Intra-ACP Climate Services and Related Applications Programme – ClimSA

WORKSHOP - SADC Region

Environmental Data Management Systems: ground legacy datasets and monitoring networks' data

Ezio CRESTAZ

South Africa, Johannesburg, June 10th – 13th 2024

Agenda

Presentations from JRC on:

- ❑ **State-of-the-art in environmental and water ground data management** (eg. location and time series as from legacy datasets, monitoring networks);
- ❑ **EMS (Environmental Monitoring System): design and implementation of an integrated Open (Source)* solution**, combina spatio-temporal database and application for effective data management.

Q&A and discussion session - Sharing experiences and views on adopted data management strategies and tools, key features and major bottlenecks

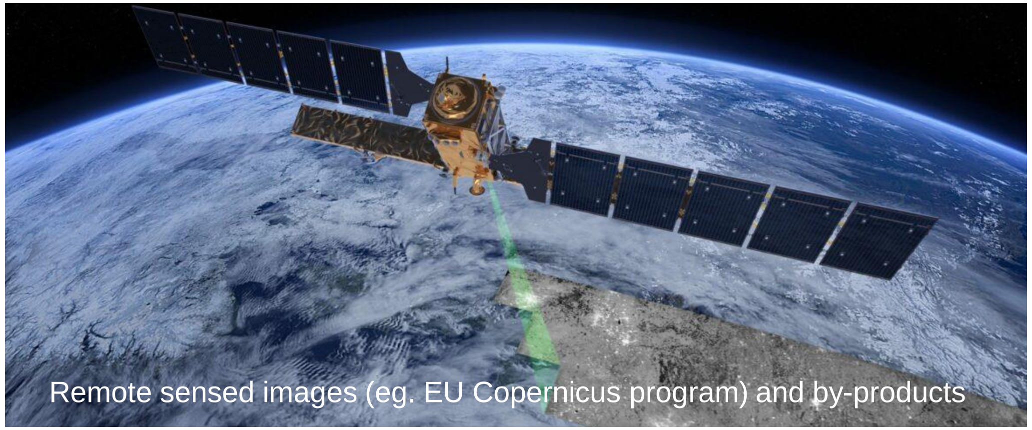
Wednesday	
9:00 - 09:45	Intoduction to WEFE
9.45 - 12:30	Case study: CV and agriculture in Southern Africa - Agriculture and WEFE - Bioenergy and Cropland allocation
12:30 - 14:00	*Lunch break*
	The EMS tool
14:00 - 17:30	Introduction: Environmental Management system
	Water data management, analysis and visualisation

* Further to being based upon Open Source technology, EMS database core component is also open to be integrated in most of leading OS and proprietary GIS, analysis and modelling environments

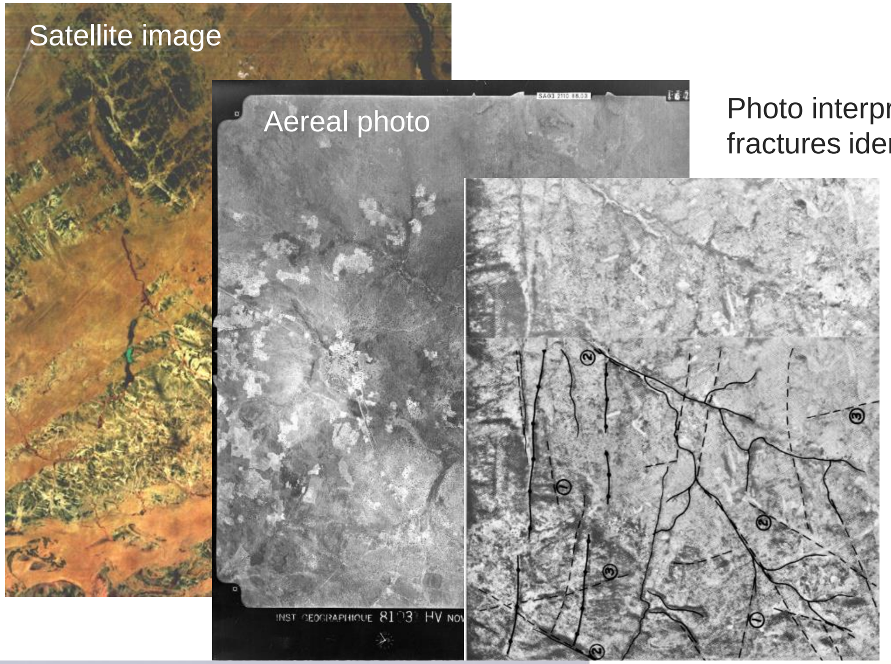
Objectives

- To introduce **challenges and opportunities in standardizing ground legacy datasets and monitoring networks' data**, to promote effective data management, analysis, modelling and visualization
- To provide an overview of **key concepts and state-of-the art in data validation, normalization and standardization**: Electronic Data Deliverables (EDDs, US/EPA - Environmental Protection Agency), data validation, data uploading to spatio-temporal database
- **To introduce the design and implementation of EMS, an Open Source Environmental Monitoring System** aimed at flexibly supporting sampling/monitoring stations and related time series (eg. meteo-climate, surface/groundwater hydrology, water withdrawal, environmental quality):
  - environmental spatio-temporal database
  - data validation, editing, visualization for spatio-temporal exploratory analysis applications

Environmental data: tensions in among scales, products and formats



Remote sensed images (eg. EU Copernicus program) and by-products



Satellite image

Aereal photo

Photo interpretation for fractures identification

Fractured rocks in sandstone and carbonate rocks, Mali and Burkina Faso

Violeto artesian well, Piauì (Brasil) 700 m deep

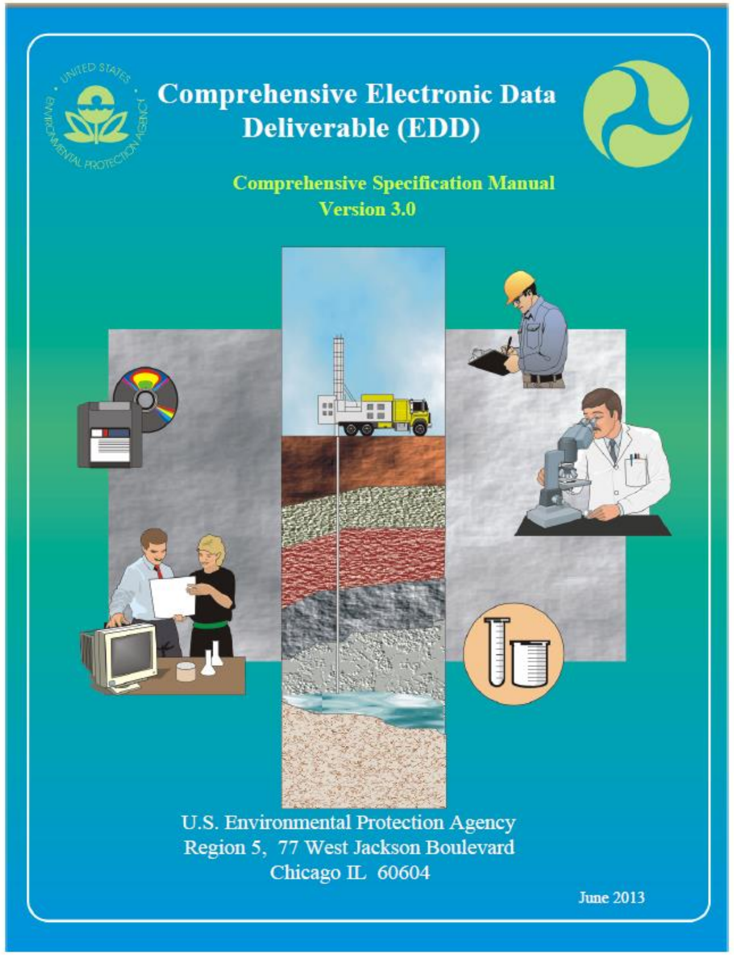
Senegal transboundary river basin WEFE nexus assessment, water sampling

Environmental data: challenges and opportunities in standardization

- ❑ Data file formats, different tools, backup issues
- ❑ Data duplication and conflicting risks
- ❑ Parameters naming conventions (eg. parameters naming, synonyms, internationalisation)
- ❑ Data formatting (eg. decimal separator, date/time standards)

Description		**P1**	**P2**	**P3**
Date d'échantillonnage		**15-02-2021**	**15-02-2021**	**15-03-2021**
Nom du paramètre	Unités			
Etilbenzene	µg/l	< 1	< 1	< 1
Stirene	µg/l	< 1	< 1	Non détecté
Toluene	µg/l	< 1	2.6	< 1
m,p-xylene	µg/l	< 0.3	< 0.3	< 0.3
metil t-butil etere (MTBE)	µg/l	19	7.8	6.8

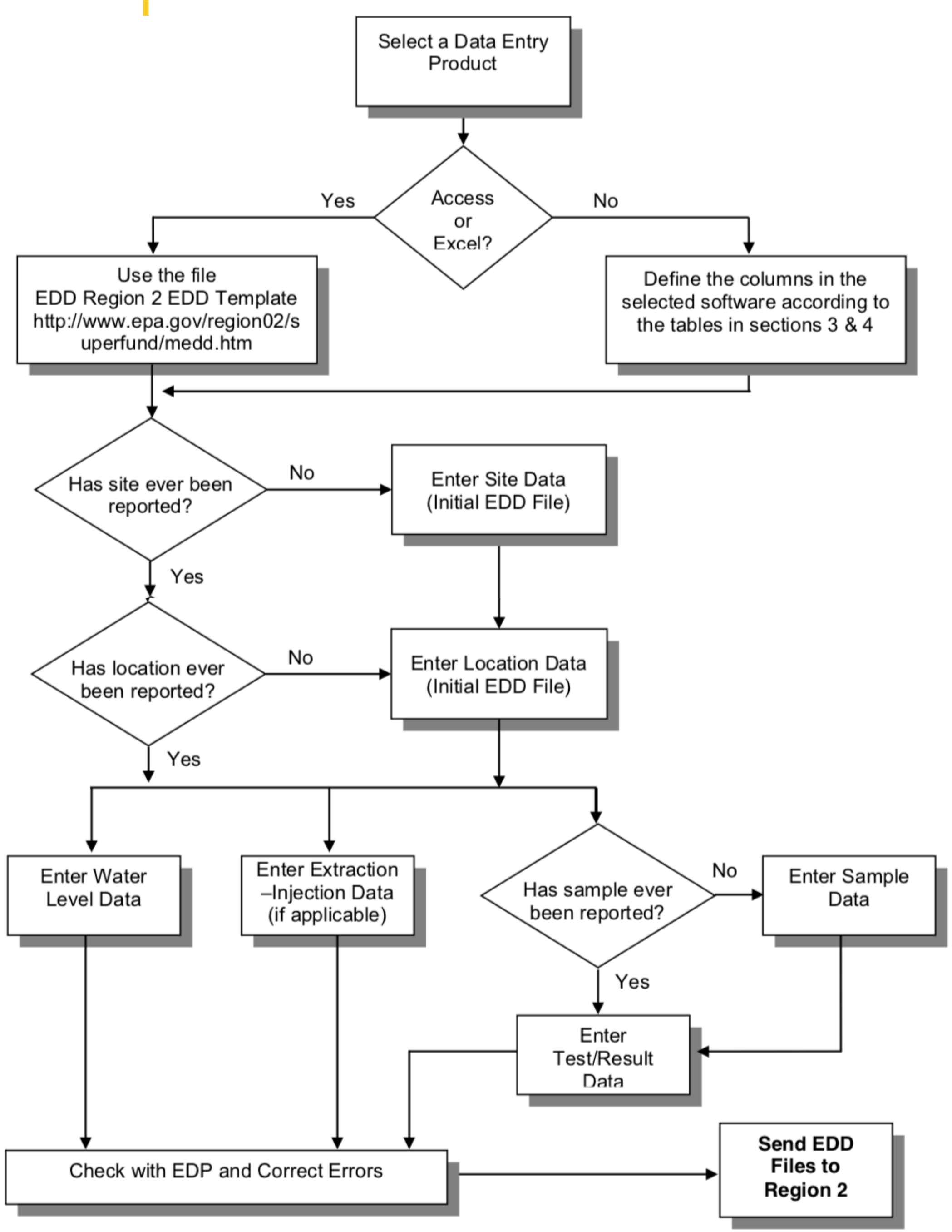
Data standardization: EDD (Electronic Data Transfer) concept



- Need for supporting **data transfer through distinct formats and information systems**
- **Use of simple text and spreadsheet** (eg. csv, MsExcel), format being standardized based on underlying database relational rules
- **Central management of constrains on acceptable values** (eg. VVLs Valid Value Lists, database domains)
- **Data validation, errors reporting and fixing**, at file level (and then at database level)
- **Flexibility in supporting a wide range of data** (meteo-climate, hydrology, hydrogeology, etc…)

https://www.epa.gov/sites/production/files/2016-01/documents/r2comprehensivemanual_sept2015.pdf

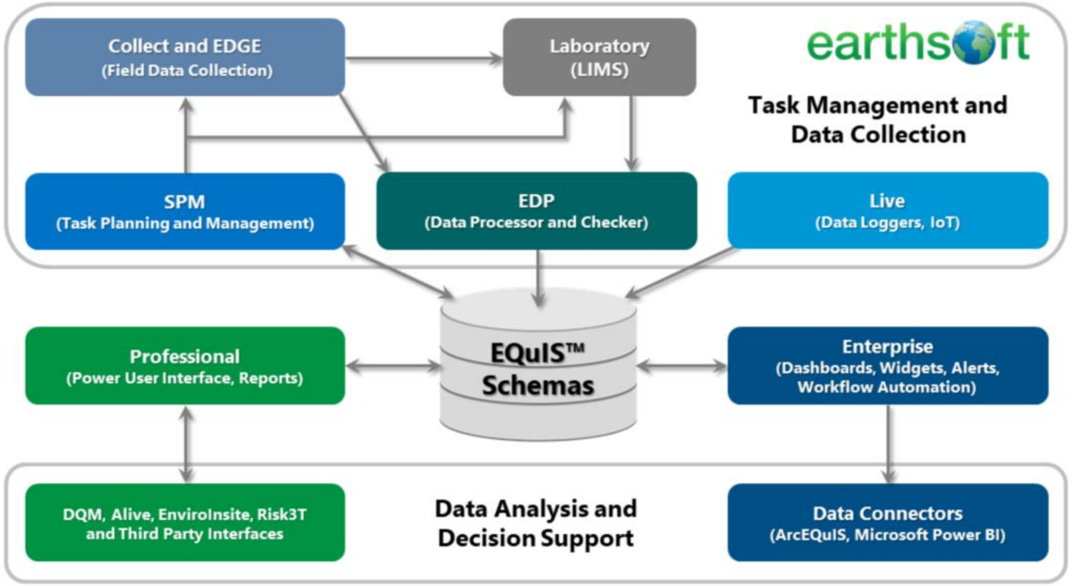
EDP – EDDs PROCESSING



- **Data compilation using standard templates and tools** (eg. multiple data formats as CSV, MsExcel, MsAccess)
- **Validate locational data** (eg. already reported? Consistent coordinates and geographic reference system?)
- **Validate time series data** (eg. already reported? Correct data formats as for dates, measures?)
- **Check/correct errors iteratively** until clean data set are consolidated
- **Validated data are transmitted and uploaded** to centralized database

https://www.epa.gov/sites/production/files/2016-01/documents/r2comprehensivemanual_sept2015.pdf

EDDs: proprietary solutions



EQUIS software suite from EarthSoft, firm specialized in data management solutions, with applications in sectors as mining, water quantity, water quality, groundwater

https://earthsoft.com/

Key features include:

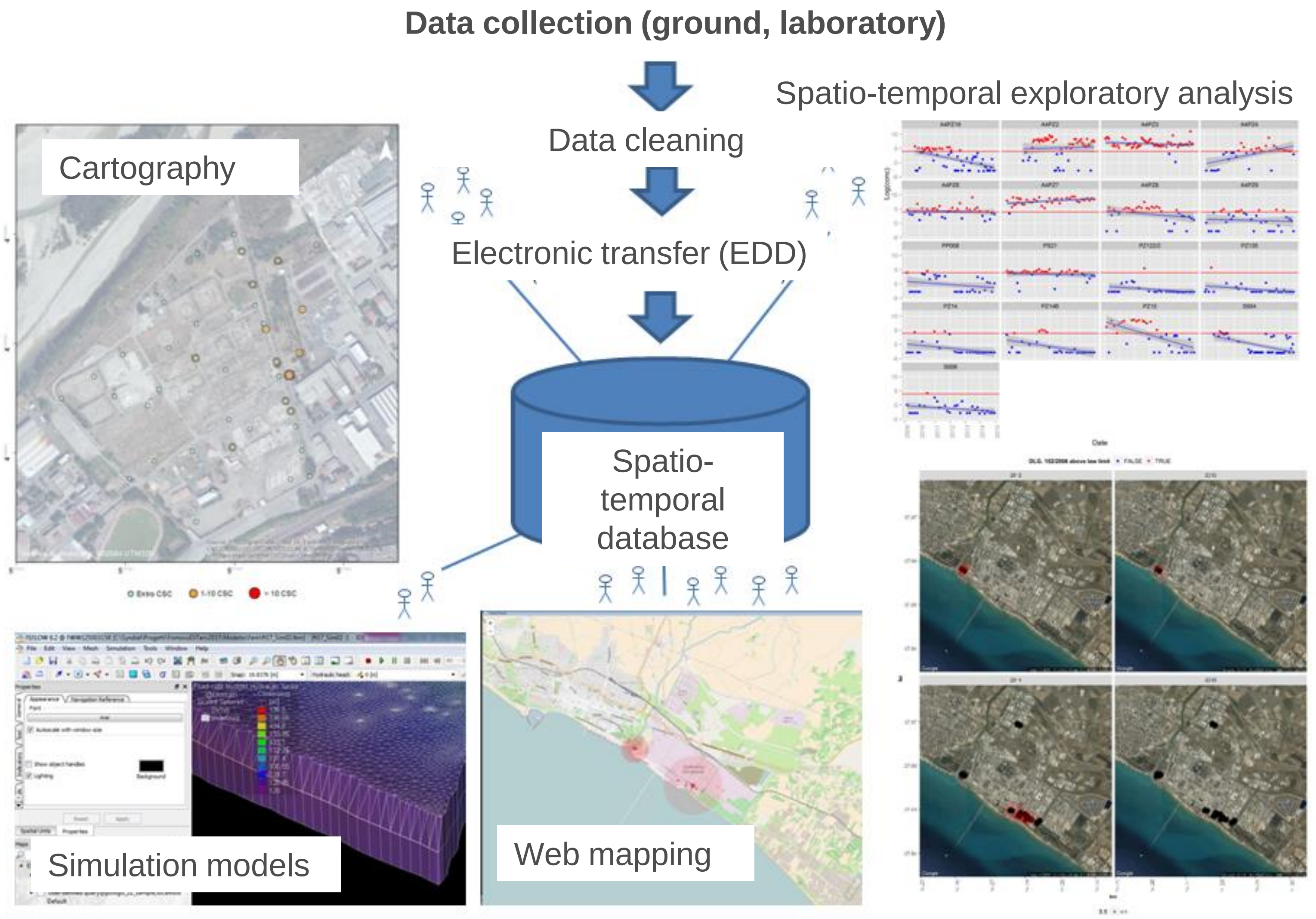
- Centralized spatio-temporal database, using ESRI geodatabase paradigms implemented on different advanced relational database platforms (eg. Oracle, PostgreSQL/PostGIS)
- Database automatically fed-in through live data generation (eg. data loggers, automated pre-processing piping)
- EDP (EDD processing) tools to validate and correct data
- Integration with professional data analysis and decision support systems, eg. GIS, dashboards, statistical packages

Q&A and discussion session – a few key questions

- How are legacy datasets and newly collected ground data validated/managed in your Organizations/Institutions? What is the process governance?
- What are adopted practices as for data cleaning, validation, standardization and management? Any major bottlenecks and improvement needs?
- Data sharing challenges: technical and beyond (OS data? confidentiality concerns? institutional mandates limiting data accessibility?)
- How databases and information systems are integrated in the climate, environmental and water analysis/modelling processes?

- Others …

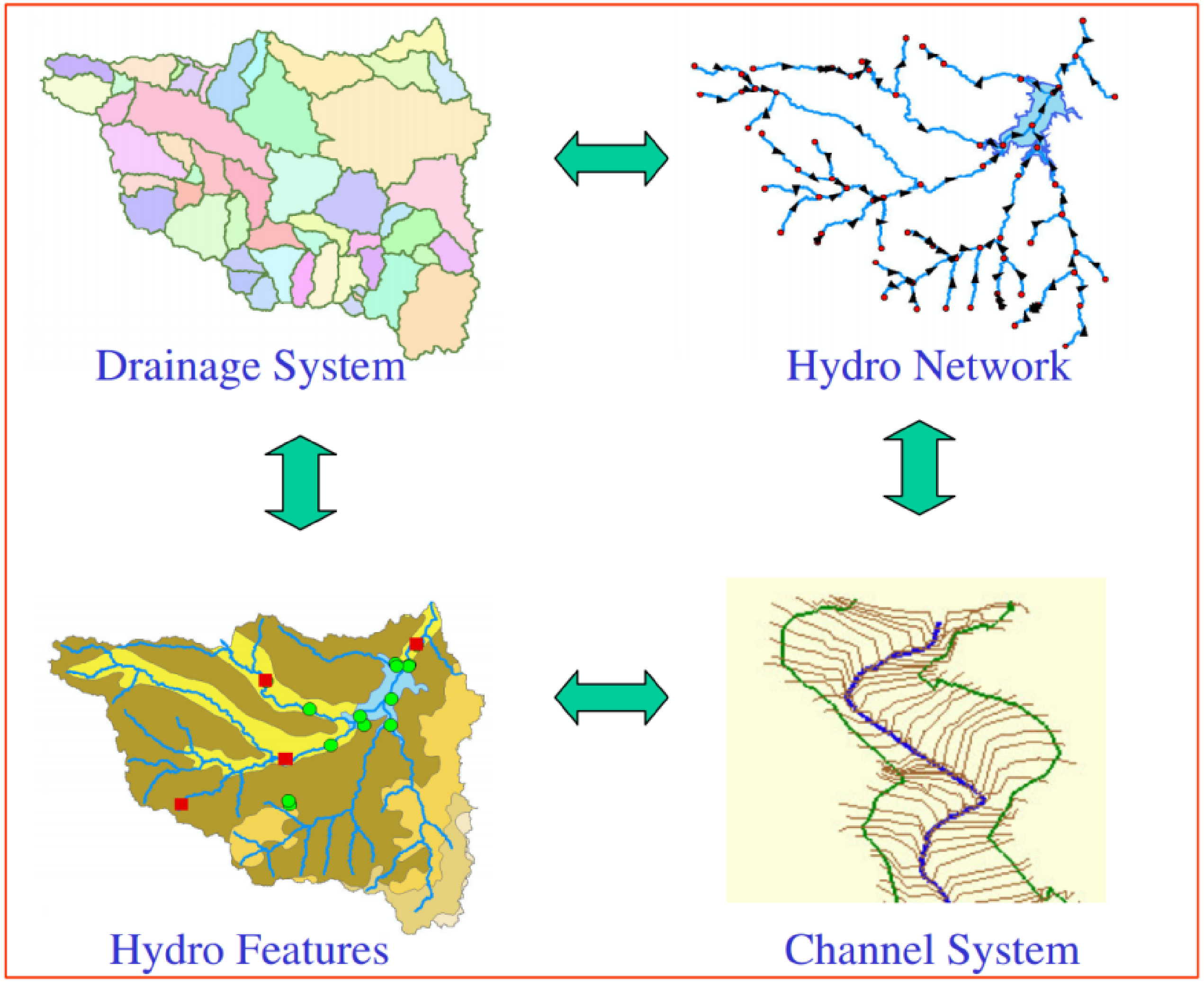
EMS (Environmental Monitoring System): general framework



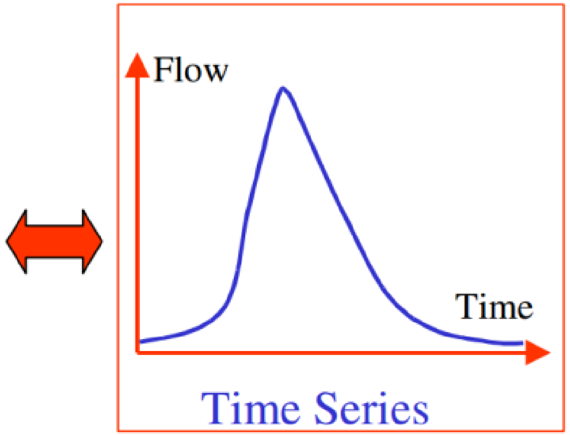
Key features include:

- ❑ Database centered system
- ❑ Integrated web apps to validate data, populate, edit and visualize database content
- ❑ Open Source technology
- ❑ Flexible, database easy to be integrated with other OS and proprietary tools (GIS, dashboards, modelling tools)
- ❑ Application installable as both desktop or web client-server

EMS database: foundation of its conceptual schema

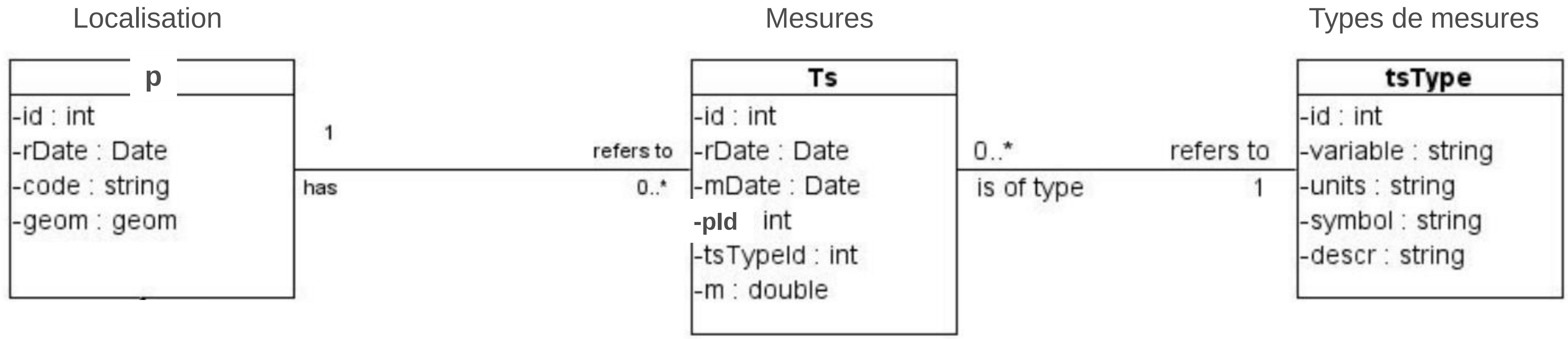


Hydro data model (CRWR, Un. of Austin, ESRI): the conceptual schema upon which the EMS database was designed and further extended



Maidment, D. R., 2002

EMS database: implementation (core tables)



- ❑ **p (monitoring geometries)**, location of any geometry type (eg. meteo-climate station, lake)
- ❑ **tsType (measurement types)**, hosting variable names, media and units
- ❑ **ts (time series)**, hosting monitoring data

Implemented using the **Open Source relational and spatial database platform PostgreSQL/PostGIS**, developed by M. Stonebraker in Berkeley (CA, USA), freely accessible since 1979, operationally used to support massive geographic databases.

Software and documentation available at https://www.postgresql.fr

EMS: database implementation (extended)

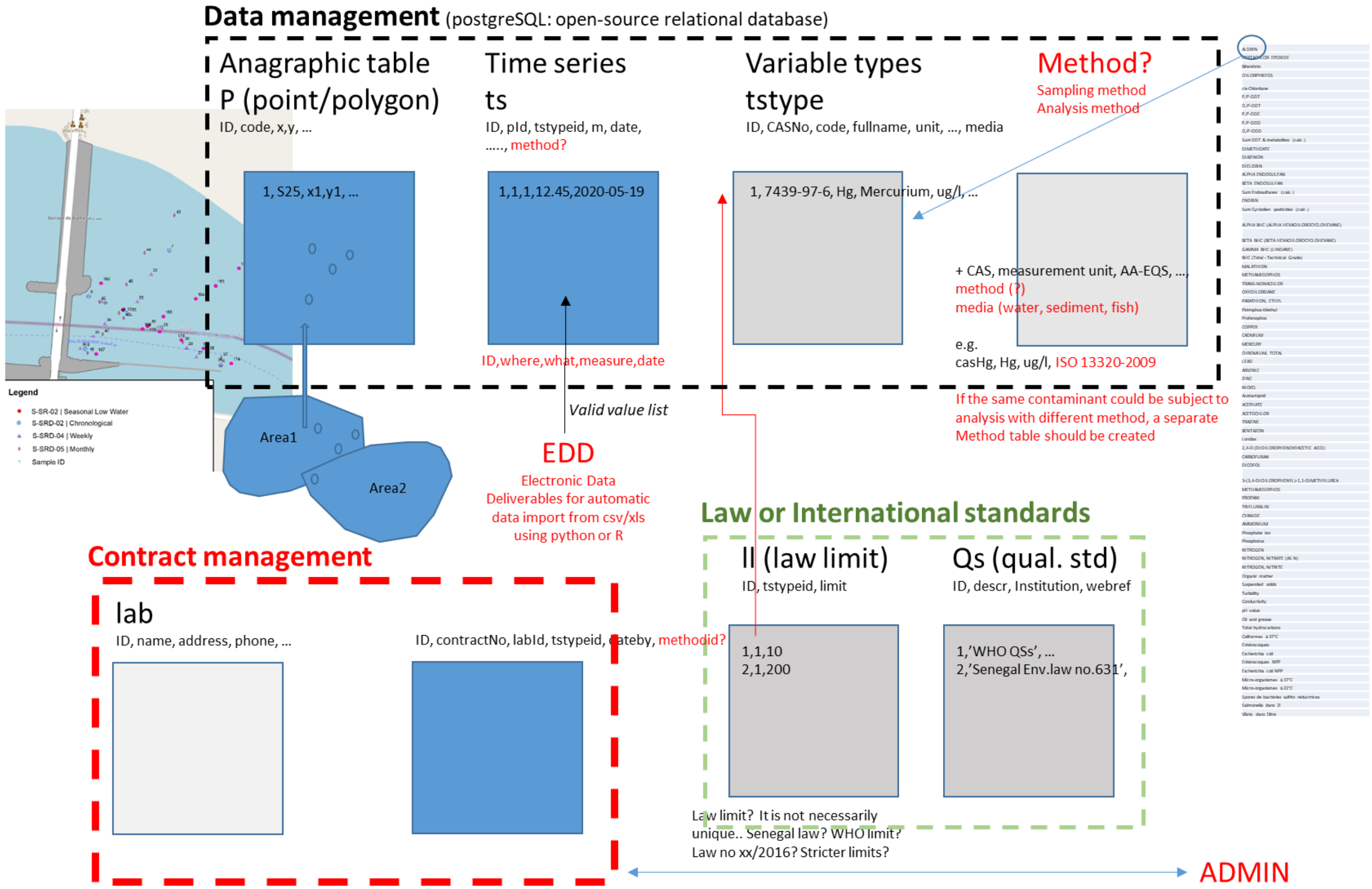
Water Quality Monitoring

<u>Key:</u>

Stand alone tool

Open Source

User-friendly



- Spatial data storage
- Flexible support to any relevant data (eg. meteo-climate, hydrology, quality)
- Support to environmental quality laws and international standards (eg. WHO, EC)
- Standardization of system tables
- Internationalization (multi-language support)
- Support long-term monitoring operationalization
- Extendable, eg. to cover contract management

EMS (Environmental Monitoring System): locational data validation

Data | Template | Result | [DB] Measurement types

Show 10 entries

Search:

	Pos	Column name	Data type	Required	Description	Valid values list
1	1	Code	TEXT (100)	Y	Monitoring point code	
2	2	X	NUMERIC	Y	x coordinate	
3	3	Y	NUMERIC	Y	y coordinate	
4	4	Srid	INTEGER	Y	Spatial reference ID (as from EPSG)	vvl_srid.csv
5	5	Type	TEXT (100)	Y	Type of monitoring point	vvl_type.csv
6	6	Note	TEXT (255)	N	Descriptive note	
7	7	Provider	TEXT(100)	Y	Locational data provider	

Showing 1 to 7 of 7 entries

Previous 1 Next

C:/Ezio/project/edd/app/cleangui - Shiny

http://127.0.0.1:7043 Open in Browser

WEFE SENEGAL

Projet d'Appui à la gestion des ressources en eau et du Nexus eau-énergie-agriculture dans le Bassin du Fleuve Sénégal

Guinée Mali Mauritanie Sénégal

OMVS

Select data file type

- Location
- Time series
- User defined

Select MsExcel data file:

Browse... No file selected

Select MsExcel template:

Browse... No file selected

Validate input file vs. template

Partial upload

Upload to the database

Data | Template

Show 10 entries

Search:

	Code	X	Y	Srid	Type	Note	Provider
1	AR0401	797306	1633620	32628	Monitoring point		CERES
2	AR0402	797283	1633652	32628	Monitoring point		CERES
3	AR0403	797232	1633741	32628	Monitoring point		CERES
4	AR0404	797219	1633795	32628	Monitoring point		CERES
5	AR0405	797085	1633954	32628	Monitoring point		CERES
6	AR0406	797063	1633958	32628	Monitoring point		CERES
7	AR0407	797209	1633914	32628	Monitoring point		CERES
8	AR0408	797208	1633876	32628	Monitoring point		CERES
9	AR0409	797225	1633810	32628	Monitoring point		CERES
10	AR0410	797236	1633727	32628	Monitoring point		CERES

Showing 1 to 10 of 338 entries

Previous 1 2 3 4 5 … 34 Next

EMS (Environmental Monitoring System): time series data validation

C:/Ezio/project/edd/app/cleangui - Shiny

http://127.0.0.1:3196 Open in Browser

WEFE SENEGAL

Projet d'Appui à la gestion des ressources en eau et du Nexus eau-énergie-agriculture dans le Bassin du Fleuve Sénégal

Choose MsExcel data file:

Browse... tmp2.xlsx

Upload complete

Select MsExcel template:

- Location [predefined]
- Time series [predefined]
- User defined

Browse... template_param.xlsx

Upload complete

Validate input file vs. template

Partial upload

Upload to the database | Preview

Data | Template | Result | [DB] Measurement types

Show 10 entries

Search:

	Pos	Column name	Data unit	Data type	Required	Description	Valid values list
1	1	Code		TEXT (20)	Y	Monitoring point code (eg. P01)	
2	2	Parameter		TEXT(100)	Y	Parameter (eg. indifferently code, cas or name)	vvl_param.csv
3	3	Media		TEXT(20)	Y	Anyone from media table (eg. Water)	vvl_media.csv
4	4	Unit		TEXT(20)	Y	Measurement unit from unit table (eg. m asl)	vvl_unit.csv
5	5	Date		DATE	Y	Measurement date and time in ISO8601 format (time being optional)	
6	6	Measure		TEXT(100)	Y	Measurement in text format. Conversion to number, if feasible, is managed internally by the database	
7	7	Note		TEXT(1000)	N	Note specific to each single measurement	
8	8	Provider		TEXT(100)	N	Provider code or name, from provider table	vvl_provider.csv

Showing 1 to 8 of 8 entries

Previous 1 Next

Data | Template | Result | [DB] Measurement t

Show 10 entries

	Code	Parameter	Media	Unit	Date	Measure	Note	Provider
1	BD0101	Temperature	Water	°C	2021-01-29	20.34		CERES
2	BD0101	pH	Water	unitless	2021-01-29	7.79		CERES
3	BD0101	Redox Potential	Water	mV	2021-01-29	147.6		CERES
4	BD0101	Suspended Matter	Water	ppm	2021-01-29	31		CERES
5	BD0101	Salinity	Water	psu	2021-01-29	0.0283		CERES
6	BD0101	Atmospheric Pressure	Water	psi	2021-01-29	14.666		CERES
7	BD0101	Dissolved Oxygen	Water	%	2021-01-29	127.8		CERES
8	BD0101	Dissolved Oxygen	Water	ppm	2021-01-29	11.51		CERES
9	BD0102	Temperature	Water	°C	2021-01-29	20.26		CERES
10	BD0102	pH	Water	unitless	2021-01-29	7.73		CERES

Showing 1 to 10 of 40 entries

Previous 1 2 3 4 Next

EMS (Environmental Monitoring System): data tables editing

Intro | Data Uploading | Data Editing | Dashboard | System Setup

Add Point | Edit Point | Delete | Refresh | Export | Info | Display selected

Monitoring Objects | TS Type | Time Series | Type | Provider | Language | Parameters | Media | Unit | Param. alternate | Family | Law | Use | Law Limit | Overall measures | Overall measures (with lawlimit)

☑ Visualise time series referring to points only

Measurement Type: ALDRIN (µg/kg) Fish

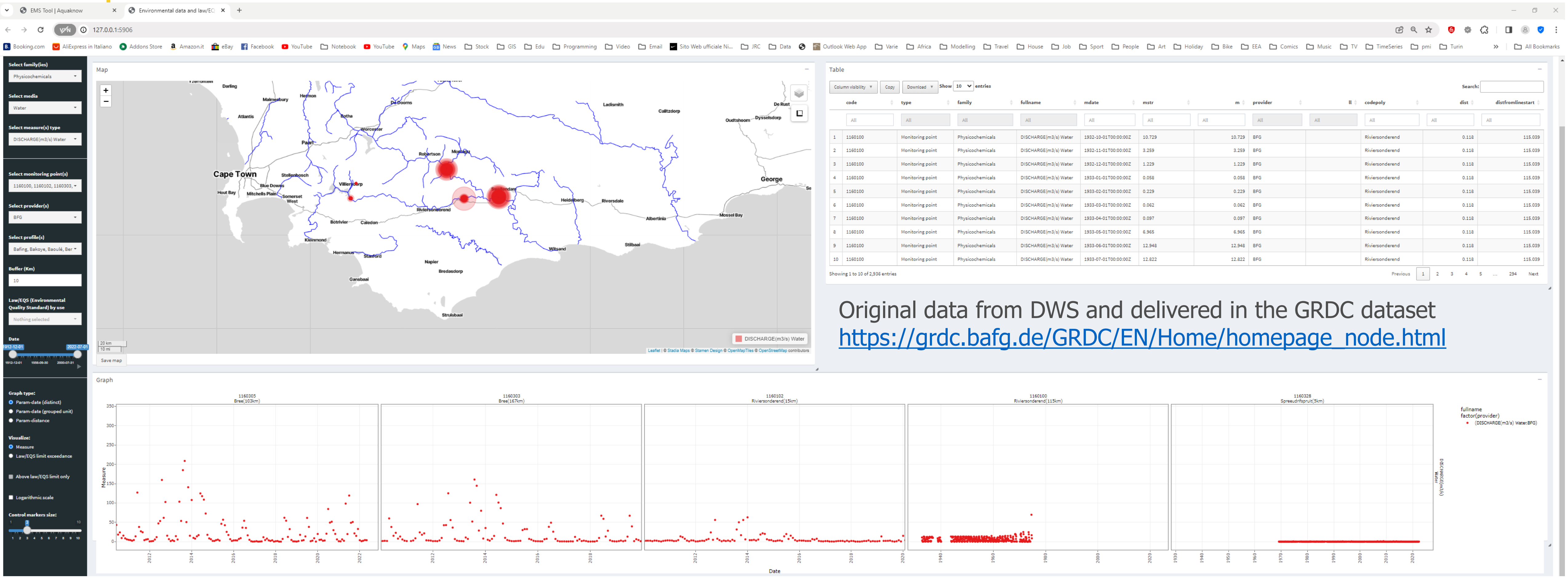
Show 10 entries — Search:

Object Code	Measurement Type	Recording Date	Measurement Date	Value str.	Value	Notes	Monitoring Frequency	Provider
All	All	All	All	All	All	All	All	All
ML09	ALDRIN (µg/kg) Fish	2024-02-24T19:08:44Z	2021-10-07T00:00:00Z	<2	1	reported code: 9 (Félou Aval); the <LOQ (Limit of Quantification) info was replaced by the LOQ-value provided in a separate table in the lab report (eg. <LOQ replaced by <0.0XX)	Day	LCV
ML07	ALDRIN (µg/kg) Fish	2024-02-24T19:08:44Z	2021-10-06T00:00:00Z	<2	1	reported code: 7 (Gourbassi); the <LOQ (Limit of Quantification) info was replaced by the LOQ-value provided in a separate table in the lab report (eg. <LOQ replaced by <0.0XX)	Day	LCV
ML12	ALDRIN (µg/kg) Fish	2024-02-24T19:08:44Z	2021-10-12T00:00:00Z	<2	1	reported code: 12 (Oualia); the <LOQ (Limit of Quantification) info was replaced by the LOQ-value provided in a separate table in the lab report (eg. <LOQ replaced by <0.0XX)	Day	LCV
ML05	ALDRIN (µg/kg) Fish	2024-02-24T19:08:44Z	2021-10-01T00:00:00Z	<2	1	reported code: 5 (Mahinamine); the <LOQ (Limit of Quantification) info was replaced by the LOQ-value provided in a separate table in the lab report (eg. <LOQ replaced by <0.0XX)	Day	LCV
ML10	ALDRIN (µg/kg) Fish	2024-02-24T19:08:44Z	2021-10-11T00:00:00Z	<2	1	reported code: 10 (Galougo); the <LOQ (Limit of Quantification) info was replaced by the LOQ-value provided in a separate table in the lab report (eg. <LOQ replaced by <0.0XX)	Day	LCV
ML03	ALDRIN (µg/kg) Fish	2024-02-24T19:08:44Z	2021-10-01T00:00:00Z	<2	1	reported code: 3 (Moussala); the <LOQ (Limit of Quantification) info was replaced by the LOQ-value provided in a separate table in the lab report (eg. <LOQ replaced by <0.0XX)	Day	LCV
ML06	ALDRIN (µg/kg) Fish	2024-02-24T19:08:44Z	2021-10-01T00:00:00Z	<2	1	reported code: 6 (Djidjan-barrage); the <LOQ (Limit of Quantification) info was replaced by the LOQ-value provided in a separate table in the lab report (eg. <LOQ replaced by <0.0XX)	Day	LCV
ML08	ALDRIN (µg/kg) Fish	2024-02-24T19:08:44Z	2021-10-07T00:00:00Z	<2	1	reported code: 8 (Ambidédi); the <LOQ (Limit of Quantification) info was replaced by the LOQ-value provided in a separate table in the lab report (eg. <LOQ replaced by <0.0XX)	Day	LCV
ML02	ALDRIN (µg/kg) Fish	2024-02-24T19:08:44Z	2021-09-30T00:00:00Z	<2	1	reported code: 2 (Daka-Saidou (Bafing Makana)); the <LOQ (Limit of Quantification) info was replaced by the LOQ-value provided in a separate table in the lab report (eg. <LOQ replaced by <0.0XX)	Day	LCV
ML01	ALDRIN (µg/kg) Fish	2024-02-24T19:08:44Z	2021-09-29T00:00:00Z	<2	1	reported code: 1 (Diangola); the <LOQ (Limit of Quantification) info was replaced by the LOQ-value provided in a separate table in the lab report (eg. <LOQ replaced by <0.0XX)	Day	LCV

Showing 1 to 10 of 24 entries — Previous 1 2 3 Next

European Commission

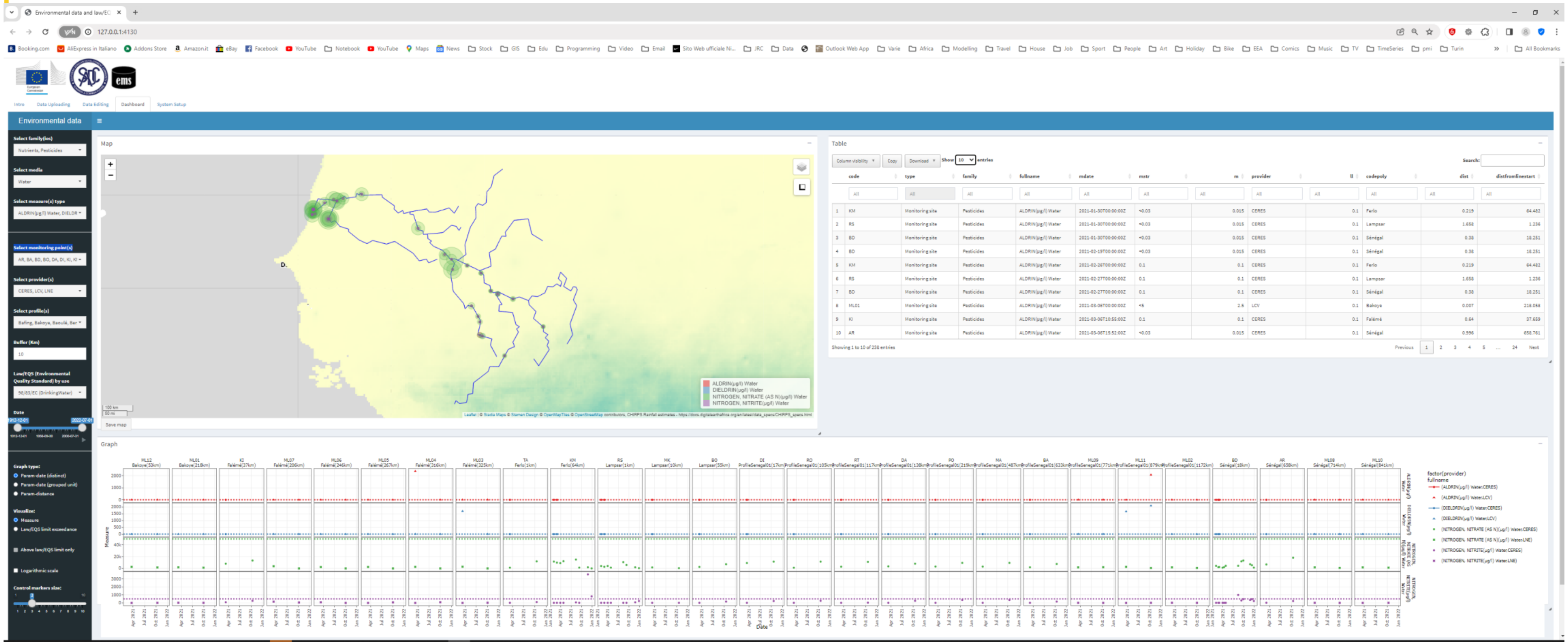
EMS: dashboard case study for hydrology in Cape Town region



Original data from DWS and delivered in the GRDC dataset
https://grdc.bafg.de/GRDC/EN/Home/homepage_node.html

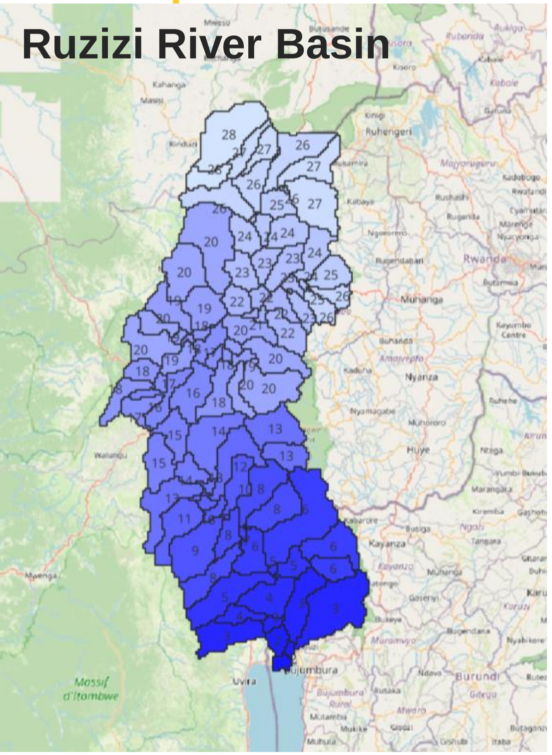
- ❑ Anomaly of low river discharges in the period pre-2018 Cape Town water crisis
- ❑ Low river discharges around Theewaterskloof dam (limited extent of upperstream watershed and dam managment practices)

EMS: dashboard case study Senegal TRB environmental quality



- ❑ Environmental quality monitoring network management and exploratory spatio-temporal analysis
- ❑ Exceedances of law/EQSs (Environmental Quality Standards)

Extending EMS: watersheds delineation and time series aggregation



Query upstream watersheds and aggregated CHIRPS from any starting watershed in QGIS or pgAdmin

1. Get watershed code from watershed100.tif

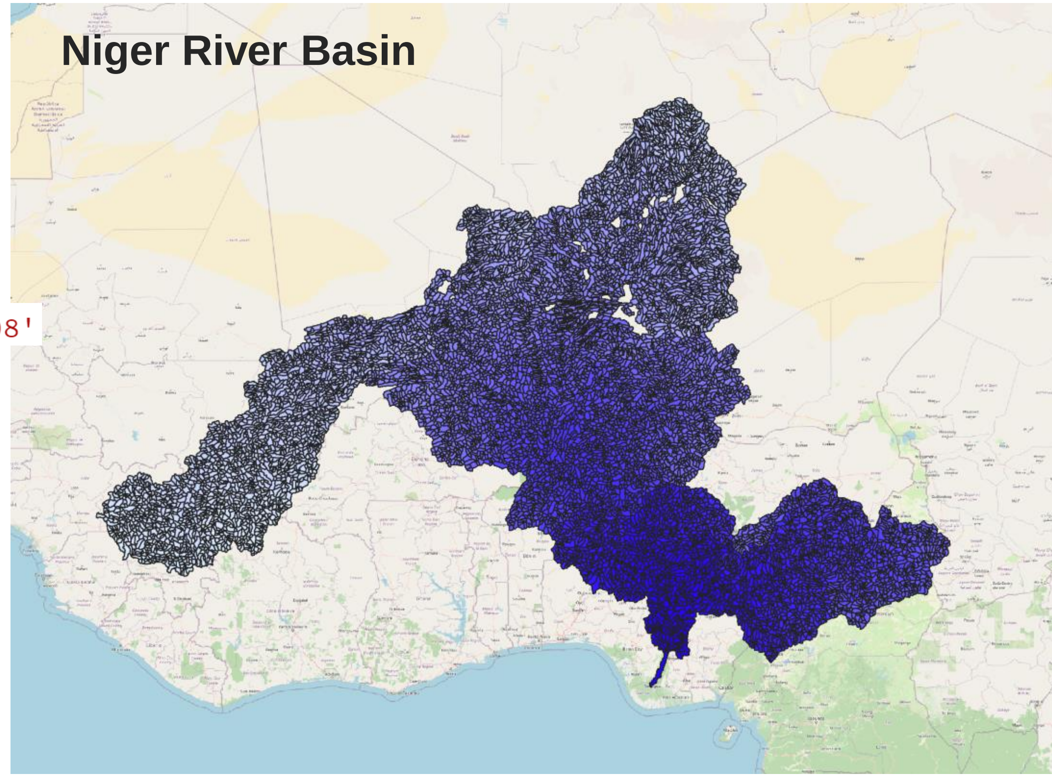
2. Get watershed id for selected watershed code
code watershed

```
SELECT * FROM wam.ws_view WHERE code LIKE 'ws149008'
```

3. Get upstream watersheds or aggregated CHIRPS for selected watershed id

```
SELECT * FROM getUsChirps(172821)
```

	Ruzizi	Niger
Watersheds	96	13290
Timesteps	495	495
Watersheds x timestep	47520	6578550

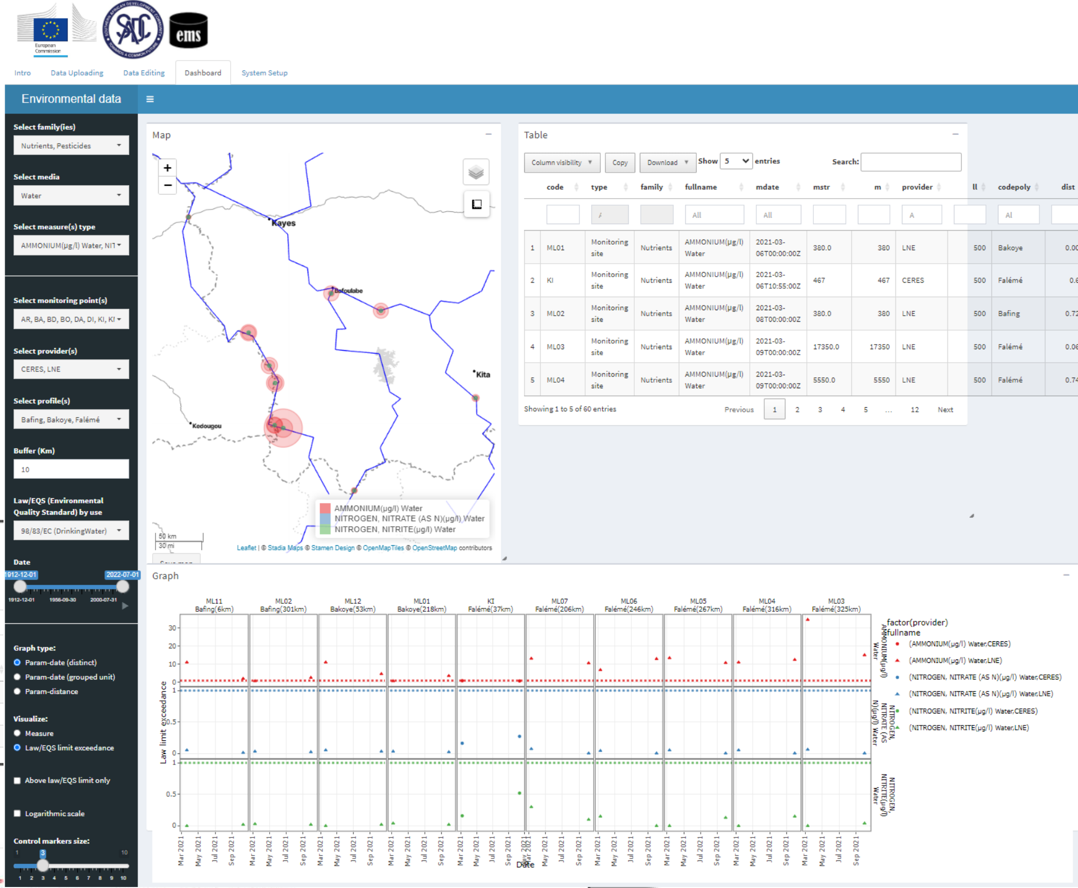


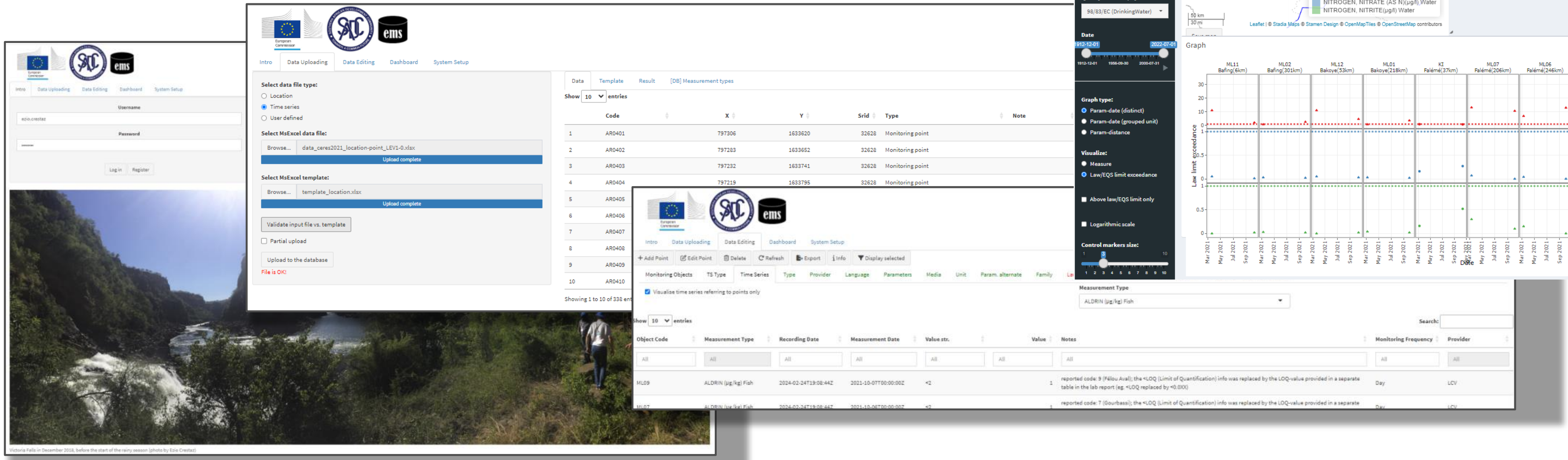
- Watersheds geometry and aggregated raster data (eg. CHIRPS) hosted in the EMS database
- Data and statistical info directly derived from database querying
- Mapping and exploratory spatial data analysis (eg. in QGIS, ArcGIS, R, dashboards): eg. rainfall distribution at a given date, at a given month through the entire dataset (mean, stddev), rainfall distribution yearly aggregated

European Commission

EMS: live demo session

- ❑ Installation/registration, data validation, editing
- ❑ Dashboard





And beyond … EMS installed to a client-server environment, main advantages being:

- ❑ Centralized database. Populate once, sharing data with all authorized people
- ❑ No need of local installation of the database nor of the application

Selected references

Crestaz E., Veneziano V. and Gibin M., 2011. Spatio-temporal Environmental Monitoring Systems. A data management and delivery approach based on PostGIS database and Google Maps mashup. To be presented at MODFLOW and More 2011: Integrated Hydrologic Modeling, June 5 - 8, 2011, Denver, Colorado

Crestaz E., 2013. Spatial Data Management in GIS and the Coupling of GIS and Environmental Models. Chapter in “GIS Fate & Transport modelling” (Eds Pistocchi A), Wiley Blackwell, Hoboken, NJ, USA

Crestaz, E., Seliger, R., Cattaneo, L., Umlauf, G., Iervolino, A., Pastori, M., Marcos Garcia, P., Cordano, E., Carmona Moreno, C., 2022. Environmental Monitoring System: An open source information system with a case study over the transboundary Senegal River Basin, JRC Technical Report, Joint Research Centre, European Commission

Kothuri, R., Godfrind, A., Beinat, E., 2004. Pro Oracle Spatial: The essential guide to developing spatially enabled business applications. APress, Berkeley, CA, USA

Kresic N. and Stevanovic Z., 2009. Groundwater hidrology of springs: Engineering, Theory, Management and Sustainability. 1st Ed., Butterworth-Heinemann

Maidment, D.R., 2002. Arc Hydro: GIS for Water Resources. ESRI Press, Redlands, CA, USA

Obe, R.O., Hsu, L.S., 2010. PostGIS in Action. Manning Publications Co. Unedited Draft, last update: 27/5/2010. Available for purchase from: http://www.manning.com/obe/

Ott T., Swiaczny F., 2001. Time-integrative Geographic Information Systems – Management and Analysis of Spatio-Temporal Data. Springer Verlag, Berlin

Strassberg, G., Jones, N.L., Maidment, D.R., 2011. Arc Hydro Groundwater: GIS for Hydrogeology. ESRI Press, Redlands, CA, USA

Worboys M.F., 1995. GIS: A Computing Perspective. Taylor and Fancis, London, UK

Worboys M.F. and Duckman M., 2004. GIS: A Computing Perspective. 2nd Ed., Boca Raton, CRC Press

Acknowledgments

EMS development benefitted of many discussions and feedback. Particular acknowledges go to:

- Farinosi F., EU Delegation, Papua New Guinea
- Gibin M., UCL, London, UK
- Veneziano V., Hertfordshire Un., UK
- Crosta A., DHI-Italy and Schätzl P., DHI-WASY, Germany
- Pappafico G. and Tramontana M., Urbino Un., Italy
- Rossetto R., Scuola Superiore Sant'Anna di Pisa, Italy
- Habashi N., Ambrosini P., Saipem (ENI), Italy; Cascelli E. and Dellasanta L., Rewind (ENI), Italy; Pellegrini M., consultant, Italy

We would also like to thank SADC and the JRC colleagues involved in the CLIMSA project for the opportunity to organize this session and for sharing valuable views.

European Commission

Thank you

EU Science Hub
joint-research-centre.ec.europa.eu

European Commission