UK Atomic Energy Authority

Making Mast Open Data Open & FAIR

Shaun de Witt

4th IAEA Technical Meeting on Fusion Data Processing, Validation & Analysis

2021-11-03

Parts of this work have been funded under the Euratom research and training programme 2014-2018 under grant agreement No 847612

This work was funded by the RCUK Energy Programme [Grant number EP/T012250/1]

CCFE CULHAM CENTRE FOR FUSION ENERGY

Agenda

- Motivation
- Where we Started
- Problems Encountered
- Where we Ended

Contributors

- UKAEA:
  - David Muir, Lucy Kogan, Rashard Hussain, Andrew Kirk
- PSNC
  - Michal Owsiak
- The FAIR4Fusion Project, Research Data Alliance, EIROforum

https://www.fair4fusion.eu/

https://www.rd-alliance.org/

https://www.eiroforum.org/

Motivation

UK Atomic Energy Authority

Funding Mandate

- EPSRC-funded research data is a public good produced in the public interest and **should be made freely and openly available with as few restrictions as possible in a timely and responsible manner.**
- EPSRC recognises that there are legal, ethical and **commercial constraints on release of research data**. To ensure that the research process (including the collaborative research process) is not damaged by inappropriate release of data, research organisation **policies and practices should ensure that these constraints are considered at all stages** in the research process.
- **Sharing research data is an important contributor to the impact of publicly funded research**. To recognise the intellectual contributions of researchers who generate, preserve and share key research datasets, **all users of research data should acknowledge the sources of their data and abide by the terms and conditions under which they are accessed.**
- **EPSRC-funded researchers should be entitled to a limited period of privileged access to the data they collect** to allow them to work on and publish their results. The length of this period will depend on the scientific discipline and the nature of the research.
- **Institutional and project specific data management policies and plans should be in accordance with relevant standards** and community best practice and should exist for all data. Data with acknowledged long term value should be preserved and remain accessible and useable for future research.
- **Sufficient metadata should be recorded and made openly available to enable other researchers to understand the potential for further research and re-use of the data**. Published results should always include information on how to access the supporting data.
- It is appropriate to use public funds to support the preservation and management of publicly-funded research data. To maximise the scientific benefit which can be gained from limited budgets, the mechanisms for managing and providing access to research data should be both efficient and cost-effective in the use of such funds.

UKRI Engineering and Physical Sciences Research Council

CCFE CULHAM CENTRE FOR FUSION ENERGY

A FEW DEFINITIONS

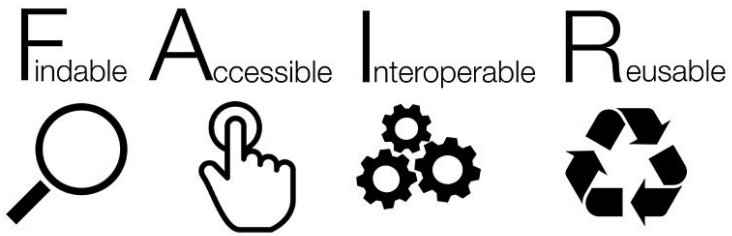
UK Atomic Energy Authority

FAIR data: Findable, Accessible, Reusable and Interoperable

- 15 core principles, but FAIR is NOT black and white
  - Practically these principles revolve around metadata, persistent identifiers, provenance, licensing, security and data access

Open Data:

- *Open means **anyone** can **freely access, use, modify, and share** for **any purpose** (subject, at most, to requirements that preserve provenance and openness).* (Open Knowledge Foundation)



Numerous scientists have pointed out the irony that right at the historical moment when we have the technologies to permit worldwide availability and distributed process of scientific data, broadening collaboration and accelerating the pace and depth of discovery ... we are busy locking up that data and preventing the use of correspondingly advanced technologies on knowledge.
— John Wilbanks, VP Science, Creative Commons

CCFE CULHAM CENTRE FOR FUSION ENERGY

Mast “Open” Data

UK Atomic Energy Authority

- MAST Data had a 3 year embargo period after which it was made ”accessible” via a web portal.
- The portal is “several” years old now (but has received minor updates). At the time (2013?) it was considered state of the art
- But how easy is it to use and how FAIR is the data?
  - Lets see how it worked…

CCFE CULHAM CENTRE FOR FUSION ENERGY

Step 1: Find the programme you are interested in

UK Atomic Energy Authority

List of MAST Experiment Programmes and Objectives

Search Programmes and Objectives using keywords (case insensitive):

To drill down to information on individual Experiments, select any table cell

Experiment date	Programme	Objective	Shot Start	Shot End	Shot Count	Analysed Data File Count	Image File Count
2009-03-04	EF correction studies	Repeat 4 shots with ELM coil configuration 15.Parasitic test of pellet microwave top cavity.	21520	21532	13	169	58
2009-03-04	Off-axis NBCD	Power scan of off-axis NBCD discharges to investigate fast-ion anomalous diffusion NBI power dependence.	21533	21545	13	323	95
2009-03-05	ELM control	Continue on testing the effect of ELM coils confg. 9 on H-mode. Assess the effect of the plasma shape.	21546	21579	34	849	287
2009-03-06	ELM control	Continue on testing the effect of ELM coils on H-mode. ELM coil in confg. 8. Assess the effect of the plasma shape.	21580	21596	17	399	144
2009-03-13	NBI test	Determine why the SS NBI breaks down in synch mode, when it doesnt in asynch	21597	21614	18	345	84
2009-03-17	Quantify TAE Damping	Measure stable mode spectrum in MAST.	21615	21628	14	292	117
2009-03-18	Fuel retention experiments	To study fuel retention during ohmic and NBI shots.	21629	21642	14	294	116
2009-03-19	Quantify TAE damping	Measure stable mode spectrum in MAST. (Cont d 17/03/2009)	21643	21665	23	531	146
2009-03-24	Quantify TAE damping	Measure stable mode spectrum in MAST.	21666	21678	13	263	83
2009-03-26	Effect of ELM coils on L-mode plasmas. Test effect of Br.	Determine the effect of Br on the density confinement in L-mode plasmas.	21679	21691	13	297	116
2009-03-31	Effect of ELM coils on L-mode discharges	To determine the effect of ELM coils in Even parity 60 degrees phasing (config. 9) on L-mode discharges	21692	21719	28	646	288
2009-04-01	Effect of ELM coils on Beam heated discharges	To determine the effect of ELM coils in even parity 60 degree phase (config 9) on beam heated discharge.	21720	21756	37	720	262

CCFE CULHAM CENTRE FOR FUSION ENERGY

Step 2: Identify which shot(s) you want

UK Atomic Energy Authority

List of individual MAST experiments

Date:

2009-06-11

Programme:

TAE damping. nu* confinement scaling.

Objective:

To obtain a two point scan in nu* of energy confinement and heat transport in ELMy H-mode.

Search the list of Experiments using keywords (case insensitive):

To drill down to Open Data, select a table cell where the Data or Image file count > 0

Experiment Number	Experiment Time	Analysed Data File Count	Image File Count
22668	10:53:00	24	9
22669	11:11:00	24	9
22670	11:26:00	0	0
22671	11:30:00	26	9
22672	11:51:00	28	9
22673	12:13:00	0	0
22674	12:21:00	28	9
22675	12:36:00	28	8
22676	13:13:00	28	9
22677	13:31:00	27	9
22678	13:49:00	27	9
22679	14:19:00	28	9
22680	14:37:00	28	9
22681		0	0

CCFE CULHAM CENTRE FOR FUSION ENERGY

Step 3: Identify which file(s) you want

UK Atomic Energy Authority

List of MAST Open Data and Image files

Shot:
22676

Date:
2009-06-11

Time:
13:13:00

Programme:
TAE damping. nu* confinement scaling.

Objective:
To obtain a two point scan in nu* of energy confinement and heat transport in ELMy H-mode.

Search the list of files using keywords (case insensitive):

To drill down to information on an Open Data file, select any table cell
Use the Download or Request Data buttons to access Open Data files

Class	Type	Description	Filename	Format	Size	Pass	Signal Count	Download or Request Data
abm	Analysed	multi-chord bolometers	abm0226.76	IDA3	3	0	21	Request Data
acd	Analysed	Carbon density	acd0226.76	IDA3	1	2	27	Request Data
act	Analysed	Time resolved Charge Exchange	act0226.76	IDA3	22	0	64	Request Data
ada	Analysed	Linear D-Alpha Camera	ada0226.76	IDA3	4	0	9	Request Data
adg	Analysed	Plasma Edge Density gradient from the linear Dalpha camera	adg0226.76	IDA3	1	0	4	Request Data
aga	Analysed	molecular deuterium pressure, neutral gas pressure, Gas Injection/Fueling	aga0226.76	IDA3	2	1	16	Request Data
ahx	Analysed	Hard X-rays	ahx0226.76	IDA3	1	0	6	Request Data
aim	Analysed	D-alpha emission and other spectral lines	aim0226.76	IDA3	2	0	6	Request Data
alp	Analysed	Langmuir Probe	alp0226.76	IDA3	4	0	116	Request Data

CCFE CULHAM CENTRE FOR FUSION ENERGY

Step 4: For each file – fill in a form

UK Atomic Energy Authority

Request Open Data

Title:

Forename:

Surname:

Contact E-mail:

Institute / Uni / Other:

Country:

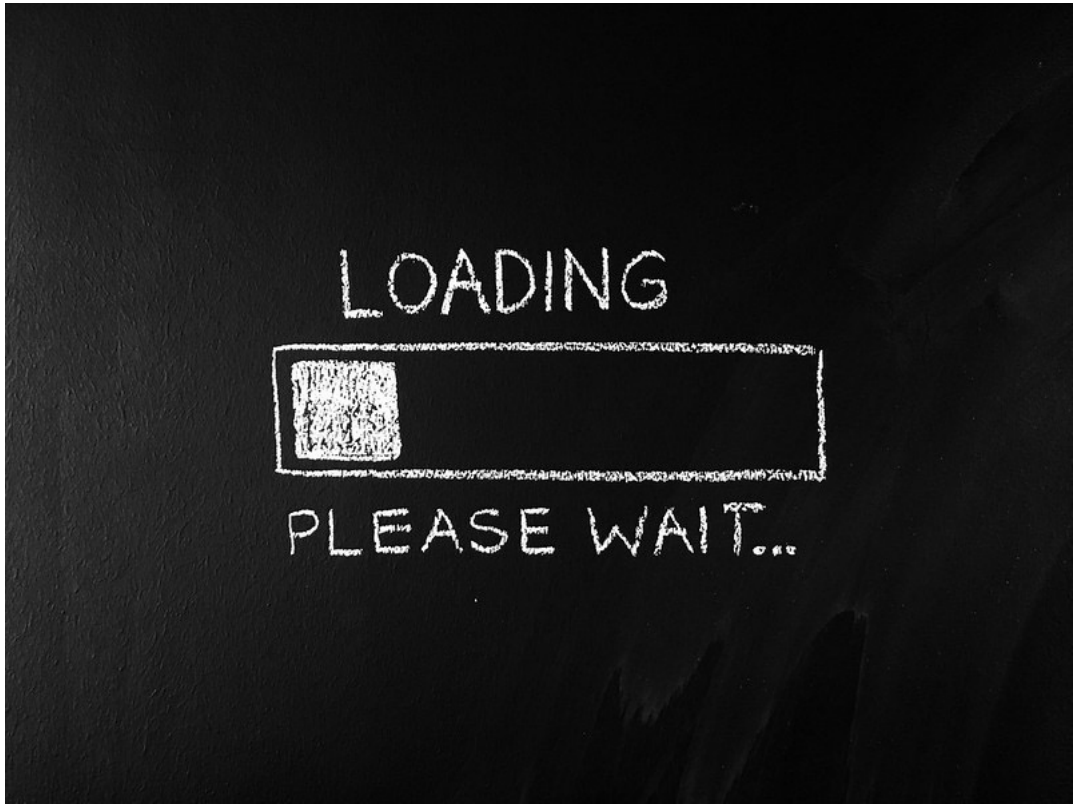
Request:

Reason for Request:

General Data Protection Regulation: The personal information that you provide to us in the form below will be held and processed by the organisation strictly in accordance with the provisions of the General Data Protection Regulation legal framework. The data will be used by the organisation to reply to your question in the most appropriate manner. We will not supply these personal details to any third party except if we are required to do so by operation of the law. We undertake to remove all of this personal information and communication every three years.

CCFE CULHAM CENTRE FOR FUSION ENERGY

Step 5: Wait...



© Jan Perseil,

21 Days Later

Problems with ”old” open data

UK Atomic Energy Authority

- Need to fill in a request for every data file is painful (HTML forms help of course)
- IDA3 format is not widely used (or used at all) outside CCFE, and the code to read it is not available without filling in another form.
  - But you can use IMAS/UDA to access it from within the community **if** your site is whitelisted
  - And for other users UKAEA kindly convert it to **excel** for you!!!
- No common metadata or ontology to allow you to identify shots of interest
  - Free text search can pull in unexpected results
- Custom license (https://opendata.ukaea.uk/license/)
- Data delivered by email after request has been assessed by a data librarian
- Scores low on the FAIRness scale
- (Also some issues around data quality but that’s a separate talk)

CCFE
CULHAM CENTRE FOR FUSION ENERGY

Making MAST Data more FAIR and Open

- Use of DataCite DOIs and metadata standards
- Metadata registered with DataCite
- Licensed moved to Creative Commons (CC-BY-NC-SA-4.0)
- Additional Provenance

The Magnifying glass, Tap, Gears set, Recycle sig, Storage, Infinity, Discussion, Shield, and Man User icons made by Freepik from www.flaticon.com are licensed by CC 3.0 BY. All other icons made by ARDC. Entire FAIR resources graphic is licensed under a Creative Commons Attribution 4.0 International License

Accessing FAIR and Open MAST Data

UK Atomic Energy Authority

List of MAST Open Data and Image files

Shot:
27461

Date:
2011-11-11

Time:
10:08:00

Programme:
ELM control with vertical kicks

Objective:
- Establish -10cm SND plasma- Perform vertical kicks to control ELMS

Search the list of files using keywords (case insensitive):

To drill down to information on an Open Data file, select any table cell
Use the Download or Request Data buttons to access Open Data files

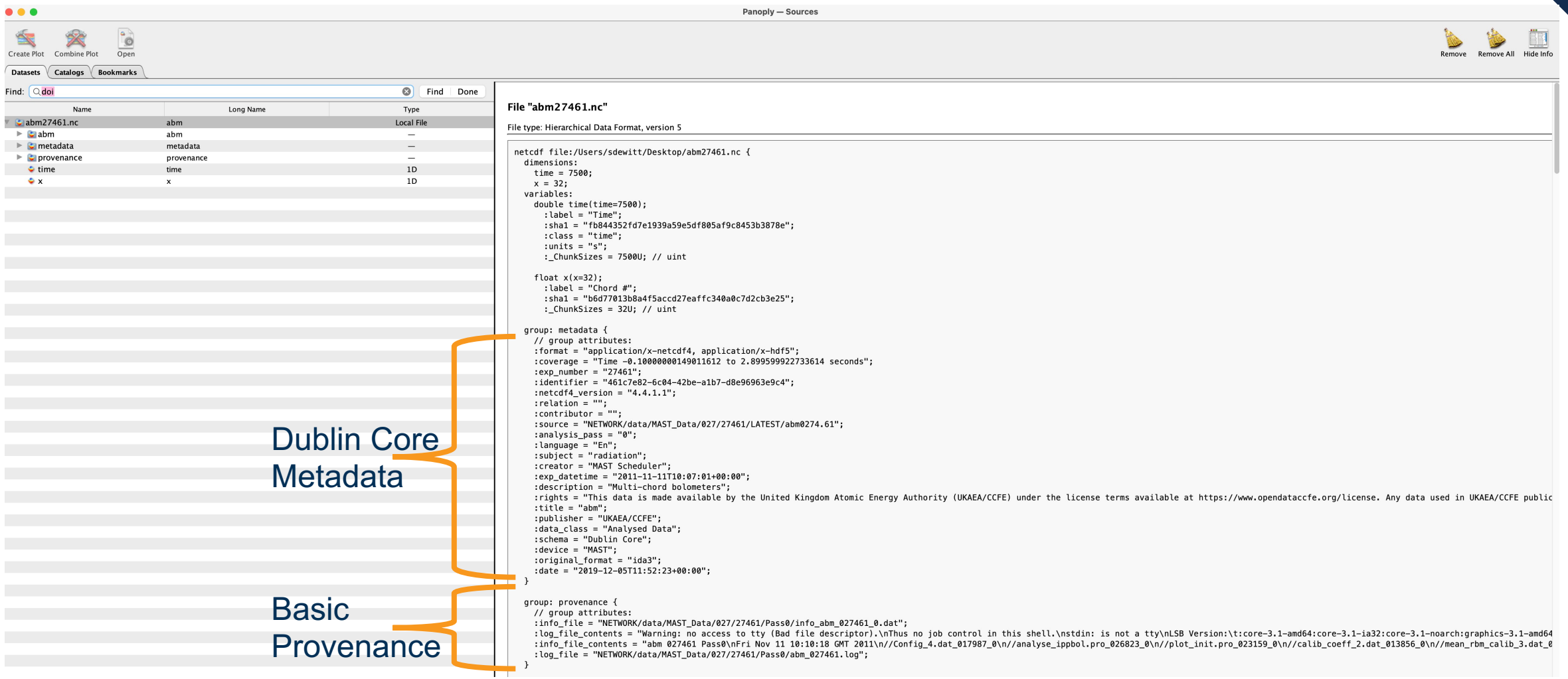
Small Change – Big Implications

Class	Type	Description	Filename	Format	Size	Pass	Signal	Download or Request Data
abm	Analysed	multi-chord bolometers	abm27461.nc	netcdf4/hdf5	3	0	21	Download
adg	Analysed	Plasma Edge Density gradient from the linear Dalpha camera	adg27461.nc	netcdf4/hdf5	1	0	4	Download
aga	Analysed	molecular deuterium pressure, neutral gas pressure, Gas Injection/Fueling	aga27461.nc	netcdf4/hdf5	2	0	15	Download
ahx	Analysed	Hard X-rays	ahx27461.nc	netcdf4/hdf5	1	0	6	Download
aim	Analysed	D-alpha emission and other spectral lines	aim27461.nc	netcdf4/hdf5	2	0	4	Download
alp	Analysed	Langmuir Probe	alp27461.nc	netcdf4/hdf5	4	0	127	Download
ama	Analysed	Magnetic Field Measurements: Analysis of Centre Column Toroidal Arrays of Bv Mirnovs	ama27461.nc	netcdf4/hdf5	15	0	9	Download

- Direct Download
- Register Once (short lived certificate generated)
- Immediate Access
- netCDF/HDF format

CCFE CULHAM CENTRE FOR FUSION ENERGY

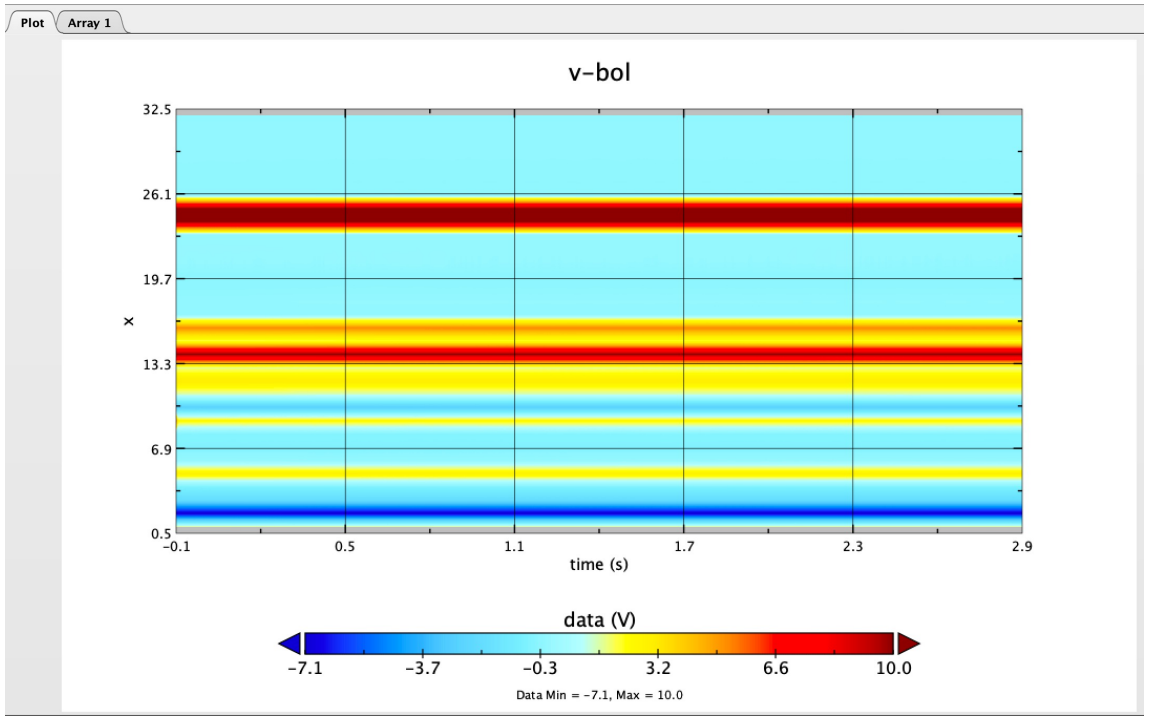
Metadata and Provenance in the data



Dublin Core Metadata

Basic Provenance

And Data



Plot | Array 1

Dataset: abm27461.nc
Variable: data, data
Units: V

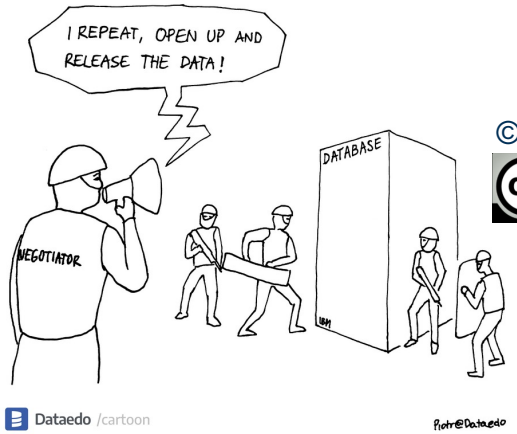
X Axis: time (s)

Y Axis: x	-0.1	-0.0996	-0.0992	-0.0988	-0.0984	-0.098	-0.0976	-0.0972	-0.0968	-0.0964	-0.096	-0.0956	-0.0952	-
1.0	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	
2.0	-7.1	-7.1	-7.1	-7.1	-7.1	-7.1	-7.1	-7.1	-7.1	-7.1	-7.1	-7.1	-7.1	
3.0	-1.9	-1.9	-1.9	-1.9	-1.9	-1.9	-1.9	-1.9	-1.9	-1.9	-1.9	-1.9	-1.9	
4.0	-0.7	-0.7	-0.7	-0.7	-0.7	-0.7	-0.7	-0.7	-0.7	-0.7	-0.7	-0.7	-0.7	
5.0	2.9	2.9	2.9	2.9	2.9	2.9	2.9	2.9	2.9	2.9	2.9	2.9	2.9	
6.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
7.0	-0.6	-0.6	-0.6	-0.6	-0.6	-0.6	-0.6	-0.6	-0.6	-0.6	-0.6	-0.6	-0.6	
8.0	-0.5	-0.5	-0.5	-0.5	-0.5	-0.5	-0.5	-0.5	-0.5	-0.5	-0.5	-0.5	-0.5	
9.0	7.3	6.3	5.6	5.1	4.6	4.3	4.0	3.7	3.5	3.4	3.2	3.1	3.0	
10.0	-2.6	-2.6	-2.6	-2.6	-2.6	-2.6	-2.6	-2.6	-2.6	-2.6	-2.6	-2.6	-2.6	
11.0	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	
12.0	3.1	3.1	3.1	3.1	3.1	3.1	3.1	3.1	3.1	3.1	3.1	3.1	3.1	
13.0	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	
14.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	
15.0	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	
16.0	5.2	5.2	5.2	5.2	5.2	5.2	5.2	5.2	5.2	5.2	5.2	5.2	5.2	
17.0	-0.1	-0.1	-0.2	-0.1	-0.1	-0.2	-0.1	-0.1	-0.2	-0.2	-0.1	-0.2	-0.2	
18.0	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	
19.0	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	
20.0	-0.3	-0.3	-0.3	-0.3	-0.3	-0.3	-0.3	-0.3	-0.3	-0.3	-0.3	-0.3	-0.3	
21.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
22.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
23.0	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	
24.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	
25.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	
26.0	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	
27.0	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	
28.0	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	

Data Format: %.1f Flip table B/T Flip table L/R Show cell indices Row/Col Header Format: %.7G

Work in Progress

- Only subset of shots are currently available (shots 27000-27999)
  - Still requires some manual inputs
- Need to complete registration with DataCite
- Programmatic Access
- Need to make searching programmatically possible
- Tested integration with FAIR4Fusion Demonstrator, but needs to be developed into a service



Questions?