World Data Centre for Aerosol: Status & News 2019

Fiebig, M.; Tørseth, K. and the EBAS-team
WDCA at the Norwegian Institute for Air Research

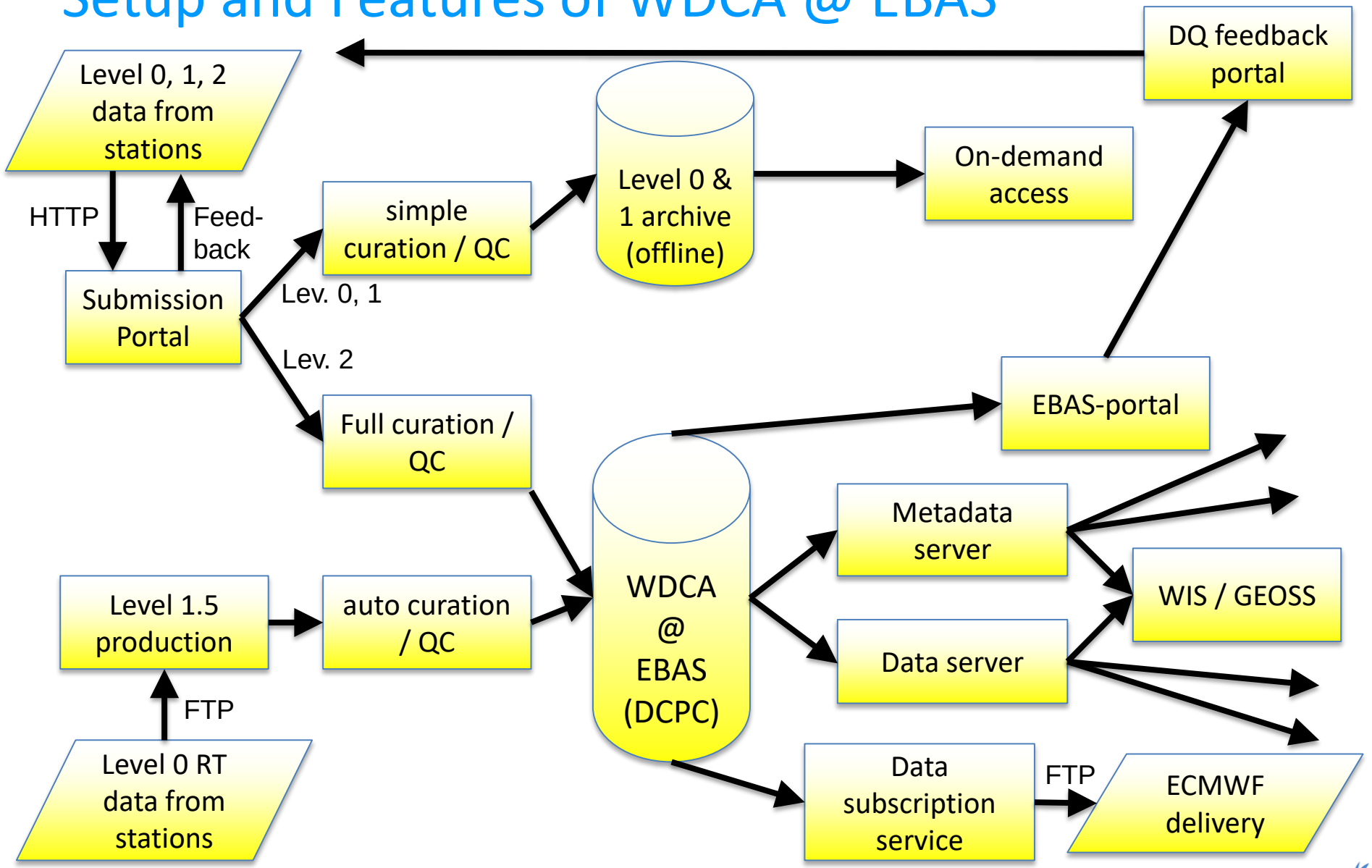
NILU Norsk institutt for luftforskning

WMO GAW

Items Covered

- WDCA setup, features, services
- WDCA Status: Key numbers and figures
  - Stations Reporting
  - Access Statistics
- Involvement in EU project ENVRI-FAIR
  - Data citation service (DOI for data)
  - Inter-operability (connection to portals, etc.)

Setup and Features of WDCA @ EBAS

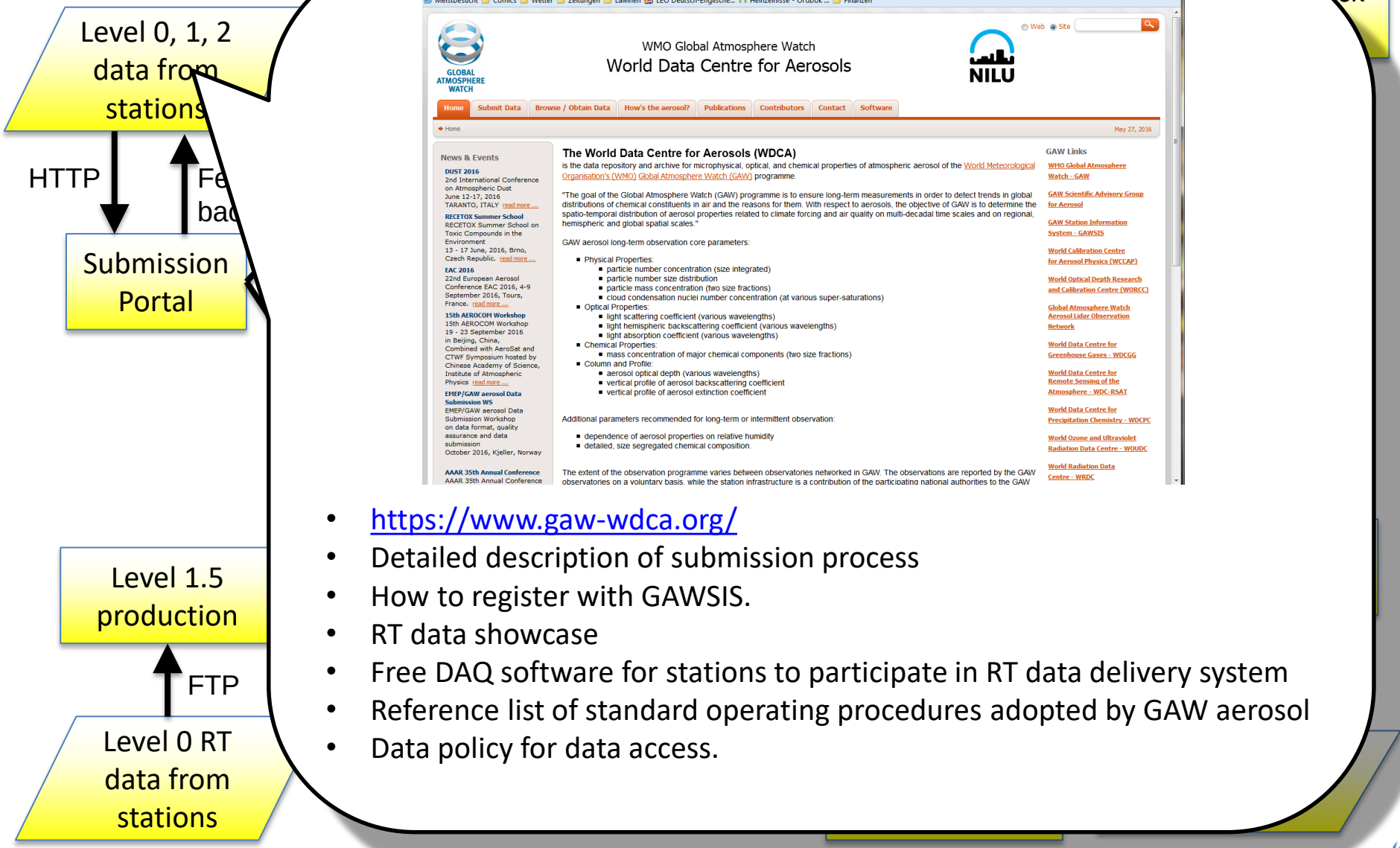


Data Levels: Implementing Traceability

Data Level	Manual QC	RRT/ auto QC	Description	
0	0a	0b	• Annotated raw data • format instrument specific • all data / information for processing to final value.	• contains all parameters provided by instrument as provided • ”native” time resolution
1	1a	1b	• processed to final parameter • invalid data removed • ”native” time resolution	• format property specific • STP correction if necessary
1.5 / 2	2	1.5	• aggregated to hourly averages • variability quantified	• format property specific

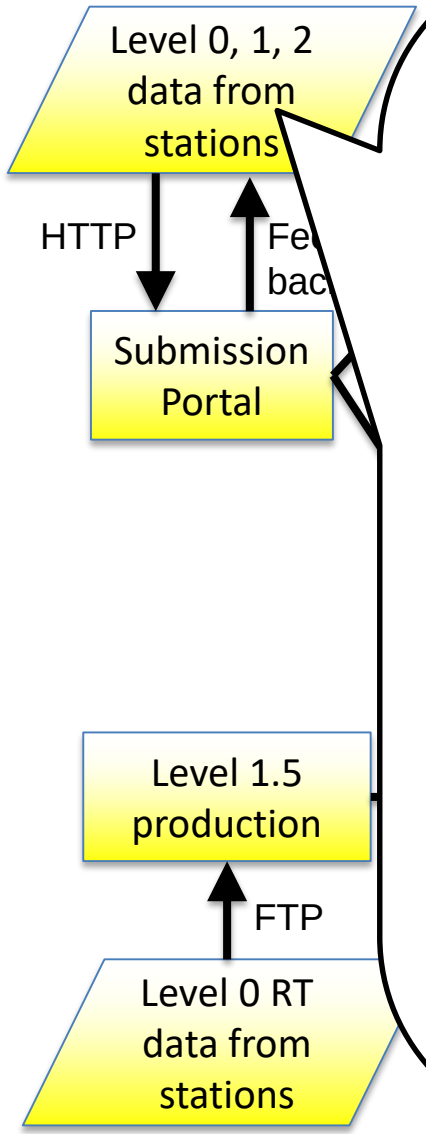
- SOP describes steps from one to the next level.

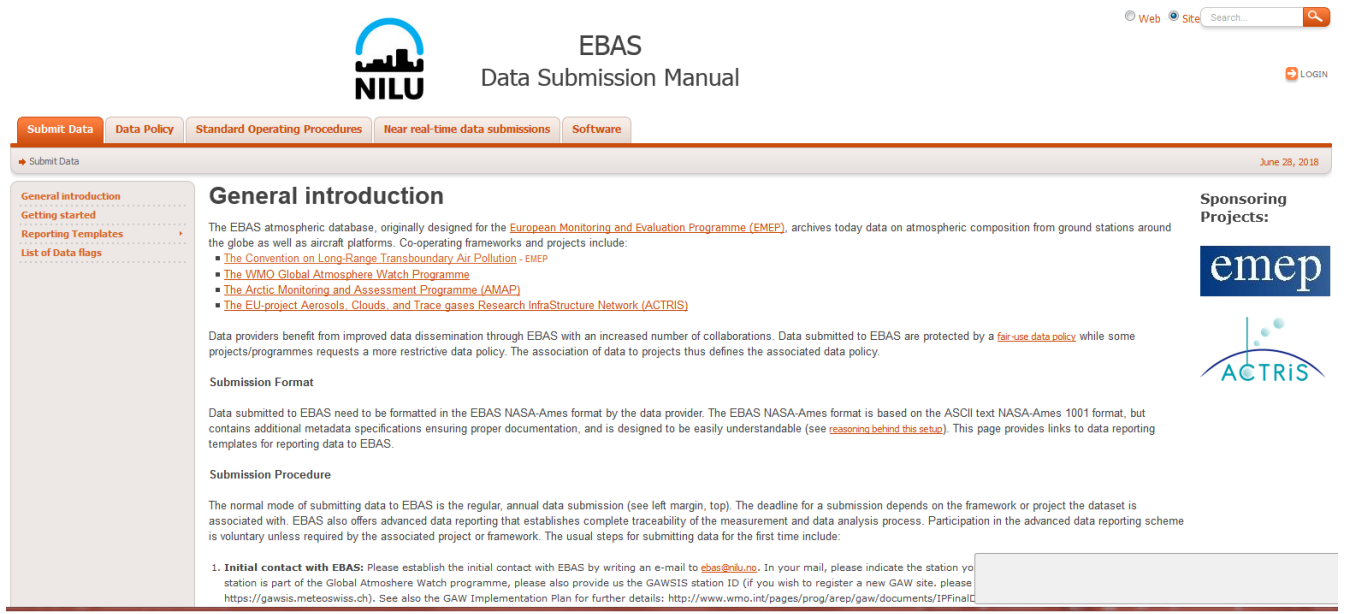
Setup and Features of WDCA @ EBAS



- https://www.gaw-wdca.org/
- Detailed description of submission process
- How to register with GAWSIS.
- RT data showcase
- Free DAQ software for stations to participate in RT data delivery system
- Reference list of standard operating procedures adopted by GAW aerosol
- Data policy for data access.

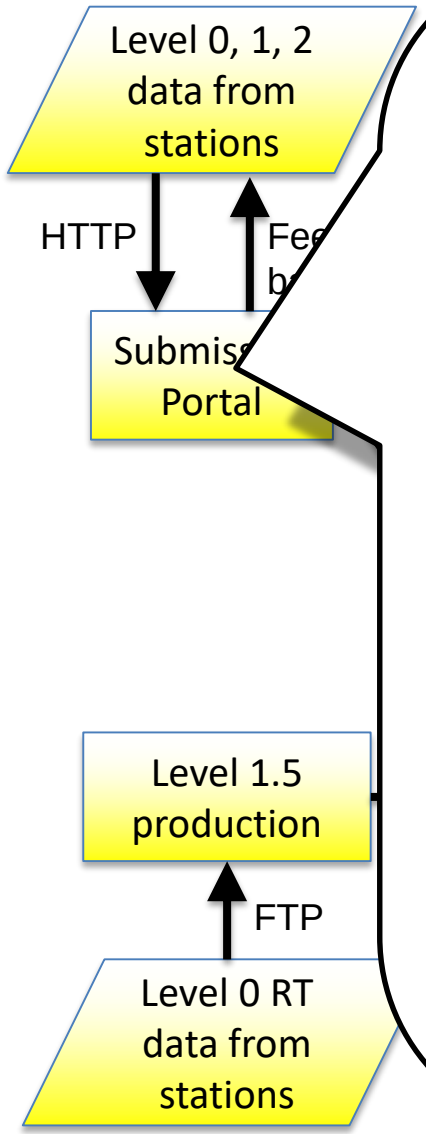
Setup and Features of WDCA @ EBAS

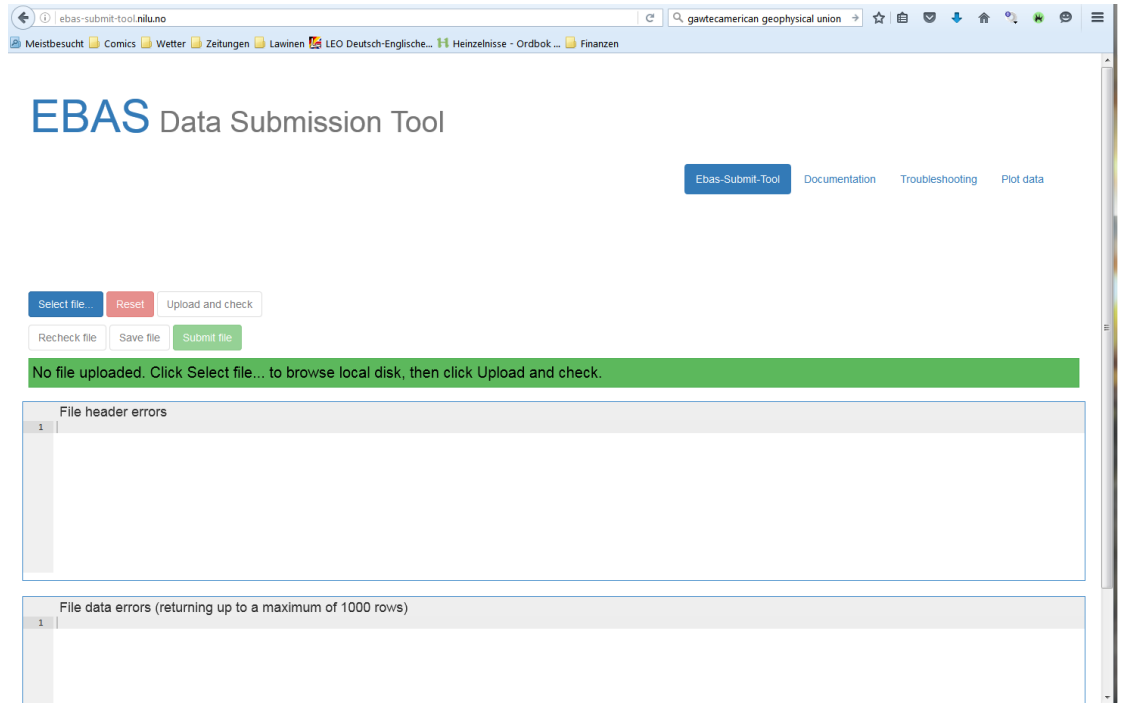




- https://ebas-submit.nilu.no/
- Repository of templates for reporting data to WDCA, co-ordinated with EMEP and ACTRIS.
- Special templates for each variable and data level.
- Rich use metadata adapted to variable and methods used, co-ordinated with expert stakeholders, e.g. systematic & statistical uncertainties, sampling conditions, sample preparation, operating procedures, QA metadata (QA measure type, when conducted, by whom, result).

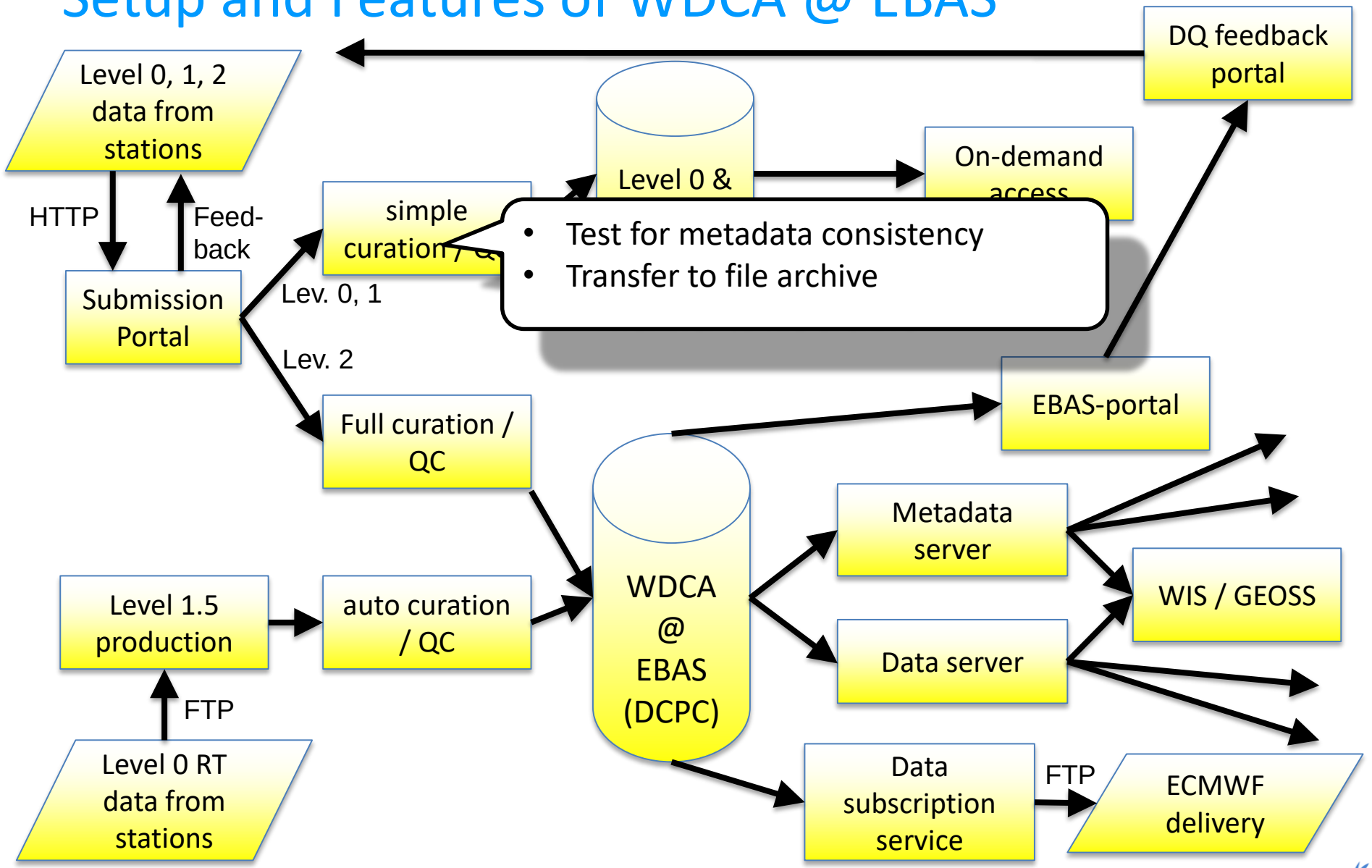
Setup and Features of WDCA @ EBAS



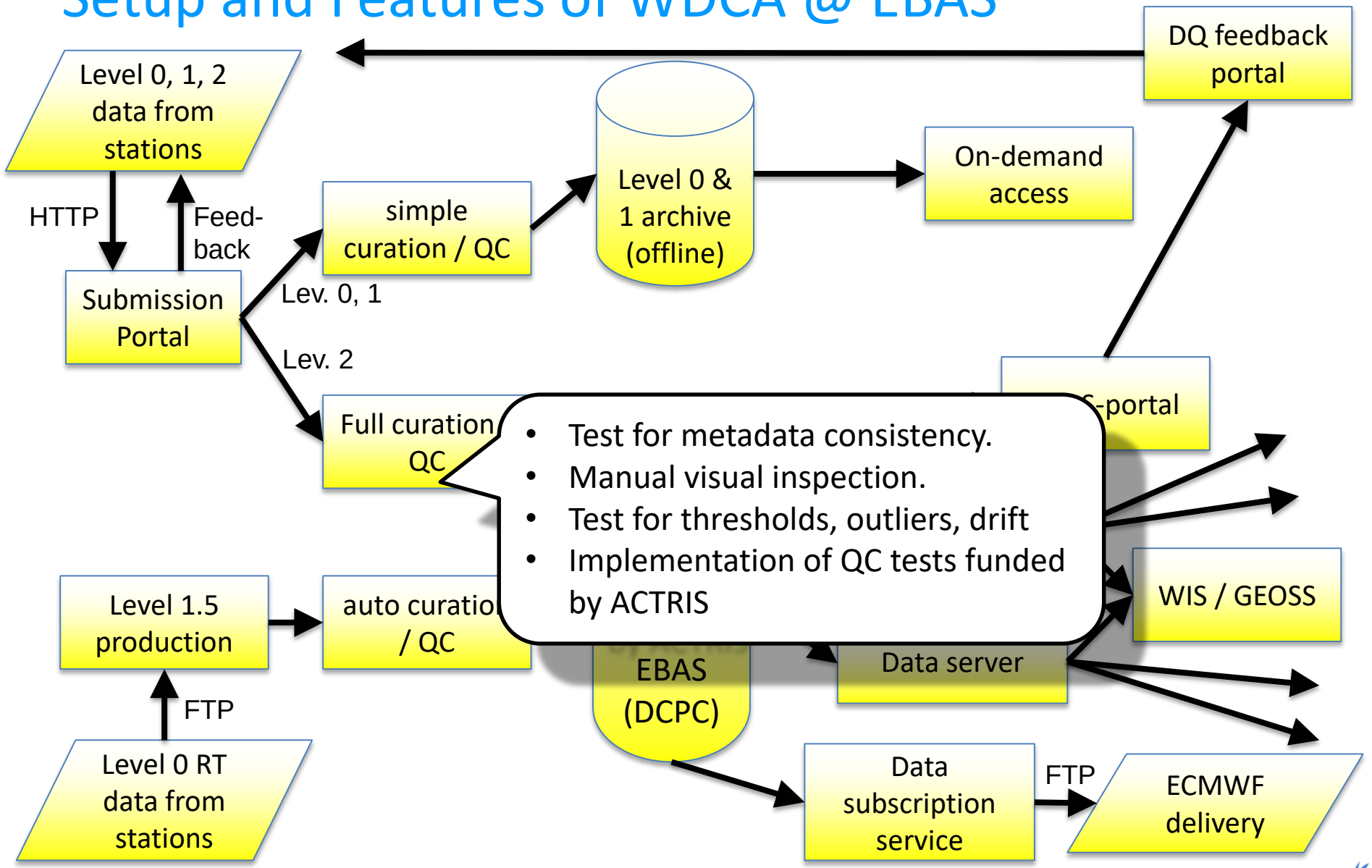


- Accepts Level 2 (GAW obl.)
- Accepts Level 0, 1 (GAW voluntary)
- Gives immediate feedback on syntax, semantic, and sanity check
- Data inspection by plotting.
- Has reduced workload for data providers and WDCA.
- Has reduced turn-around time for data curation significantly
- Funded by ACTRIS.

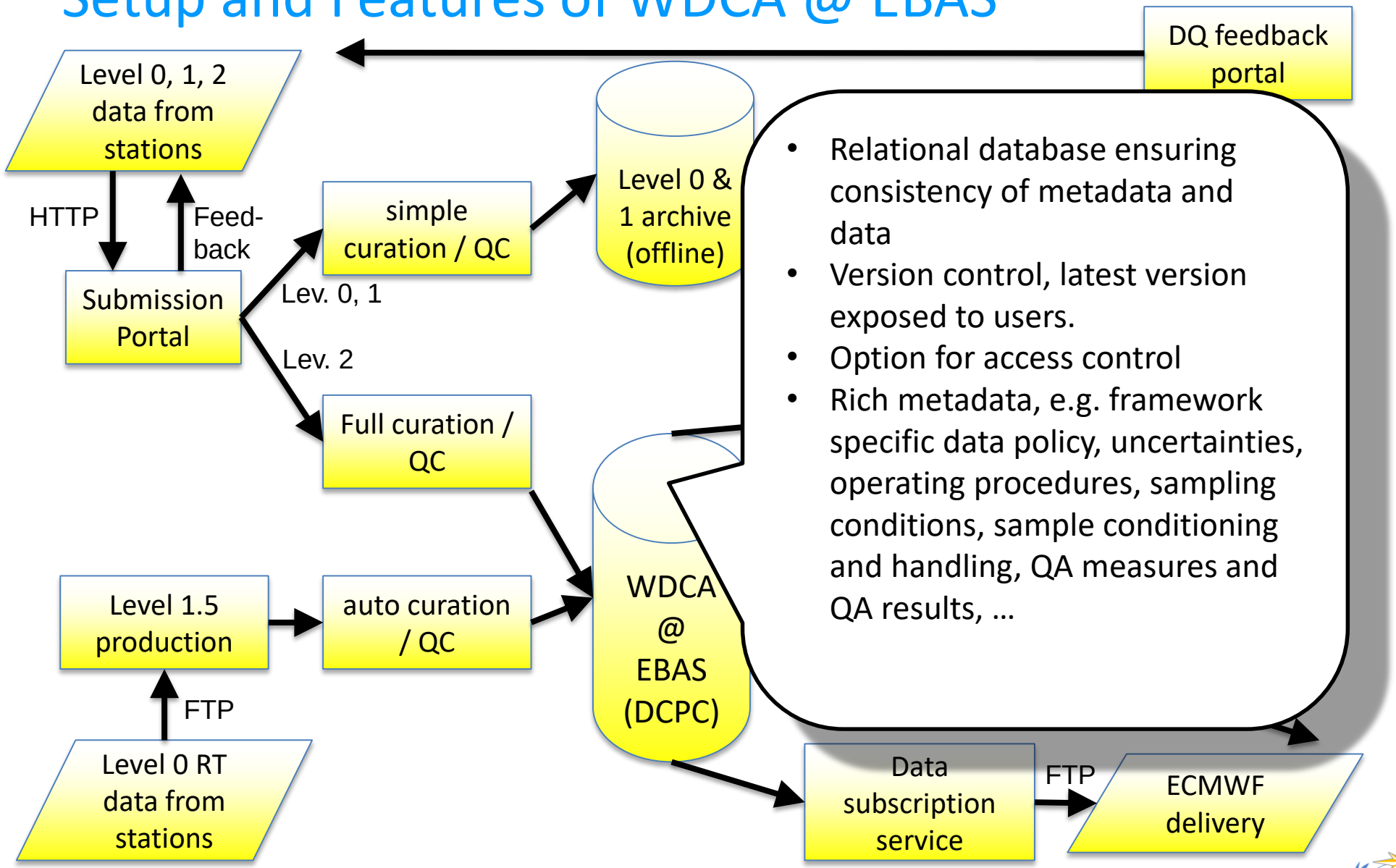
Setup and Features of WDCA @ EBAS



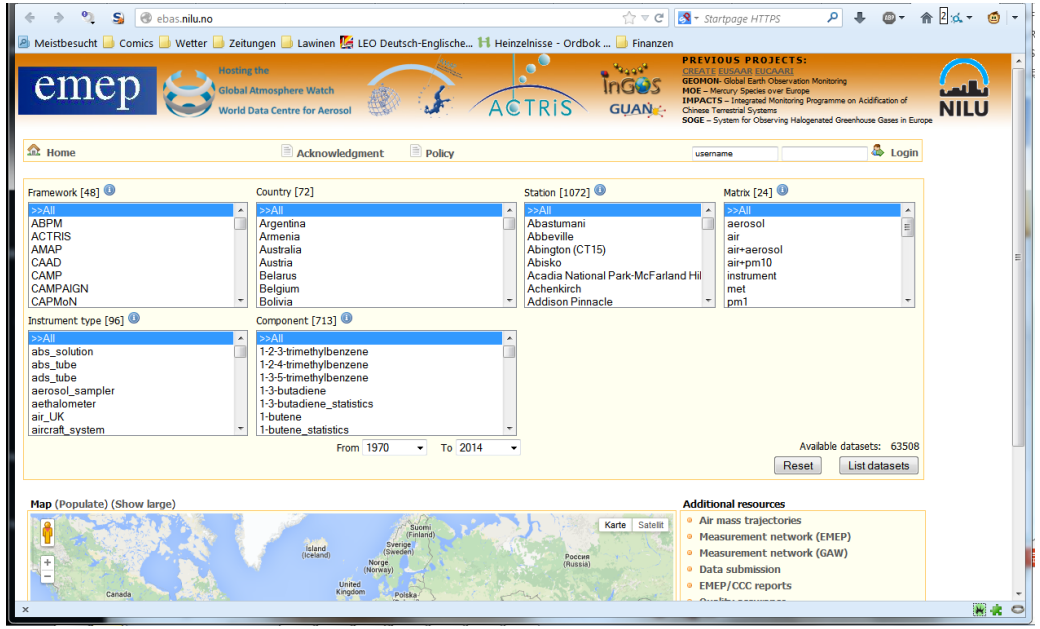
Setup and Features of WDCA @ EBAS



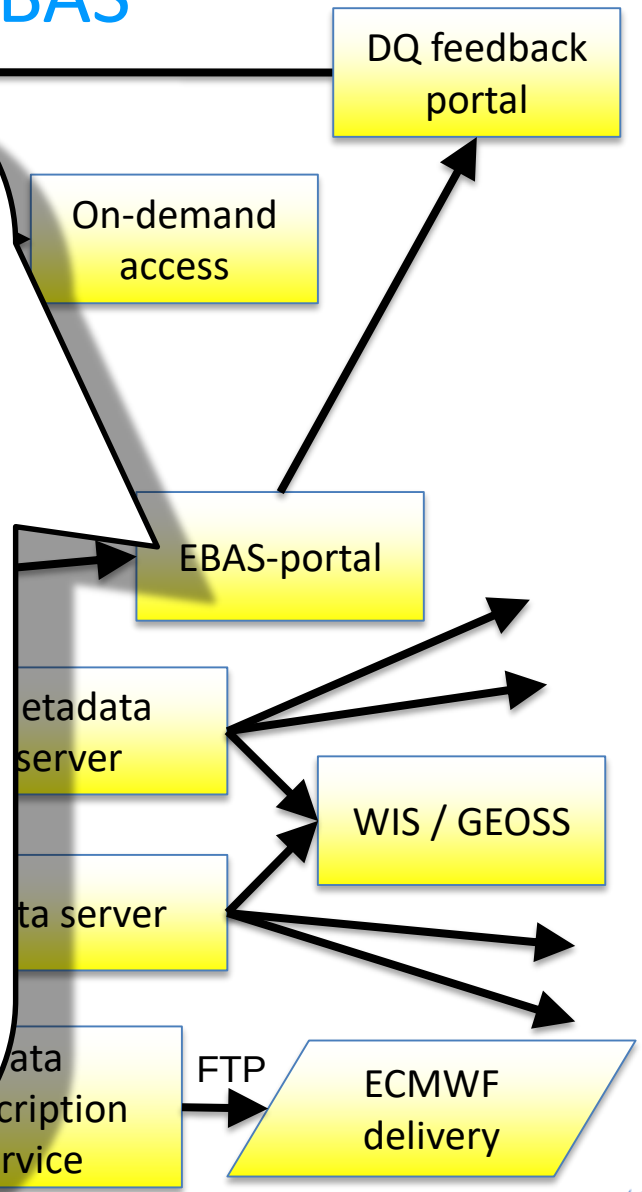
Setup and Features of WDCA @ EBAS



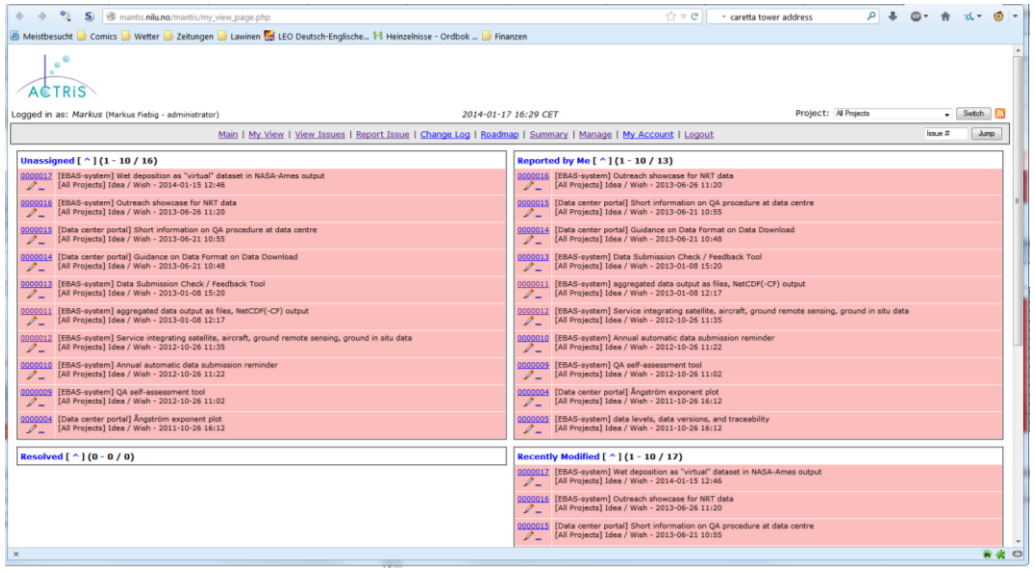
Setup and Features of WDCA @ EBAS



- Structured search of data repository
- Map visualisation of search results
- Graphical display and download of data
- Selection in search result for display or download.
- Direct access to search results by direct links, suitable as landing pages.
- Option for access control, e.g. RT data.

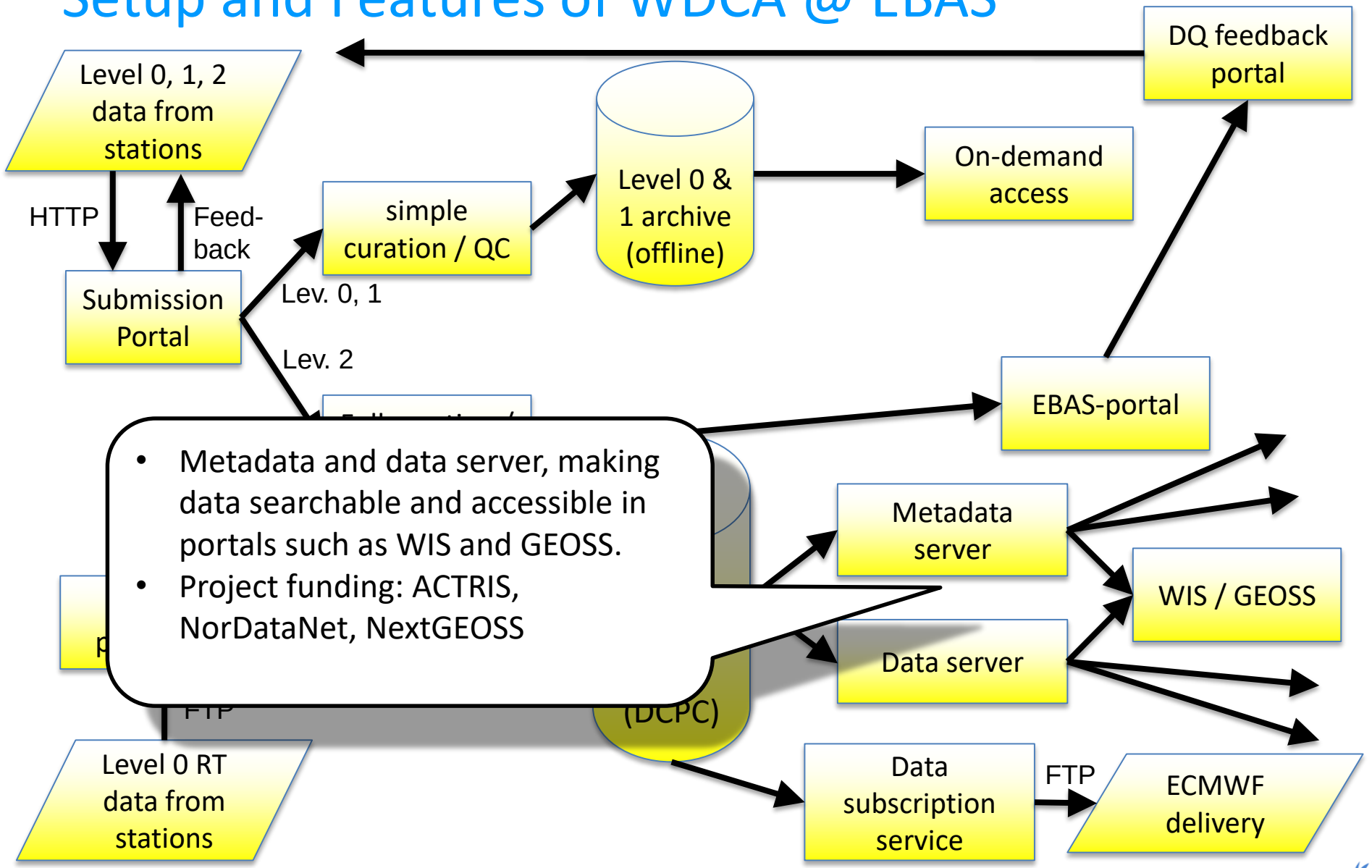


Setup and Features of WDCA @ EBAS

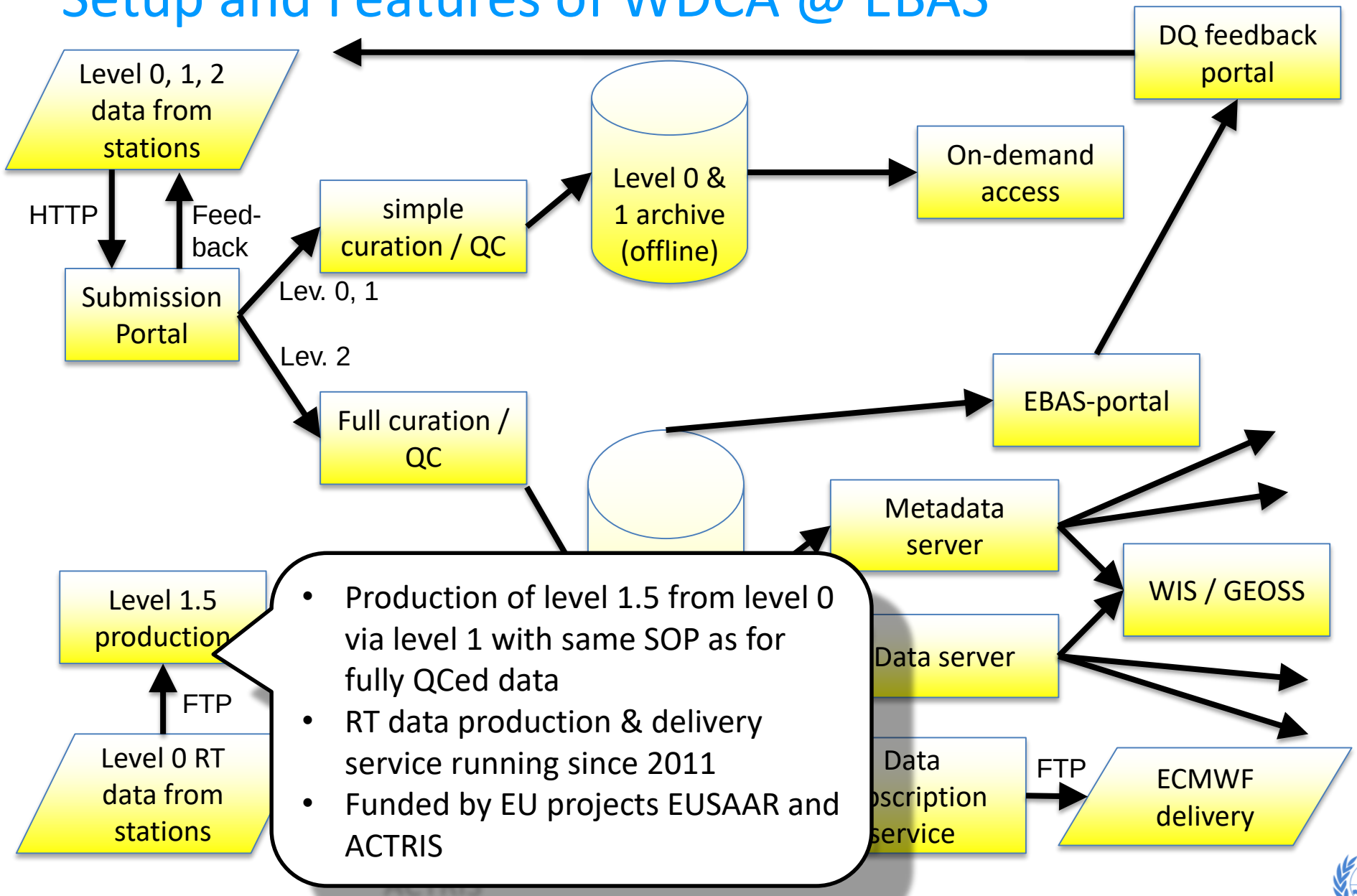


- Records general user feedback, relays to data provider, and tracks correction of data.
- Records data quality issues found while using the data, facilitates follow-up.

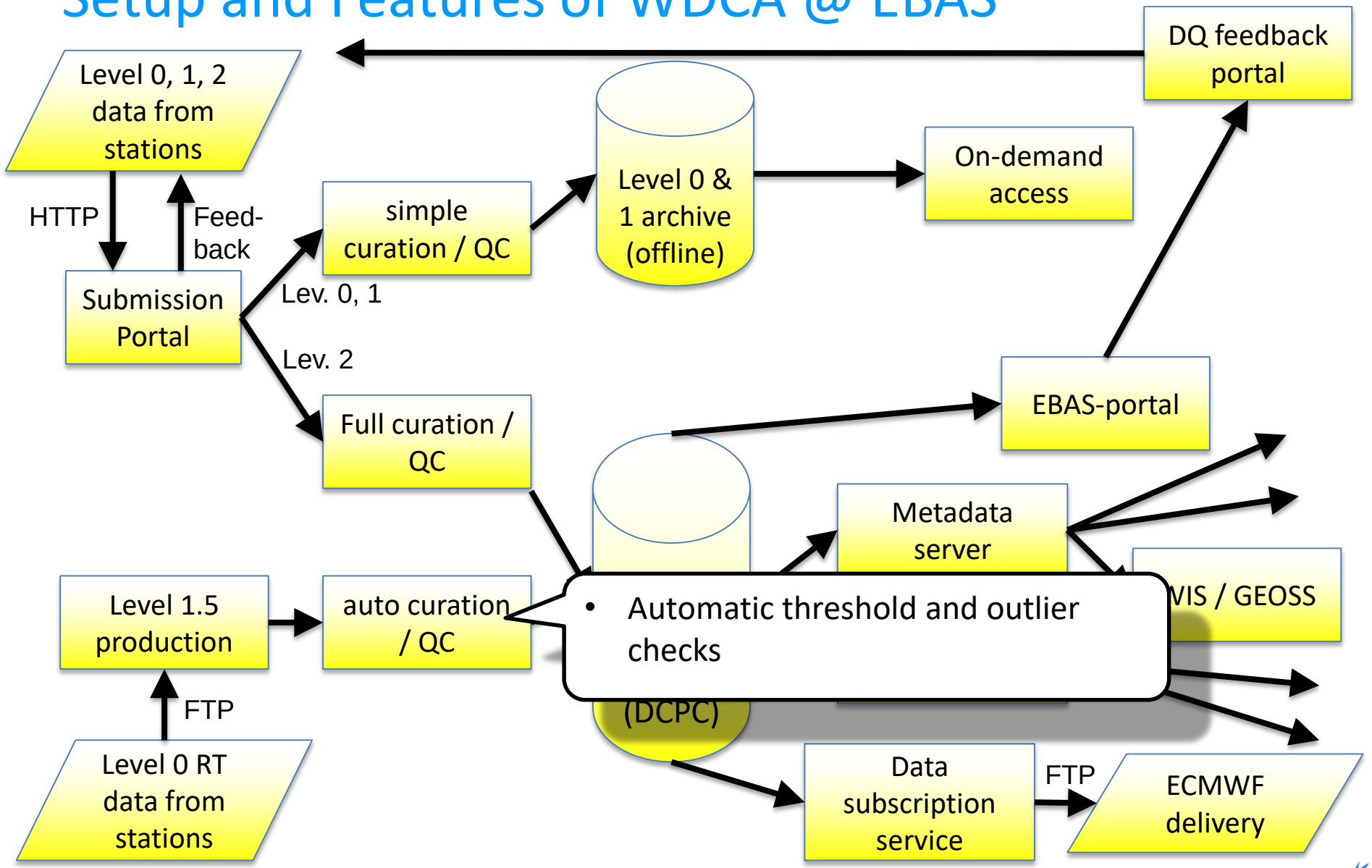
Setup and Features of WDCA @ EBAS



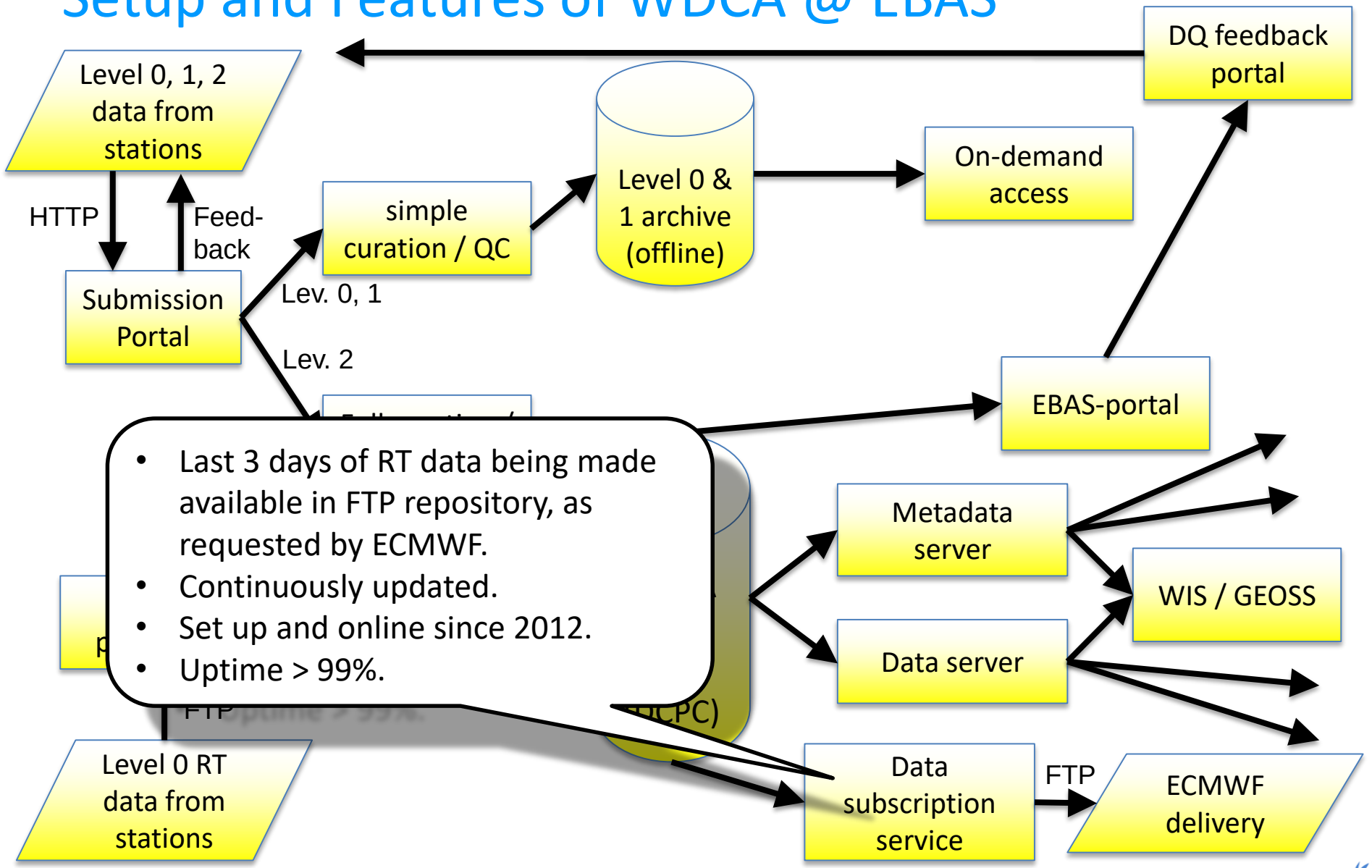
Setup and Features of WDCA @ EBAS



Setup and Features of WDCA @ EBAS

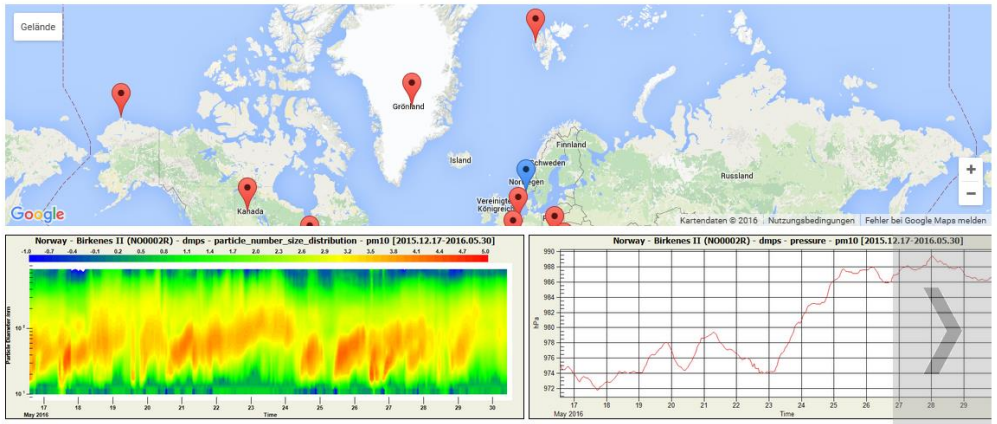


Setup and Features of WDCA @ EBAS



Setup and Features of WDCA @ EBAS

Latest Near-Real-Time Data



This service has been funded or supported by the Norwegian Institute for Air Research (NILU), the EU research infrastructure ACTRIS (Aerosols, Clouds, and Trace gases Research InfraStructure), the European Monitoring and Evaluation Programme (EMEP), and the WMO Global Atmosphere Watch (GAW) programme.

GLOBAL ATMOSPHERE WATCH

NILU Norsk institutt for luftforskning Norwegian Institute for Air Research

ACTRiS

- Graphical display of last 2 weeks of data for all RT datasets.
- Selectable by station on a map.
- Whole page, or single graphs, can be linked to for use on other websites.

Observations with Reporting Support

Regular / Traceable:

- Particle number concentration
- Particle number size distribution (sub-micron) **(NRT)**
- Cloud Condensation Particle Number Conc. / Size dist.
- Scattering Coefficient **(NRT)**
- Absorption Coefficient **(NRT)**

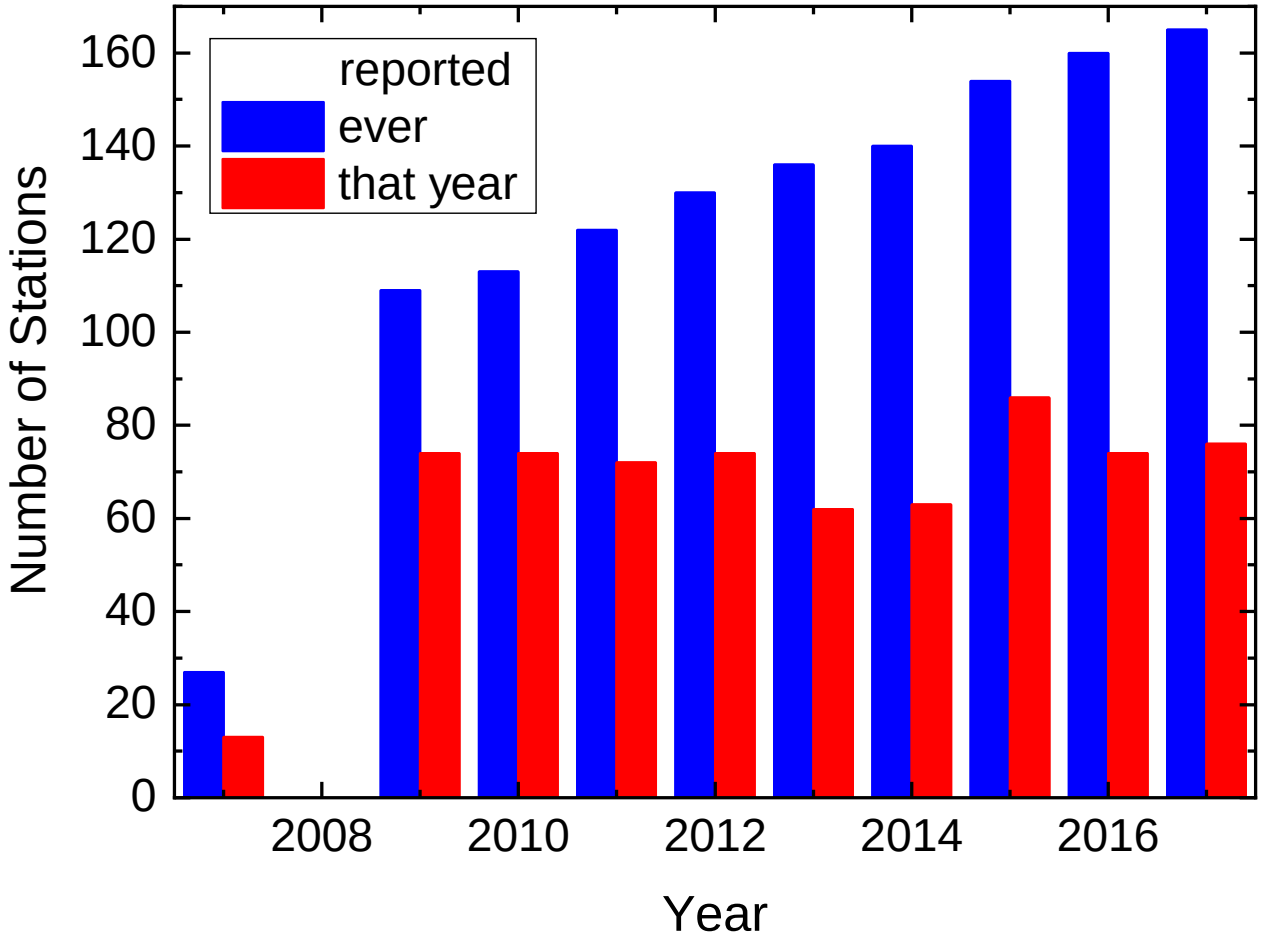
Regular only:

- Aerosol optical depth **(NRT)**
- PM mass (gravimetric)
- PM mass (online)
- Particle number size distribution (super-micron, OPC, APS)
- Aerosol Chemical Composition (GAW standard, co-ordinated with EMEP)
- Aerosol Chemical Speciation (online, AMS / ACSM)
- OC / EC concentration

recently added:

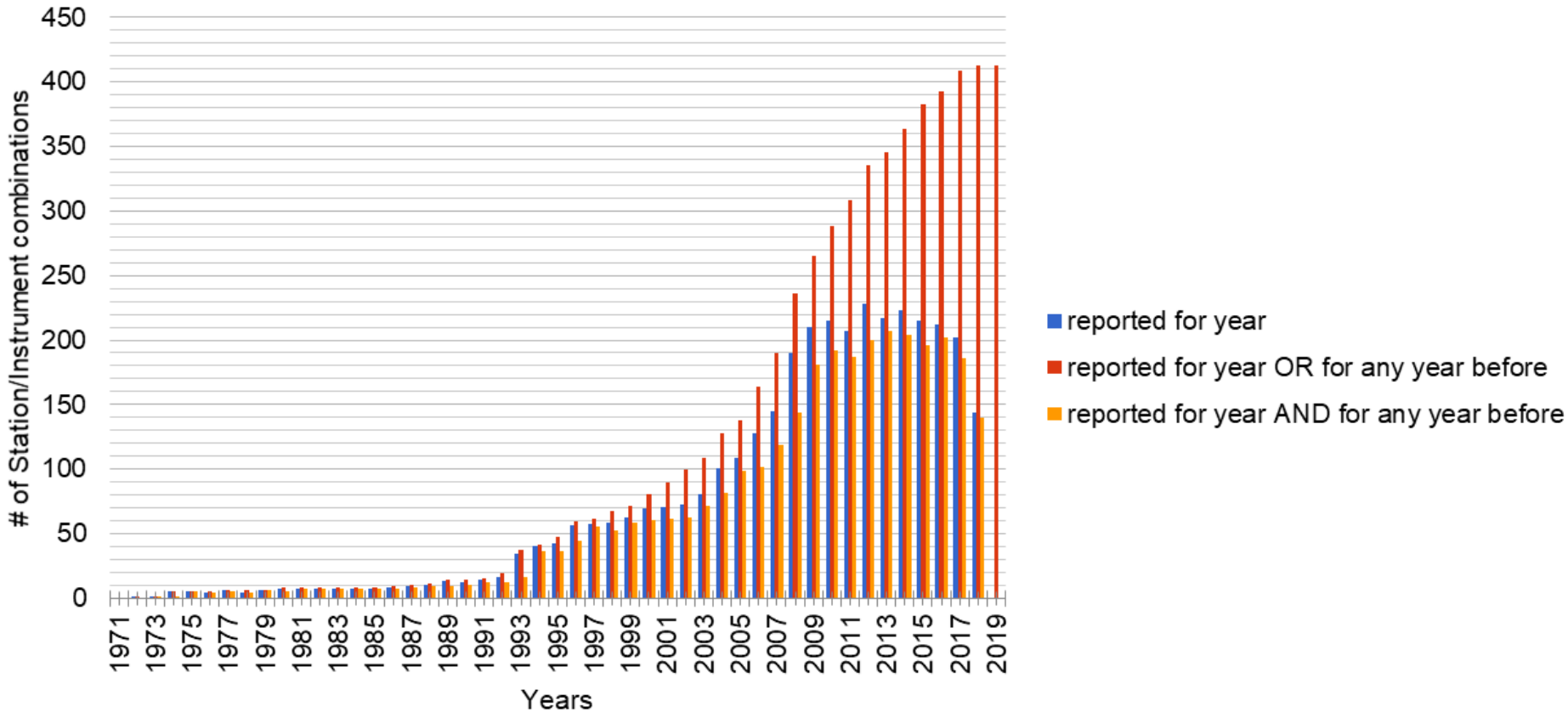
- Met. base paramaters

Stations Reporting, by Year



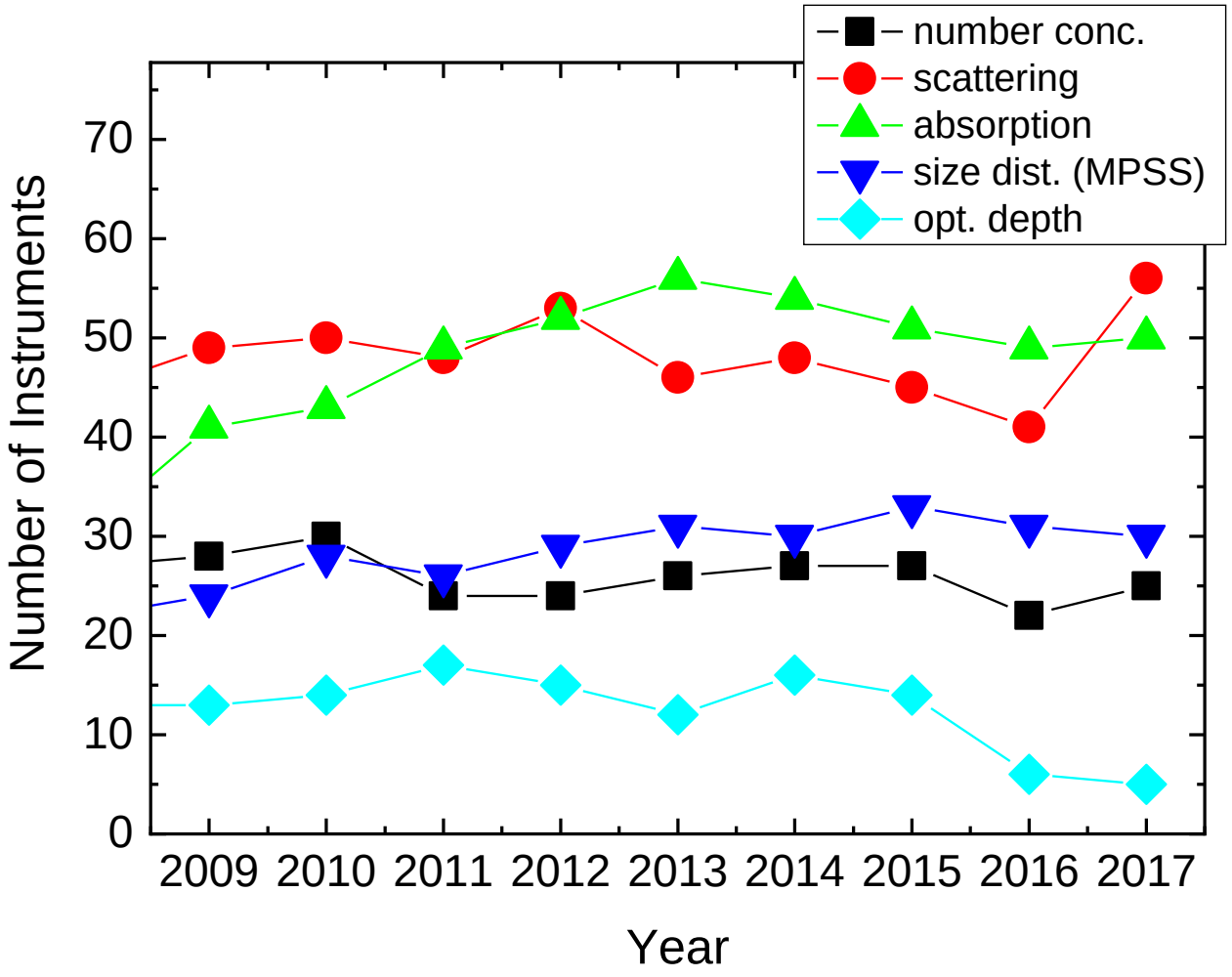
- Stable basis of ≈ 80 reporting stations
- Some fluctuation in stations

Station / Instruments Reported, by Year



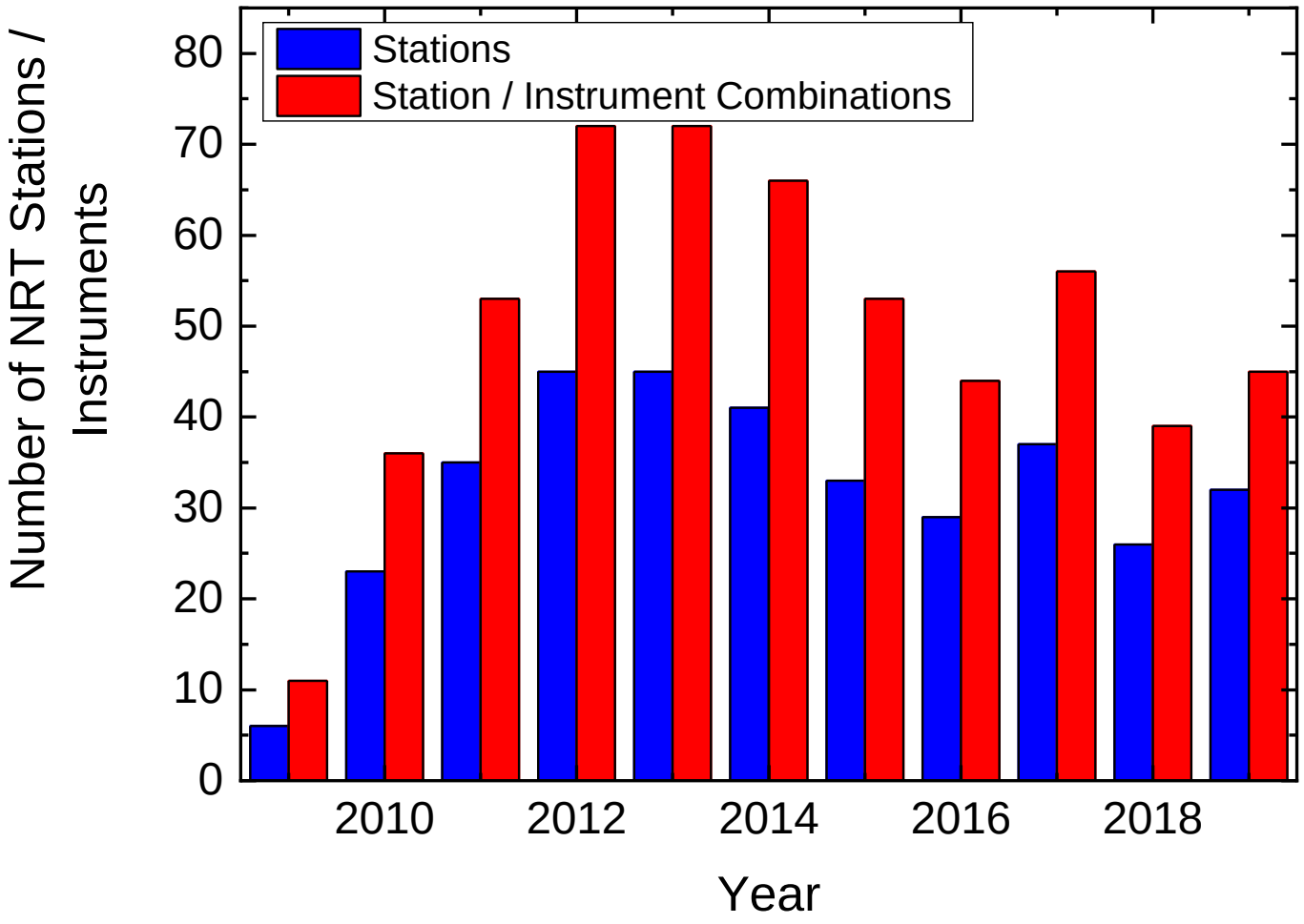
Stable at about 200 Instruments reporting per year.

Instruments Reporting, by Year



- Most frequent instrument types measure particle scattering and absorption, i.e. optical aerosol properties.
- Reflects GAW aerosol focus on climate effects.
- Stable numbers.

Number of RT Stations & Instruments, by Year

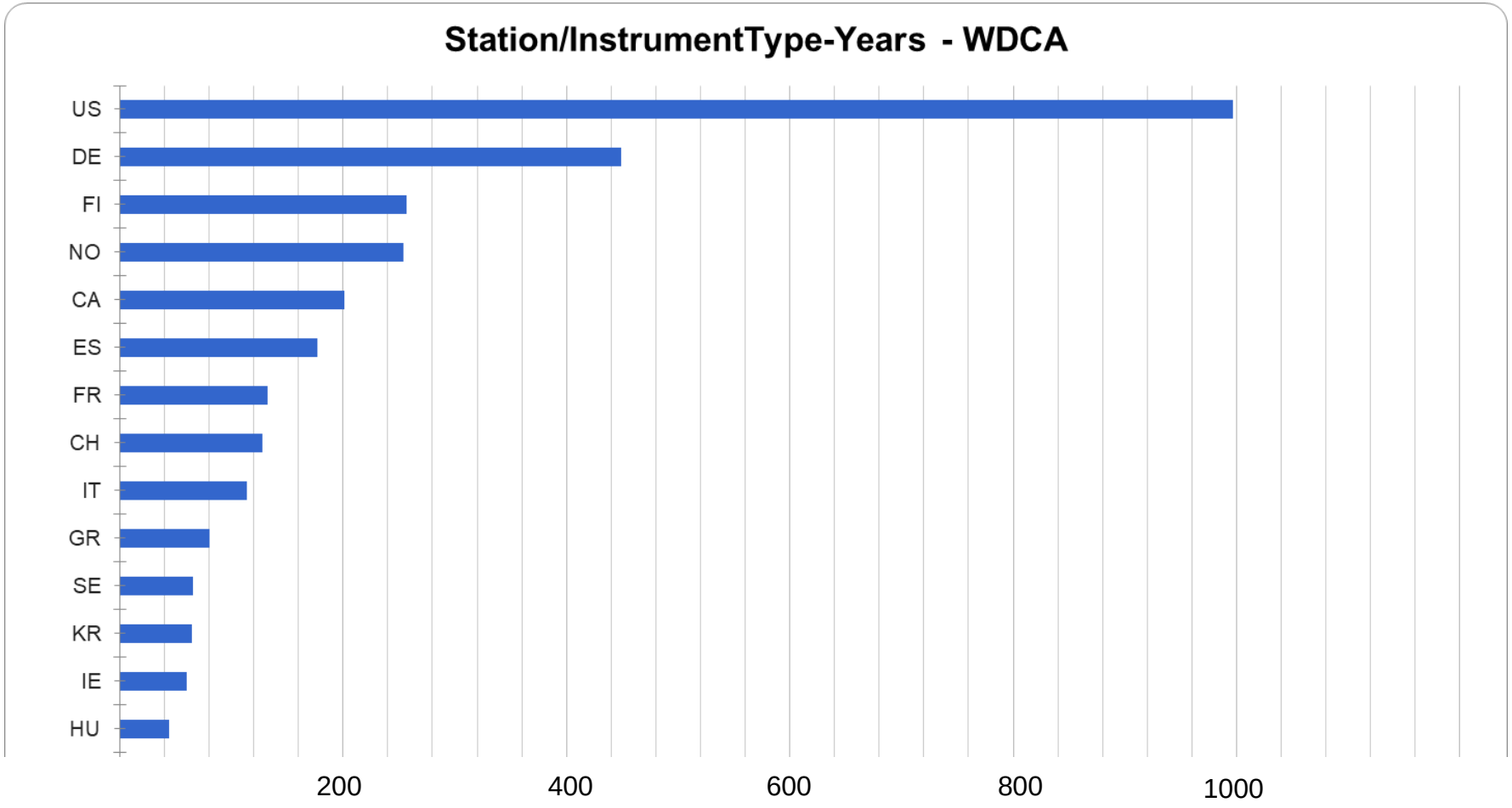


- WDCA has been providing a stable RT data service for 10 years.
- Stable number of participating stations and instruments.
- Decrease in 2015 due to GAW AOD instruments.

Coverage for: RRT Instruments, Active Stations

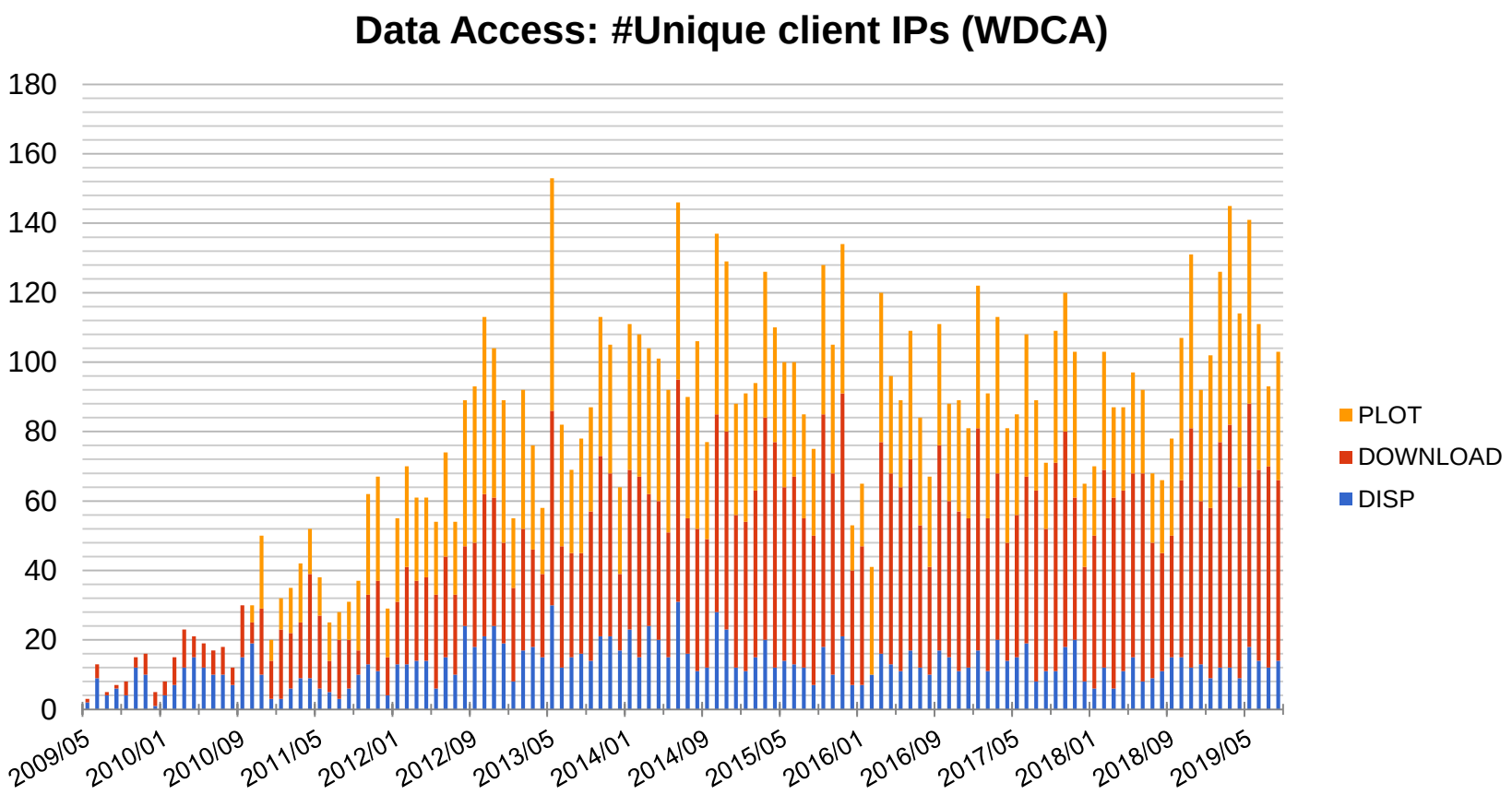
- Global coverage with surface in situ aerosol observations.
- 45 instruments at 32 stations
- “White areas” in South America, Africa, Central and South-West Asia.

Station / Instrument Years Provided, per Country



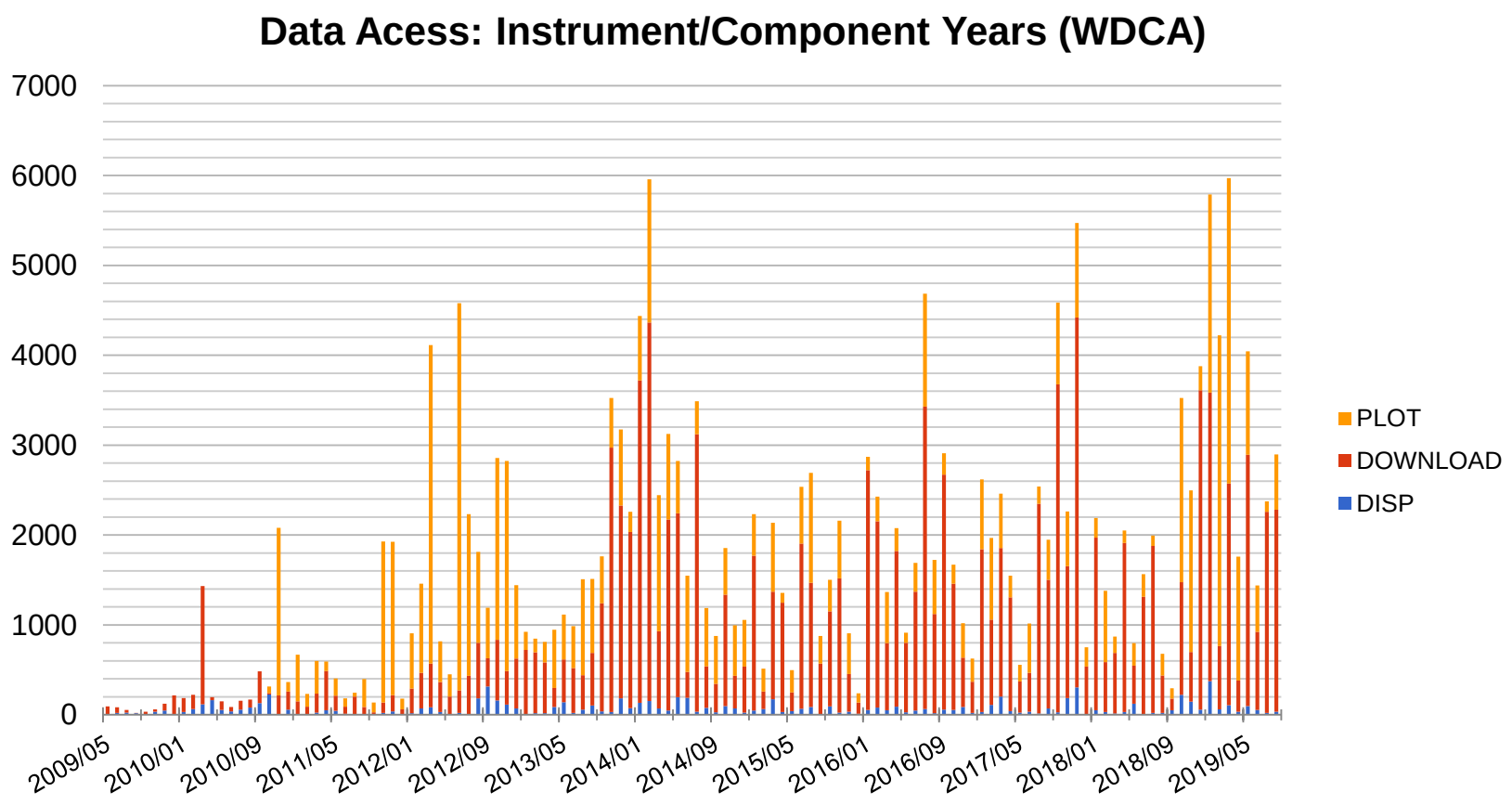
- Graph shows only part of 44 countries in total.
- Largest contribution by US (NOAA), 997 instrument years.
- Due long-running own station network.
- European countries contribute ≈2000 instrument years.
- In total 3544 instrument years.

Data Access: Unique Client IPs per Month



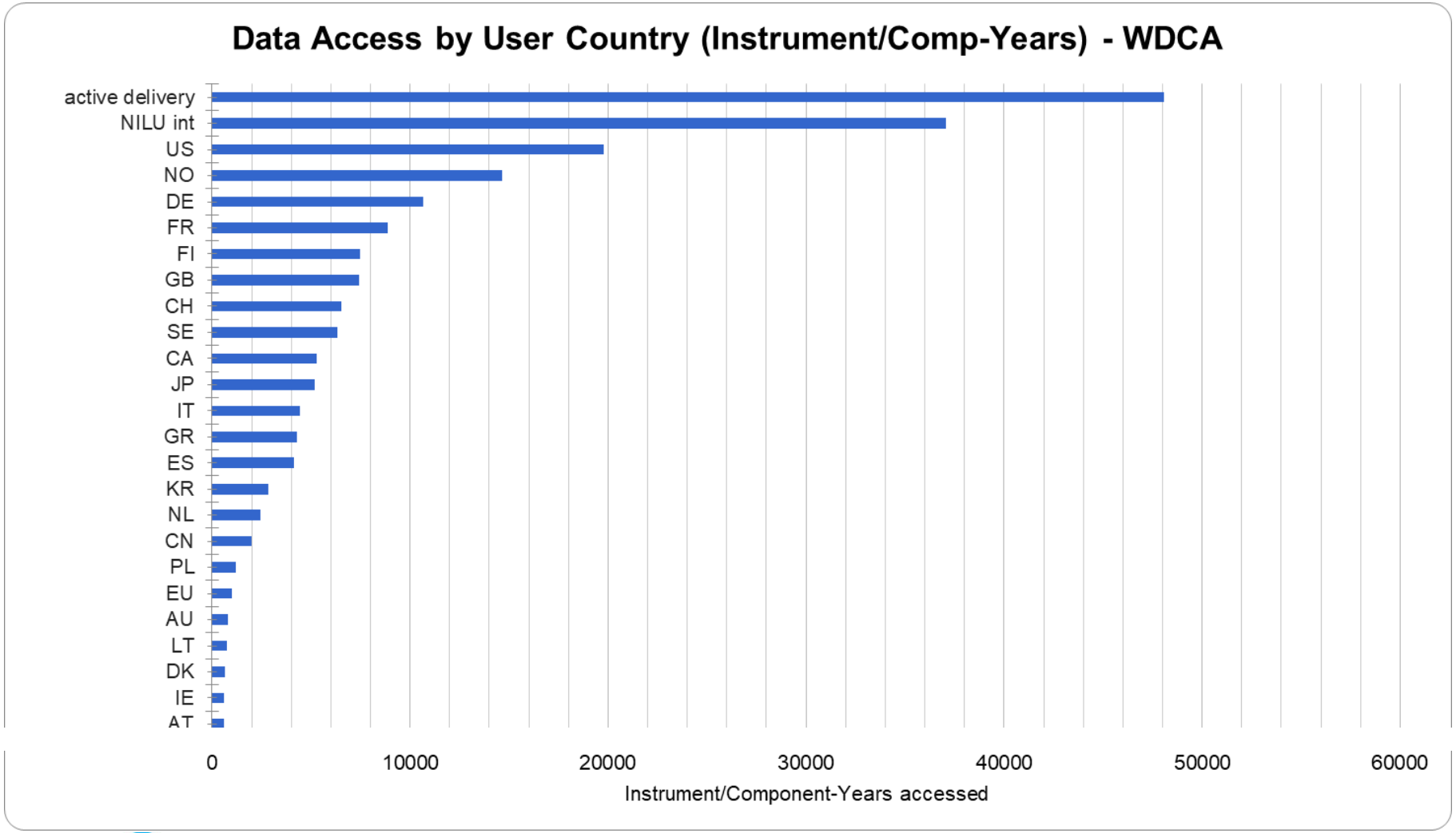
- Client IPs closest proxy to number of users w/o requiring login.
- One IP may hide 10s++ users (e.g. NOAA has 1 IP)
- One user may use different IPs (office, home, ...)
- Between 80-120 unique IPs per month.

Data Access: Instrument/Component Years per Month



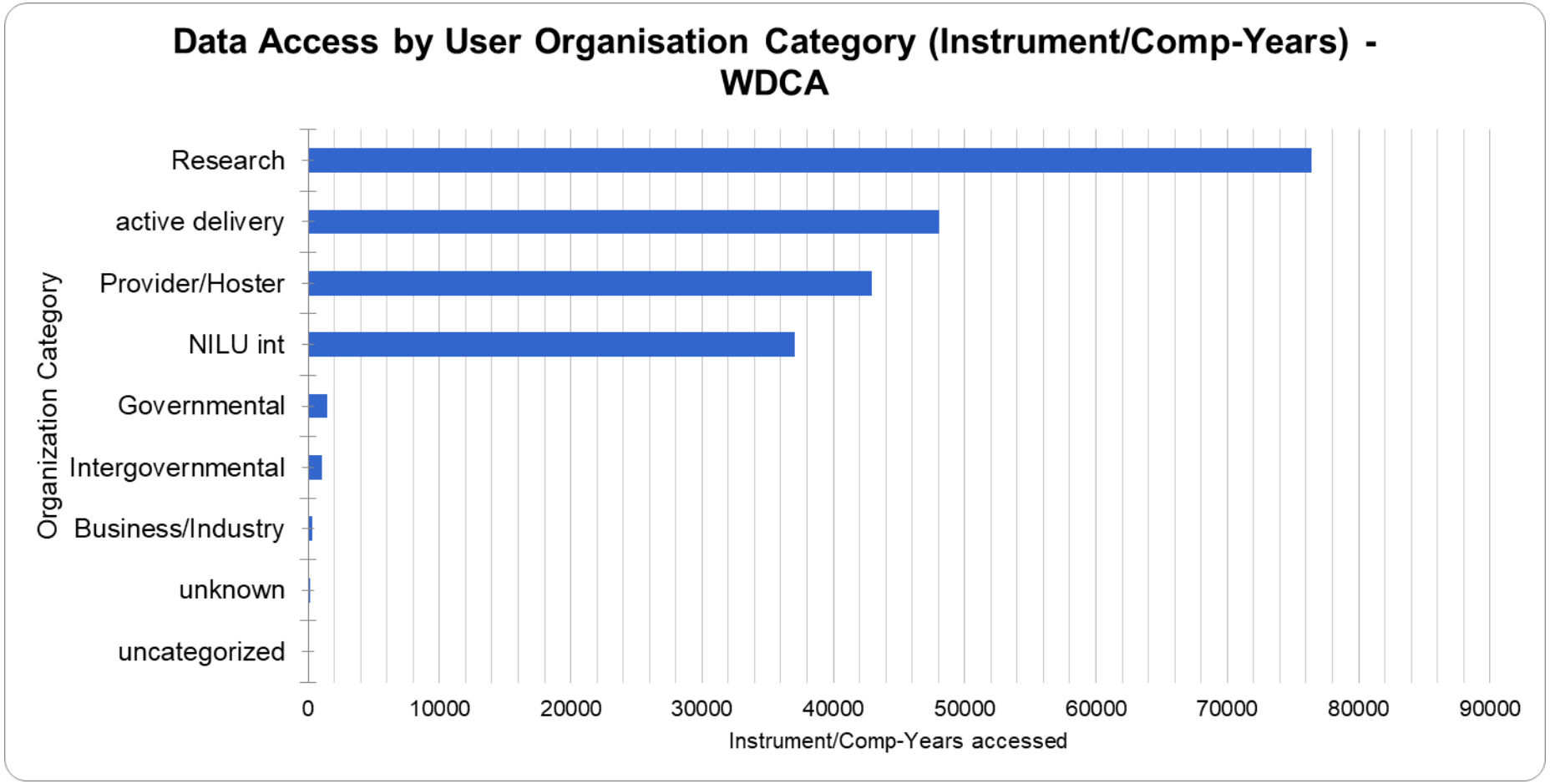
- On average, 1700 instrument years accessed each month.
- Modulated by holidays (lows) and large climatology studies (highs).

Data Access: Instrument/Component Years by Country Ever



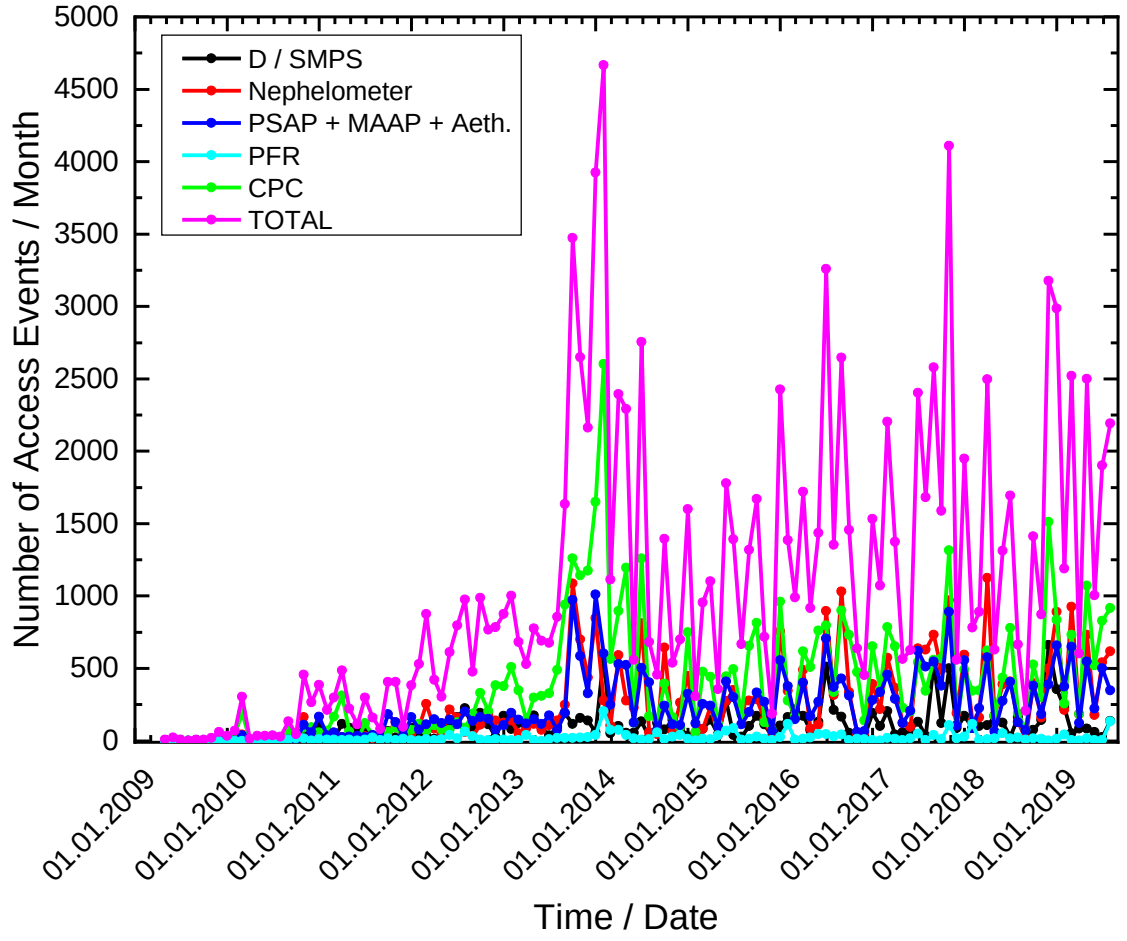
- Active delivery: delivery to collaborating initiatives, e.g. AeroCom
- Largest providers are also largest users.
- 72 user countries in total.

Data Access: Instrument/Component Years by Organisation Type



- Research largest user group.
- Provider / hoster: private access points
- Active delivery and provider / hoster also mostly reasearch

Data Access: Access Events per Month by Instrument Type



- Most access events for CPC (particle number concentration) and nephelometer (scattering coeff.)

Involvement in Ongoing EU-Projects

ENVRI-FAIR:

- Umbrella project for European environmental research infrastructures (RIs).
- Collaboration between topical data managers and information scientists.
- Objective to implement data FAIRness across Ris.

NextGEOSS:

- EU project providing European contribution to GEOSS.
- European GAW components participate.
- Work on inter-operability, i.e. standards for sharing metadata and data (interfaces, vocabulary, …).
- Work on pilot of a Cold Region data portal.

What Are the FAIR Principles?

Designed by FORCE 11 (open data advocate-) group to describe the general requirements that “Open Data” should meet.

Consist of four main points. Data should be:

- **F**indable:
  - **(meta)data have globally unique and persistent identifier**
  - **data are described with rich metadata**
  - **(meta)data are indexed in a searchable resource**
  - **metadata specify the data identifier**
- **A**ccessible
  - **(meta)data are retrievable by their identifier using a standardized communications protocol**
  - **metadata are accessible, even when the data are no longer available**.
- **I**nteroperable
  - **(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.**
  - **(meta)data use vocabularies that follow FAIR principles.**
  - **(meta)data include qualified references to other (meta)data**
- **R**e-usable
  - **(meta)data are with data usage license.**
  - **(meta)data are associated provenance.**
  - **(meta)data meet domain-relevant community standards**

ENVRI-FAIR Atospheric Subdomain Implementation Plan

Tasks for immediate implementation:

1. Consolidation of consistent use of PIDs throughout data production workflow.
   - DOIs for final data products
   - Persistent Identifiers for eResearch (ePIC) PIDs for other identifiable items
2. Common standard interfaces for metadata and data
   - At interface level!
   - **Metadata:** 1) Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) interface serving metadata in ISO19115 format with WMO Information System (WIS) profile; 2) • Open Geospatial Consortium (OGC) Catalogue Service for the Web (CSW)
   - **Data**: depending on data type: OPeNDAP, OGC-WCS, OGC-WMS
3. Indexing of data resources in WIS, GEOSS
   - Priority to these, as compared to e.g. Google Dataset Search.

ENVRI-FAIR Atospheric Subdomain Implementation Plan

Tasks for immediate implementation planning:

1. Domain vocabulary / ontology for observed parameters, discovery and use metadata.
   - Dataset-centric scope
   - Use existing, widely used, well-maintained vocabulary wherever possible
   - Identify gaps and extend where necessary
2. Common use of authentication schemes
   - Candidates: Google ID, eduGAIN (GÉANT), B2ACCESS (EUDAT), ORCiD
3. Consistent documentation of provenance throughout data production workflow
   - Based on deliverable of ENVRI+ predecessor project.
4. Recommendations for licenses on metadata and data
   - Provide overview of (meta)data licenses currently in use in atm. subdomain
   - Provide recommendation.
5. Semantic search for atmospheric ENVRI RI user interfaces
   - Describing and reviewing the state of the art in semantic search interfaces
   - Describing typical use case scenarios and their demands
   - Reviewing and giving recommendations on implementation technologies.

An Approach To Data Identification and Use Quantification

DIOs can be used in 2 ways (at least):

1. Identification of all data archived at fixed granularity in primary archive: provides quantifiable credit to data provider.
   ISSUE: Granularity varies
2. Identification of user selected data collections:
   provides ease of use of data

Approach in 3 Functionalities:

Primary identification of all data in a repository with homogeneous granularity, resolution fine enough to resolve single data originator

Type 2 DOIs need to link correctly to type 1 DOIs to facilitate correct accounting of data use – link in metadata.

Data use accounting service provided by primary archive by resolving collection DOIs for budgeting data use on primary DOI granularity.