DAS data management: Metadata and data access

Javier Quinteros, Submerse and Geo-INQUIRE DAS group

Summer School, Lisbon, Portugal. August 2025

GFZ Helmholtz Centre for Geosciences

Geo-INQUIRE

HELMHOLTZ

FDSN Web Services

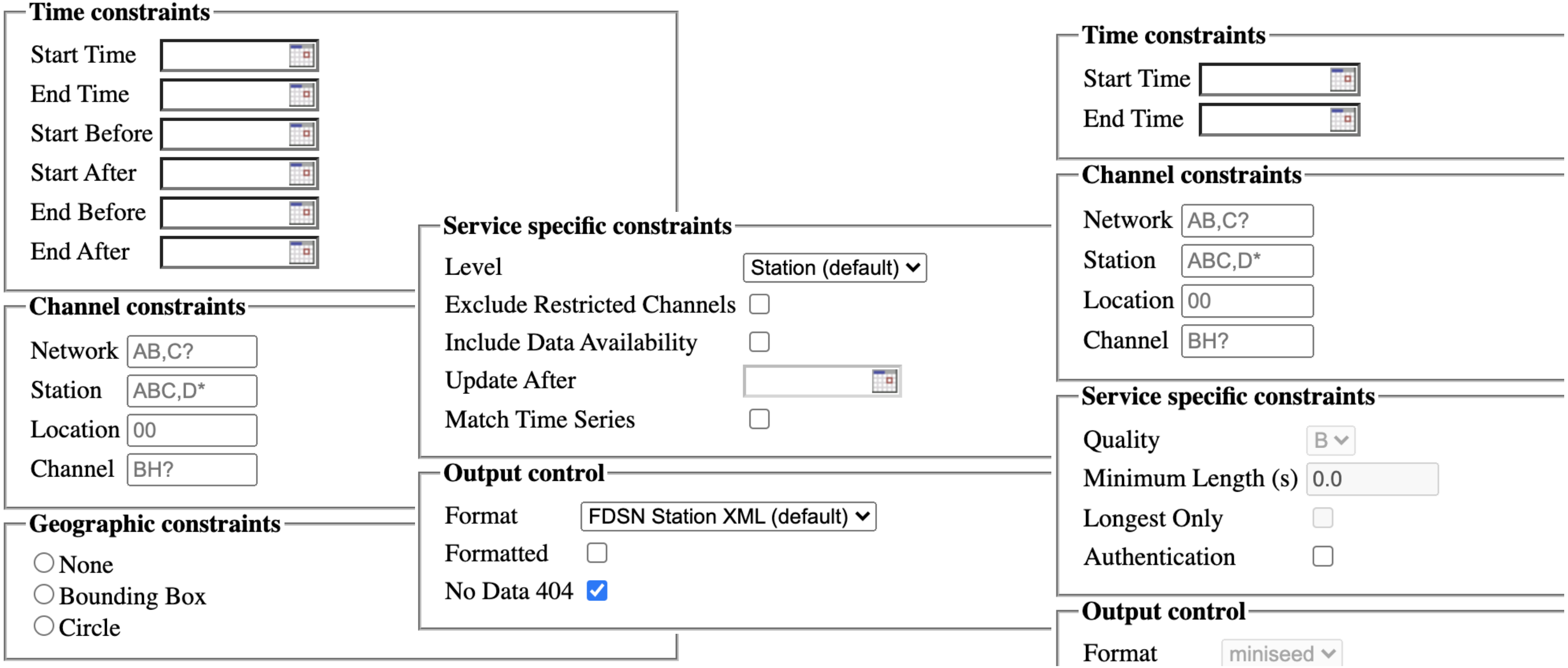
Interfaces for the exchange of time series data, related metadata, event parameter. Namely,

- **Station**: metadata in StationXML and alternate formats.
- **Dataselect**: raw time series data in miniSEED format.
- **Event**: parametric data for events in QuakeML and alternate formats.

General workflow to get data:

1. Station discovery
2. Request metadata
3. Select and download data

FDSN Web Services



Station: http://www.orfeus-eu.org/fdsnws/station/1/builder

Dataselect: http://www.orfeus-eu.org/fdsnws/dataselect/1/builder

GFZ Helmholtz Centre for Geosciences

FDSN Web Services

`<site>/fdsnws/<servicename>/<majorversion>/`

<site>: domain name of the hosting WS (top institutional level),

<servicename>: name of the service, as in the previous slide,

<majorversion>: first number describing the WS version.

For instance, https://geofon.gfz.de/fdsnws/dataselect/1/

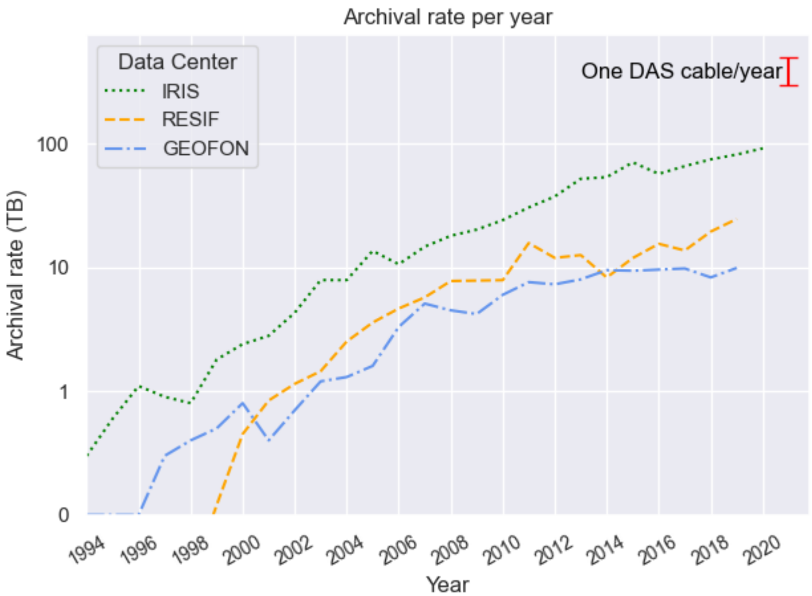
Station(?) discovery

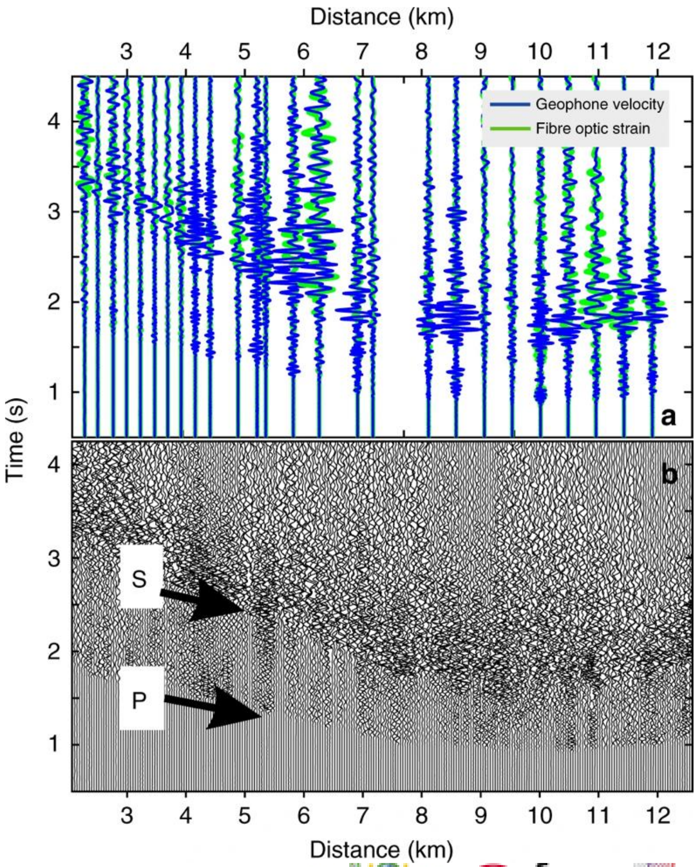
https://geofon.gfz.de/fdsnws/station/1/query?net=3U&format=text

```
#Network|Station|Latitude|Longitude|Elevation|SiteName|StartTime|EndTime
3U|A0905|52.38517750|13.01958146|32.0|Station A0905, GER|2023-02-01T00:00:00|2023-02-28T23:59:59
3U|A0915|52.3852445|13.01930937|32.00612208892051|Station A0915, GER|2023-02-01T00:00:00|2023-02-28T23:59:59
3U|A0925|52.3853191|13.01904268|32.76337562531175|Station A0925, GER|2023-02-01T00:00:00|2023-02-28T23:59:59
3U|A0935|52.3853849|13.01877011|33.49665967486104|Station A0935, GER|2023-02-01T00:00:00|2023-02-28T23:59:59
3U|A0945|52.3853983|13.01848139|32.56618372835495|Station A0945, GER|2023-02-01T00:00:00|2023-02-28T23:59:59
3U|A0955|52.3854171|13.01818943|32.24139172904475|Station A0955, GER|2023-02-01T00:00:00|2023-02-28T23:59:59
3U|A0965|52.3854388|13.01789807|32.2381928738347|Station A0965, GER|2023-02-01T00:00:00|2023-02-28T23:59:59
3U|A0975|52.385461|13.01760691|32.13179102218067|Station A0975, GER|2023-02-01T00:00:00|2023-02-28T23:59:59
3U|A0985|52.3854839|13.01731568|32.17124585346524|Station A0985, GER|2023-02-01T00:00:00|2023-02-28T23:59:59
3U|A0995|52.3854946|13.01702282|32.21960064446748|Station A0995, GER|2023-02-01T00:00:00|2023-02-28T23:59:59
3U|A1005|52.3854984|13.01672927|32.205174931704505|Station A1005, GER|2023-02-01T00:00:00|2023-02-28T23:59:59
3U|A1015|52.3854951|13.01643572|32.21736902331234|Station A1015, GER|2023-02-01T00:00:00|2023-02-28T23:59:59
```

GFZ Helmholtz Centre for Geosciences

Roadmap for DAS standardization





Top: A simple DAS experiment can generate more data than the one archived in a whole year at any data centre in the world. Right: Comparison of the resolution between broad band sensors and a DAS interrogator.

Quinteros et al. (SRL, 2021); Jousset et al. (Nature, 2018)

GFZ Helmholtz Centre for Geosciences

Geo-INQUIRE Orfeus EarthScope Consortium

Roadmap for DAS standardization

- Metadata: community and seismologica
- Data format(s)
- Long-term archival
- Data provisioning
- Real-time transmission
- Processing
- Ethical issues, or related to privacy and security

Exploring Approaches for Large Data in Seismology: User and Data Repository Perspectives

Javier Quinteros*[1], Jerry A. Carter[2], Jonathan Schaeffer[3], Chad Trabant[2], and Helle A. Pedersen[3,4]

Abstract

New data acquisition techniques are generating data at much finer temporal and spatial resolution, compared to traditional seismic experiments. This is a challenge for data centers and users. As the amount of data potentially flowing into data centers increases by one or two orders of magnitude, data management challenges are found throughout all stages of the data flow.

The Incorporated Research Institutions for Seismology—Réseau sismologique et géodésique français and GEOForschungsNetz data centers—carried out a survey and conducted interviews of users working with very large datasets to understand their needs and expectations. One of the conclusions is that existing data formats and ser-

Quinteros et al. (SRL, 2021) doi:10.1785/0220200390

DAS Metadata

Toward a Metadata Standard for Distributed Acoustic Sensing (DAS) Data Collection

Voon Hui Lai[*1], Kathleen M. Hodgkinson[2], Robert W. Porritt[2], and Robert Mellors[3]

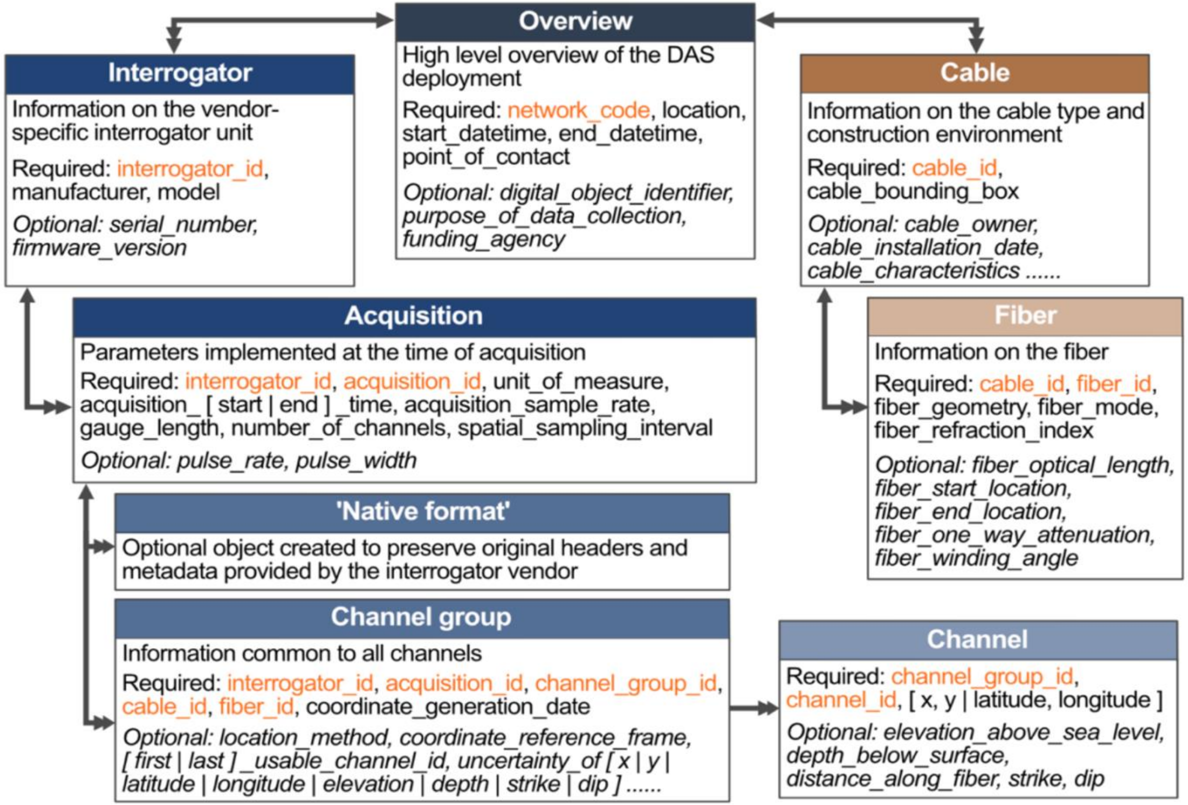
Abstract

With increasing geophysical applications using distributed acoustic sensing (DAS) technology, there is a need to implement a metadata standard specifically for DAS to facilitate the integration of DAS measurements across experiments and increase reusability. We propose a metadata standard intended primarily for the DAS research community, which fully describes the five key components of a DAS experiment: (1) interrogator; (2) data acquisition; (3) channels; (4) cable; and (5) fiber. The proposed metadata schema, which is the overall structure of the metadata, is hierarchical based, with a parent "overview" metadata block describing the experiment, and two main child

Voon Hui Lai, et al. (2024),
doi:10.1785/0220230325

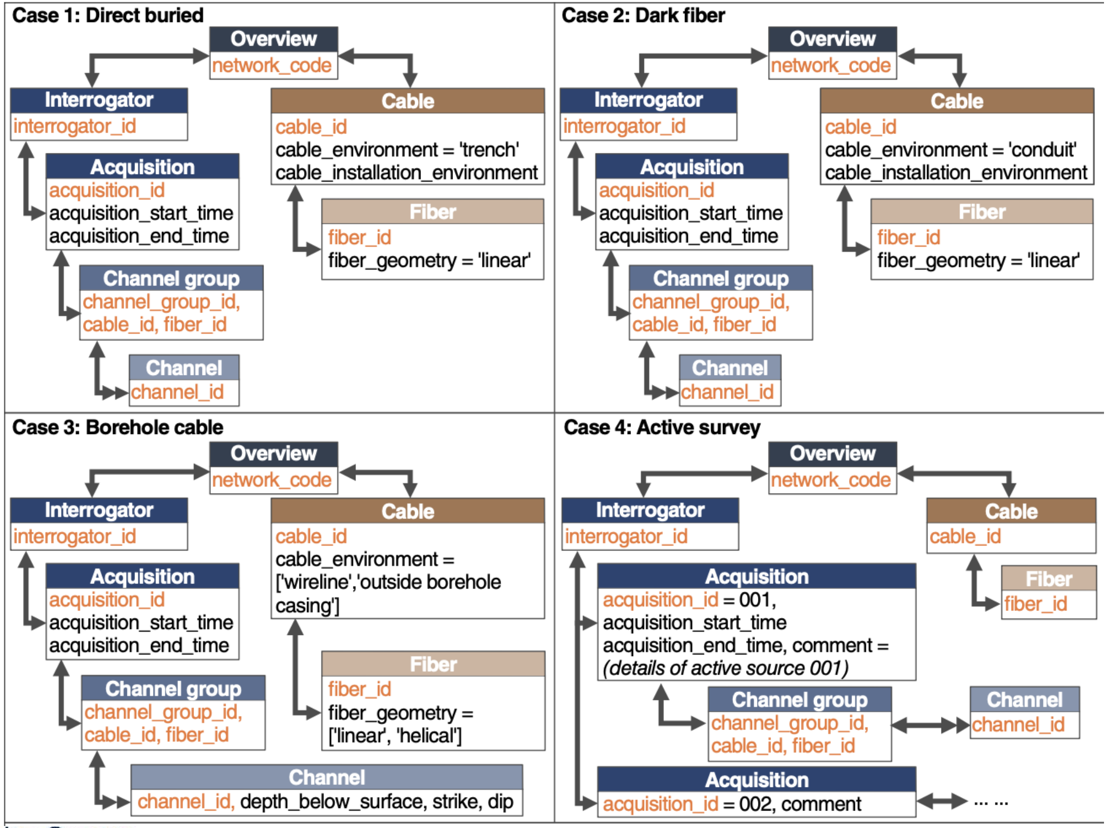
- DAS-RCN Data Management Working Group proposed a starting point for a common DAS metadata standard for archival purposes and to guide data collection at experiments.
- The specification was published after 2 years of discussion within the community.
- ORFEUS and EarthScope did a first evaluation of the schema and met the authors to prepare a proposal for FDSN evaluation and adoption.

DAS Metadata



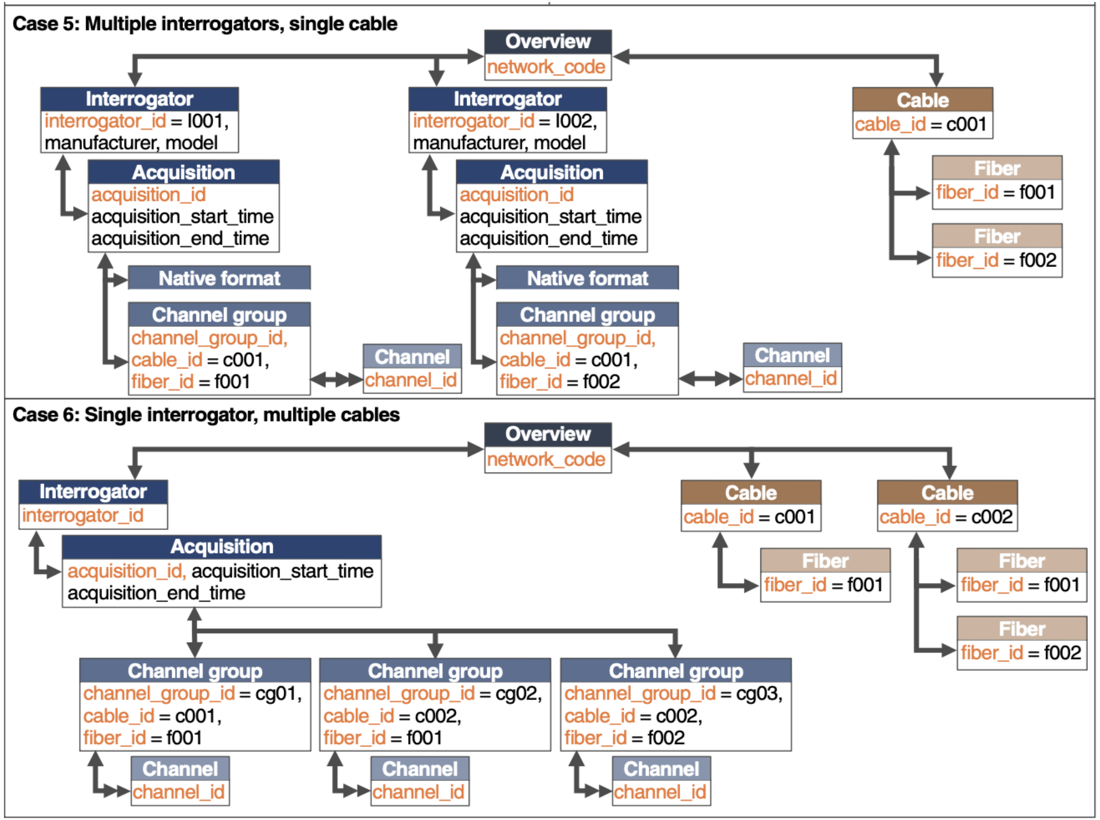
GFZ Helmholtz Centre for Geosciences

DAS Metadata



GFZ Helmholtz Centre for Geosciences

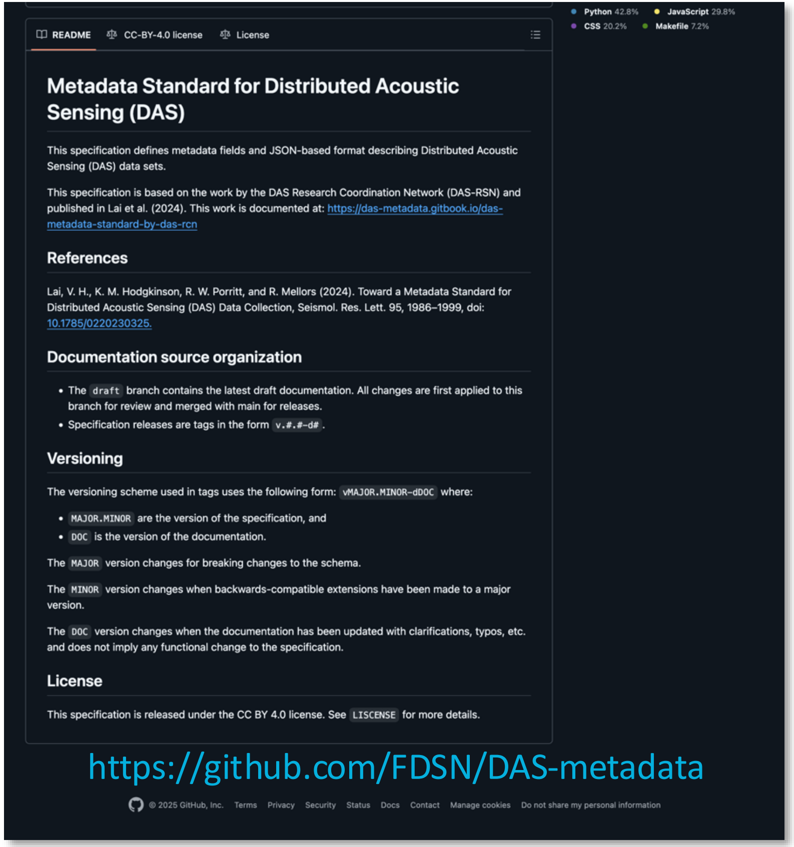
DAS Metadata



GFZ Helmholtz Centre for Geosciences

Metadata schema as FDSN standard

- DAS-RCN authors, EarthScope, GFZ submitted a FDSN proposal (as presented in the paper), FDSN Github repository ready
- Formal JSON schema included with minor improvements and examples
- Approval by WG2 expected this Summer 2025
- Build (or adapt) our software ecosystem on top of that (started!)



README · CC-BY-4.0 license · License

Python 42.8% · JavaScript 29.8% · CSS 20.2% · Makefile 7.2%

Metadata Standard for Distributed Acoustic Sensing (DAS)

This specification defines metadata fields and JSON-based format describing Distributed Acoustic Sensing (DAS) data sets.

This specification is based on the work by the DAS Research Coordination Network (DAS-RSN) and published in Lai et al. (2024). This work is documented at: https://das-metadata.gitbook.io/das-metadata-standard-by-das-rcn

References

Lai, V. H., K. M. Hodgkinson, R. W. Porritt, and R. Mellors (2024). Toward a Metadata Standard for Distributed Acoustic Sensing (DAS) Data Collection, Seismol. Res. Lett. 95, 1986–1999, doi: 10.1785/0220230325.

Documentation source organization

- The `draft` branch contains the latest draft documentation. All changes are first applied to this branch for review and merged with main for releases.
- Specification releases are tags in the form `v.#.#-d#`.

Versioning

The versioning scheme used in tags uses the following form: `vMAJOR.MINOR-dDOC` where:

- `MAJOR.MINOR` are the version of the specification, and
- `DOC` is the version of the documentation.

The `MAJOR` version changes for breaking changes to the schema.

The `MINOR` version changes when backwards-compatible extensions have been made to a major version.

The `DOC` version changes when the documentation has been updated with clarifications, typos, etc. and does not imply any functional change to the specification.

License

This specification is released under the CC BY 4.0 license. See `LISCENSE` for more details.

https://github.com/FDSN/DAS-metadata

© 2025 GitHub, Inc. Terms Privacy Security Status Docs Contact Manage cookies Do not share my personal information

GFZ Helmholtz Centre for Geosciences · Geo-INQUIRE · Orfeus · EarthScope Consortium

Data Management (in the mean time)

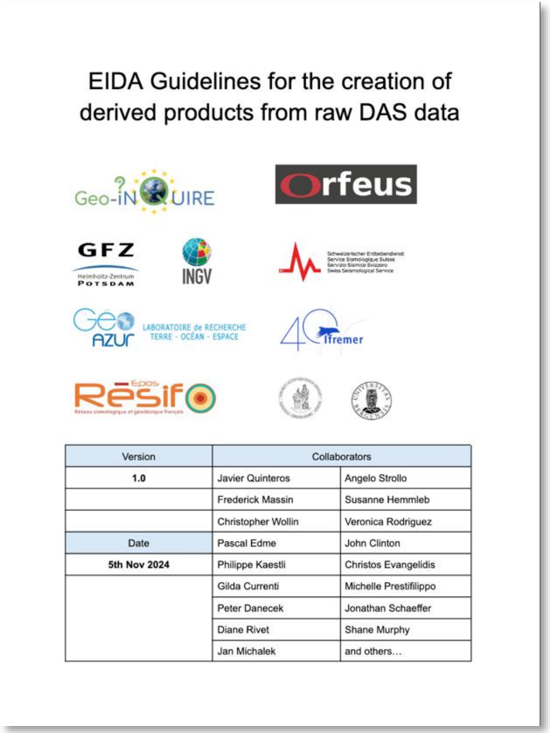
- Until new standards are developed and adopted the community needs a seamless way to integrate the DAS datasets using current seismic standard formats (e.g. miniseed for data, StationXML for metadata)
- Strategy to standardize these datasets by downsampling them and creating a basic standard metadata (StationXML) mapped from raw data and extra information provided by the PI. The result is ready to be archived in a standard way
- Decimated real-time Miniseed derived products already flowing

Guidelines: derived products from DAS data

First version agreed with Geo-I partners

Including the following points:

- Data Management
- How to subsample
- Channel Naming
- Miniseed technicalities
- New DAS metadata (Voon Hui Lai et al, 2024)
- DataCite metadata

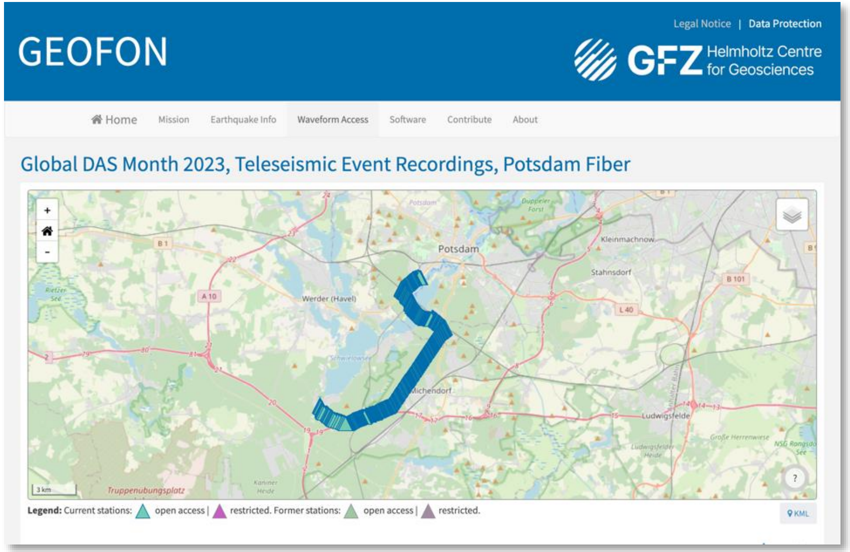


EIDA Guidelines for the creation of derived products from raw DAS data

Version	Collaborators	
1.0	Javier Quinteros	Angelo Strollo
	Frederick Massin	Susanne Hemmleb
	Christopher Wollin	Veronica Rodriguez
Date	Pascal Edme	John Clinton
5th Nov 2024	Philippe Kaestli	Christos Evangelidis
	Gilda Currenti	Michelle Prestifilippo
	Peter Danecek	Jonathan Schaeffer
	Diane Rivet	Shane Murphy
	Jan Michalek	and others...

https://orfeus.readthedocs.io/en/latest/das_guidelines.html

Archiving consistently through EIDA



https://dx.doi.org/10.5880/GFZ.2.2.2023.001

Cite as Wollin, Christopher; Ehsaninezhad, Leila; Hart, Johannes; Rodríguez Tribaldos, Verónica; Krawczyk, Charlotte M. (2023): Global DAS Month 2023, Teleseismic Event Recordings, Potsdam Fiber. GFZ Data Services. Dataset/DAS Data. doi:10.5880/GFZ.2.2.2023.001.

Identifier 10.5880/GFZ.2.2.2023.001

FDSN network code 3U

DataCite metadata HTML | JSON | XML | INSPIRE | Local XML source file

Terms/rights Creative Commons Attribution 4.0 International Available since 2023-04-01

Creator(s) Wollin, Christopher [a]; Ehsaninezhad, Leila [a]; Hart, Johannes [a]; Rodríguez Tribaldos, Verónica [a]; Krawczyk, Charlotte M. [a]

[a] GFZ German Research Centre for Geosciences, Potsdam, Germany

Description (Abstract) The here referenced dataset provides event-based Distributed Acoustic Sensing (DAS) recordings made with an approximately 22 km long dark telecommunication fiber lying in urban Potsdam and surroundings. For each of 164 M>=5 earthquakes occurring in February 2023 and listed by the USGS, one hour of data is provided starting with the event's origin time. Additionally, the whole day of February 14 is provided in hourly files. The data was recorded in the frame of the global DAS month, an initiative to collaboratively record and share simultaneously recorded DAS data from all over the world (https://www.norsar.no/in-focus/global-das-monitoring-month-february-2023). DAS is an emerging technology increasingly used by seismologists to convert kilometer long optical fibers into seismic sensors.

Title Global DAS Month 2023, Teleseismic Event Recordings, Potsdam Fiber

Publisher GFZ Data Services

Publication Year 2023

Resource Type Dataset / DAS Data

Dates Collected 2023-02-01/2023-03-01
Accepted 2023-03-01
Created 2023-05-25
Available 2023-04-01

Place(s) Study area southwest of Potsdam (Germany)

Contributor(s) Project manager: Wollin, Christopher (GFZ)
Project members: Ehsaninezhad, Leila (GFZ); Hart, Johannes (GFZ); Rodríguez Tribaldos, Verónica (GFZ)
Project leader: Krawczyk, Charlotte M. (GFZ)
Contact person: Wollin, Christopher (GFZ)
Hosting institution: Deutsches GeoForschungsZentrum GFZ
Data manager: GEOFON Data Centre
Sponsor: Geo-Inquire

Subjects SeisData terms: Array DAS Geophysics MiniSEED Strain Temporary
GCMD keywords: Solid earth Geophysical stations/networks
Other terms: Monitoring system; fibre optics

Funding Reference(s) 1. Helmholtz-Zentrum Potsdam - Deutsches GeoForschungsZentrum GFZ

Language en

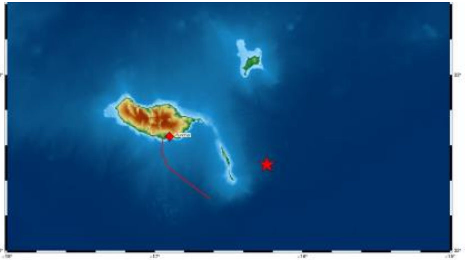
Sizes 78 GB

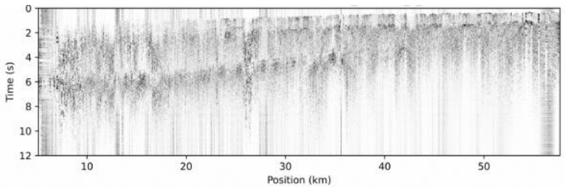
Reference(s)
1. Object Storage with raw data and documentation. WWW
2. Wuestefeld, A.; Spica, Z. J.; Aderhold, K.; Huang, H.; Ma, K.; Lai, V. H.; Miller, M.; Urmantseva, L.; Zapf, D.; Bowden, D. C.; Edme, P.; Kiers, T.; Rinaldi, A. P.; Tuinstra, K.; Jestin, C.; Diaz-Meza, S.; Jousset, P.; Wollin, C.; Ugalde, A.; Ruiz Barajas, S.; Gaite, B.; Currenti, G.; Prestifilippo, M.; and Araki, E.; Tonegawa, T.; de Ridder, S.; and Nowacki, A.; and Lindner, F.; and Schoenball, M.; Wetter, C.; Zhu, H.; Baird, A. F.; Rørstadbotnen, R. A.; Ajo-Franklin, J.; Ma, Y.; Abbott, R. E.; Hodgkinson, K. M.; Porritt, R. W.; Stanciu, C.; Podrasky, A.; Hill, D.; Biondi, B.; Yuan, S.; Luo, B.; Nikitin, S.; Morten, J. P.; Dumitru, V.; Lienhart, W.; Cunningham, E.; Wang, H. (2023). The Global DAS Month of February 2023. Seismological Research Letters. doi:10.1785/0220230180 DOI
3. GLOBUS Server (DAS): folder overview with raw data (registration required). WWW

GFZ Helmholtz Centre for Geosciences

Example of DAS data GeoLab (restricted)

Afonso Loureiro (2024): GeoLab. GFZ Data Services. Dataset/Seismic Network. [doi:10.14470/8K802502](https://doi.org/10.14470/8K802502)

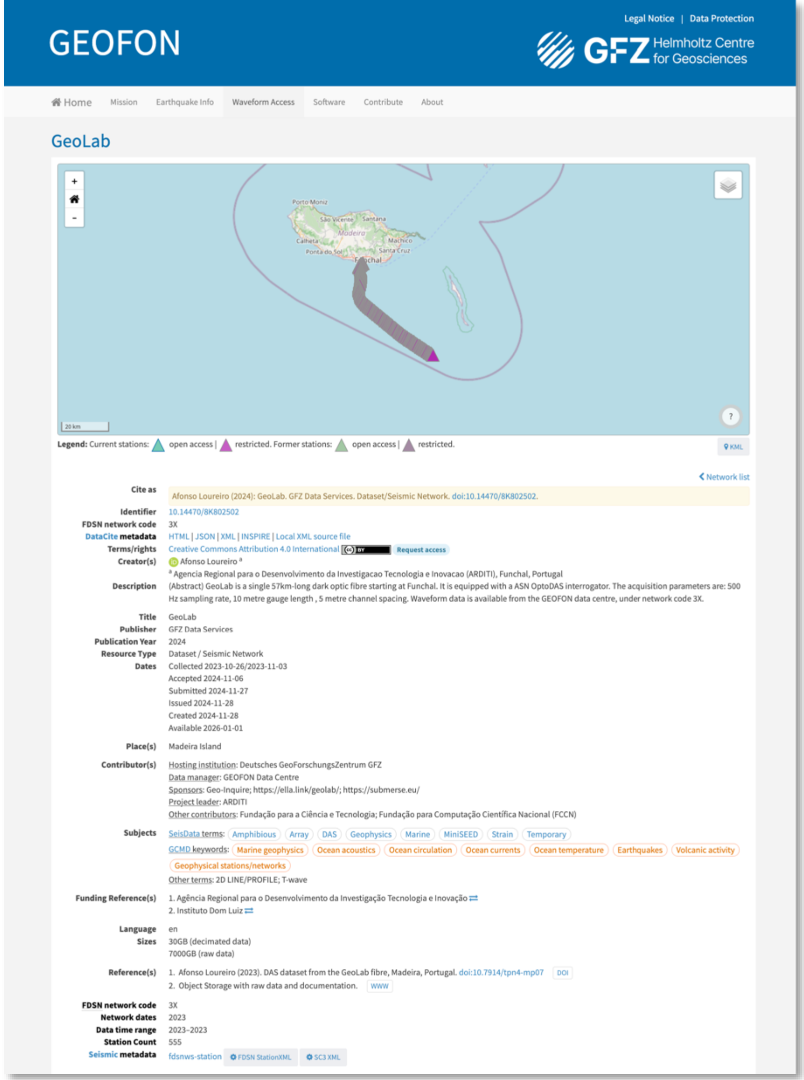




GFZ Helmholtz Centre for Geosciences

EIDA Guidelines for the creation of derived products from raw DAS data

Version	Collaborators	
1.0	Javier Quinteros	Angelo Strollo
	Frederick Massin	Susanne Hemmleb
	Christopher Wollin	Veronica Rodriguez
Date	Pascal Edme	John Clinton
5th Nov 2024	Philippe Kaestli	Christos Evangelidis
	Gilda Currenti	Michelle Prestifilippo
	Peter Danecek	Jonathan Schaeffer
	Diane Rivet	Shane Murphy
	Jan Michalek	and others...



Data formats

Proprietary formats

- TDMS (Silixa)
- HDF5 (OptoDAS, Silixa v2, others)

Community:

- Seg-Y (some manufacturers)

Other solutions:

- Ad-hoc user-tailored formats (usually HDF5-based)
- miniSEED

Candidates to be the next ‘de-facto’ (?) standard:

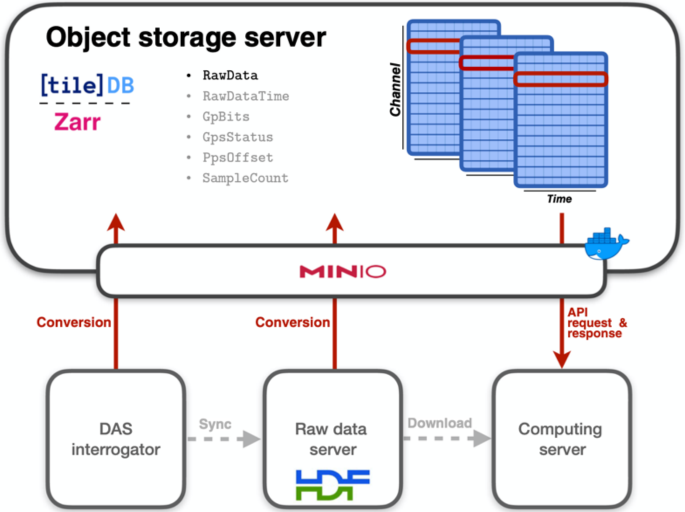
- Something based on HDF5 ❌
  - Known in the community
  - Not well suited for multithread/multiprocess
- Zarr ✅
  - RW multithread/multiprocess
  - Cloud is supported
- TileDB ✅
  - Full multi-threaded implementation
  - Different storage solutions supported natively
  - Versioning

GFZ Helmholtz Centre for Geosciences

Geo-INQUIRE Orfeus EarthScope Consortium

Data formats

- HDF5 can dramatically slow down DAS data sharing in modern distributed computing environments
- Ni et al. show a promising proof-of-concept based on TileDB, S3 and MinIO
- TileDB natively parallelizes I/O
- Parallelize I/O with MPI for TileDB scales very well up to 16 concurrent processes

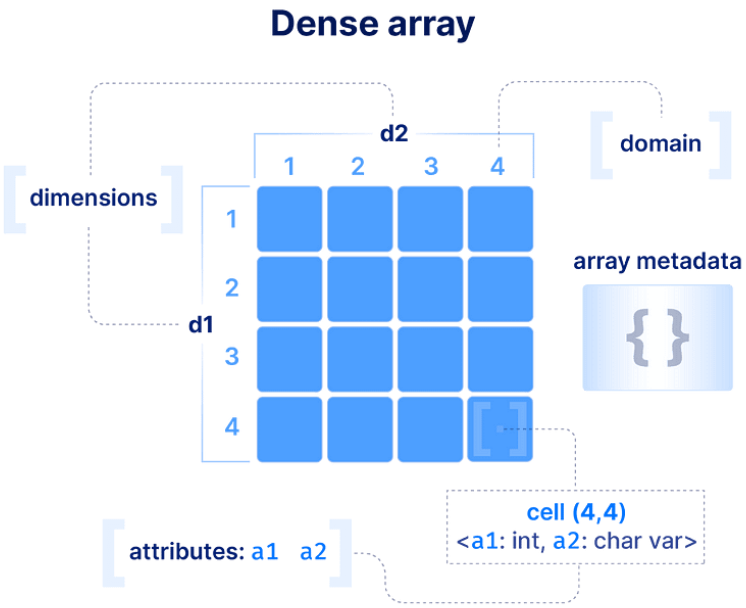


Ni, Y. et al. (2023) SRL, doi:10.1785/0220230172

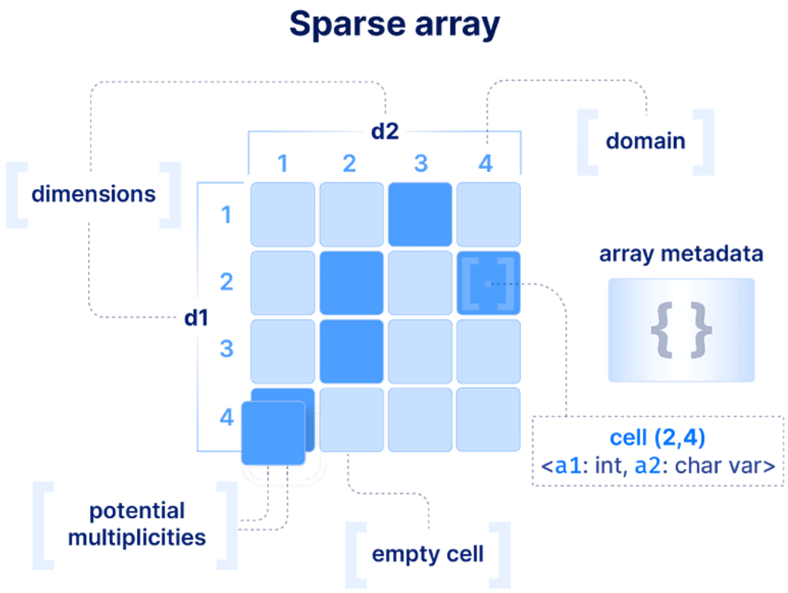
TileDB

- Support for both dense and sparse arrays
- Support for dataframes and key-value stores
- Optimized for object stores (AWS S3, Google Cloud Storage, Azure Blob Storage)
- Chunked (tiled) arrays
- Tiling and compression
- Parallel IO
- Data versioning (rapid updates, time traveling)
- Groups
- Arbitrary metadata
- APIs from most typical programming languages

TileDB



Dense array
d2
1 2 3 4
dimensions
d1
1 2 3 4
domain
array metadata
{ }
cell (4,4)
<a1: int, a2: char var>
attributes: a1 a2

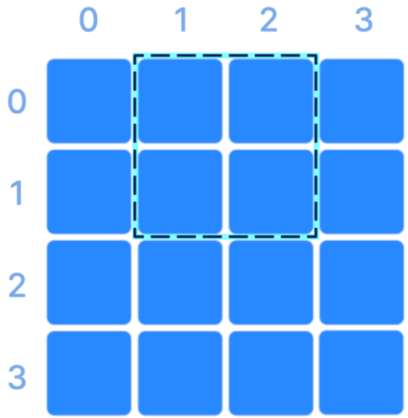


Sparse array
d2
1 2 3 4
dimensions
d1
1 2 3 4
domain
array metadata
{ }
cell (2,4)
<a1: int, a2: char var>
potential multiplicities
empty cell

TileDB

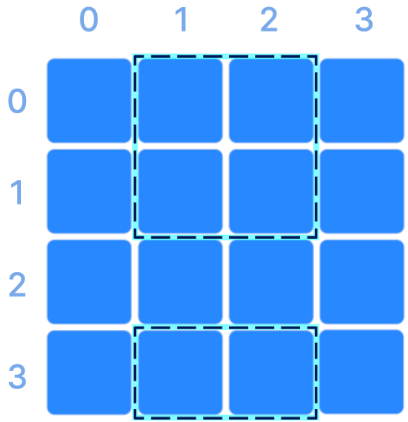
Slicing

```
A [0:2, 1:3]

SELECT attr FROM A
WHERE d1>=0 AND d1<=1 AND
      d2>=1 AND d2<=2
```

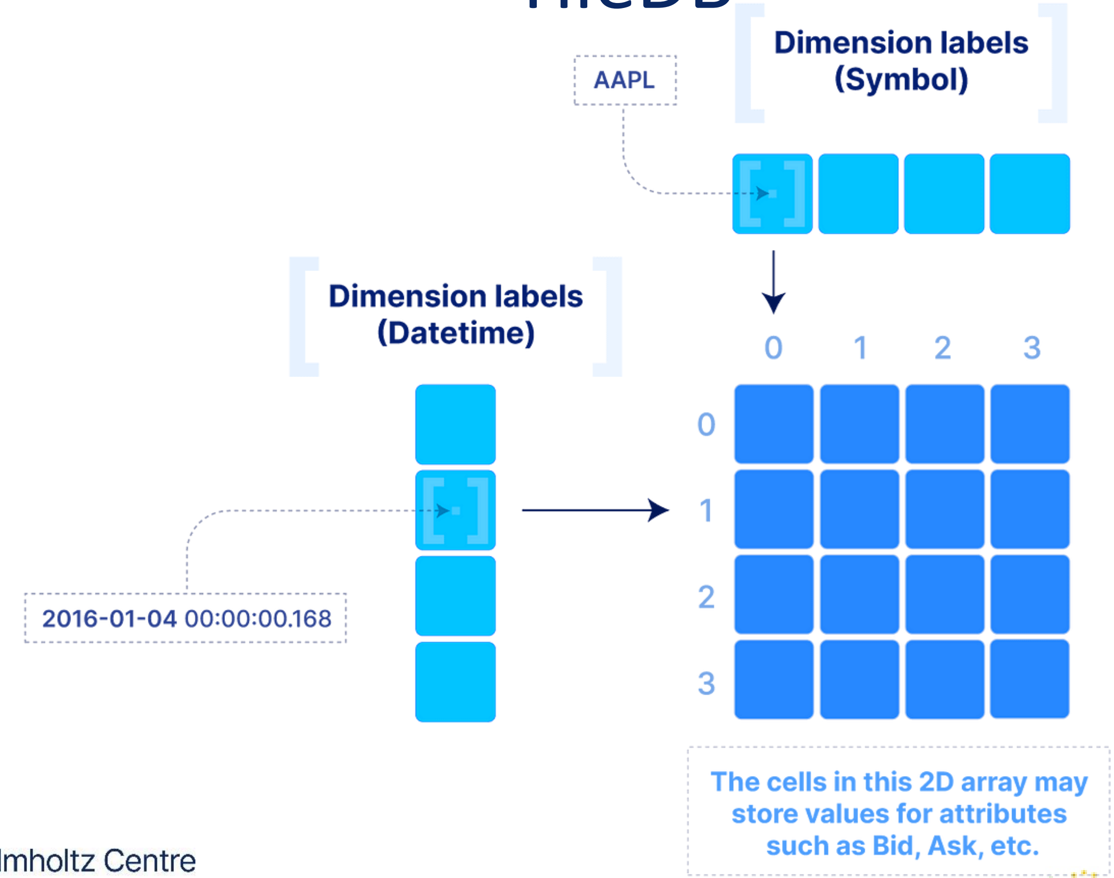


Multi-range Slicing

```
A [[0,1,3], 1:3]

SELECT attr FROM A
WHERE ((d1>=0 AND d1<=1) OR
        d1==3) AND
      d2>=1 AND d2<=2
```



TileDB



GFZ Helmholtz Centre for Geosciences

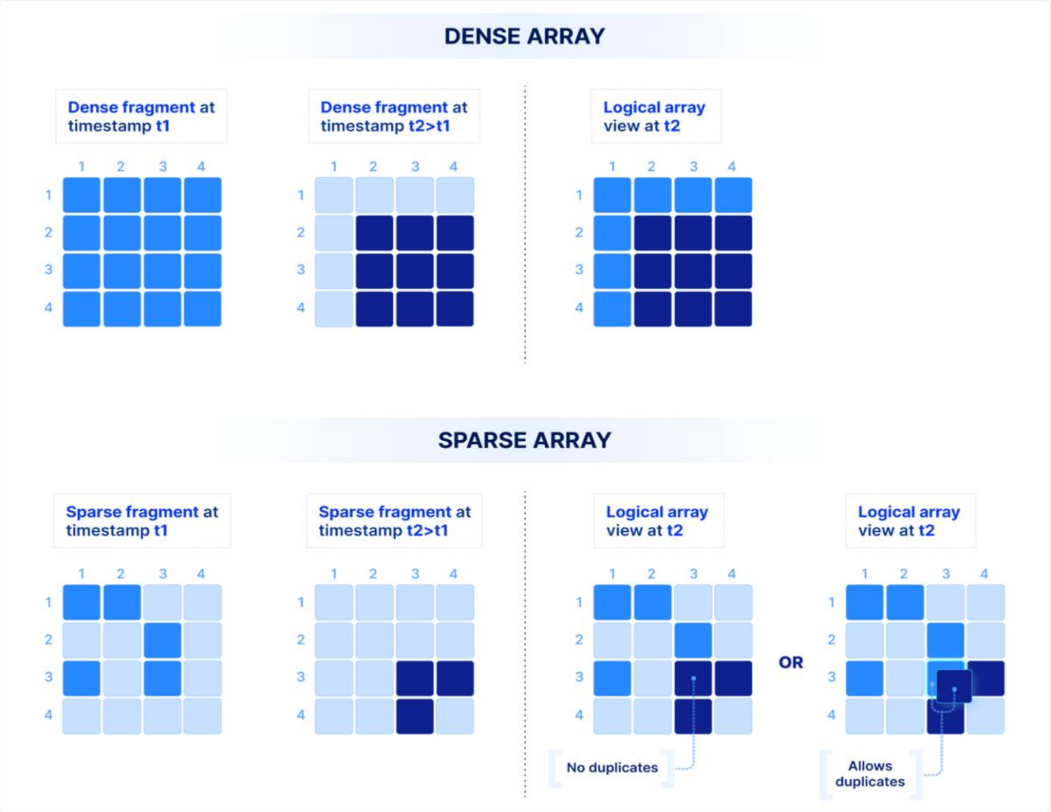
Geo-INQUIRE Orfeus EarthScope Consortium

TileDB

Copy

```
my_array                              # array directory
  ├── ...
  ├── __fragments
  │     └── <timestamped_name>        # fragment directory
  │           ├── __fragment_metadata.tdb  # fragment metadata
  │           ├── a0.tdb              # fixed-sized attribute
  │           ├── a1.tdb              # var-sized attribute (offsets)
  │           ├── a1_var.tdb          # var-sized attribute (values)
  │           ├── ...
  │           ├── a2_validity.tdb     # validity of fixed- or var-sized attribute
  │           ├── ...
  │           ├── d0.tdb              # fixed-sized dimension
  │           ├── d1.tdb              # var-sized dimension (offsets)
  │           ├── d1_var.tdb          # var-sized dimension (values)
  │           └── ...
  └── ...
```

TileDB


DENSE ARRAY
Dense fragment at timestamp t1
Dense fragment at timestamp t2>t1
Logical array view at t2
SPARSE ARRAY
Sparse fragment at timestamp t1
Sparse fragment at timestamp t2>t1
Logical array view at t2
Logical array view at t2
OR
No duplicates
Allows duplicates

Data distribution (PubDAS)

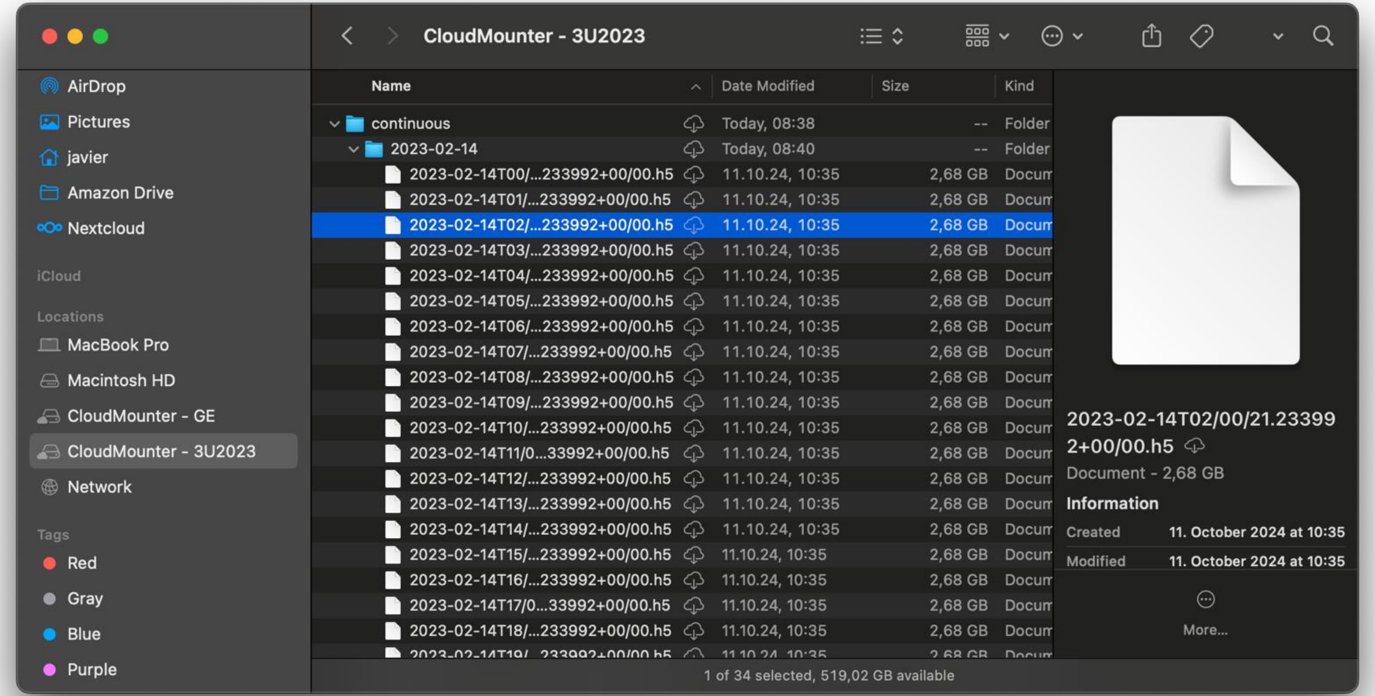
PubDAS was the first rough approach using Globus.

Name	IU	T. span (d)	Format	Sps (hz)	Vol. (Gb)	GL (m)	CL (m)	CS (m)	units
Fairbanks	iDAS	$59^{\star}$	TDSM	1,000	10,441	10	4,000	1	$\dot{\epsilon}$
FORESEE	iDAS-v2	365	HDF5	$125^{\div}$	29,338	10	4,900	2	$\dot{\epsilon}$
FOSSA	iDAS-v2	7	TDSM	500	11,680	10	23,300	2	$\dot{\epsilon}$
LaFarge	iDAS	$2^{\star}$	SEG-Y	1,000	45	10	1,120	1	$\dot{\epsilon}$
Stanford-1	ODH3	940	SEG-Y	50	18,908	7.14	2,500	8.16	ϵ
Stanford-2	ODH3	14	SEG-Y	250	2,887	20	10,200	8.16	ϵ
Stanford-3	ODH4	6	SEG-Y	$\sim$	92	$\sim$	2,500	8.16	ϵ
Valencia	A1-R	7	HDF5	$250^{\div}$	3,213	30.4	50,000	16.8	$\dot{\epsilon}$

Table 1. List of the data sets currently available on PubDAS and their main characteristics. IU: Interrogator Unit; T. Span: Time span in days; Sps: Samples Per Second in Hertz; Vol.: Volume in Gigabytes; GL: Gauge Length in meters; CL: Cable Length in meters; CS: Channel Spacing in meters; $\dot{\epsilon}$: strain rate; ϵ: strain; A $\star$ means data contain active sources. A $\sim$ means that this value may vary; $^{\div}$: means the dataset is downsampled. Name abbreviations are the same as in Fig. 1.

Ni, Y. et al. (2023) SRL, doi:10.1785/0220230172

S3 Bucket with DAS data and metadata



Endpoint: https://s3.gfz-potsdam.de/ Bucket: gc.3u2023 Size: 468 GB. Objects: 195

GFZ Helmholtz Centre for Geosciences

AWS client

- The AWS command line client can be installed by the Python pip utility.
- It requires Python 3.9 or a more recent version.
- To do that execute the following command (check if you can do that as a normal user or if you need sudo):

```
$ python3 -m pip install awscli
```

Data distribution

List the objects:

```
$ aws --no-sign-request --endpoint-url https://s3.gfz-potsdam.de s3 ls s3://gc.3u2023
                           PRE continuous/
                           PRE event_based/
2025-02-24 11:47:40     383235 3u2023.json
2025-02-24 13:40:32        752 README.txt
...
```

Synchronize the bucket:

```
$ aws --no-sign-request --endpoint-url https://s3.gfz-potsdam.de s3 sync s3://bucketname/ .
```

Data distribution

- S3 buckets as the first step.
- To provide the curated raw data (original format, could be non-standard).
- For instance, one (or more) bucket for a full resolution dataset.
- Allow sync/replication of buckets to your S3 provider if needed.
- Your institution to AWS, AWS to Google, and so on.
- We have an example you can try:
  - 3U-2023. Check our landing page of the network.
- One of the main objectives is to provide computation on top of the data.

File Storage System

- Storage of files is a traditional method of saving files.
- Organize data in structured form to make use of files and folders.
- Each file is stored with a specific name and location on a storage device.
- Group data into files and files structured in a directory hierarchy of folders and subfolders.

Object Storage System

- Data is stored independently in structures referred to as objects.
- For each object, it maintains data, metadata, and a unique identifier.
- Rank all objects directly in a flat format for ease of handling large quantities of unstructured data.
- Rich metadata for easier search and retrieval of particular data.
- Objects are identified by a unique ID. Users do not need to know where the object is stored in physical space.
- Access directly through an API.

Object Storage System

- The typical cloud storage is organized in buckets.
- Inside these buckets we store the objects.
- One can attach metadata to objects or buckets.

GFZ Helmholtz Centre for Geosciences

File vs Object Storage System

	File Storage	Object Storage
Data Structure	Hierarchical file and folder system	Individual objects in a flat structure
Metadata	Limited metadata	Rich metadata
Scalability	Problems as data grows	Infinite scalability for handling large datasets
Accesibility	Through network-based file systems	Through API, often used in cloud environments
Performance	Slow down with large workloads or many files	Optimized for large datasets Latency for small tasks

GFZ Helmholtz Centre for Geosciences

DAS data management activities in progress

Geo-INQUIRE

- development of **Data Management Policies**, software, services and tools for DAS data (including proposal, discussion and adoption of standards @FDSN)

SUBMERSE (**inheriting** part of the **Data Management Policies** from Geo-INQUIRE)

- **real-time streaming**
- data scrubbing/vetting
- definition of use cases (earthquakes, volcanoes, tsunamis, biology, etc);
- define and extract observables in real-time where possible

GFZ Helmholtz Centre for Geosciences

Data Management Policies for DAS and standardization in seismology

1 — 60% — Global community discussions: standardization

2 — 60% — Data and metadata formats for cloud and for integration

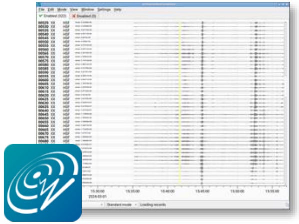
3 — 100% — First pilot datasets archived following common policies: DAS-RCN metadata format

4 — 50% — Prototyping and testing the S3 interface (cloud integration) with GFZ computing center (RZ)

5 — 100% — Real-time streaming of DAS data to SeisComP®

6 — 10% — Testing scrubbing algorithms based on actual use cases

n — Adding scrubbing and other custom filters in real-time possibly on-site



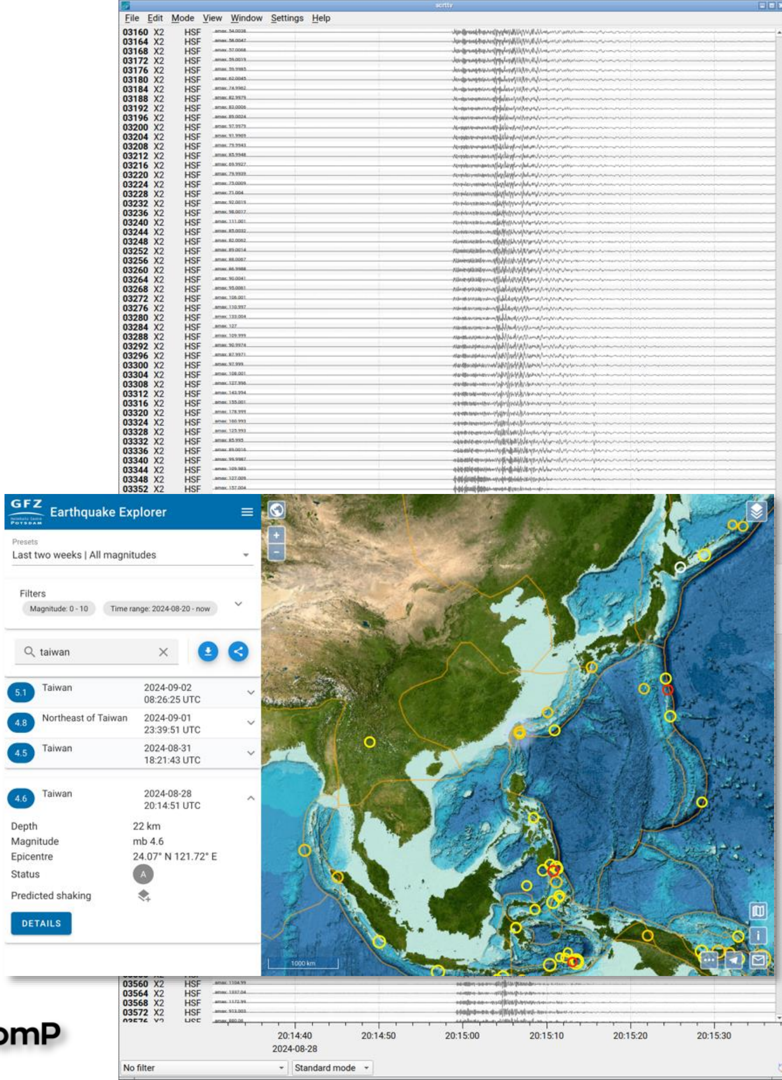
SeedLink in action with **OptoDAS**

Tested streaming with data from Black Forest Observatory and GFZ campus in Germany (being used in Portugal-Madeira, Preveza-Greece).

Acquired decimated data from Taiwan with SeedLink plugin and server running on the OptoDAS on site (allows to recover gaps)

- Original sample rate 12500 Hz => decimated by 125 to get 100 Hz via SeedLink
- Total number of channels: 7253 ~30 km fiber
- SeedLink channels: 1464 (spacing 4) ~9 GB/day

https://www.seiscomp.de/
https://www.seiscomp.de/doc/apps/seedlink.html#seedlink



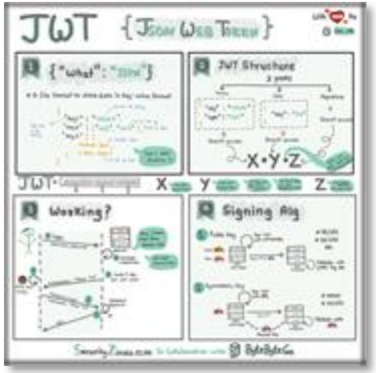
GFZ Helmholtz Centre for Geosciences

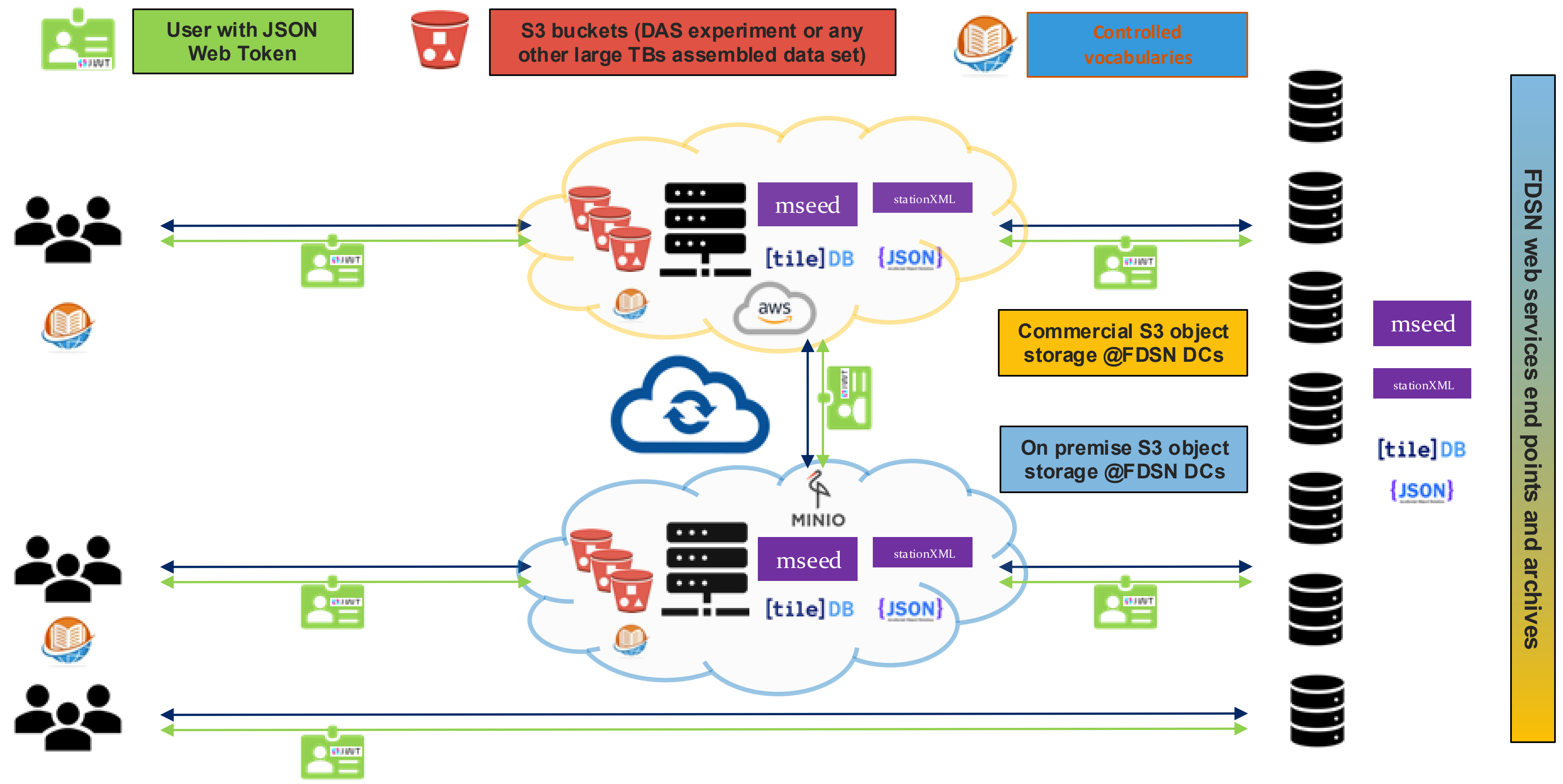
ALCATEL SUBMARINE NETWORKS

SeisComP

Next FDSN standard for AAI

- Adoption of common AAI approach within FDSN federated data centres to seamlessly identify users based on **Json Web Tokens** (https://jwt.io/);
- Approach usable by all FDSN standard webservices, including seedlink (real-time streaming protocol
- EarthScope and EIDA-ORFEUS already using a token based system AAI for the waveform service (for ORFEUS powered by Juelich/GFZ)
- Proposal for standardization @FDSN being finalized



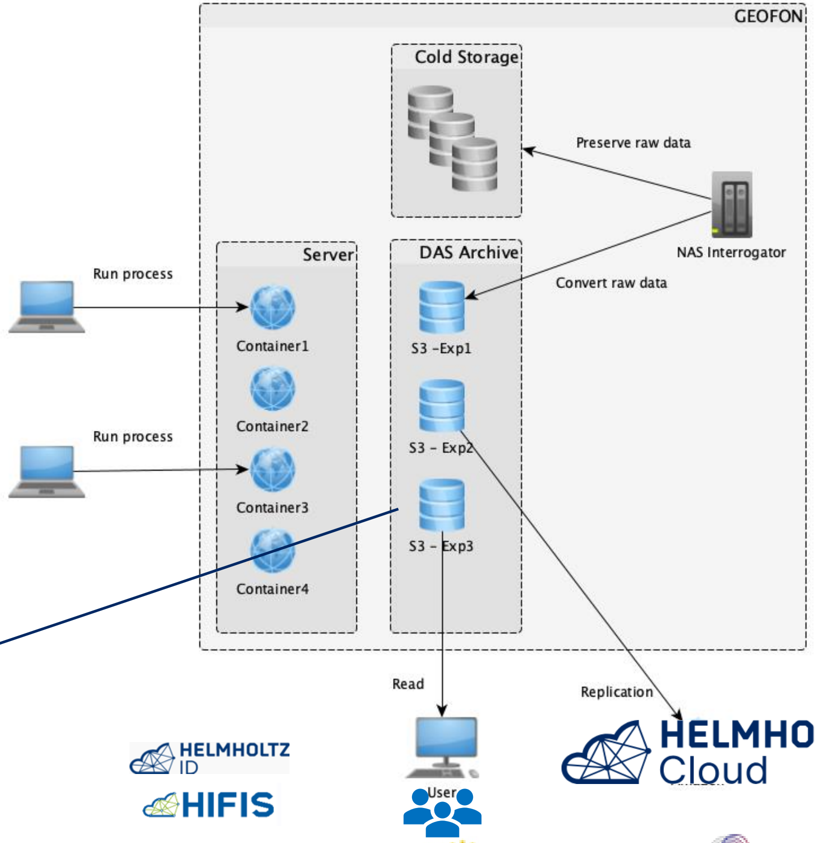


GFZ Helmholtz Centre for Geosciences

Geo-INQUIRE Orfeus EarthScope Consortium

Schematic view of a data centre

- Raw data preserved in cold storage (tapes?).
- Full resolution data staged on S3 buckets.
- Open to users and ready to synchronize with other providers.
- Some users could run their codes in containers with fast (local) access to data.



Outlook

- Expecting DAS metadata proposal becoming a standard in 2025
- Open science to co-exist with access control => Common global AAI solution for federated FDSN data centres needed!
- Cable operators, data owners and national authorities should agree on a policy document before they come to us to manage their data
- Possibility to restrict access to metadata (latitude/longitude)
- Proposal for new data format and provisioning via S3 buckets
- Access control on top of S3
- SAFATOR building on top of this towards provision of world class services for DAS, similar data types and use cases

GFZ Helmholtz Centre for Geosciences

Summary

- Different lines to work on:
  - Metadata format
  - Data format
  - Data provisioning
  - Real-time transmission
- Discussion within FDSN about the ”easy” topics:
  - Metadata format
  - Mapping to StationXML. Both will coexist for some time.
  - Channel Naming (Source Identifier?)

GFZ Helmholtz Centre for Geosciences

Summary

- More complex issues:
  - Standard format for TileDB/Zarr? Or standard interface (API)? This is critical for ML!
  - Abandon the synchronous behavior?
  - Real-time transmission (no Miniseed?)
  - Data provisioning via S3 buckets
  - Computation on top-of-the-data
- If the approach succeeds we can expect a slow migration of standard seismic data to this new solution.

dasmetadata in dastools

- Basic operations to create DAS metadata under different conditions.

```
$ dasmetadata --help
Usage: dasmetadata [OPTIONS] COMMAND [ARGS]...

Commands:
  add-coords       Complete coordinates in the DAS metadata from a CSV or...
  add-datacite     Complete the DAS metadata with information from...
  create           Create DAS metadata or StationXML for a DAS dataset
  json2datacite    Convert JSON metadata to Datacite
  json2stationxml  Convert JSON metadata to StationXML
  showraw          Show raw metadata from the headers of the raw data
```

GFZ Helmholtz Centre for Geosciences

dasmetadata in dastools

```
$ dasmetadata create --help
Usage: dasmetadata create [OPTIONS]

  Create DAS metadata or StationXML for a DAS dataset

Options:
  --experiment TEXT               Experiment to read and process
  --directory TEXT                Directory where files are located (default: ".")
  --start TEXT                    Start of the selected time window
  --end TEXT                      End of the selected time window
  --inputfmt [OptoDAS|TDMS]       Format of the data files
  --outfile FILENAME              Filename to save the output
  --outputfmt [json|stationxml]   Format of the output
  --empty                         Create empty metadata in standard format
```

GFZ Helmholtz Centre for Geosciences

dasmetadata in dastools

```
$ dasmetadata showraw --help
Usage: dasmetadata showraw [OPTIONS]

  Show raw metadata from the headers of the raw data

Options:
  --experiment TEXT          Experiment to read and process
  --directory DIRECTORY      Directory where files are located (default: ".")
  --start TEXT               Start of the selected time window
  --end TEXT                 End of the selected time window
  --inputfmt [OptoDAS|TDMS]  Format of the data files
  --outfile FILENAME         Filename to save the output
```

GFZ Helmholtz Centre for Geosciences

dasmetadata in dastools

```
$ dasmetadata add-datacite --help
Usage: dasmetadata add-datacite [OPTIONS] DOI

  Complete the DAS metadata with information from Datacite metadata via a DOI

Options:
  --infile FILENAME    Input file in JSON format as proposed by the DAS-RCN
                       group.
  --outfile FILENAME   File where the modified JSON metadata should be saved
  --help               Show this message and exit.
```

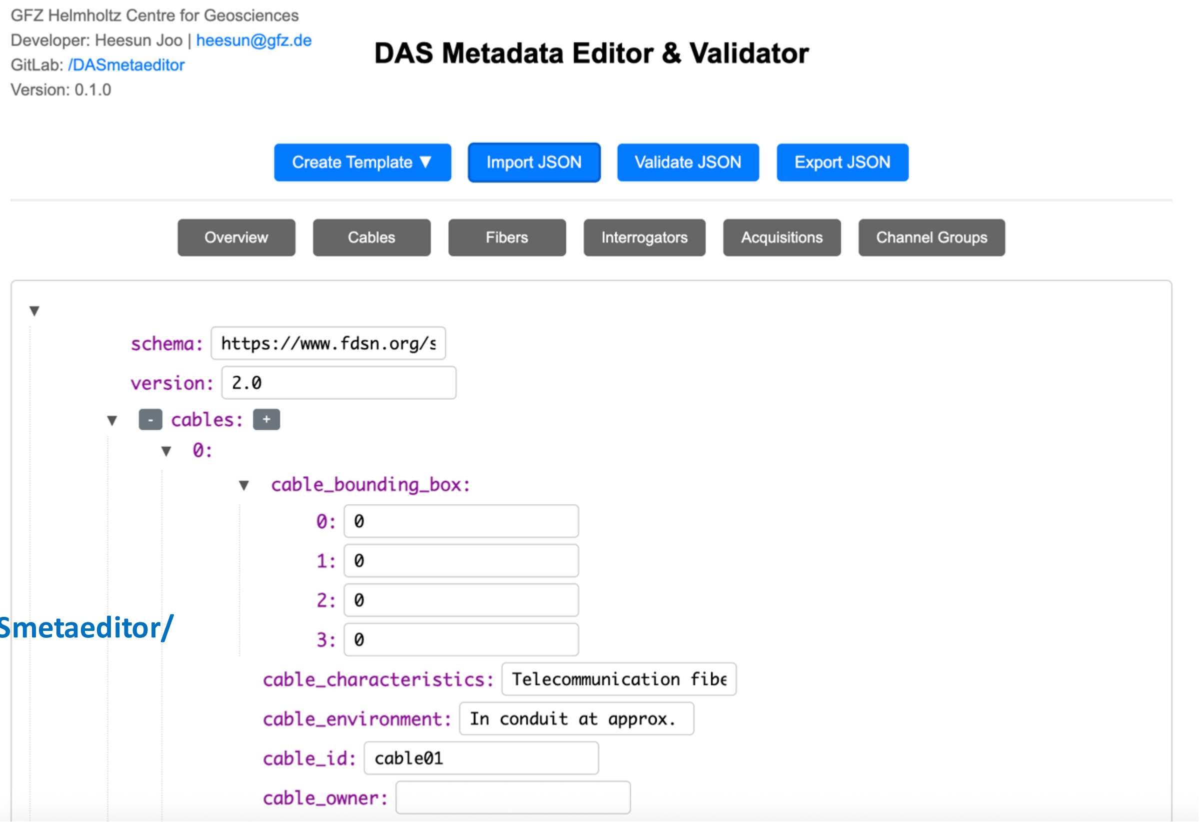
GFZ Helmholtz Centre for Geosciences

dasmetadata in dastools

```
$ dasmetadata add-coords --help
Usage: dasmetadata add-coords [OPTIONS] COORDS

  Complete coordinates in the DAS metadata from a CSV or KML file

  The expected structure of the CSV file is the following: 1) channelID as
  expected in the output metadata (e.g. A1234, B4321). No more than 5
  characters. 2) latitude as a float 3) longitude as a float 4) elevation in
  meters as a float (optional) 5) distance along fiber in meters as a float
  (optional)

Options:
  --infile FILENAME   Input file in JSON format as proposed by the DAS-RCN
                      group.
  --outfile FILENAME  File where the modified JSON metadata should be saved
```

DAS Metadata Editor

- Offline web page to edit DAS metadata
- You can import existing metadata, complete and validate it



https://git.gfz-potsdam.de/geofon/DASmetaeditor/

GFZ Helmholtz Centre for Geosciences

Curious to explore more?
Visit our joint booth (Stand #1) for Lunchtime presentations, Quiz games and little surprises…