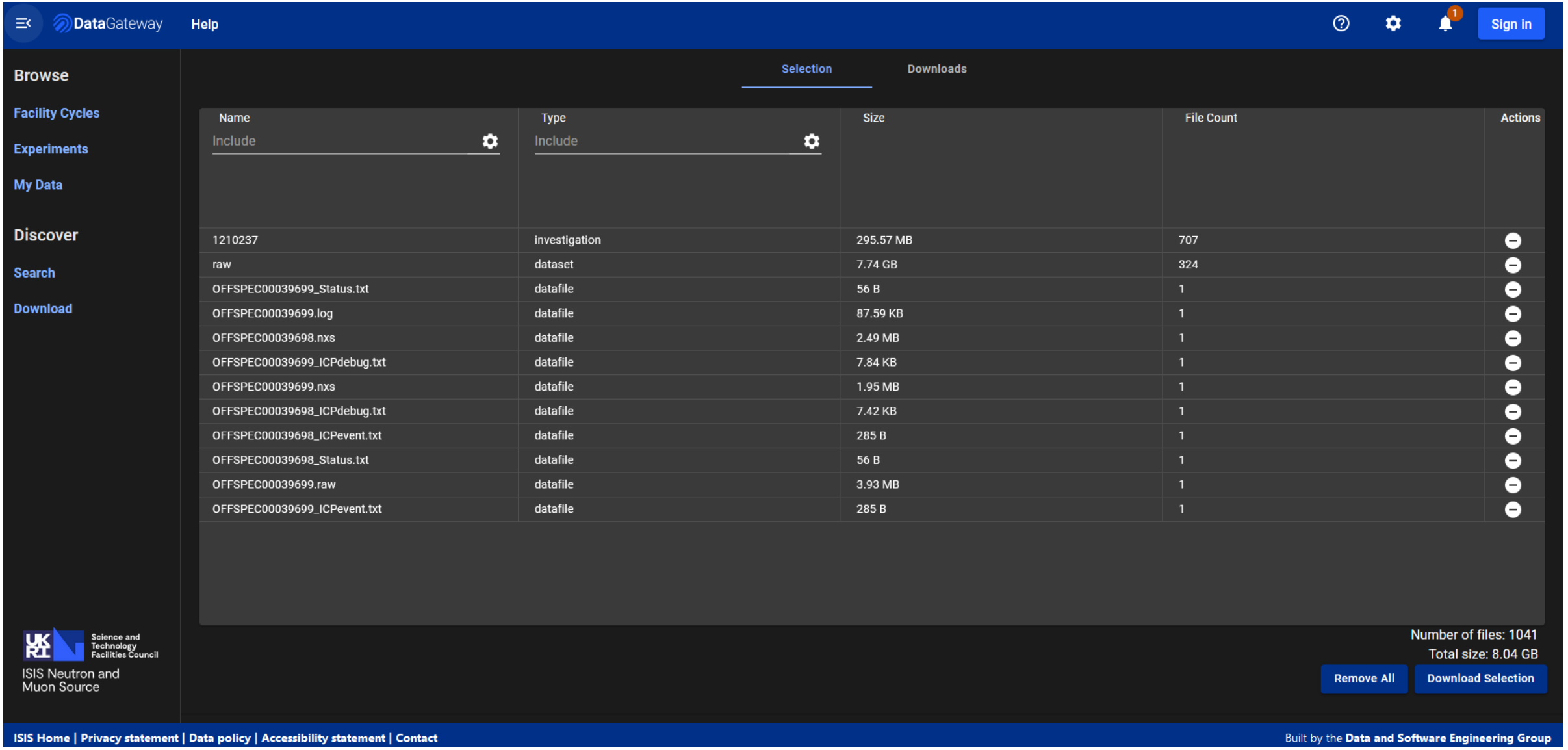
Screenshots!



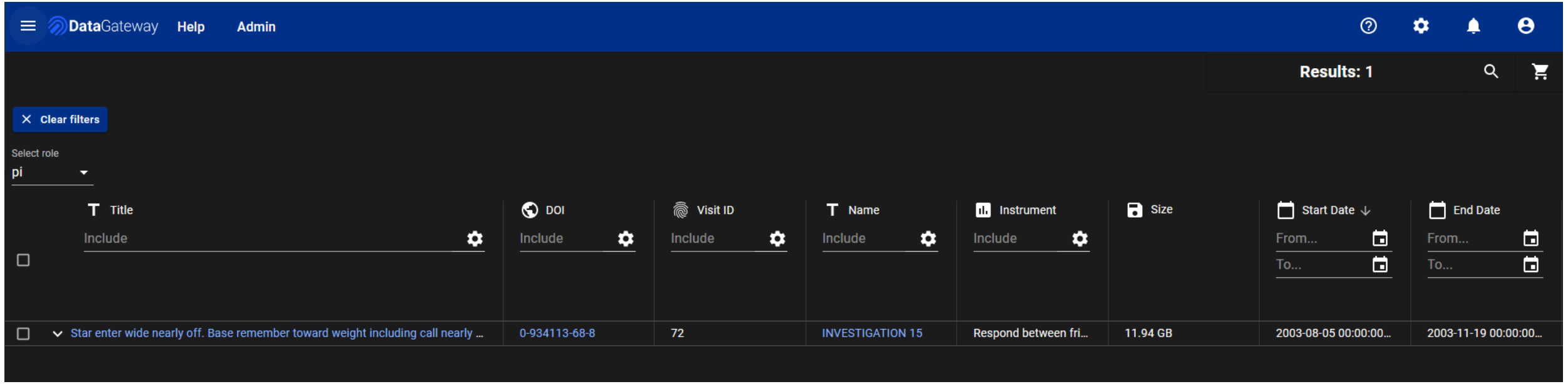
Screenshots!



Screenshots!



Screenshots!



Science and Technology Facilities Council

Science and Technology Facilities Council

New Features

Science and Technology Facilities Council

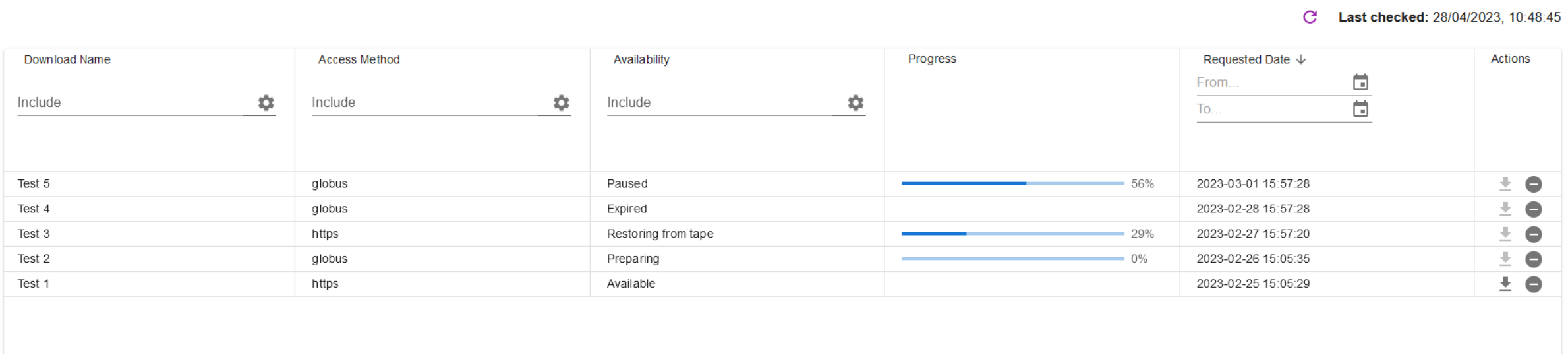
Misc “new” features

- Automatic anonymous login
- Citation formatter
- Role selector in My Data view
- Dark mode & high contrast mode(s)
- Card view

And many more minor improvements & features

Download progress bars

- Requested by Diamond
- Kevin made new IDS – ids.r2dfoo which provides the progress information
- Visible both to users and to facility admins



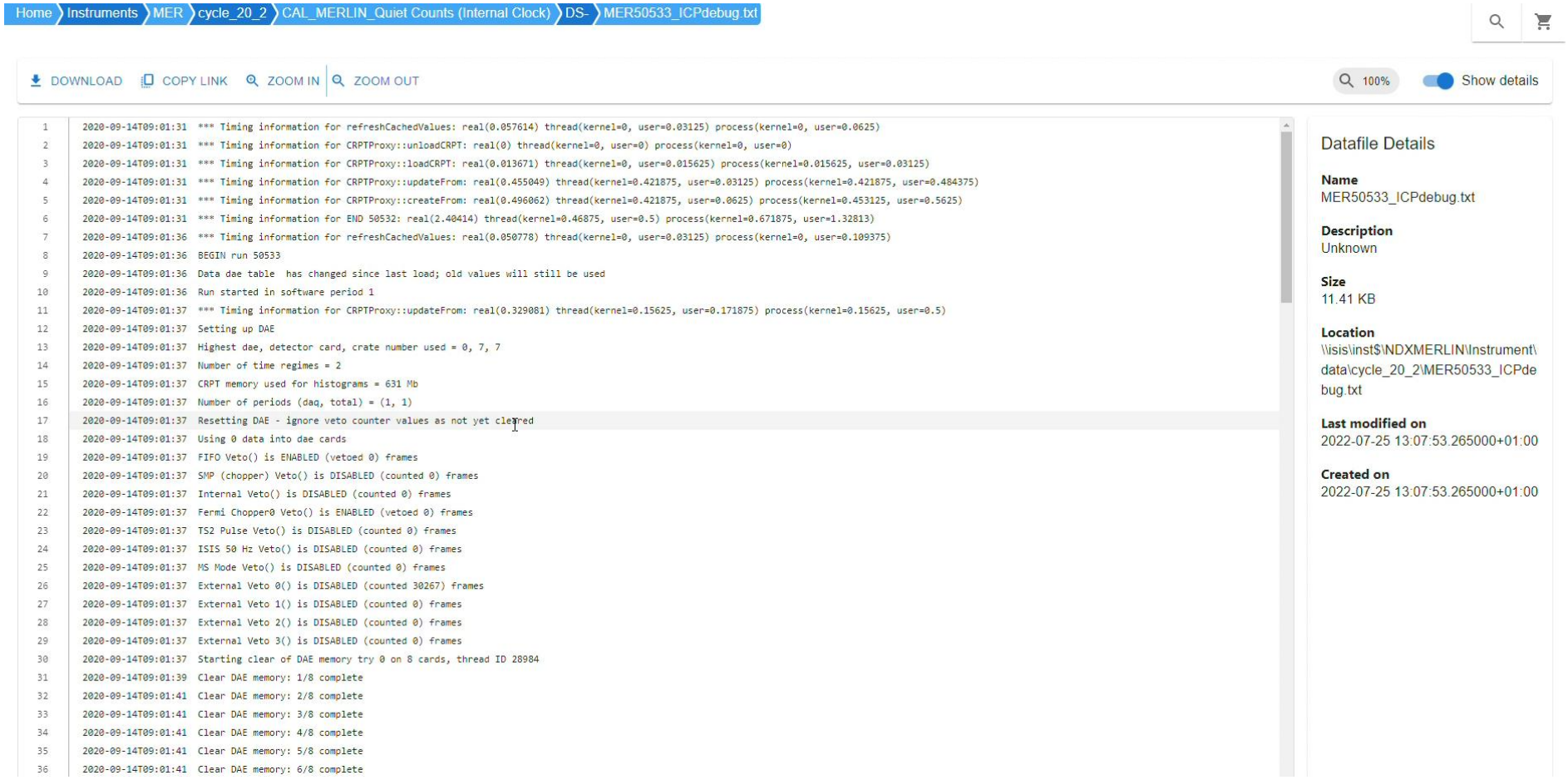
Science and Technology Facilities Council

File previews

Feature to view files directly in the data catalogue

Currently focusing on plain text files (.e.g .txt & .log)

Plans to integrate a HDF viewer to visualise .nxs files



Science and Technology Facilities Council

Science and Technology Facilities Council

Works in Progress

- Data publications
- Search improvements
- Etherpad
- Machine learning explorations

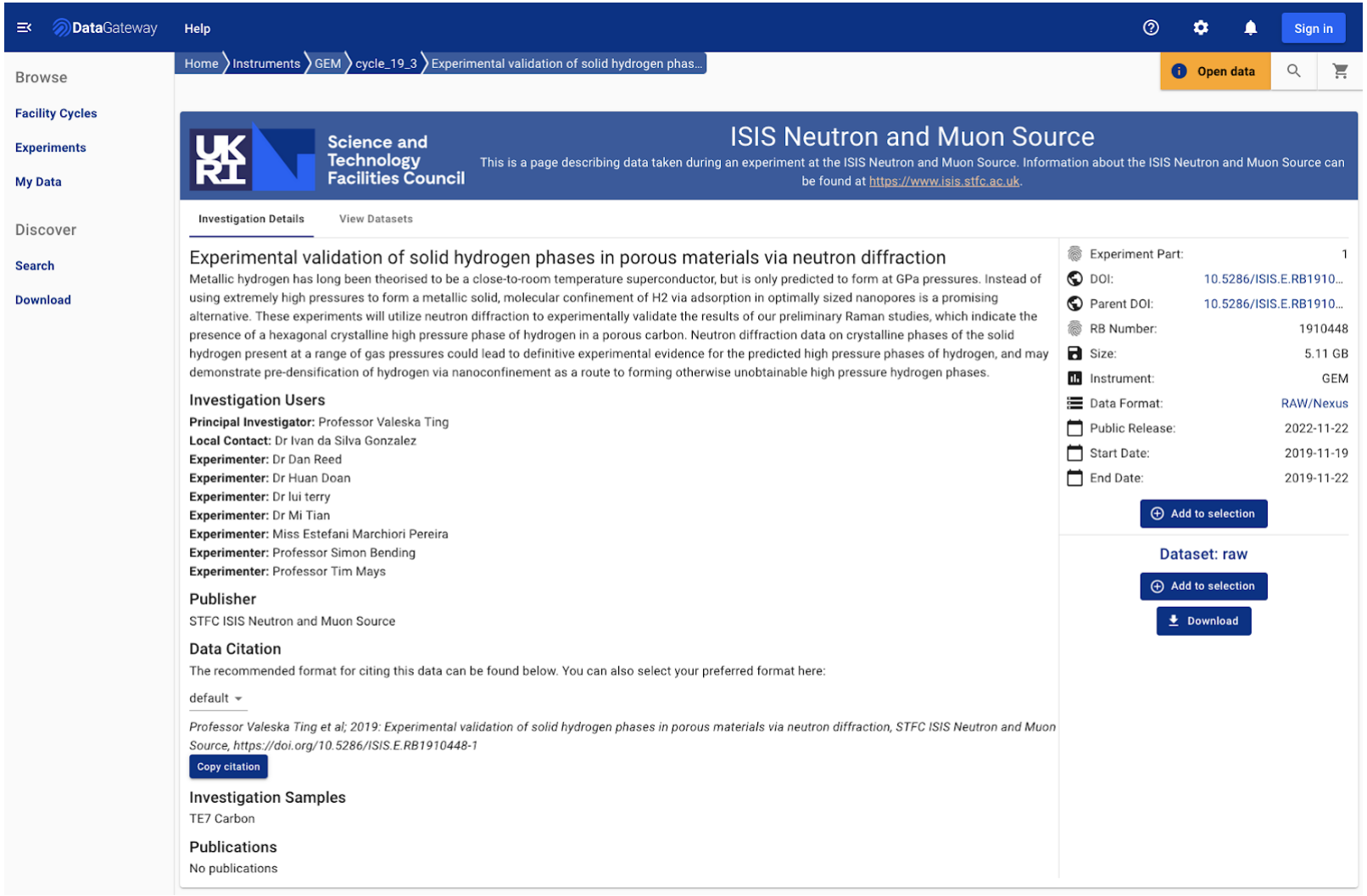
Science and Technology Facilities Council

Data Publications

Metadata changes enable use of integrated DOI landing pages in DataGateway

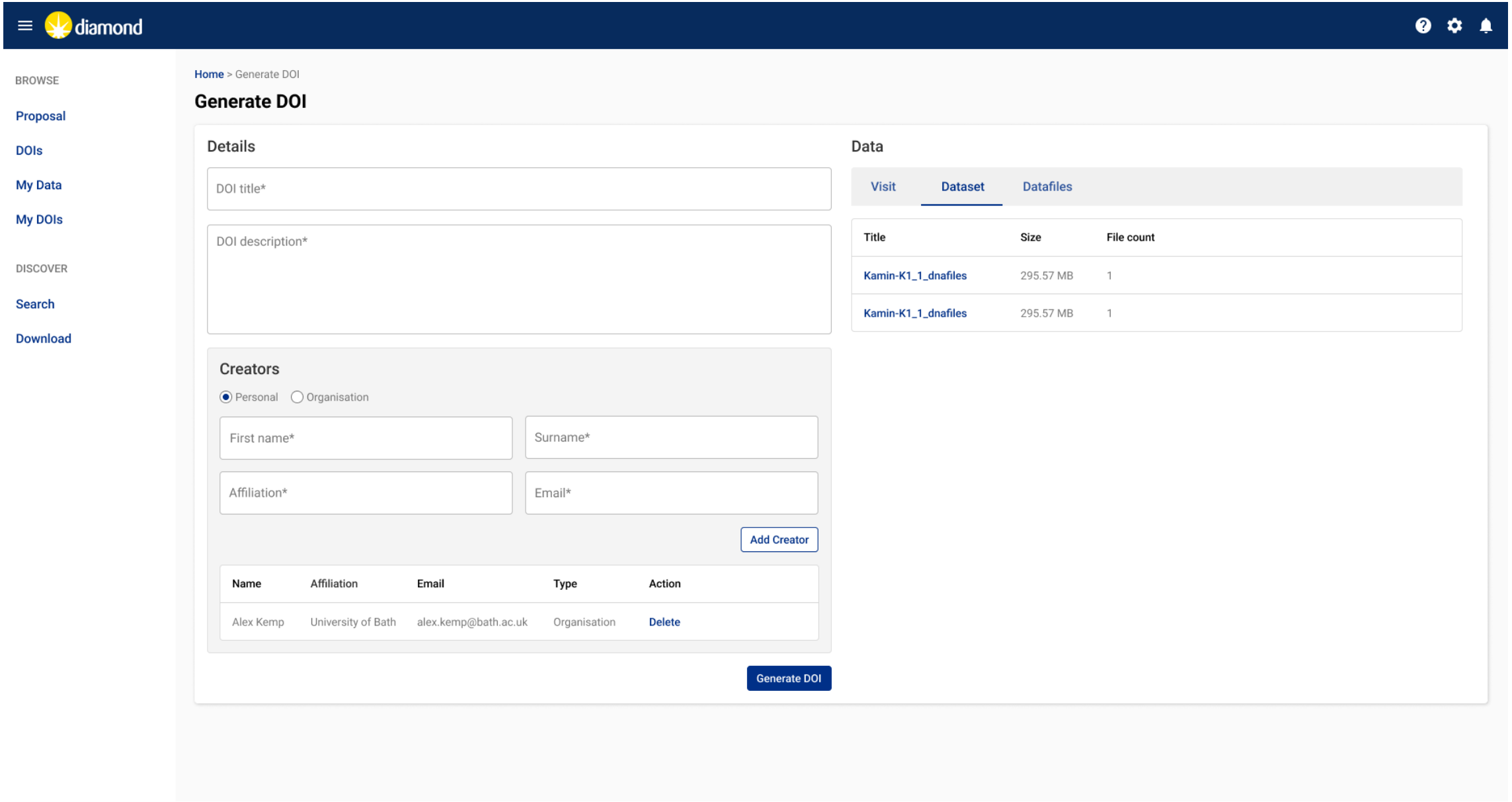
Will be able to better explore published data in DG

Plans to enable DOI minting of user defined selections of data (see Alex' talk)



Science and Technology Facilities Council

User Defined Data Publications



Search improvements

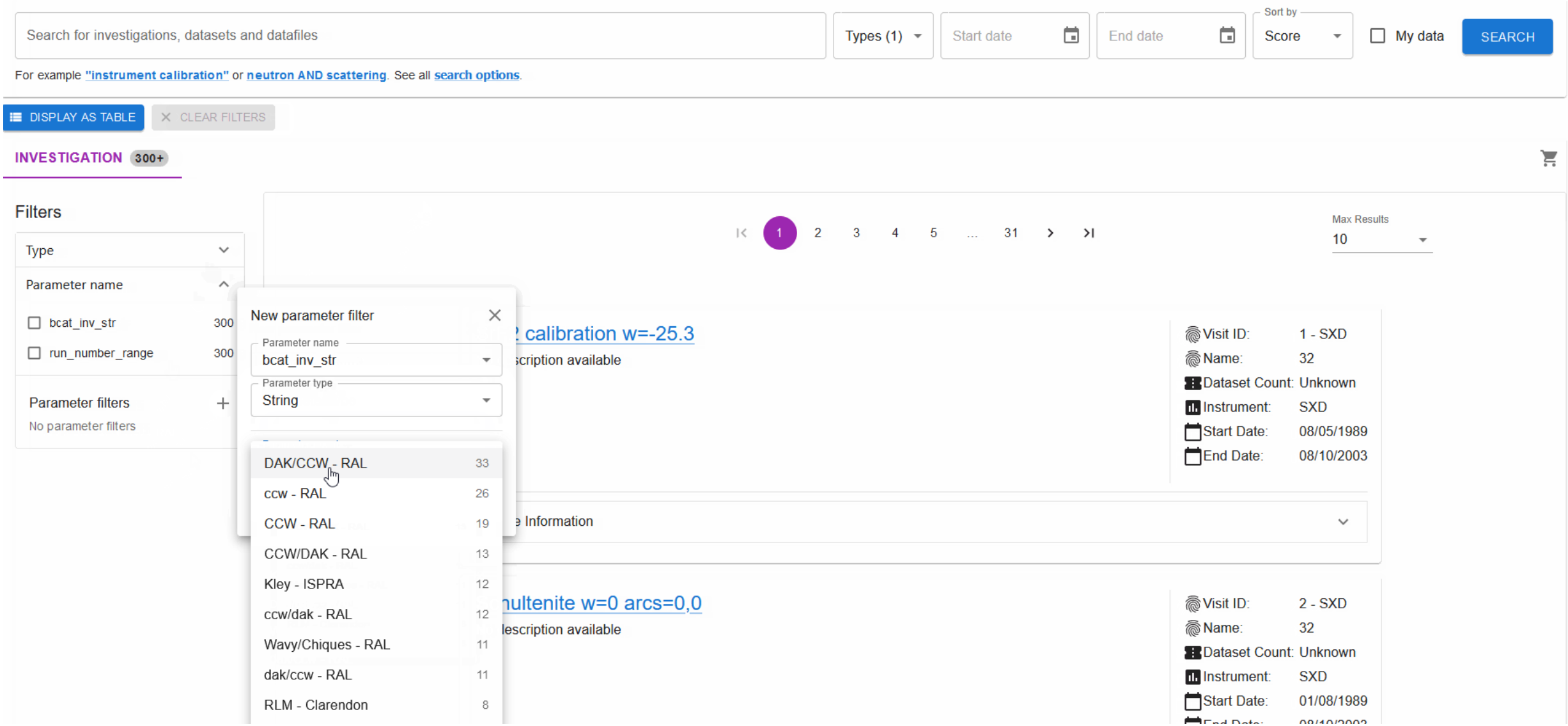
Search functionality in DG is being improved:

- Sort by search relevance
- Filter by parameters & samples
- Search more than top 300 results
- More fields are searchable & you can now search on specific fields
- Synonyms support

Science and Technology Facilities Council

See Patrick’s talk for more details!

Search improvements



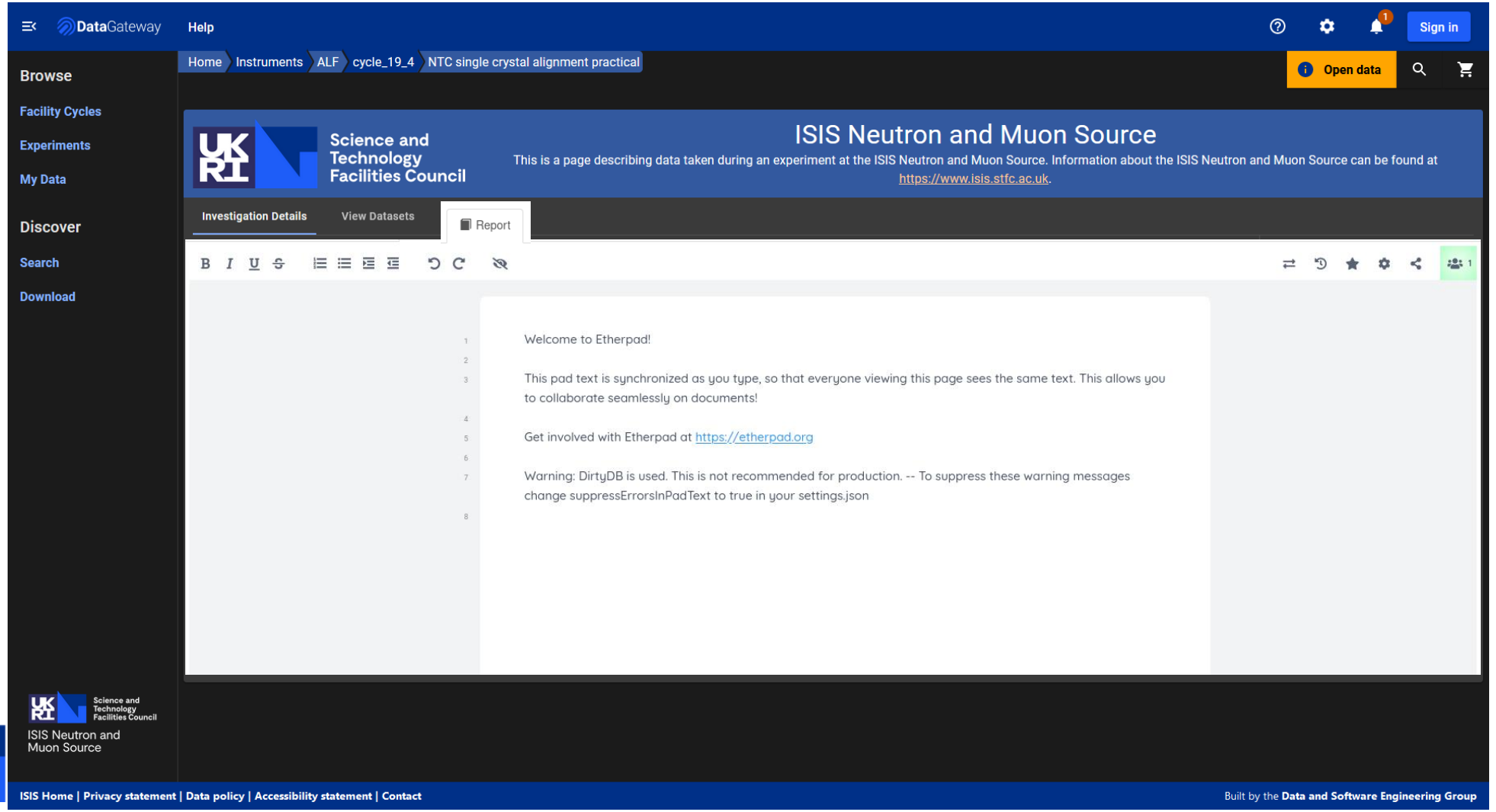
Etherpad in DataGateway

Based on discussion with Alex de Maria about ESRF's setup, will need:

- Etherpad (lite) instance
- Auth plugin to Etherpad
- Changes to DataGateway

Will Edwards currently working on this at STFC as part of 6-month apprenticeship rotation

DataGateway Implementation?

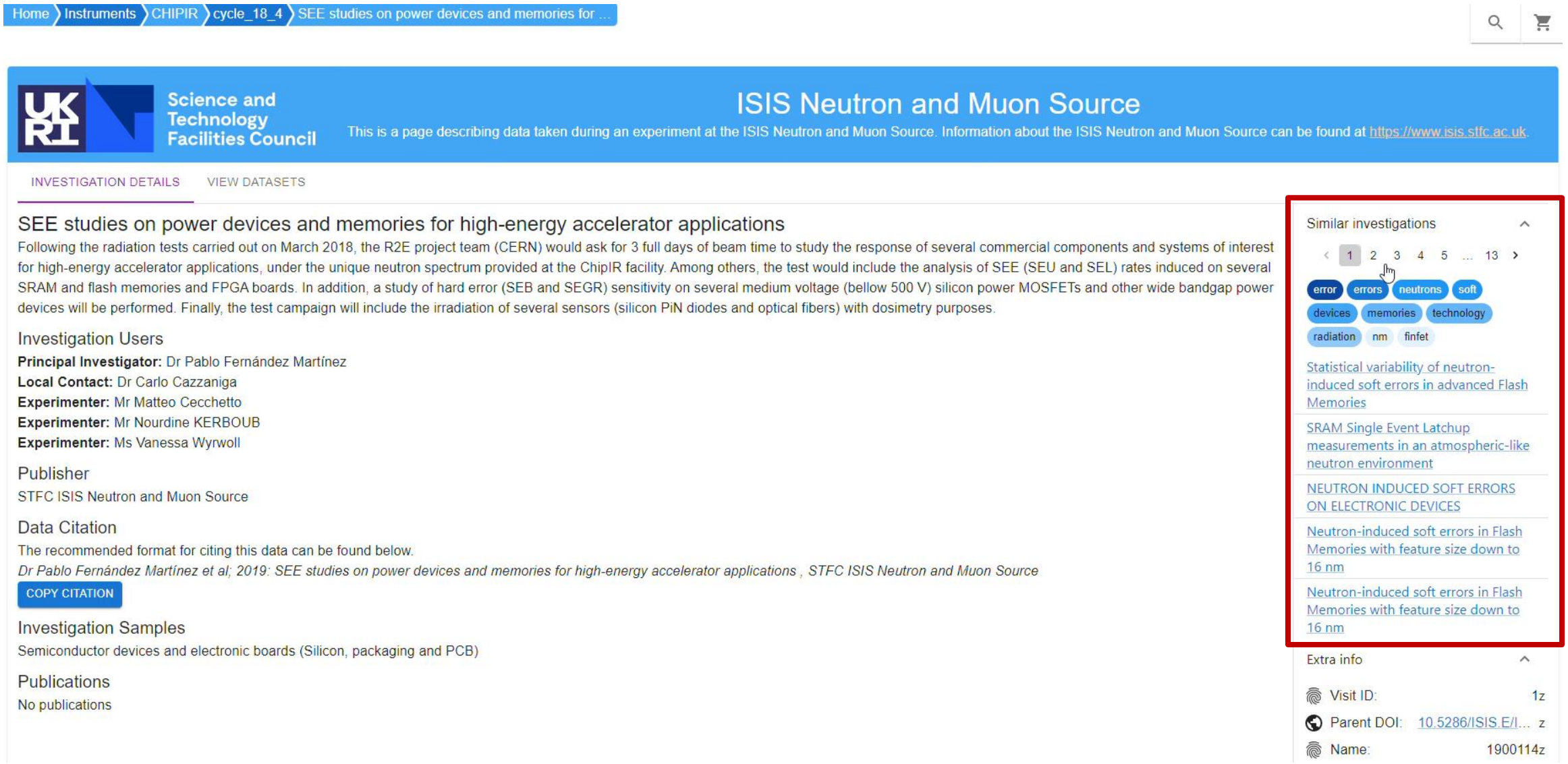


Machine learning features

- New backend app has been developed using ML techniques to analyse ICAT data and provide insights
- First feature is recommending similar investigations based on the summary & title
- Future features around clustering & integrating with search are also being explored

Science and Technology Facilities Council

Machine learning features



Science and
Technology
Facilities Council

Questions?

UKRI Science and Technology Facilities Council

Thank you

Science and Technology Facilities Council @STFC_matters Science and Technology Facilities Council