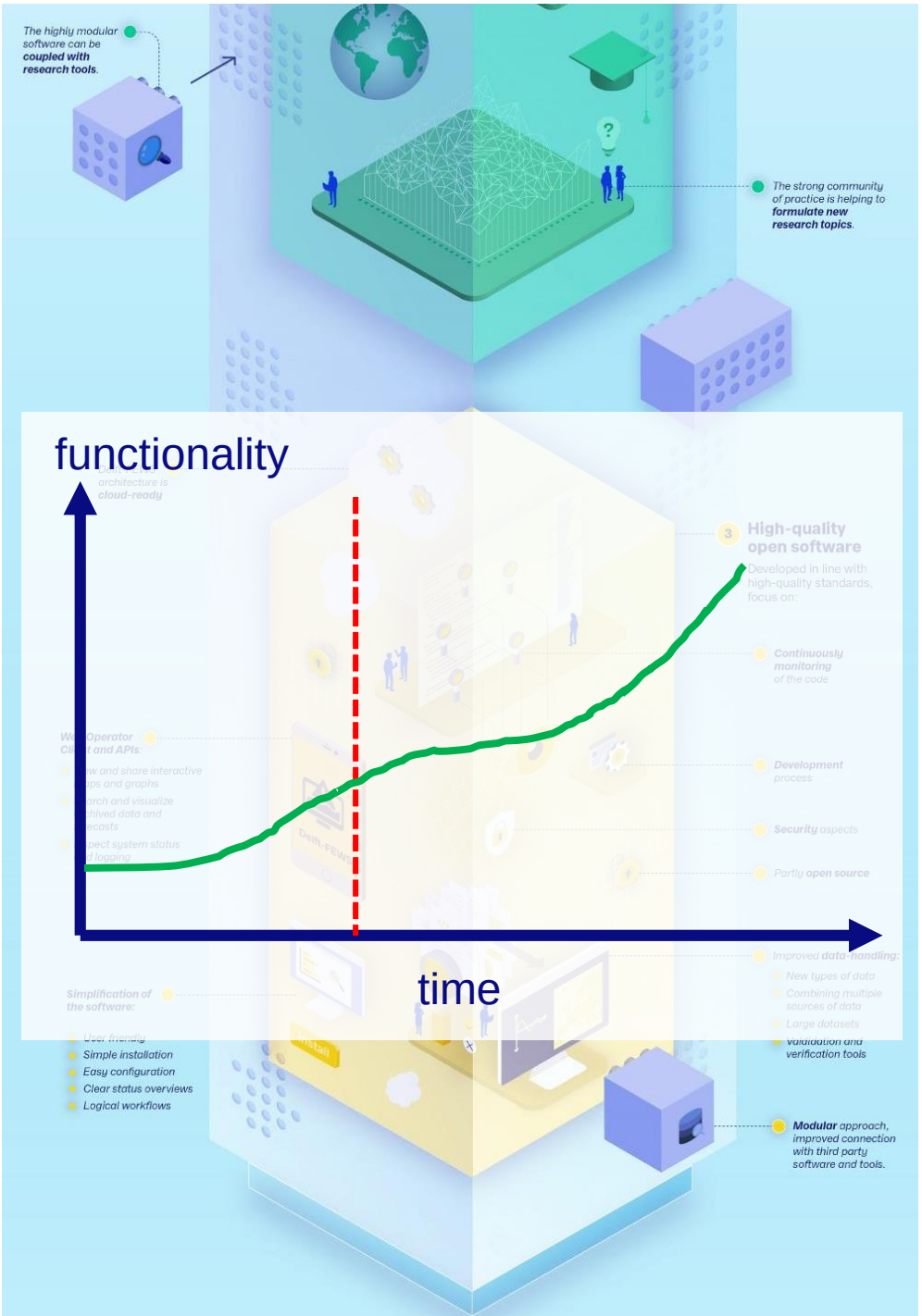


Deltares

Delft-FEWS and Open Archive

- past, present, and future -

Delft-FEWS International User Days

Maarten Smoorenburg, Gabriel Miller, Gerben Boot, Andre Grijze, Tom Bogaard, Peter Gijsbers, Marc van Dijk

…. and many, many more

10 November 2022

First, a quick inventory

Who of you stores FEWS results in an Open Archive?

… or has set up a FEWS system that does?

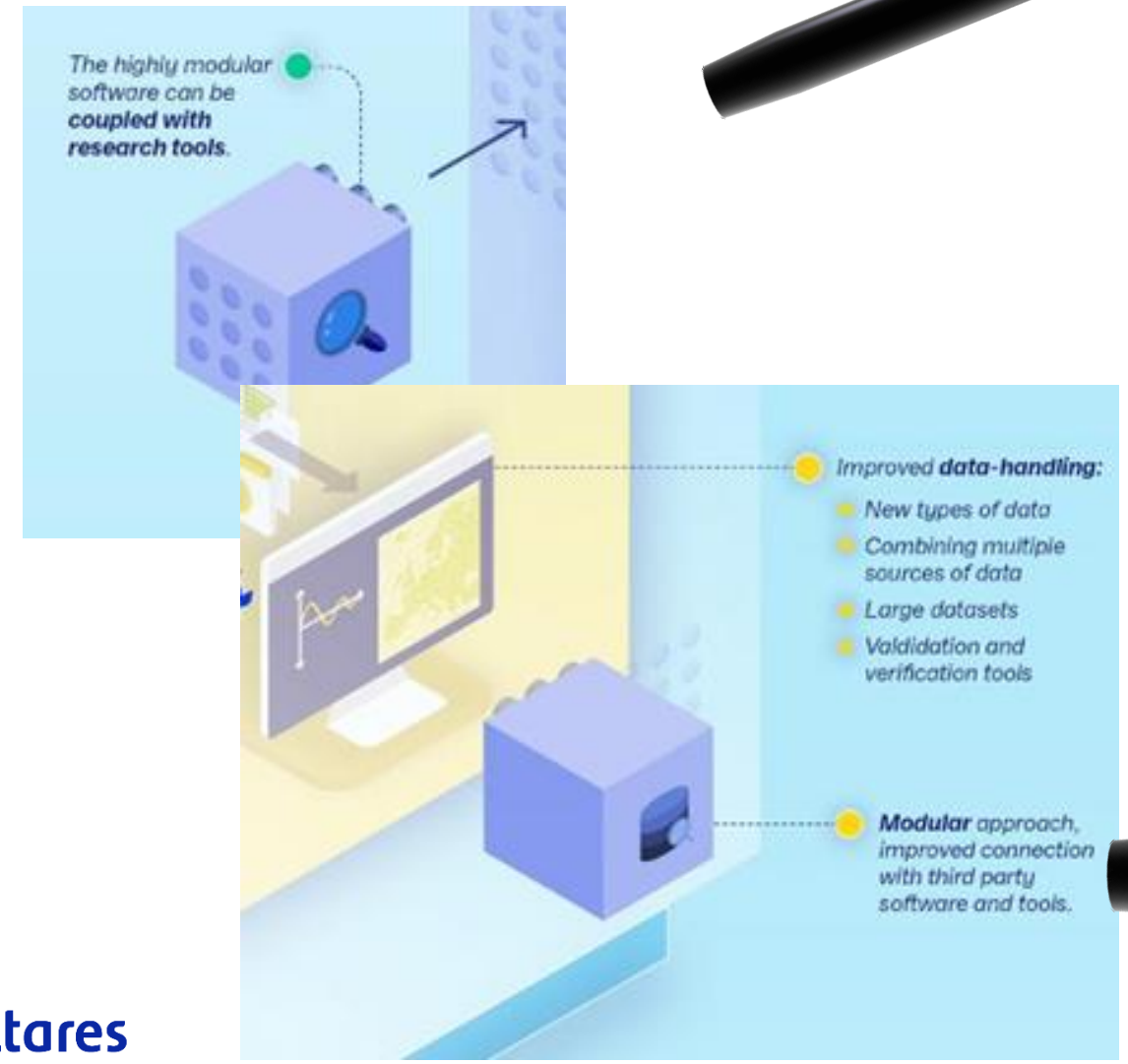
Who of you use webservices to disseminate data stored in the Open Archive?

Who of you stores <u>validated</u> measurement data in the Open Archive?

Contents

- Introduction – links to FEWS road map 2025
- FEWS Open Archive
  - Why?
  - What's in a name?
  - How? …some new developments
  - (Future)
- Open Archive Database with MongoDB (Gabriel Miller)
- Open Archive Vision/Roadmap ToC (Gerben Boot)

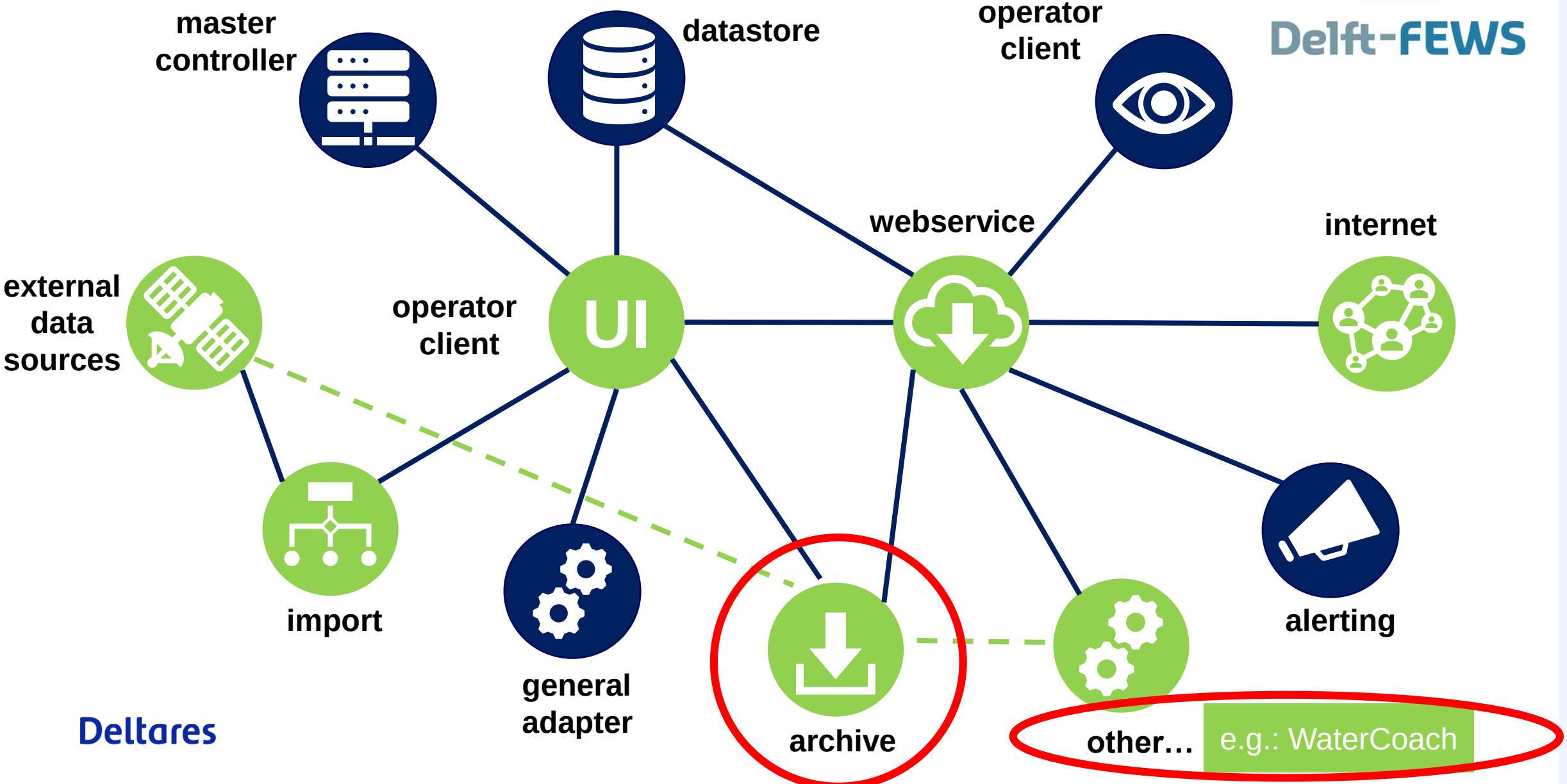
Open Archive - modularity



Deltares

Open Archive – interaction

Delft-FEWS *Overview*

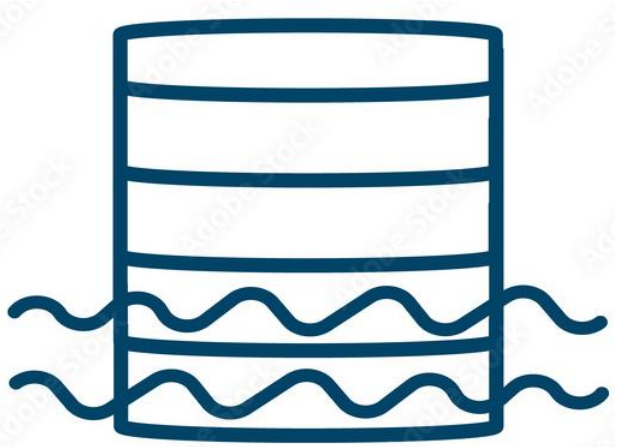


Deltares

Open Archive: What's in a name?

Archive (dusty?)

(Big) Data Lake
(or data graveyard?)



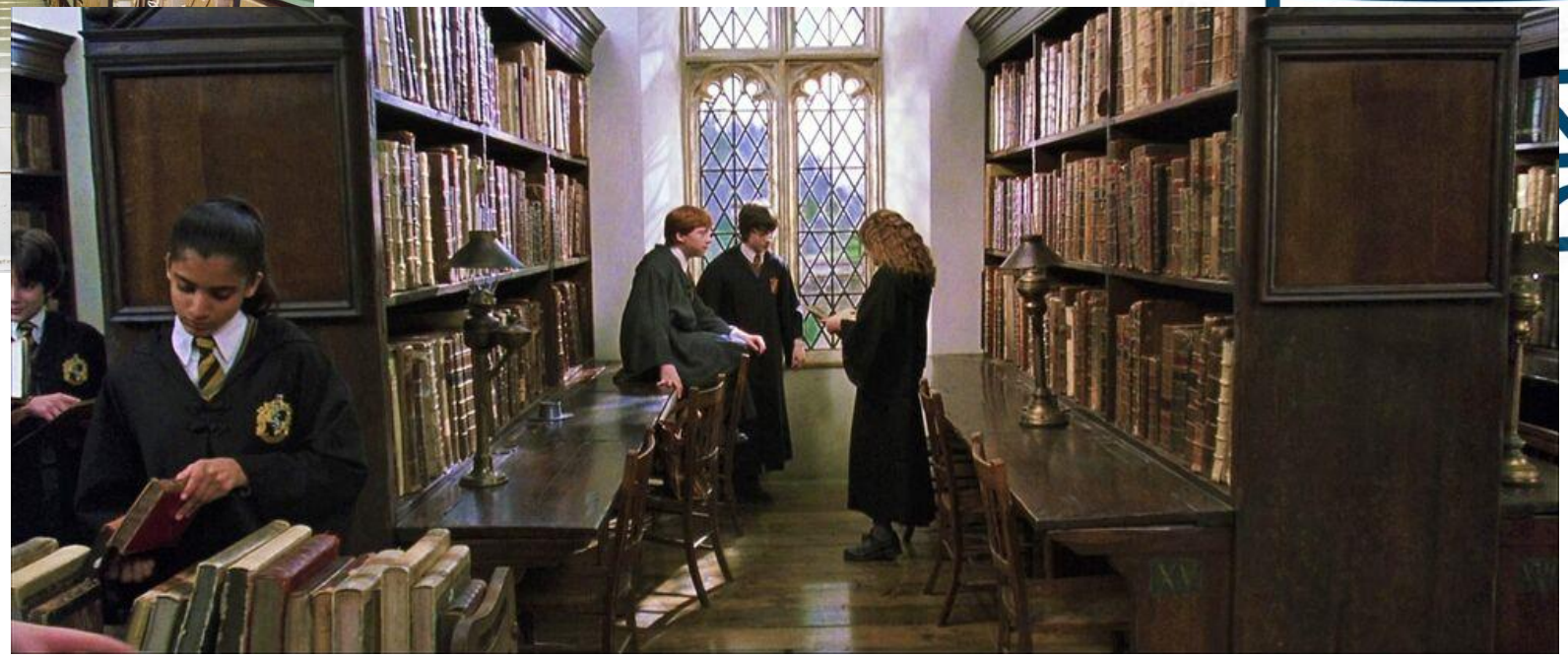
Deltares

Open Archive: What's in a name?

- **Archive (dusty)**

- **(Big) data lake (or graveyard?)**

Library with magic

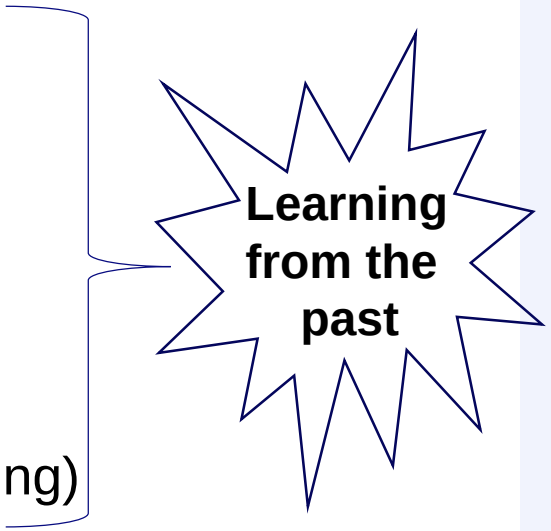


Deltares

The FEWS Open Archive – Why?

- **Accountability:** (raw) timeseries data, forecaster notes, database snapshots, etc.
- **Dissemination** to outside world → via FEWS webservices

- **Event analysis**
  - In OC by displaying historical floods a alongside current forecasts
  - Post-event analysis: FEWS Stand Alone or FEWS WaterCoach
- **Training** software and forecasting → FEWS WaterCoach
- **Model development** and scenario analysis → FEWS Comp. Framework
- **Evaluation** of models, forecasts and services
- **Forecast calibration** (statistical pre- and postprocessing, machine learning)



Open Archive: What's in a name?

Why is it called **Open** Archive?

- Started as FEWS's archive solution, filled by FEWS' exported files and **openly accessible** with FEWS' webservices ***(if desired)***
- Then already possible to **store different data types** *(timeseries, states, reports, configuration, forecaster notes, database snapshots, rating curves)*

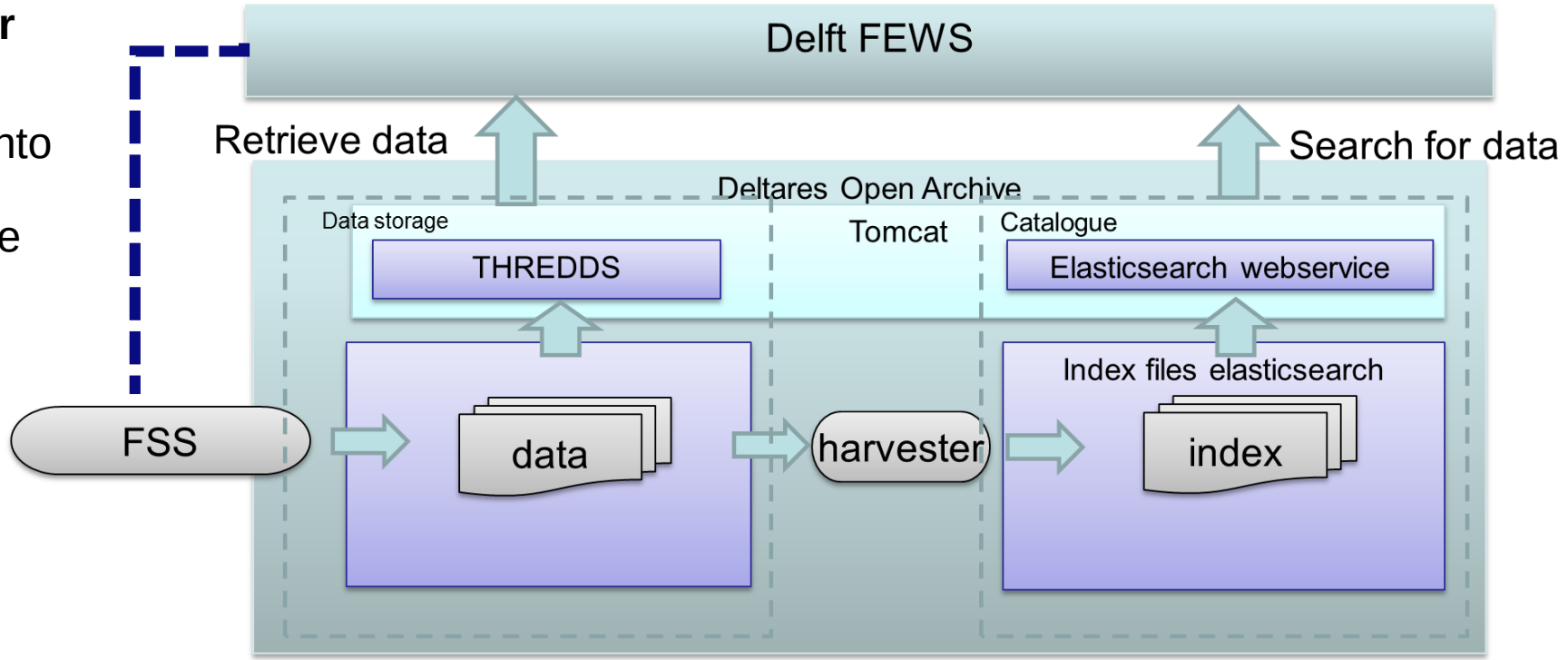
Now **opened up** more:

- catalogue can deal with **NetCDF files produced by other systems**
- connects also to **external archives via OpenDAP**
- **data can be archived** using the **webservice 'post' call**
- extended with **archive database(s) like MongoDB**

How? → Exporting

Past + Present

- **FEWS exports** files with **accompanying XML for metadata**
- Data can be **imported** into a FEWS Stand Alone or Operator Client using the **Archive Catalogue Display**



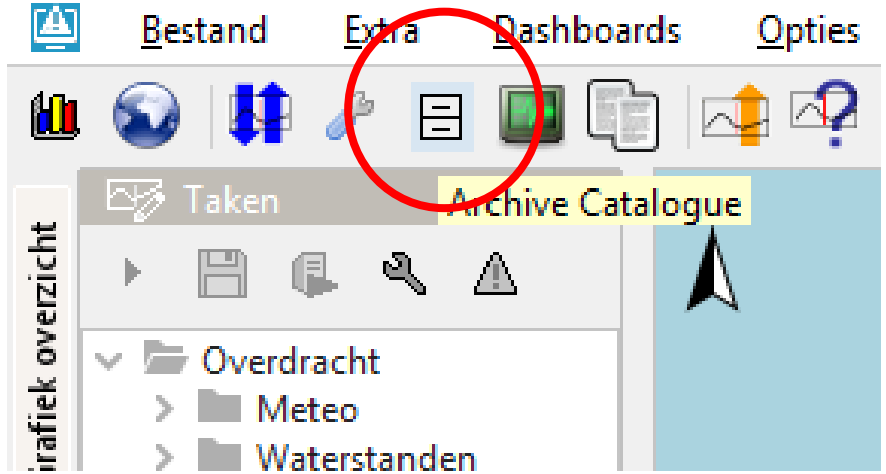
- Source: FEWS public wiki

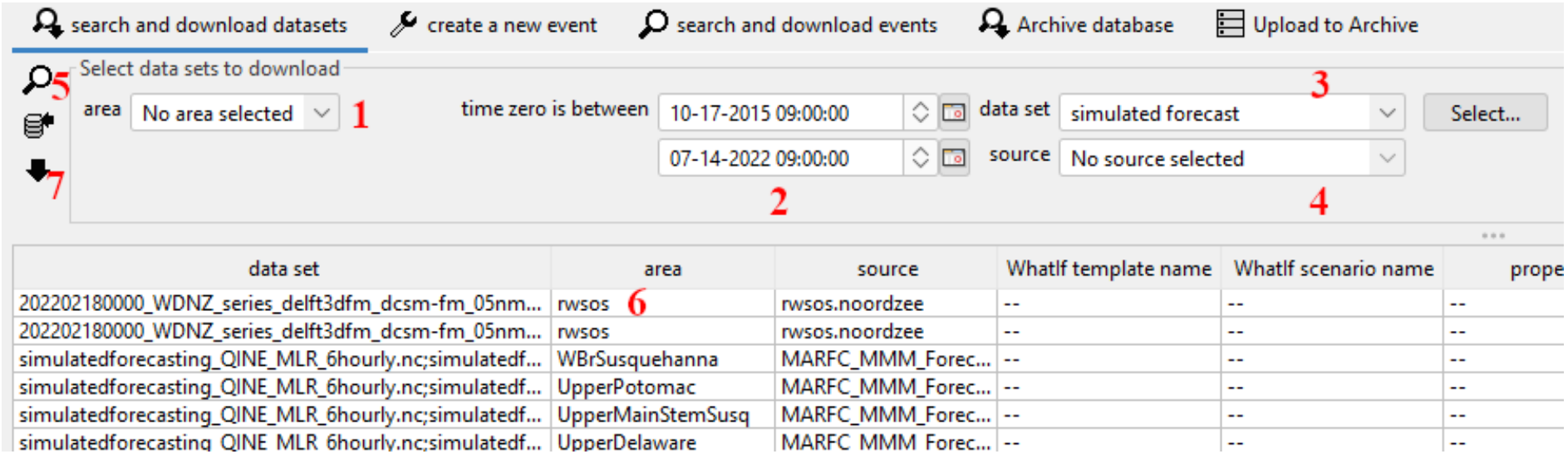
powered by open source:

How? → Importing back into FEWS

Past + Present

- **FEWS exports** files with **accompanying XML for metadata**
- Data can be **imported** into a FEWS Stand Alone or Operator Client using the Archive Catalogue Display

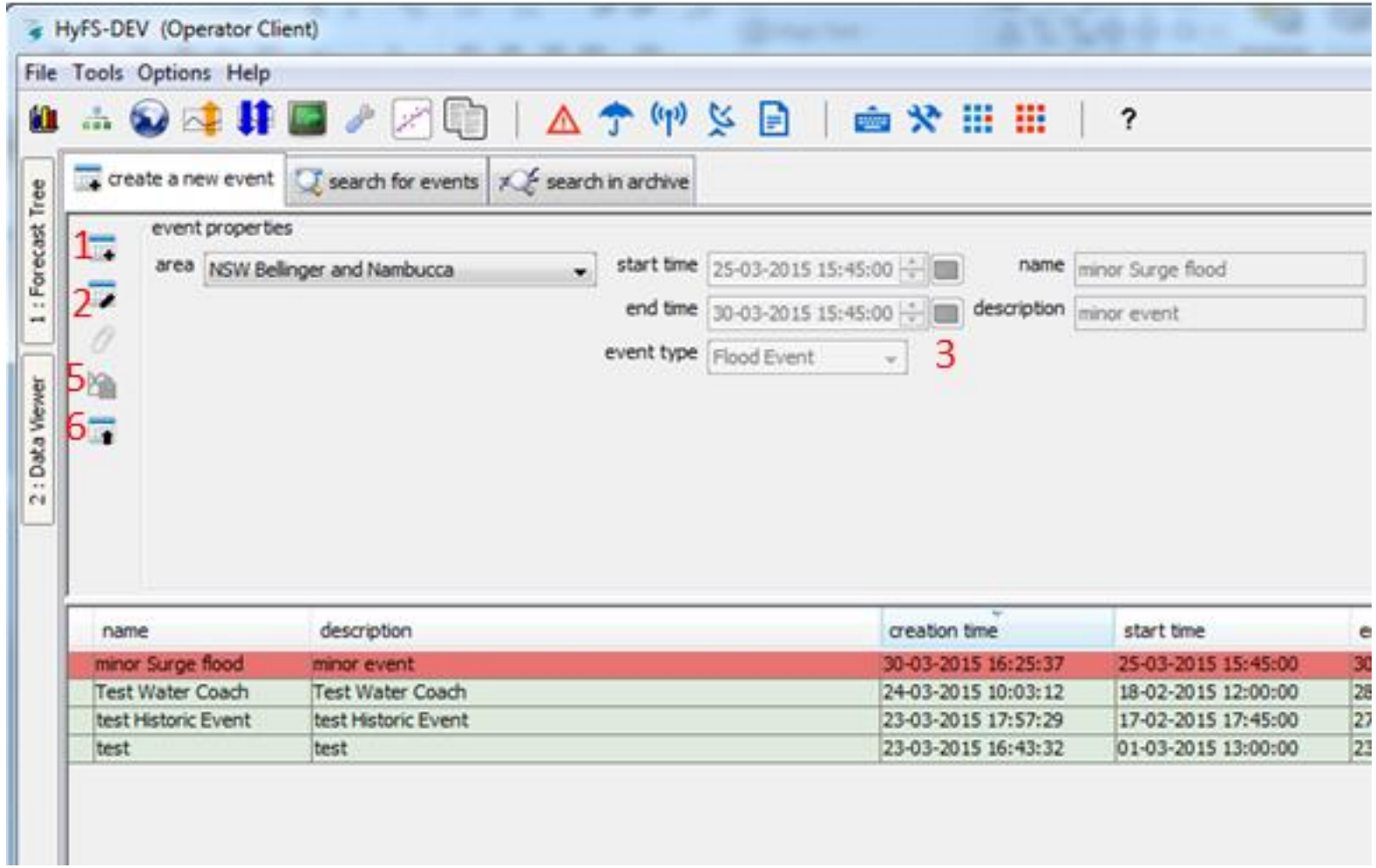




How? → Tagging events

Past + Present

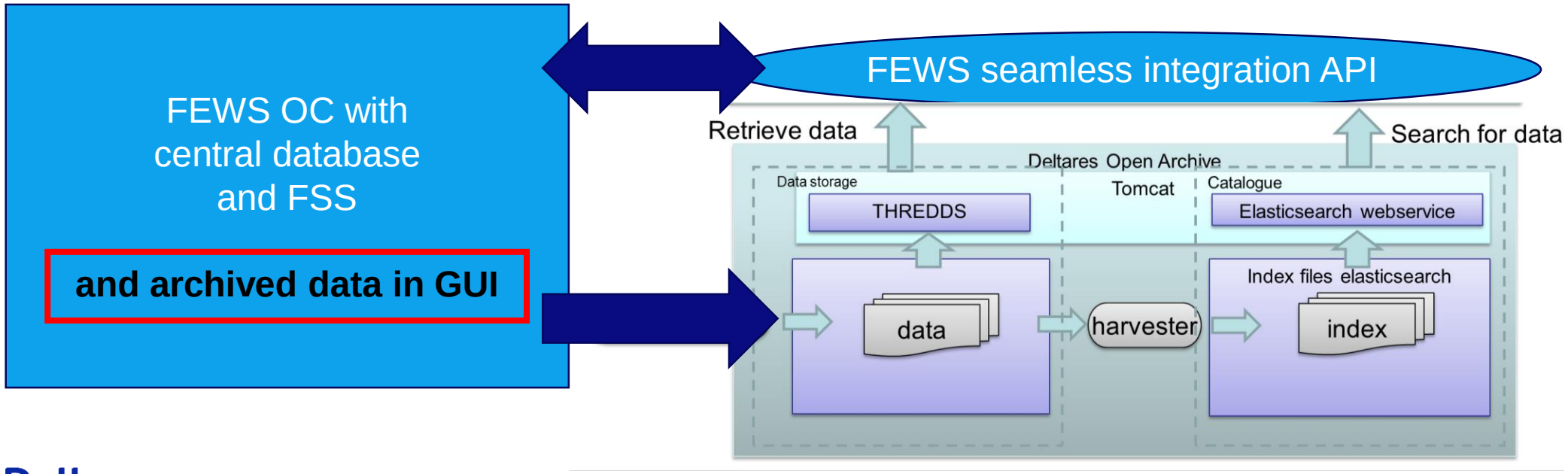
- **FEWS exports** files with **accompanying XML for metadata**
- Data can be **imported** into a FEWS Stand Alone or Operator Client using the Archive Catalogue Display
- Same display for tagging event data for historic events and later use (e.g. WaterCoach)



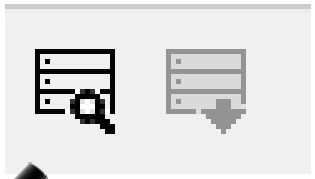
How? → Seamless integration

Past + Present

Seamless integration of
measurement data visible in the OC

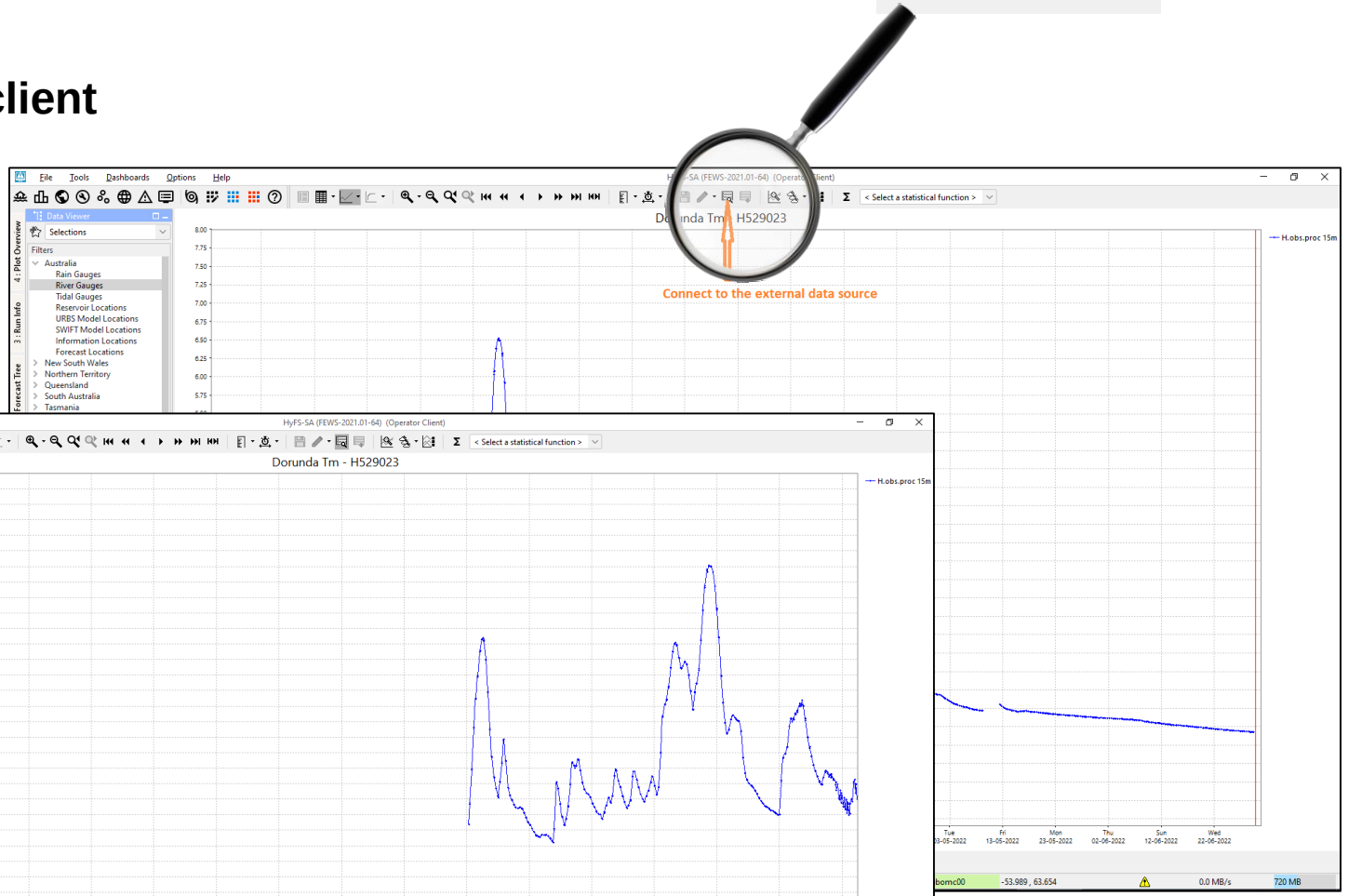


How? → Seamless integration



Seamless integration with the **FEWS client**

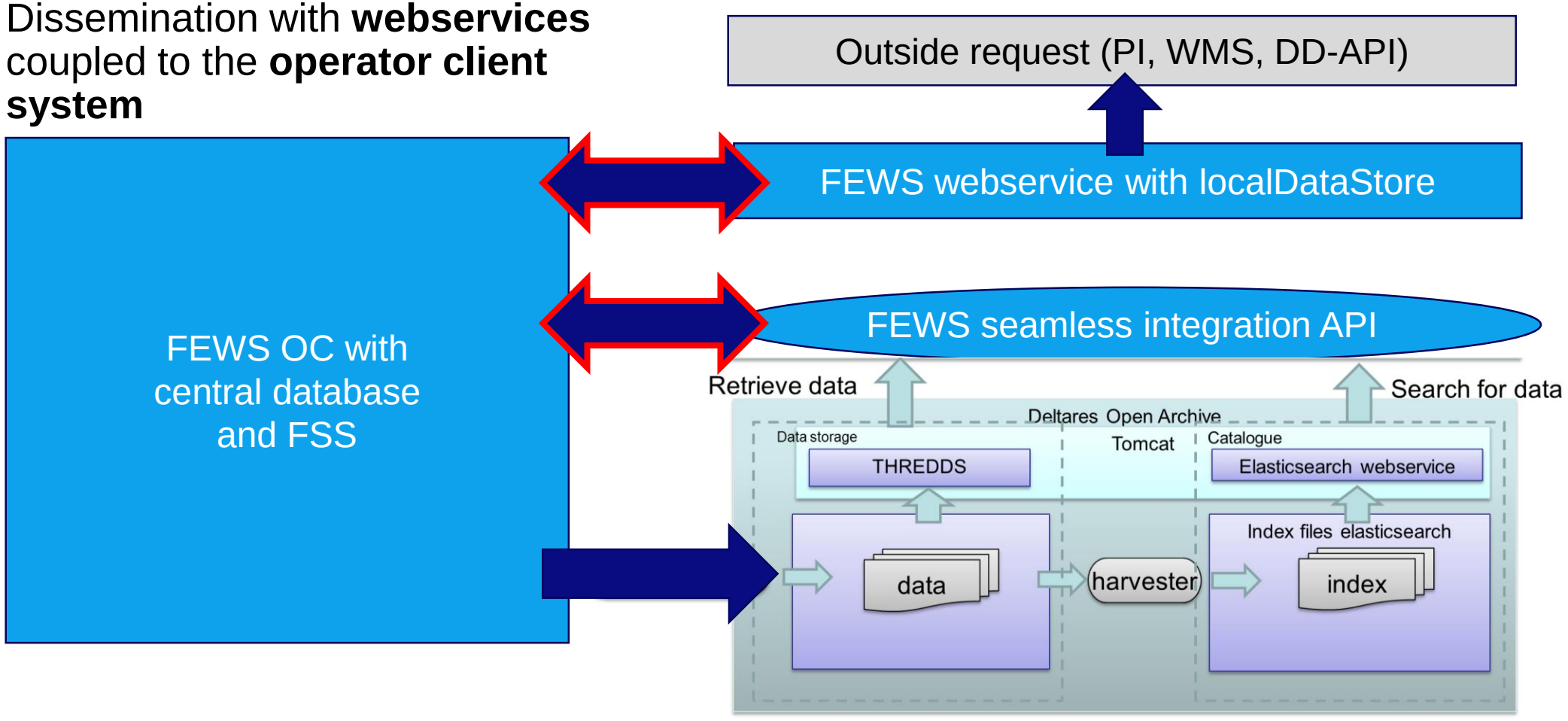
- Time series display
- Grid Display
- Workflows
  (downloadMissingDataFromArchive)



How? → Webservices via OC

Past + Present

Dissemination with **webservices** coupled to the **operator client system**



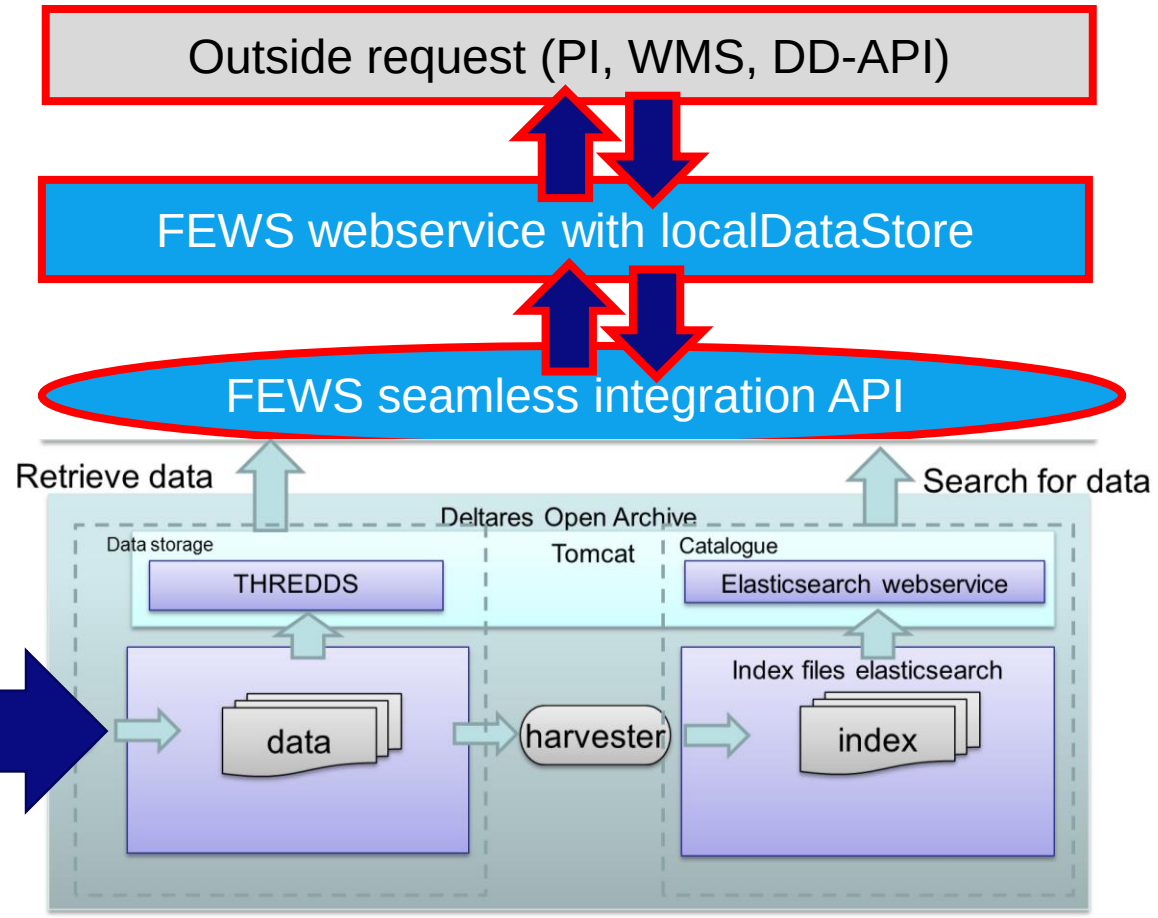
How? → Webservices direct

Past + Present

Dissemination with **webservices** coupled **directly** to Open Archive

Not only **read**, but also **write** to the archive

FEWS OC with central database and FSS



How? → Webservices POST() call

Present

Third party applications can get data into the archive with webservices POST command

FEWS Product Service

POST

/archive/products

file:

HTML

parameters:

key	value
timeZero	2021-06-21T08:00:00Z
attribute(product_id)	Hoogwaterbericht Maas
attribute(source_id)	RWS
attribute(status)	draft
…	

Product

HTML DOC PDF JPG SVG PNG

Metadata

- onderdeel van Fews Archief
- centrale plaats voor aanlevering
- gearchiveerd / traceerbaar

GET

/archive/products

- timeZero (tijd van uitgifte)
- Laatste actuele bericht
- Specifieke versie van een eerder bericht

GET

/archive/productsmetadata

- Lijst met beschikbare producten
- Binnen een bepaalde tijdperiode
- Filteren op metadata

How? → Webservices POST() call

Present

Third party applications can get data into the archive with webservices POST command

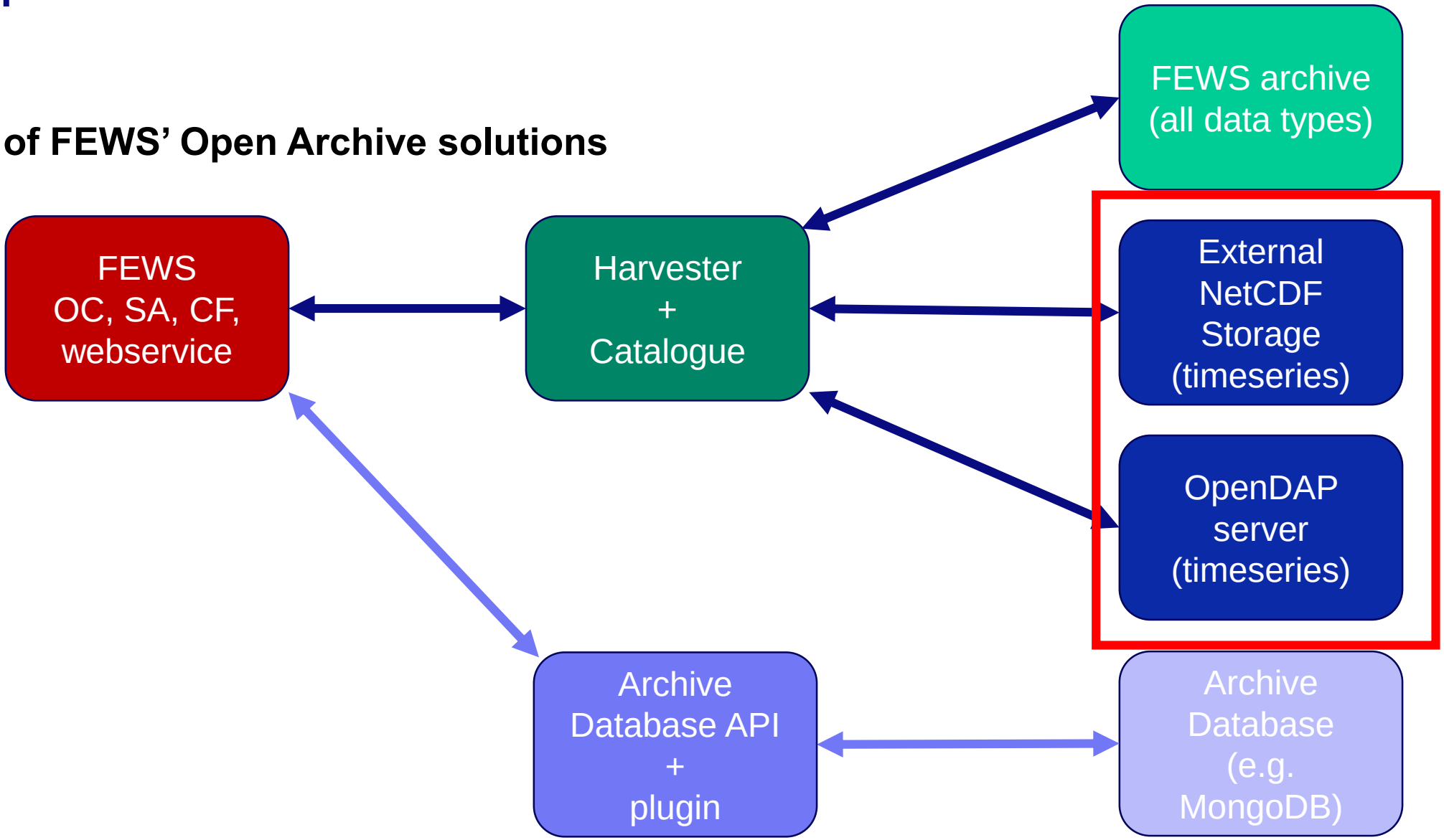
Why?

Rijkswaterstaat: Automatic and collaborative reporting

1) ***Interactive forecast reports***
2) ***Verification reports***
3) ***…***

The Open Archive – How?

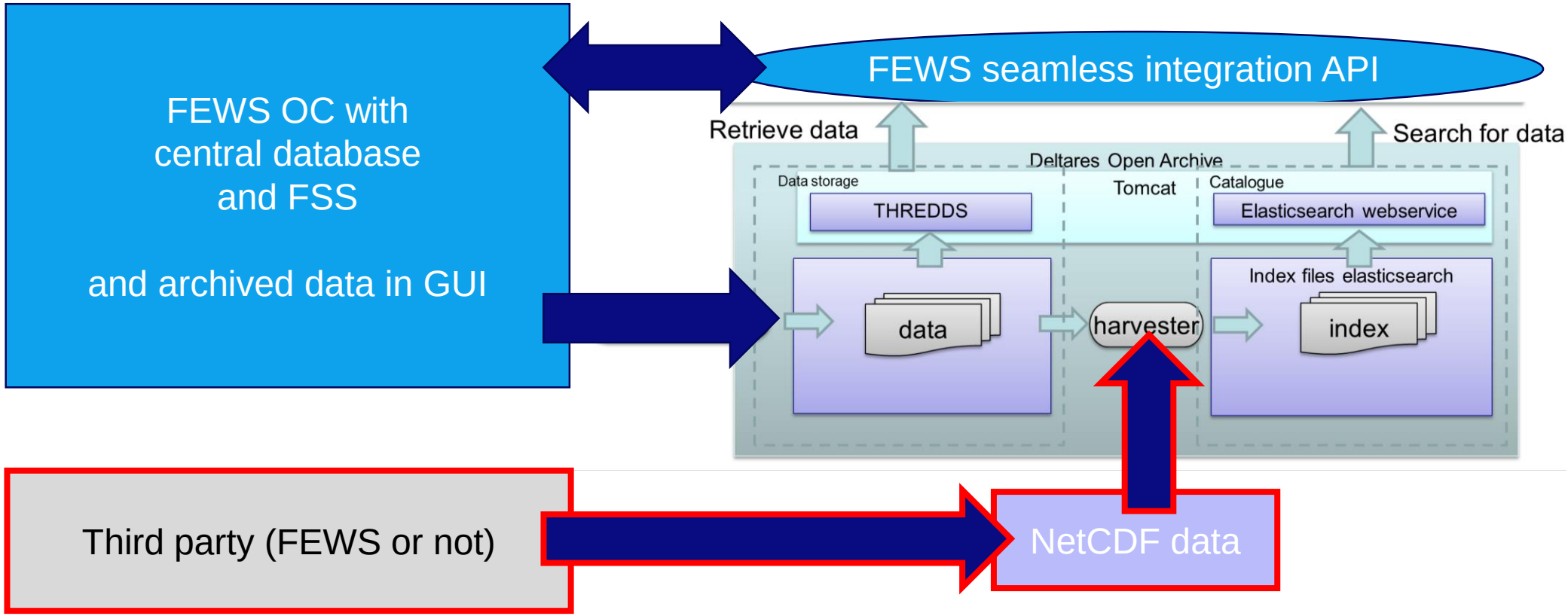
Summary of FEWS' Open Archive solutions



How? → External NetCDF storage

Present

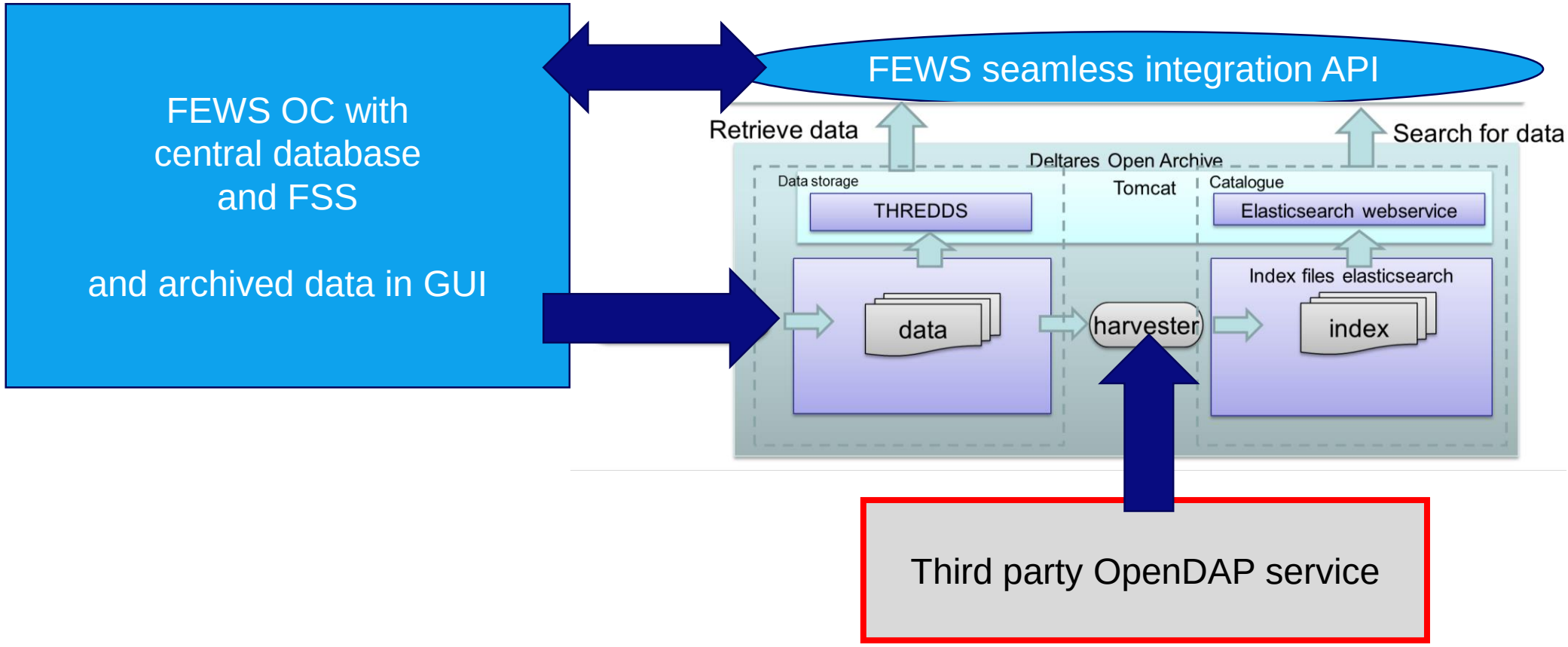
Recent developments: **external NetCDF storage**



How? → Harvesting any OpenDAP

Present

Recent developments: **leverage OpenDAP**

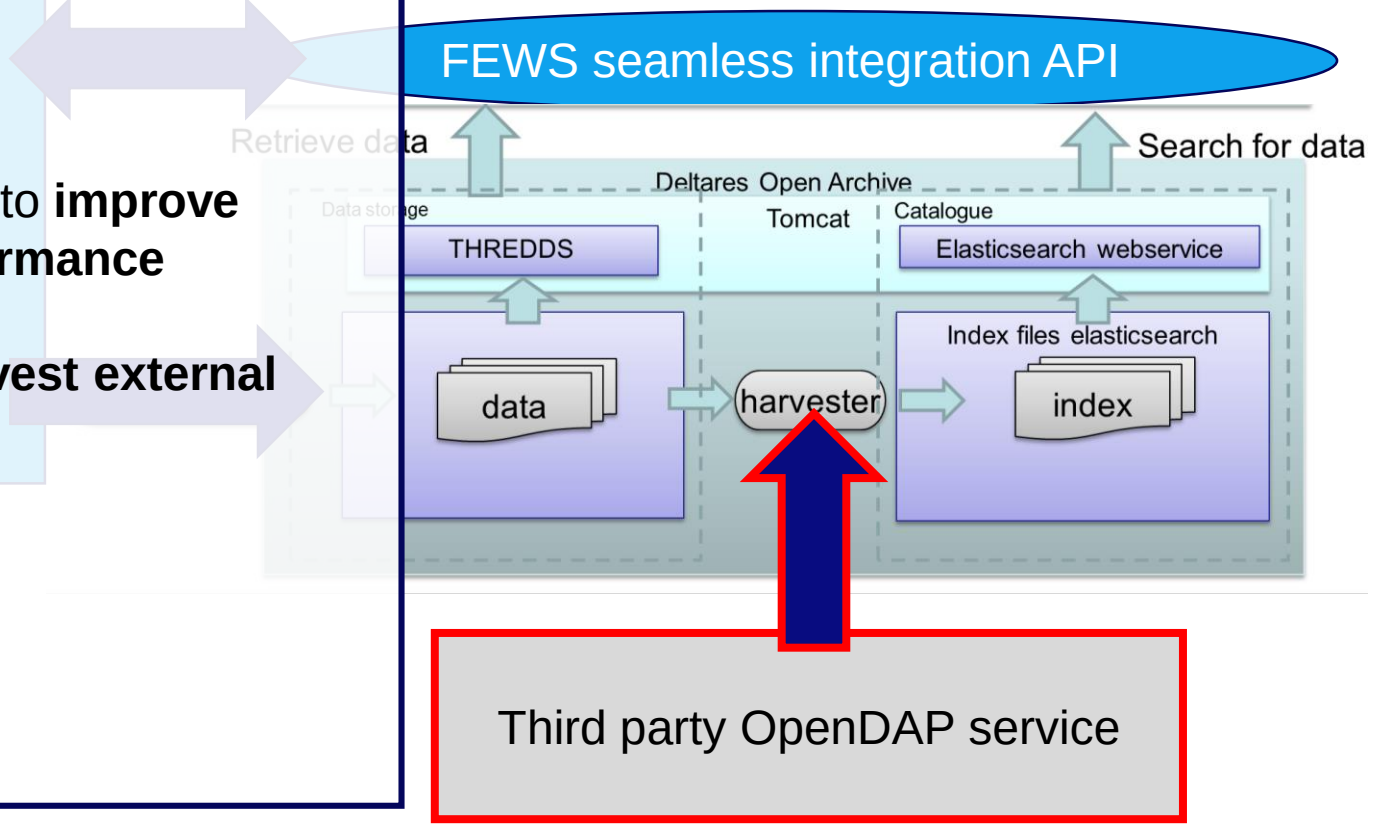


How? → Harvesting any OpenDAP

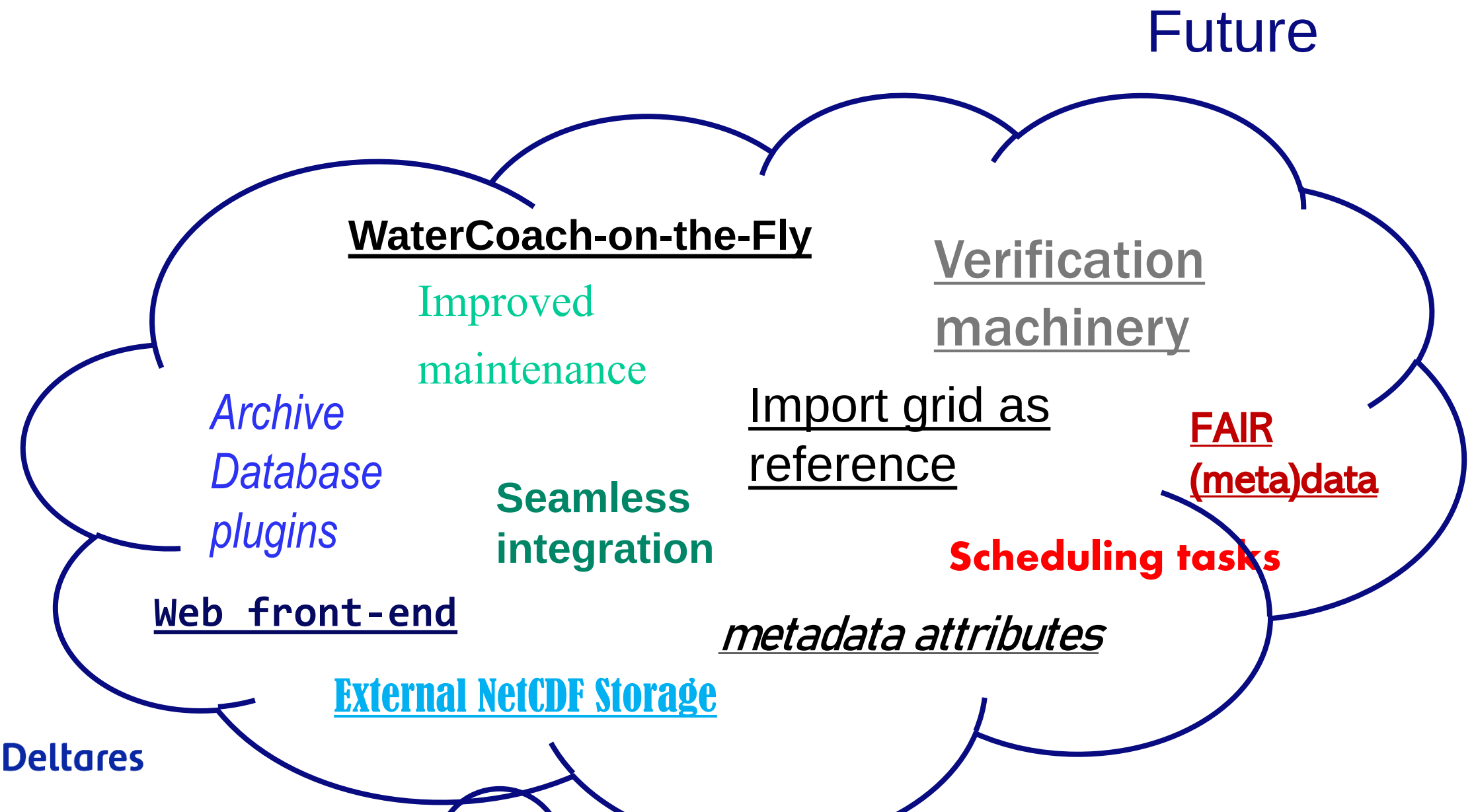
Present

Recent developments: **leverage OpenDAP**

- **OpenDAP** was implemented to **improve download and import performance**
- By-product: possibility to **harvest external OpenDAP services**



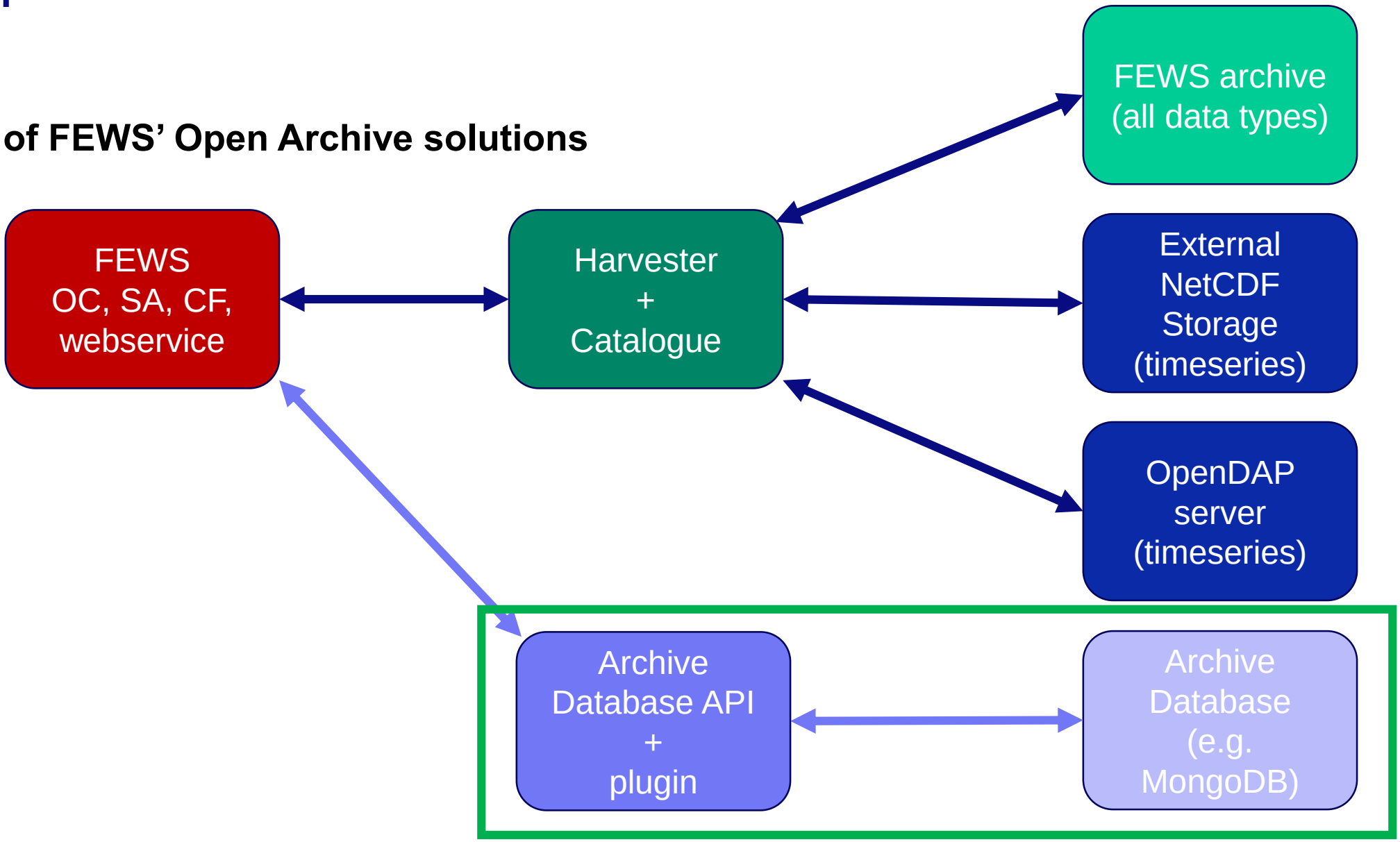
The Open Archive



Deltares

The Open Archive – How?

Summary of FEWS' Open Archive solutions



Open Archive Database
MongoDB/NoSQL

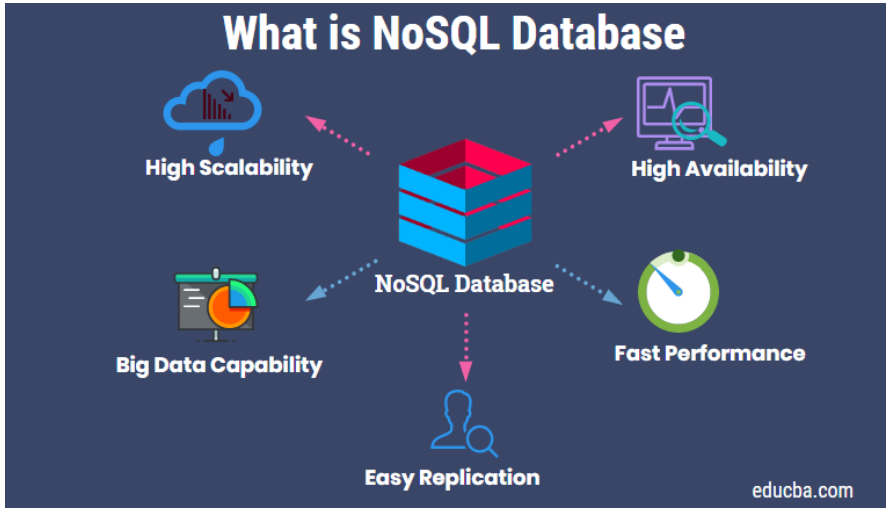
Gabriel Miller, P.E.

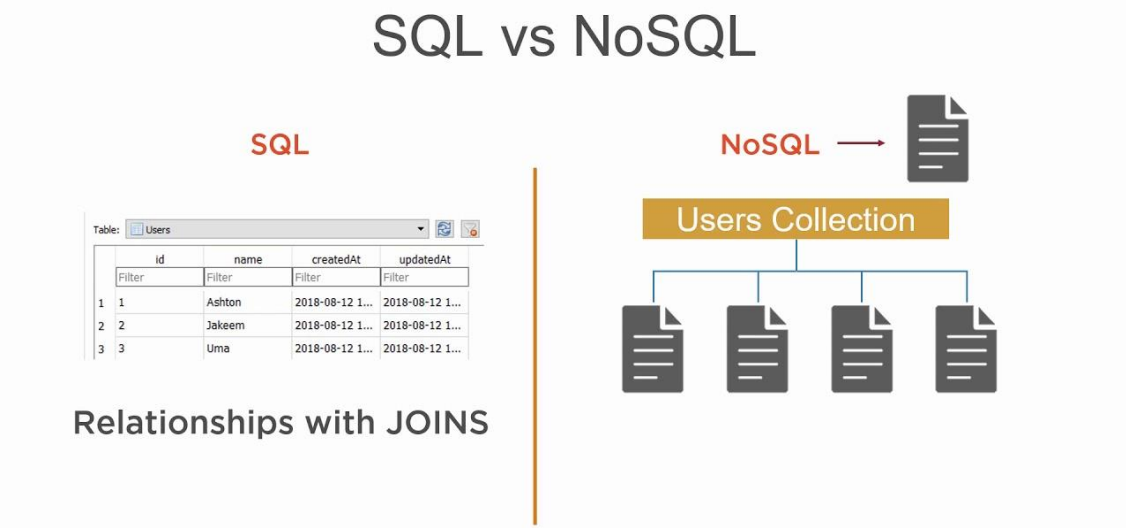
Tennessee Valley Authority

TVA TENNESSEE VALLEY AUTHORITY

What is NoSQL?

- Non-relational database
- Features
  - Flexible schemas
  - Horizontal Scaling
  - Fast queries due to data model
  - Ease of use for developers
- Using document type storage with MongoDB/FEWS





Deltares

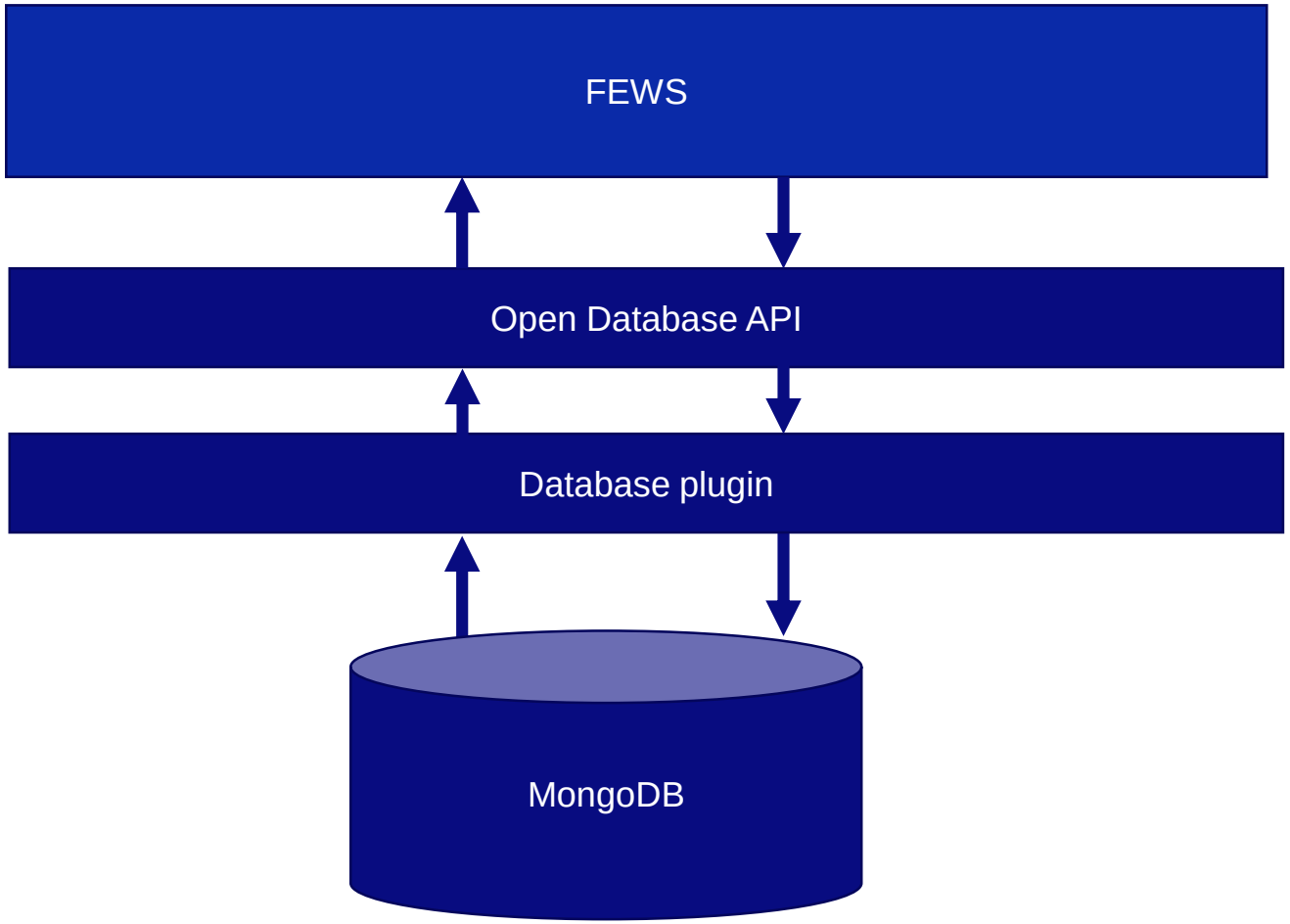
MongoDB

- Tested many NoSQL databases and MongoDB was the best suited for timeseries archive
- Document Style Storage in Binary JSON Format
- Community Edition is Free
- Very fast
- Developed simple migration tools (In FEWS) to transition from NetCDF files to the MongoDB


```
{
    "_id" : ObjectId("62663b38le8c6e0dd95cl2a7"),
    "timeSeriesType" : "external historical",
    "moduleInstanceId" : "CalculateSpill_Recalculation",
    "locationId" : "GUH-OG-08",
    "parameterId" : "QO",
    "qualifierIds" : [],
    "qualifierId" : "[]",
    "encodedTimeStepId" : "SETS60",
    "startTime" : ISODate("2020-03-01T00:00:00.000Z"),
    "endTime" : ISODate("2020-12-31T23:00:00.000Z"),
    "committed" : true,
    "bucketSize" : "YEAR",
    "bucket" : NumberLong(2020),
    "metaData" : {
        "sourceId" : "scalar_externalhistorical_QO",
        "areaId" : "scalar",
        "unit" : "CMS",
        "displayUnit" : "CFS",
        "locationName" : "GUH-OG-S_Embank_SY",
        "parameterName" : "Other Spill Flow",
        "parameterType" : "instantaneous",
        "timeStepLabel" : "hour",
        "timeStepMinutes" : 60,
        "localTimeZone" : "CST",
        "archiveTime" : ISODate("2022-04-22T01:04:14.334Z"),
        "modifiedTime" : ISODate("2022-04-22T01:04:14.334Z")
    },
    "timeseries" : [
        {
            "t" : ISODate("2020-03-01T00:00:00.000Z"),
            "lt" : ISODate("2020-02-29T18:00:00.000Z"),
            "v" : null,
            "dv" : null,
            "f" : 8
        },
        {
```

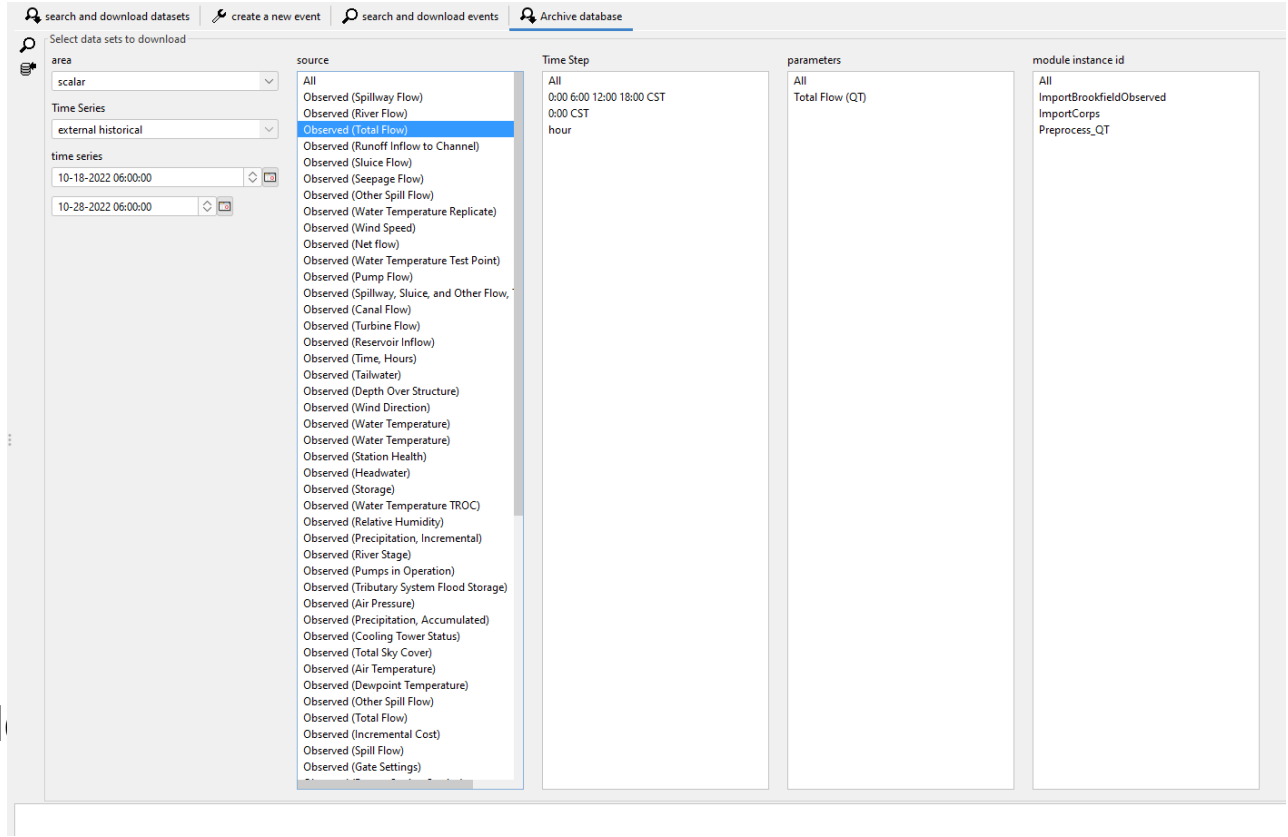
Integration MongoDB Deltares Open Archive

How to integrate MongoDB into the Deltares Open Archive?



Direct Data Access

- Many third party tools can hook directly into MongoDB with BI Connector (very fast)
  - PowerBI
  - Tableau
  - Excel
- Easy programmatic access with MongoDB and programming languages (Python, Java, etc)
- Archive catalogue GUI allows for more flexibl access to data



Query Language, In-Database Analytics & Data Integrity

- Simple query language allows for the ability to:
  - Query and view data in archive
  - Make data quick updates
  - Update metadata if location names/parameters change
  - Easy programmatic access to data
- Data Integrity
  - Unique keys/index enforcement assures data integrity and confidence
  - Found multiple instances of non-unique data in transition to MongoDB

```
db.getCollection('ExternalHistoricalScalarTimeSeries').find({
    moduleInstanceId:'Preprocess_HP',
    locationId:'HADT1',
    parameterId:'HP',
    encodedTimeStepId:'SETS60'})
```

ExternalHistoricalScalarTimeSeries 2.08 sec.

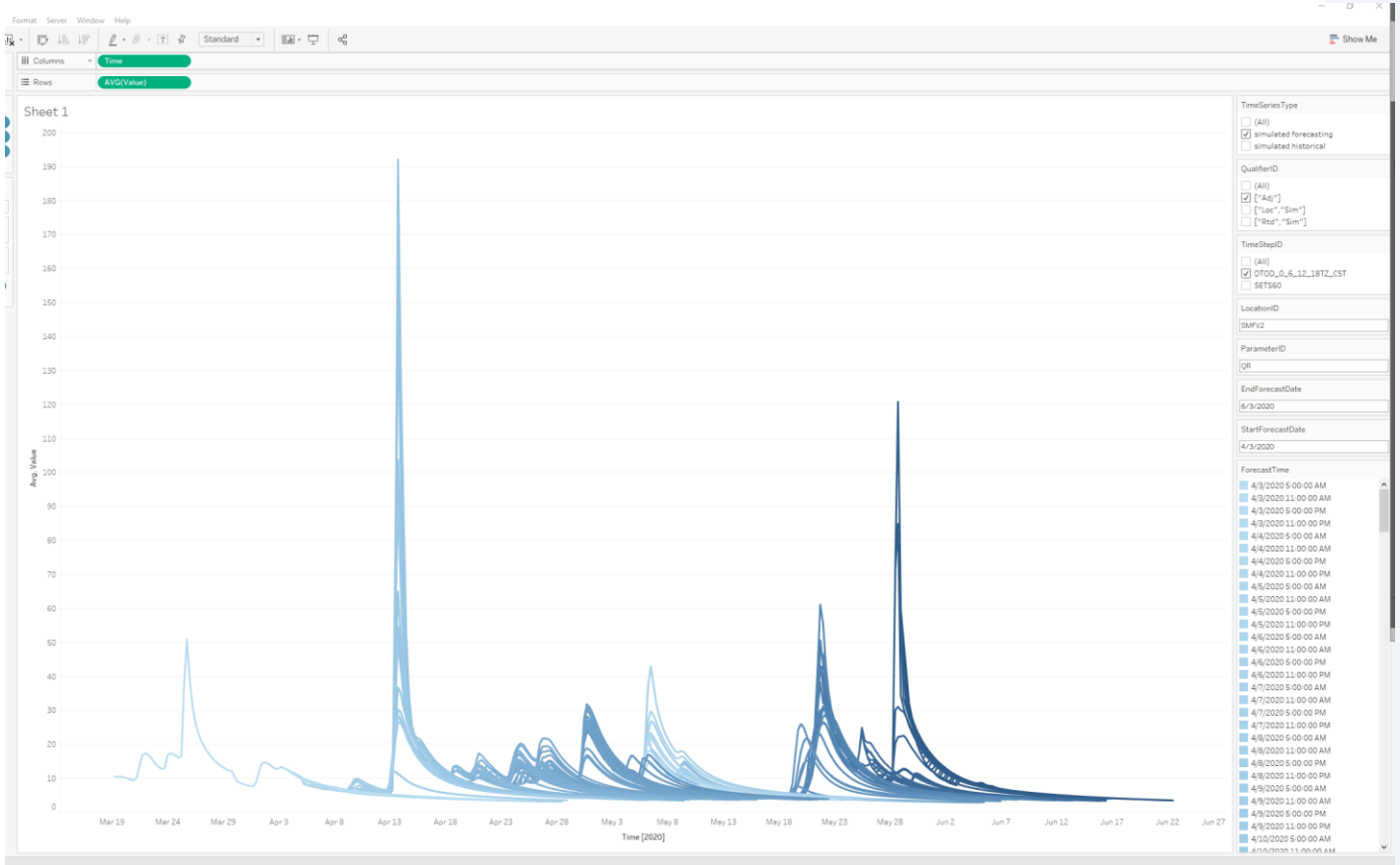
Key	Value
(1) ObjectId("62663fb81e8c6e0dd95d66ce")	{ 15 fields }
_id	ObjectId("62663fb81e8c6e0dd95d66ce")
timeSeriesType	external historical
moduleInstanceId	Preprocess_HP
locationId	HADT1
parameterId	HP
qualifierIds	[1 element]
qualifierId	["Editable"]
encodedTimeStepId	SETS60
startTime	2016-09-01 00:00:00.000Z
endTime	2016-12-31 23:00:00.000Z
committed	true
bucketSize	YEAR
bucket	2016
metaData	{ 12 fields }
timeseries	[2928 elements]
(2) ObjectId("62663fb81e8c6e0dd95d66cf")	{ 15 fields }
(3) ObjectId("62663fb81e8c6e0dd95d66d0")	{ 15 fields }
(4) ObjectId("626641f81e8c6e0dd95df2b4")	{ 15 fields }

Performance - 2020.02 vs 2021.02 with MongoDB

- Workflow Integration
  - ~100x improvement (8s vs 15 min 26s in test)
- Seamless Integration in Plots
  - One location, many parameters – 10x improvement
  - Many locations, single parameter – Roughly equivalent
- Large webservice pulls
  - 10-30x improvement
- Downloading scalar data from archive
  - 5-20x improvement

Simplification/Instant Availability

- Archive export simplified
  - No need to specify NetCDF filename,
  - No need to split files based on parameters, timestep, etc
  - Easier to update data
- No amalgamation
  - MongoDB bucket size based on timestep
- No catalogue harvesting (elastic search)
  - Data is instantly available upon export
  - No downtime from catalogue rebuilding



TVA
TENNESSEE VALLEY AUTHORITY
Deltares
INFISYS
Manitoba Hydro
Questions?

Deltares

Open Archive Vision/Roadmap Table of Contents

First ideas…

Delft-FEWS International User Days 2022

November 2022

Open Archive Vision - Rationale

- **Growing** demand…
- **Multiple** (read: many) **types of users** and **different use cases** for **storing** and **retrieving** data…
- **Strategy** is needed to accommodate the use cases…

A **vision** helps…

- **Mechanism** to assess, channel and prioritize the development requests…
- **Agreement** on common/overlapping objectives and (future) use cases…
- Way to attach **funding**…

To achieve…

- **Robustness** and **maintainability** (of underlying software and tools)…
- **Easy** to deploy, use, support, maintain, upgrade…
- **Understand** and **value** your **archived resources**…

Open Archive Vision – stakeholders/contributors

- Global Delft-FEWS Community, but represented by
- Delft-FEWS International Community Strategy Board (Gabe, Justin)
- Deltares Department OWM (Marc, Peter, Maarten, Indra)
- Deltares Department HAF (Tom)
- Delft-FEWS Development/ICT team (Andre, Orshi, Onno, Bart, Jitka, Gerrit/Andre)
- Delft-FEWS Product management (Nadine, Marcel, Gerben)

Open Archive Vision
working title

Relevant Records for Retrieval and Re-use,
a vision for the Delft-FEWS Open Archive

or

Relevant Records Ready to Retrieve
a vision for the Delft-FEWS Open Archive

or …

Open Archive Vision
table of contents

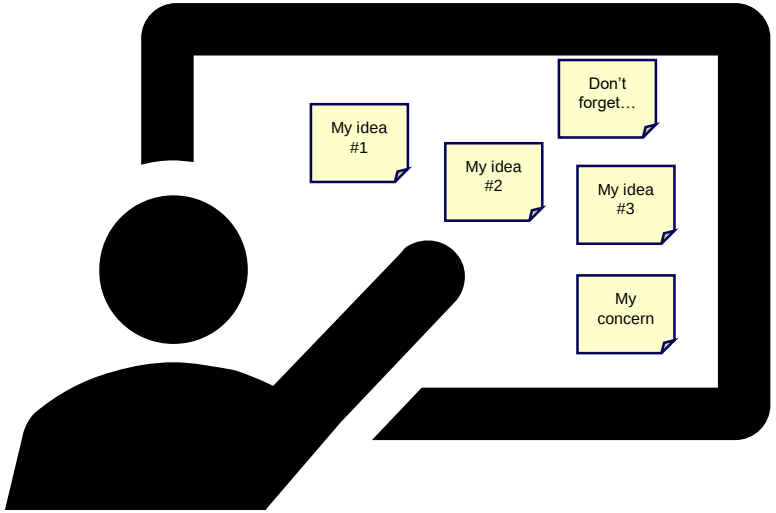
Chapter	Sections	Goal
Introduction	• Rationale and *story* for having an archive • Rationale and *story* for this vision • Typical End Users (personas) and their requirements • Timeline of the vision • Related vision documents • Relevant (internal, external) developments	Explain the **WHY**
Current Situation	• General overview, objectives and main workflows of the Open Archive • NetCDF file-based storage of Delft-FEWS produced timeseries • NetCDF file-based storage of externally produced timeseries in (self-describing) NetCDF • MongoDB Database storage of timeseries for processing in BI tooling • Elastic Search based catalogue, Thredds, OpenDAP • Interaction with the Delft-FEWS Client and Server components (displays, web services, workflows, data viewer/web oc, water coach) • Supporting mechanism: seamless integration	Explain the **NOW** to bring everyone on the same page

Open Archive Vision
table of contents

Chapter	Sections	Goal
Vision elements	• The basics (FOUNDATION FOR THE FUTURE) • New directions (NEW BUILDING BLOCKS) • Ready for Removal (WHAT TO DUMP) • Visualizing the vision…(VISUALIZE)	Explain the future **WHAT**
Planning & Deliverables	• Planning & milestones • Definition of (testable) deliverables • Release schedule	Explain the **WHAT & WHEN**
Finances	• Estimates (COSTS) • Funding (CONTRIBUTIONS, CONTRIBUTORS)	Explain the **HOW MUCH**

In the Open Archive break-out…

- Besides interesting demos and discussions of current (and near future) usage and use cases…
- Please write down your ideas, suggestions, do's & don't's and concerns related to the Vision on the Open Archive on the flip-overs



Contact

www.deltares.nl

@deltares

linkedin.com/company/deltares

info@deltares.nl

@deltares

facebook.com/deltaresNL