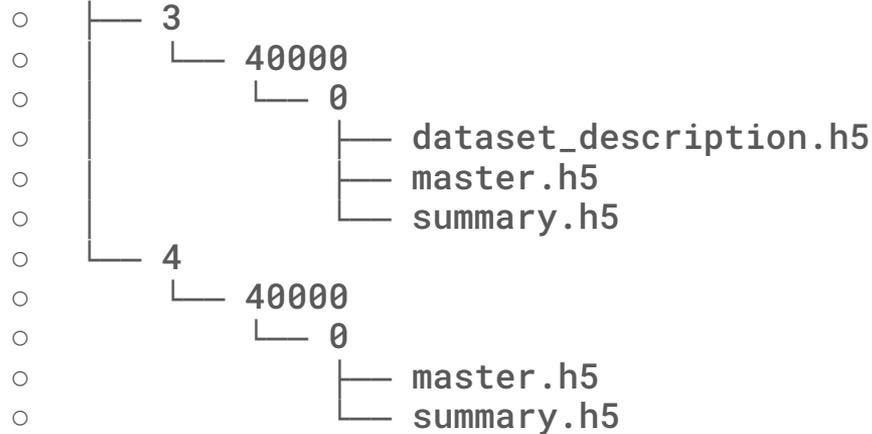# DMP Implementation Status of IPP

2025-04-01

Update of the AUG SUMMARY IDS production process

- "fill_summary.py" has been updated
  - to use **imas_python** rather than **AL-PYTHON**
    - to produce SUMMARY and DATASET_DESCRIPTION IDSes in **DD 3.42.0**
    - and also to produce SUMMARY IDS in **DD 4.1.0**
- Test output for shot 40000 can be found at
  - `/afs/eufus.eu/user/g/g2dpc/public/aug-dmp/`
  - `├── 3`
  - `│   └── 40000`
  - `│       └── 0`
  - `│           ├── dataset_description.h5`
  - `│           ├── master.h5`
  - `│           └── summary.h5`
  - `└── 4`
  - `    └── 40000`
  - `        └── 0`
  - `            ├── master.h5`
  - `            └── summary.h5`
  -
  - `6 directories, 5 files`

Update of the AUG SUMMARY IDS production process

```
summary
├── ids_properties
│   ├── comment: 'Using fill_summary.py'
│   ├── homogeneous_time: 1
│   ├── provider: 'dpc'
│   ├── creation_date: '2026/03/27'
│   ├── ids_properties/version_put
│   │   ├── data_dictionary: '4.1.0'
│   │   ├── access_layer: '5.6.0'
│   │   └── access_layer_language: 'IMAS-Python 2.2.0'
├── type
│   ├── name: 'experimental'
│   ├── index: 1
│   └── description: 'Data coming from an experiment (contains only data from an experimental database)'
├── machine: 'ASDEX Upgrade'
├── pulse: 40000
├── pulse_time_begin: '2022-01-13T16:38:00Z'
├── tag
│   ├── name:
│   │   '{"description": "** AT pedestal at different beta_pol and fueling\\nchanged beta_pol FB, same n0div - worked well\\nHappy anniversary ASDEX
│   │   Upgrade!!!"}'
│   └── comment: 'Additional data for which there is no home in the summary IDS.'
├── global_quantities
│   ├── global_quantities/ip
│   │   ├── value: array([507070.4048, 336141.3362, 365356.8183, ..., 218076.179 , 163851.2793, 100095.5337], shape=(63,))
│   │   ├── value_error_upper: array([22183.9713, 12710.8831,  8745.2933, ..., 15263.3538, 16016.9171, 15832.5637], shape=(63,))
│   │   └── source: 'AUGD/FPC/IpiFP/1'
│   ├── global_quantities/v_loop
│   │   ├── value: array([ 1.8923,  1.5538,  1.6285, ...,  0.2671, -0.0677, -0.2575], shape=(63,))
│   │   ├── value_error_upper: array([2.4623, 0.3042, 0.1769, ..., 0.1262, 0.0897, 0.0542], shape=(63,))
│   │   └── source: 'AUGD/TOT/u_loop/2'

├── heating_current_drive
│   ├── heating_current_drive/power_ec
│   │   ├── value: array([3677.7567, 3487.975 , 3414.3834, ..., 3184.4584, 3092.7008, 3042.212 ], shape=(63,))
│   │   ├── value_error_upper: array([1429.837 , 1421.831 , 1421.8188, ..., 1410.0454, 1350.9939, 1339.879 ], shape=(63,))
│   │   └── source: 'AUGD/ECS/PECRH/1'
│   ├── heating_current_drive/power_nbi
│   │   ├── value: array([0., 0., 0., ..., 0., 0., 0.], shape=(63,))
│   │   ├── value_error_upper: array([0., 0., 0., ..., 0., 0., 0.], shape=(63,))
│   │   └── source: 'AUGD/NIS/PNI/1'
├── time_width: array([0.1, 0.1, 0.1, ..., 0.1, 0.1, 0.1], shape=(63,))
├── code
│   ├── name: 'fill_summary.py'
│   ├── commit: '0.0.0-5-g452fbc4-dirty'
│   ├── version: '0.0.0-5-g452fbc4-dirty'
│   ├── repository: 'git@gitlab.com:fair-for-fusion/aug-summary-ingestion.git'
│   └── output_flag: array([0, 0, 0, ..., 0, 0, 0], shape=(63,), dtype=int32)
└── time: array([0.17, 0.27, 0.37, ..., 6.17, 6.27, 6.37], shape=(63,))
```

Update of the AUG SUMMARY IDS production process

- Do we want a new set of run 0 cases to replace the currently existing data?
  - Or should I produce a new set of run 1 cases to add to the existing data
    - **This is, I think, an important policy decision about how we handle updates in the catalogue system**

- Do we want the new data in 3.42.0 or 4.1.0?

Long lead time items

No	Title	Description	Expected date
1	MA. 1	Provide scenario A data using DD 3	2026/Q1
2	MA. 2	Provide a list of recommended signals for scenario A	2026/Q2
3	MA. 3	Provide updated scenario A data using DD 4	2026/Q4
4	MA. 4	Include modelling and simulation data in metadata Catalogue	2027/Q2
5	MB. 1	Provide a list of recommended signals from first use cases (live document)	2026/Q1
6	MB. 2	Provide scenario B data using DD 3	2026/Q2
7	MB. 3	Strategy for Migration DD 4	2026/Q2
8	MB. 4	Implement AAI access to data (on all sites?)	2026/Q4
9	MB. 5	Provide scenario B data using DD 4	2027/Q2
10	MC. 1	Strategy for PID integrations	2026/Q3
11	MC. 2	Release of data with enhanced provenance	2026/Q4
12	MC. 3	Release of data with PIDs	2027/Q2
13	MD. 1	Plan for open data releases	2026/Q4
14	MD. 2	Release of open data from at least one device	2027/Q4

- MA.4 best done with SimDB with parallel UDA server
- MB.4 requires UDA docker images with AAI functionality
- MB.5 requires UDA docker image supporting DD4 (?)
  - Move to IMAS-python?
- MC.3 requires PID minting from experiments and SimDB
  - **Could we store the SUMMARY IDSes on LTDSF SimDB and just have PIDs minted by SimDB?**

Update of the AUG SUMMARY IDS production process

- Could / should we use the DATASET_FAIR **identifier** field to store version 7 UUIDs until we get PIDs working?
  - This would allow for setting **replaces** field and inferring **is_replaced_by**
- What should be stored in dataset_fair/rights_holder
  - Text name?
    - "Max Planck Institute for Plasma Physics" or
    - "Max-Planck-Institut für Plasmaphysik"?
  - **ROR**: https://ror.org/03taest98
- Do we have agreed "License(s) under which the data is made available (license description or, more convenient, publicly accessible URL pointing to the full license text)"?

Update of the AUG SUMMARY IDS production process

```
dataset_fair
└── ids_properties
    ├── comment: 'Using fill_summary.py'
    ├── homogeneous_time: 2
    ├── provider: 'dpc'
    ├── creation_date: '2026-04-01T09:07:00Z'
    └── ids_properties/version_put
        ├── data_dictionary: '3.42.0'
        ├── access_layer: '5.6.0'
        └── access_layer_language: 'IMAS-Python 2.2.0'
├── identifier: '069cce04-7928-74bd-8000-e5aae54e9810'
├── valid: '2026-04-01/'
└── rights_holder: 'https://ror.org/03taest98'
```

```
dataset_fair
└── ids_properties
    ├── comment: 'Using fill_summary.py'
    ├── homogeneous_time: 2
    ├── provider: 'dpc'
    ├── creation_date: '2026-04-01T09:05:00Z'
    └── ids_properties/version_put
        ├── data_dictionary: '4.1.0'
        ├── access_layer: '5.6.0'
        └── access_layer_language: 'IMAS-Python 2.2.0'
├── identifier: '069ccdfd-c7d0-7514-8000-bf6d8554f499'
├── valid: '2026-04-01/'
└── rights_holder: 'https://ror.org/03taest98'
```

- Possible example dataset_fair IDSes (DD 3.42.0 and 4.1.1)
  - Chose to fill
    - identifier with uuid7
   - rights_holder with Research Organization Registry (ROR) https reference

Also explored imas-python + UDA

- Pip installation of imas-python, idstools
  - After some slack exchanges, now also have UDA support
    - Possible to run (from home)
      - `setenv IMAS_VERSION 3.42.0`
      - `time plotkineticprofiles -u 'imas://uda.ipp.mpg.de:56565/uda?path=/root/public/imasdb/aug/3/32408/12&backend=hdf5' --save`
      - `11.274u 4.187s 6:50.98 3.7%    0+0k 0+0io 2050pf+0w`
    - Need to set IMAS_VERSION otherwise
      - ` File "/Users/dpc/IMAS/venv/lib/python3.12/site-packages/idstools/domain/kineticprofiles.py", line 251, in fill_idses`
      - `   self.common_time_length = len(self.edge_profiles.time)`
      - `                             ^^^^^^^^^^^^^^^^^^^^^^^^`
      - `AttributeError: 'NoneType' object has no attribute 'time'`
    - Need to fix plotkineticprofiles changing np.NaN to np.nan
      - ` File "/Users/dpc/IMAS/venv/lib/python3.12/site-packages/idstools/domain/kineticprofiles.py", line 1143, in get_v_phi_profile`
      - `   vphioid = np.asarray([np.NaN] * self.nrho)`
      - `                         ^^^^^^`
      - ` File "/Users/dpc/IMAS/venv/lib/python3.12/site-packages/numpy/__init__.py", line 778, in __getattr__`
      - `   raise AttributeError(`
      - `AttributeError: ` np.NaN` was removed in the NumPy 2.0 release. Use `np.nan` instead.`

Also explored mdsthin

- https://github.com/MDSplus/mdsthin
  - A Python MDSplus Thin Client Implementation
    - pip install mdsthin

IMAS versions updated on the TOK and Citrix systems

| DD	AL	TOK cluster	Horizon system
3.38.1	4.11.4		2023-06-08
3.39.0	4.11.5	2023-07-26	2023-07-31
3.39.0	5.0.0	2023-10-10	
3.41.0	4.11.10	2024-04-17	
3.41.0	2024.07-foss-2023b		2024-08-11
3.42.0	2024.08-foss-2023b		2024-08-13
3.42.0	2024.08.1-foss-2023b	2024-08-23	2024-08-22
3.42.0	2024.08.1-intel-2023b	2024-08-23	2024-08-22
3.42.0	4.11.10		2024-10-09
3.42.0	2024.09-foss-2023b	2024-10-21	2024-10-21
3.42.0	2024.09-intel-2023b	2024-10-21	2024-10-21
4.0.0	2024.12-foss-2023b	2025-01-22	2025-01-22
4.0.0	2024.12-intel-2023b	2025-01-22	2025-01-22

- Entries in red have a problem after the cluster upgrade
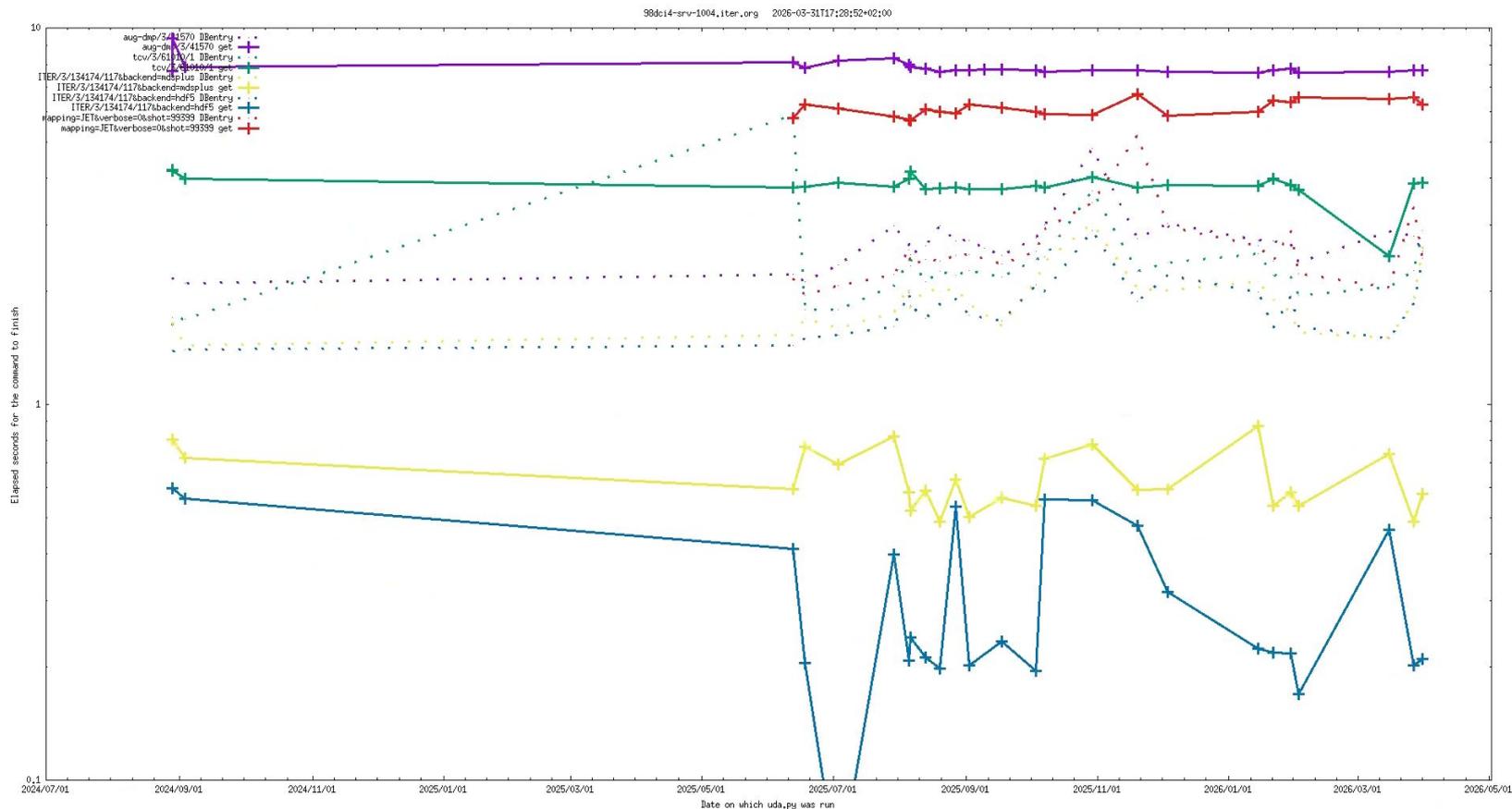 - Underlying python changed, together with numpy
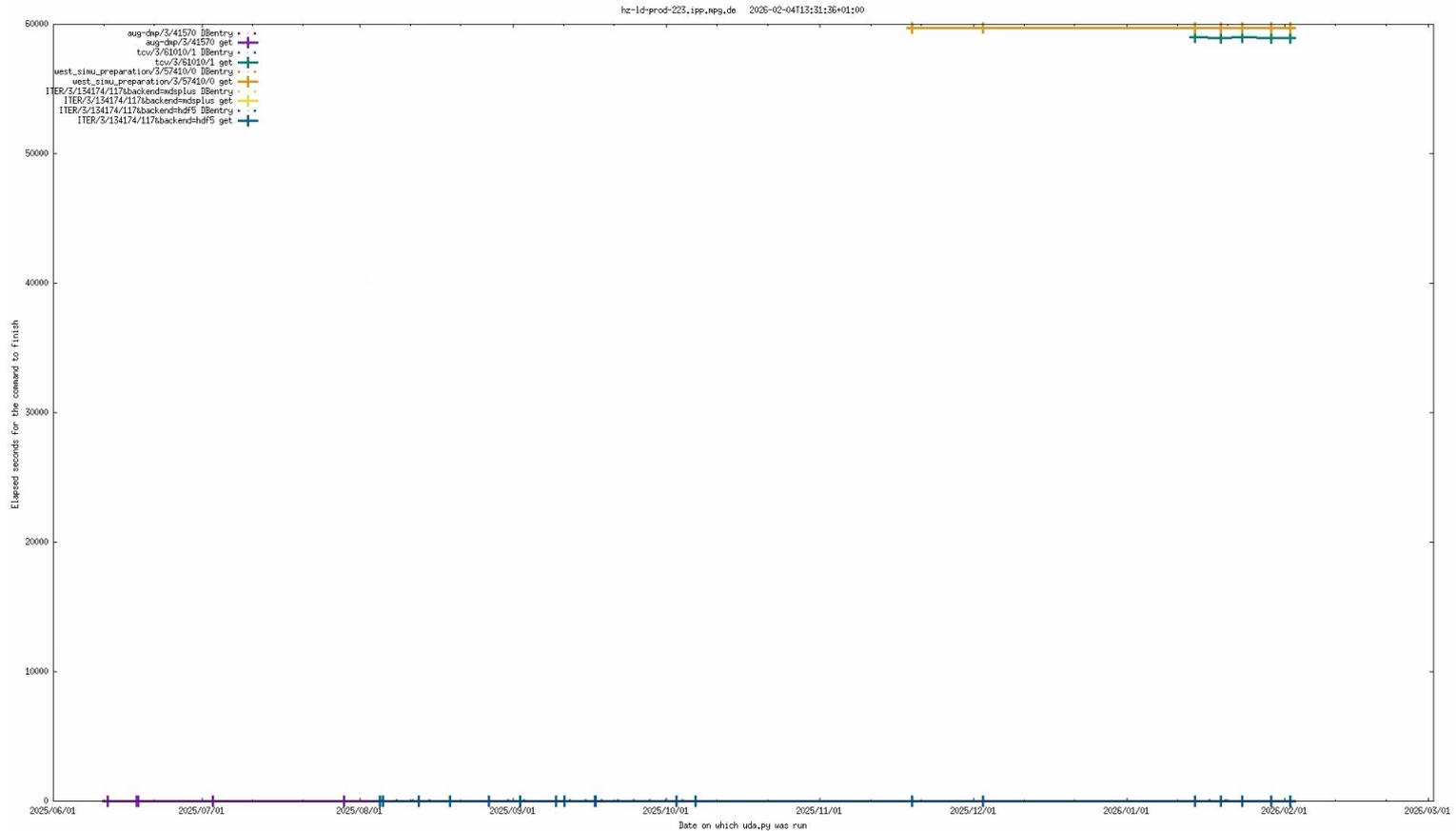   - Remove?

Data access times: run on Gateway
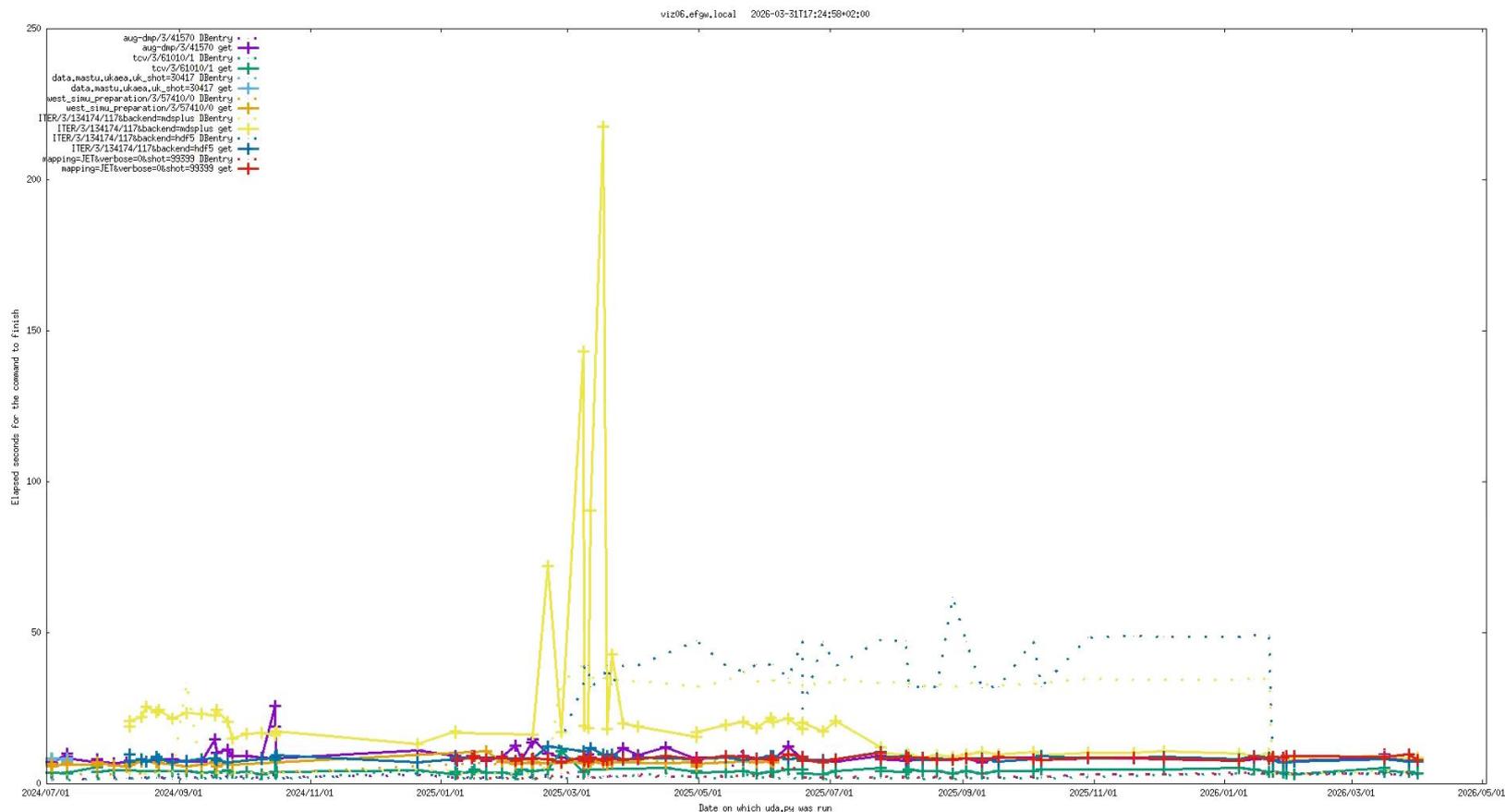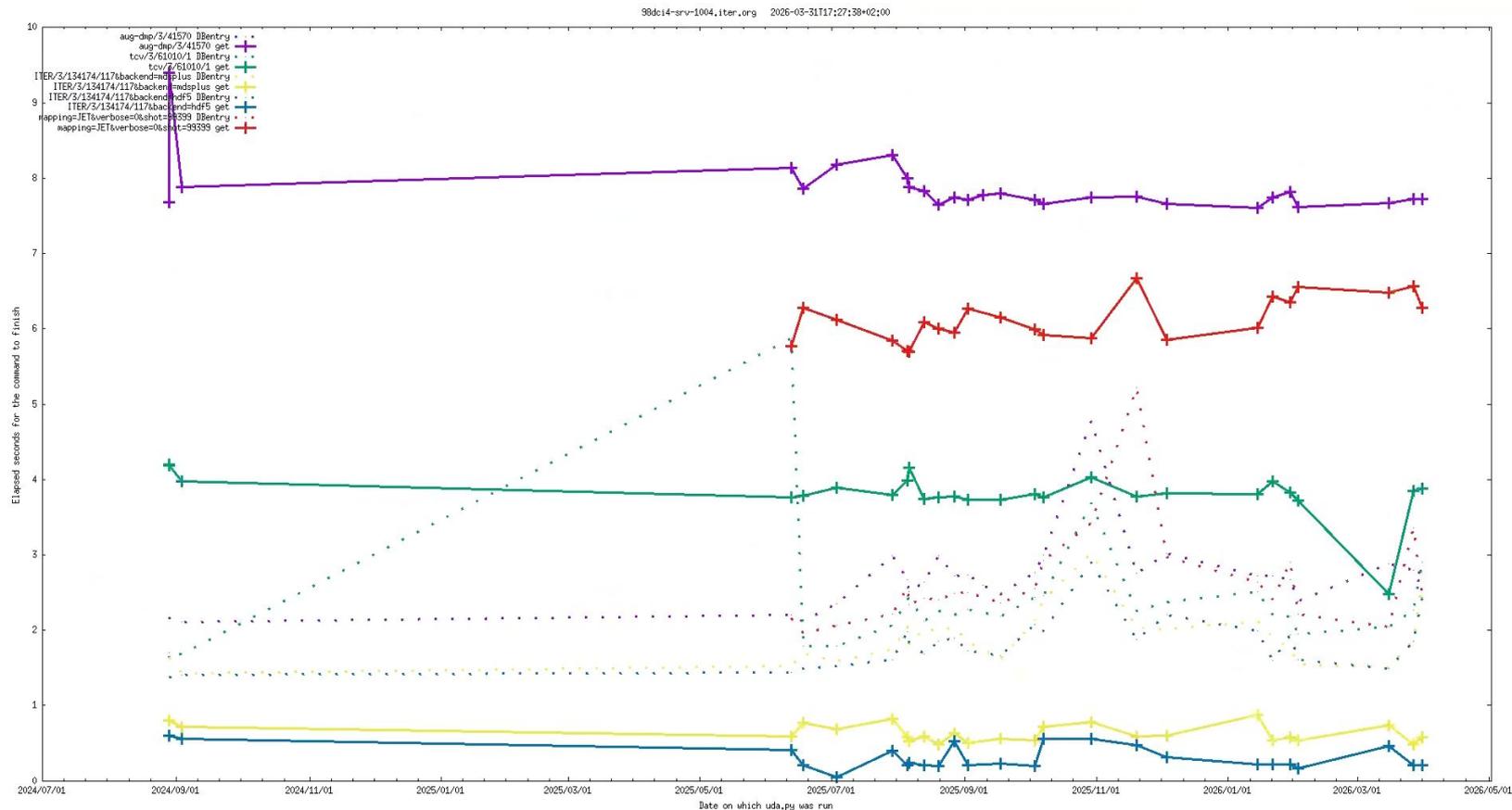
Data access times: run on Gateway

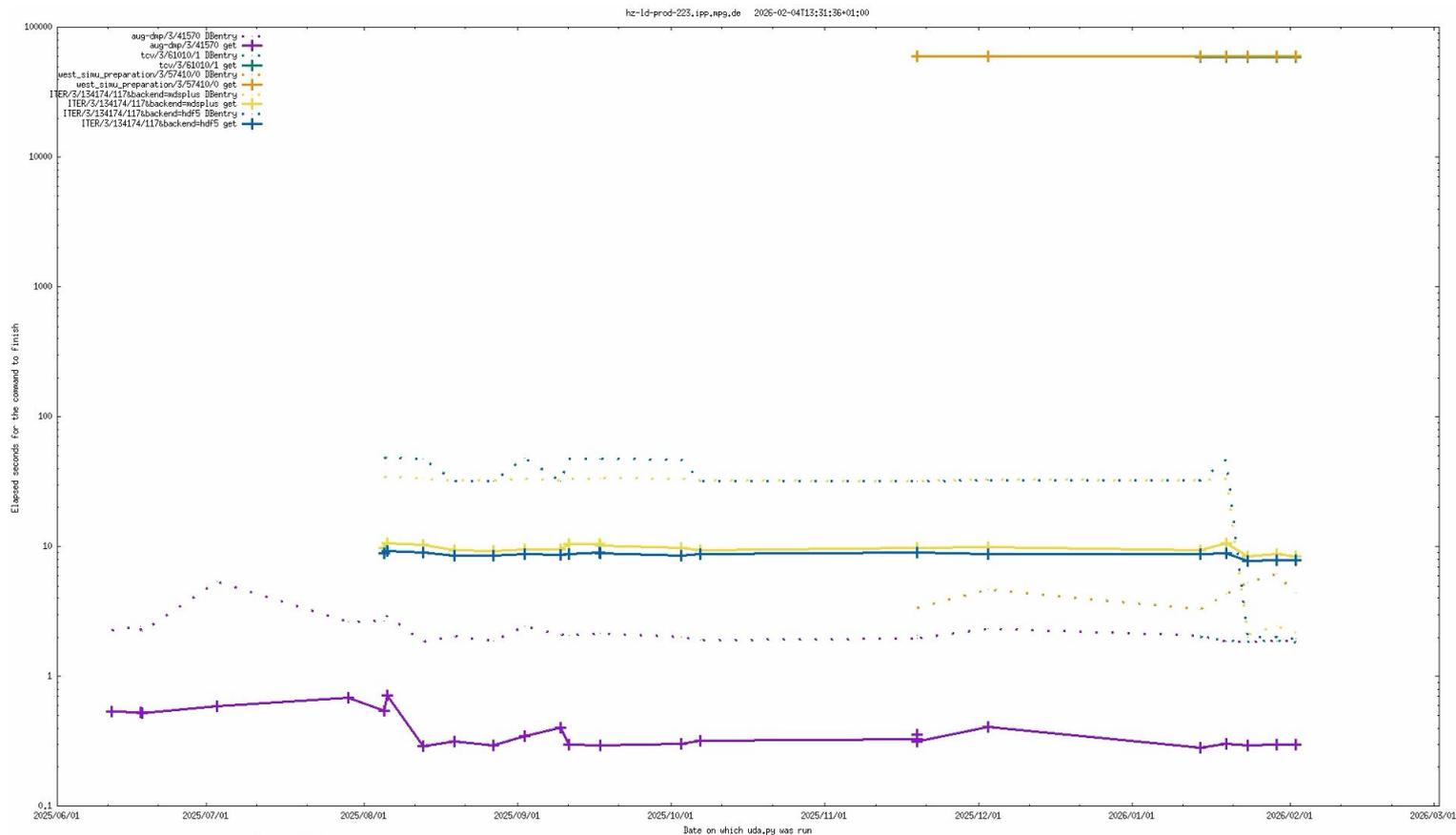# Data access times: run at ITER
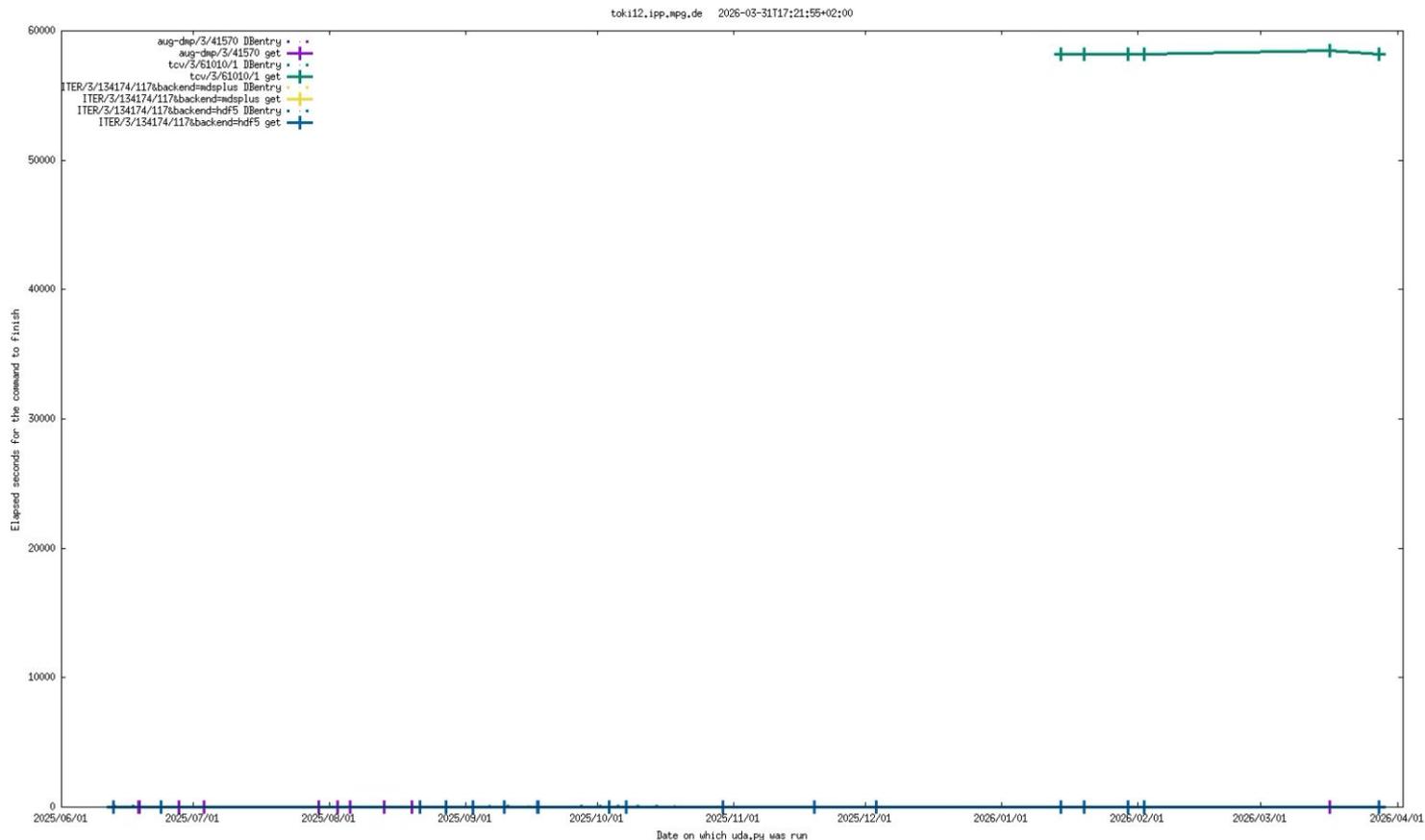
Data access times: run at ITER

Data access times: run at IPP (horizon)

Data access times: run at IPP (horizon)

Data access times: run at IPP (TOK cluster)

Data access times: run at IPP (TOK cluster)