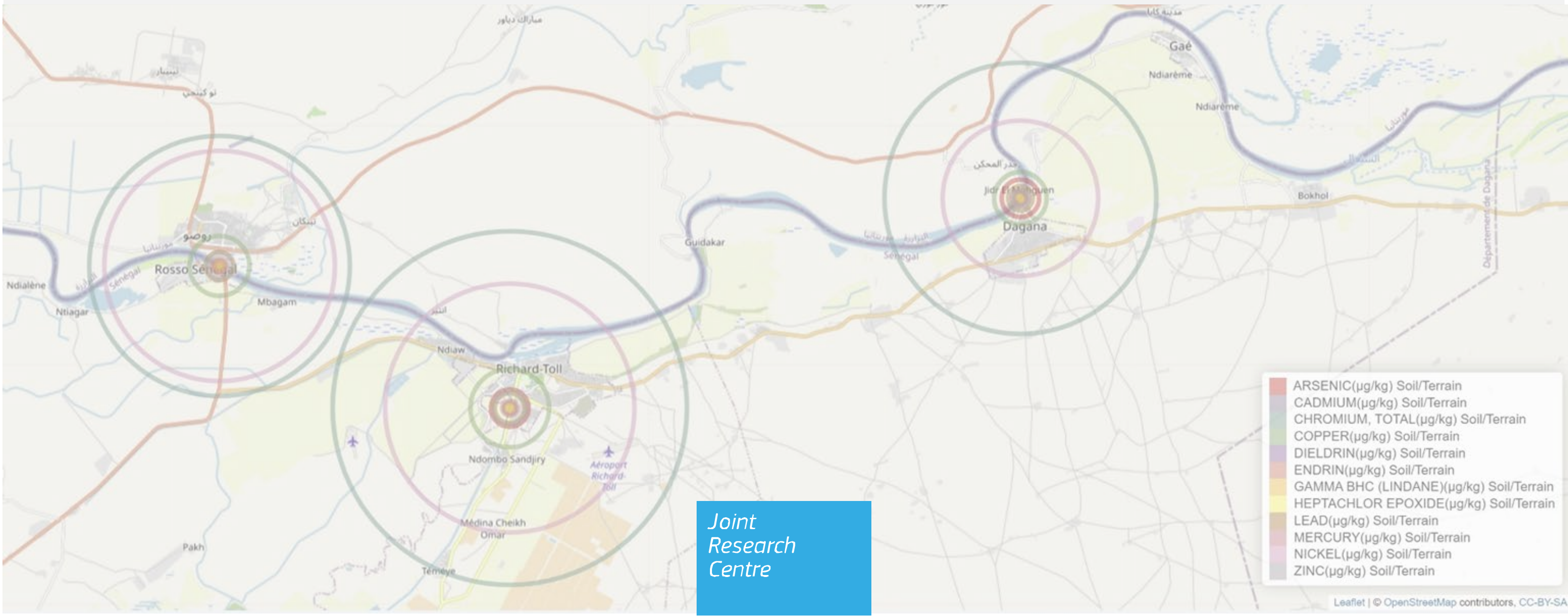
JRC TECHNICAL REPORT

Environmental Monitoring System

An open source information system with a case study over the transboundary Senegal River Basin

Crestaz, E., Seliger, R., Cattaneo, L., Umlauf, G.,
Iervolino, A., Pastori, M., Marcos Garcia, P., Cordano, E.,
Carmona Moreno, C.

2023


Rosso Sénégal
Ndialène
Ntiagar
Mbagam
Ndiaw
Richard-Toll
Ndombo Sandjiry
Aéroport Richard-Toll
Médina Cheikh Omar
Pakh
Témeye
Guidakar
Dagana
Gaé
Ndiarème
Bokhol
Département de Dagana
ARSENIC(µg/kg) Soil/Terrain
CADMIUM(µg/kg) Soil/Terrain
CHROMIUM, TOTAL(µg/kg) Soil/Terrain
COPPER(µg/kg) Soil/Terrain
DIELDRIN(µg/kg) Soil/Terrain
ENDRIN(µg/kg) Soil/Terrain
GAMMA BHC (LINDANE)(µg/kg) Soil/Terrain
HEPTACHLOR EPOXIDE(µg/kg) Soil/Terrain
LEAD(µg/kg) Soil/Terrain
MERCURY(µg/kg) Soil/Terrain
NICKEL(µg/kg) Soil/Terrain
ZINC(µg/kg) Soil/Terrain
Leaflet | © OpenStreetMap contributors, CC-BY-SA

Joint
Research
Centre

This publication is a Technical report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The contents of this publication do not necessarily reflect the position or opinion of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication. For information on the methodology and quality underlying the data used in this publication for which the source is neither Eurostat nor other Commission services, users should contact the referenced source. The designations employed and the presentation of material on the maps do not imply the expression of any opinion whatsoever on the part of the European Union concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

Contact information
Name: Cesar CARMONA MORENO (DG JRC.D2)
Address: Via Fermi, 2749 I-21027 ISPRA (VA) ITALY
Email: cesar.carmona-moreno@ec.europa.eu
Tel.: +39.0332.789654

EU Science Hub
https://joint-research-centre.ec.europa.eu

JRC130504

Ispra: European Commission, 2023

How to cite this report: Crestaz, E., Seliger, R., Cattaneo, L., Umlauf, G., Iervolino, A., Pastori, M., Marcos Garcia, P., Cordano, E. and Carmona Moreno, C., *Environmental Monitoring System*, European Commission, Ispra, 2023, JRC130504.

Contents

Abstract

Management, analysis and visualisation of huge and complex datasets is a challenge. An Open Source (OS) Environmental Monitoring System (EMS) has been developed to support effective management of the entire process of data collection, data cleaning/tidying and validation, the establishment of a mature spatio-temporal database, data uploading, integration with spatial visualisation and analysis tools, and easy access through dashboards supporting explorative research. The single components consist of: (i) a concurrent multi-user spatio-temporal database, implemented using the leading PostgreSQL/PostGIS platform, (ii) an application aimed at supporting EDD (Electronic Data Deliverables), and (iii) a dashboard, aimed at supporting exploratory data analysis in space and time. The spatio-temporal database can be easily accessed through state-of-the-art OS and proprietary GIS tools, (as QGIS and ArcGIS), (geo)statistical platforms and any programming languages (as R, Python, C), in order to fully leverage its added value. The WEFE Senegal nexus project "*Support for water resources management and the Water-Energy-Agriculture Nexus in the Senegal River Basin*" served as an EMS case study for designing and implementing an environmental quality monitoring network over the transboundary Senegal River Basin (SRB), involving various laboratories in Senegal, Mali and the Netherlands.

Keywords: Environmental Monitoring System, data validation, data management, data visualisation, water quality, WEFE-Nexus, Senegal River Basin, Open Source

Acknowledgements

The design and implementation of the EMS tool benefited of many different contributions, reasoning and discussions. The authors would like to particularly thank Farinosi, F., a former JRC colleague now at the Delegation of the European Union to Papua New Guinea for the fruitful exchanges.

Special thanks go to:

Veneziano V. from the Hertfordshire Un. (UK) and to Gibin M., a former JRC colleague at JRC now at UCL (University College of London, UK); the former for having supervised the MSc in Computer Science dissertation that somehow inspired the EMS database schema (Crestaz, 2011) and to both for the valuable discussions that led, among others, to the publication of a conference paper (Crestaz et. al., 2015).

Pistocchi A. from thre JRC for having included in his contribution to the geospatial analysis of chemical fate (Pistocchi, 2014) a chapter on spatial databases (Crestaz, 2014), addressing both the theoretical background and practical issues that inspired the EMS design.

Schätzl P., DHI-WASY (Germany) and Crosta A., DHI-Italia, for the joint efforts that, in the framework of my former manager role in the ENI Group (Italian Oil Group), led to: (i) the design and implementation of a groundwater modelling information system, encompassing the FEFLOW flow&transport numerical engine within ArcGIS with a personal geodatabase largely drawn upon the concepts discussed in this report; and (ii) the support for both proprietary and OS native spatial databases (Oracle, PostgreSQL/PostGIS) within the FEFLOW tool (since version 6.2).

Pappafico G. and Tramontana M., from the University of Urbino (Italy) for the joint implementation of a former water quality information system devoted to environmental consultancies focused on contaminated sites.

Rossetto R., from the Scuola Superiore Sant'Anna di Pisa (Italy), for the valuable exchanges about the choices adopted in the framework of the development of the FREEWAT groundwater modelling system.

The former colleagues at different branches of the ENI Company, namely Habashi N. and Ambrosini P. (Saipem SpA), Cascelli E., Dellasanta L., Tatangelo F. and Pellegrini M. (ENI Rewind SpA), for the many joint collaboration efforts in the framework of both Corporate R&D and operational activities, that led to the development and testing of both EDP tools and advanced groundwater information systems.

Authors

Ezio CRESTAZ

Roman SELIGER

Luigi CATTANEO

Gunther UMLAUF

Andrea IERVOLINO

Marco PASTORI

Patricia MARCOS GARCIA

Emanuele CORDANO

César CARMONA-MORENO

1 Introduction

Projects aimed at environmental and water quality characterization, analysis and modelling, whether at regional or local scale, often demand for the setup of a network of ground monitoring stations leading to the collection of relevant datasets (e.g. monitoring stations locations, contaminants concentration time series). Such datasets can quickly turn to be huge and complex, demanding a structured and efficient management approach capable to effectively addresses data QA/QC (Quality Assurance and Quality Control), advanced spatio-temporal data storage, models coupling and visualisation for different audiences (experts, policy makers, large public).

Examples of existing tailored geospatial solutions include standardised geodatabase frameworks (e.g. hydro and groundwater database models for ESRI ArcGIS; Maidment, 2002; Strassberg et al., 2011) and commercial solutions as EQuIS (Earthsoft, 2022), RockWorks and PetraSim (RockWare, 2022), and Leapfrog Geo (Seequent, 2022), specialized in specific domains as the groundwater and geothermal characterisation and modelling. These tools are pretty advanced but they generally demand for costly commercial licenses, that limit their widespread adoption while focusing on specific domains as the mining, Oil and Gas and, at a minor extent, the water sector. Licensing costs turn to be a major bottleneck when coming to environmental information systems adoption in developing countries and particularly in Africa (Chen et al., 2010; Silva et al., 2023; Gillwald et al., 2022). Another concern attains at usability, as most of the aforementioned tools demand for strong domain expertise and advanced knowledge of GIS, spatial databases and problems related to coupling of information systems with modelling tools. Limited IT resources (hardware, software, internet connectivity) can further hamper efficient access and use of tools.

Other Open Source (OS) solutions, as the FREEWAT (Rossetto et al., 2018) in the groundwater hydrology domain, provide advanced integrated modelling solutions (e.g. support to USGS finite difference codes as MODFLOW; USGS, 2022), still relying upon loosely data management (e.g. traditional unstructured shapefiles or SQLite; Hipp Wyrick and Company, 2022), while providing limited features for the foundation tasks of data QA/QC and validation.

At the authors knowledge, a comprehensive and easy-to-use OS solution specialized on the above mentioned issues, facilitating data QA/QC, storage and visualization for both experts, policy makers and larger public, is still lacking. The OS solution “Environmental Monitoring System” (hereafter referred to with the acronym EMS) presented in current report aims at addressing the foundation tasks above, building upon former experience in the framework of environmental, surface water and groundwater data management for regional scale water supply and local scale protection and remediation at contaminated industrial sites in the framework of both internationally funded projects and the Oil and Gas Industry (Crestaz, 2014; Crestaz et al., 2015).

From a broader perspective the tool can complement larger environmental information systems through improving data quality, effective data management and advanced visualization dashboards addressing the needs of both specialised users and policy makers and providing an efficient solution for developing countries.

Among such larger environmental information systems and related studies, we mention here some examples:

- A Shared Environmental Information System (SEIS, Hřebíček and Pillmann, 2009) and the foundation of an Infrastructure for Spatial Information in Europe (INSPIRE, Directive 2007/2/EC, 2007) has been developed by the European Commission together with EU member states to tackle data heterogeneity, limited use of standards, limited data access and lack of trust. It is used to improve access and usability of spatial data on noise and ensure high quality noise reporting in support to Environmental Noise Directive (Abramic et al., 2017; Blanes et al., 2021).
- Studies on spatial data analytics (e.g. Alam et al., 2021) contribute to consistent and fast processing and dissemination of spatio-temporal data needed for e.g. near-real-time georeporting for adequate emergency management (Alamouri et al., 2021), groundwater monitoring and modelling applications (Crestaz et al., 2015) or monitoring of emissions to air and water from industrial installations (Brinkmann et al., 2018).
- Joint initiatives such as SERVIR between NASA and USAID (United States Agency for International Development) and leading geospatial organizations have been developing innovative user-centred earth observation, geospatial applications and services to support sustainable resource management at different scales (Bajracharya et al., 2021).

- Free and OS QGIS modules such as FREEWAT AkvaGIS aim to facilitate storage, management, analysis, modelling and visualisation of hydrochemical and hydrogeological data (Criollo et al., 2019; De Filippis et al., 2020)
- Initiatives like the UN Open GIS initiative foster the creation of open-source extended spatial data infrastructures, giving attention to the software product, data and people behind them (Brovelli et al., 2021)

The EMS has been developed in the framework of the WEFE-Senegal project, providing a perfect opportunity to test its benefits in managing a complex environmental quality dataset. In fact, an environmental quality monitoring network has been designed and implemented over the transboundary Senegal River Basin, in the framework of the project “Appui à la gestion des ressources en eau et du Nexus Eau-Energie-Agriculture dans le bassin du fleuve Sénégal”, hereafter referred to as Senegal WEFE (Water-Energy-Food-Ecosystems)[1] nexus project. 27 sampling sites have been identified along the medium and lower part of the river basin, in Senegal and Mali, accounting for a total of 338 monitoring points. Locally clustered, the analyses at monitoring points aim at reporting about mean qualitative conditions at 27 sites, their evolution in time, while capturing local variability and extremes in different media as surface water, sediments and fish. Various laboratories, in Senegal, Mali and in the Netherlands, have been involved in the sampling and/or analytical determinations, based on their specific expertise and with cross-checking and validation objectives in mind:

- CERES, Centre Régional de Recherches en Ecotoxicologie et Sécurité Environnementale, Senegal
- DGPRE, Ministère de l'Eau et de l'Assainissement - Direction de la Gestion et de la Planification des Ressources en Eau, Senegal
- IPD, Institute Pasteur de Dakar, Senegal
- DNH, Direction Nationale de l’Hydraulique, Mali
- LCV, Laboratoire Central Vétérinaire, Mali
- LNE, Laboratoire National des Eaux, Mali, and
- The laboratory of VUA, Vrije Universiteit Amsterdam

The field and laboratory activities led to the production of huge and complex datasets, covering a large variety of parameters and analytical determinations; their interpretation and the evaluation of related risks is even more challenging in the light of the different law limits of both national and international legislation and EQS (Environmental Quality Standards) guidelines.

Hence, the recognised need for the availability of an environmental information system to support effective data collection, data cleaning/tidying and validation, spatio-temporal data management, advanced visualisation and analysis, supporting experts in conducting spatio-temporal exploratory research and top-level managers and senior officers in fast querying and visualisation. An environmental information system is key to the implementation and management of transboundary, regional and local scale projects, WEFE nexus assessment, environmental pollution control and remediation, and advanced modelling. It further contributes to promoting a clear splitting of data management and data analysis tasks, which generally demand for quite different background and expertise. The advancements in environmental database design, data validation and cleaning (namely Electronic Data Deliverables (EDD) and EDD Data Processing (EDP)) and dedicated geospatial tools largely inspired the activities hereafter described.

Given the framework above, an EMS has been designed and developed. The current report provides a detailed description of its main components and underlying foundation theory, namely (i) the spatio-temporal database, and (ii) the user front-end application for database editing, data files validation and massive database uploading, dashboards for spatio-temporal exploratory data analysis and access to the PostgreSQL pgAdmin[2] web tool.

[1] The WEFE nexus approach integrates management and governance across the multiple complex and inextricably entwined sectors of food, energy, water, and ecosystems.

[2] PostgreSQL is an advanced free and Open Source Relational DataBase Management System (RDBMS). It can be managed and administered through free desktop and web-based platforms, as pgAdmin versions 3 and 4. PostGIS is a PostgreSQL spatial extension providing a rich toolset of spatial functions in addition to spatial data types, being a powerful tool for storing, processing, querying and analysing spatial data.

A web application, integrated in (and accessible from the main page of) the JRC water KMS (Knowledge Management System) Aquaknow (https://aquaknow.jrc.ec.europa.eu/), has also been developed in parallel using Drupal framework. The application provides basic spatio-temporal visualisation features, including mapping of monitoring points location and graphs of multi-parameters temporal trends, based upon dynamic selection of a point of interest. The current version of the web application relies upon properly formatted CSV datasets, as exported from the spatio-temporal database. Future developments will address a more advanced integration with the underlying database.

The report concludes with an overview of the application of the EMS tool in the framework of the Senegal WEFE nexus project, supporting the subsequent activities of chemical data interpretation.

The spatio-temporal database can be easily accessed through state-of-the-art OS and proprietary GIS tools, as QGIS and ArcGIS, (geo)statistical platforms, scripting and programming languages as R and Python, and professional dashboard platforms (e.g. Tableau), in order to fully leverage its added value while limiting the inherent risks of data duplication and consistency failure.

2 Scope of the work

Scope of the work is to present the design and implementation of a EMS web Information System to support the collection, validation, spatio-temporal management, visualisation, and analysis of water and environmental quality data collected in the framework of the Senegal WEFE nexus project, over the transboundary river basin.

The report provides a detailed description of the tool architecture, starting from the foundations of state-of-the-art best practices and models, and then focuses on its operationalisation. A brief introduction to a parallel web development in the framework of the JRC Water Portal Aquaknow is also made.

The environmental quality monitoring network setup over the Senegal River Basin and preliminary quantitative analysis outcomes are briefly reviewed, the reader being directed towards other project contributions for the details.

The EMS is intended as a general purpose system, to be installed on single user computers or in client-server environments for multi-users concurrent access. As such, any project leading to the collection and analysis of whatever complex environmental datasets (e.g. groundwater hydrology, climatology) is expected to strongly benefit from this tool and the overall approach detailed in the current document.

3 Spatio-temporal database

3.1 Theoretical framework

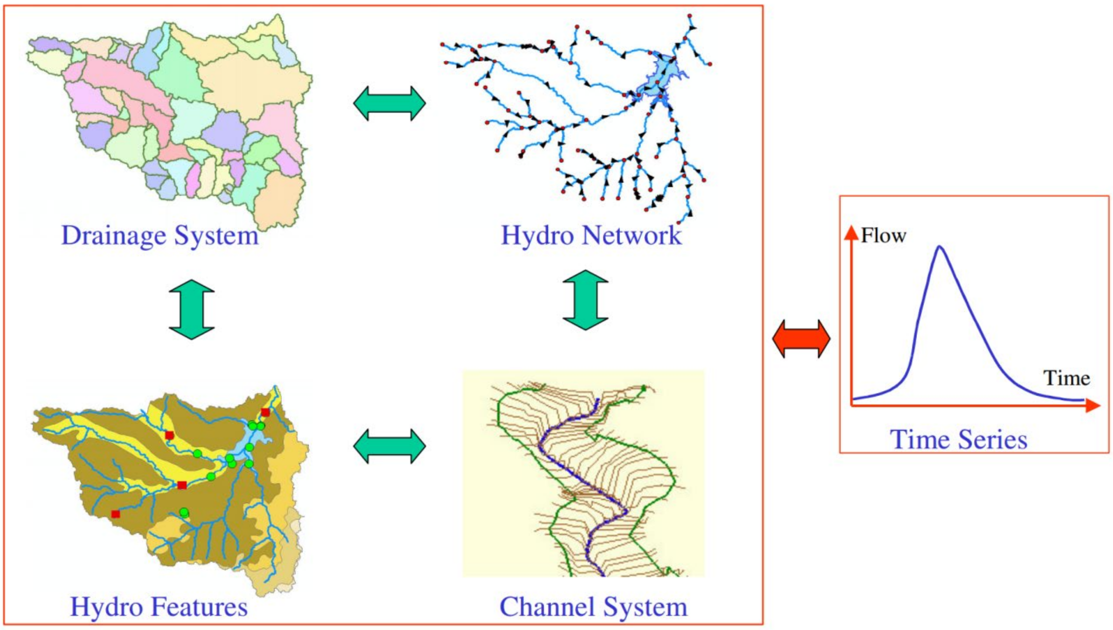
The spatio-temporal database has been designed after the ESRI and CRWR Hydro Data Model (Maidment, 2002) and marginally influenced by its later extension to groundwater hydrology, the Groundwater Hydro Data model (Strassberg et al., 2011), as well as by other ESRI geodatabase models ([3]).

The basic foundation pillars of the Hydro Data Model are:

1. Any relevant hydro feature, whether an occasional point of measurement, a human-made infrastructure (e.g. well, gauging and meteo-climate station, channels) or a natural feature (e.g. water basin, lake, pond), is stored in a table and uniquely identified through its code and spatial location (coordinates and related spatial reference system).
2. Time series data (e.g. piezometric heads, contaminant concentrations, abstractions) are stored in a unique table and linked to the hydro feature at which they were collected.

The model is conceptually captured here below (**Figure 1**). The original ESRI database schema for ArcGIS can be accessed from the company web site.

Figure 1. Hydro Data Model framework



Source: CRWR, 2021

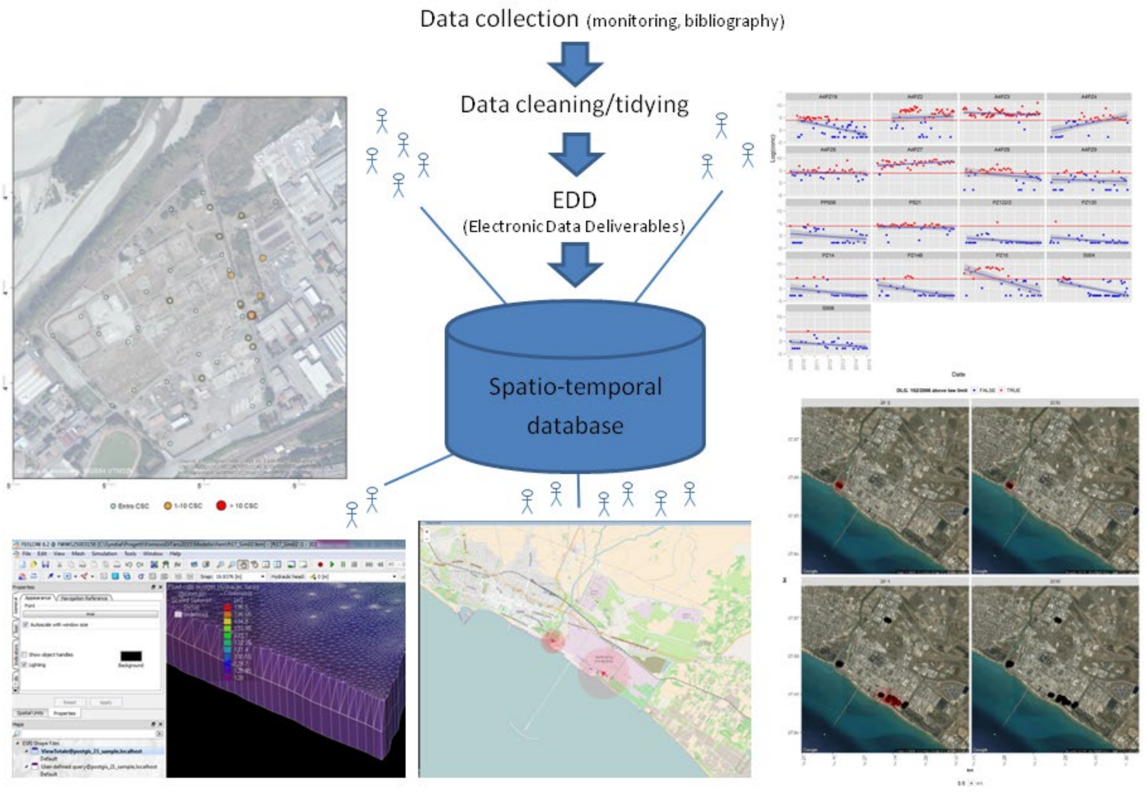
The spatio-temporal database of the EMS system has been implemented using the OS (Open Source) PostgreSQL 13.x, a state-of-the-art platform, widely used since early 70s and nowadays increasingly used for geospatial applications thanks to its extension PostGIS (now version 3.x; Obe and Hsu, 2011). PostgreSQL and PostGIS combination turns to be the most powerful Open Source multi-user geospatial platform, highly flexible, supporting the storage of both vector and (more recently) raster datasets, spatial indexing for fast search, spatial querying extending the traditional SQL features and, last but not least, functions for exchanging (import/export) spatial data from/to standard formats (e.g. ESRI shape file, Well Known Text format - WKT). Besides, the platform also provides plenty of spatio-temporal functions to manipulate, query

[3] ESRI geodatabase conceptual models provide a solution for effective data management in such different domains as clima, forestry, geology and marine applications. They can be downloaded and implemented within the ArcGIS platform. For a brief overview, reference can be made to: https://gisandscience.com/2009/06/24/more-than-30-essential-data-models-available-for-arcgis/

and analyse the data. The combination of PostgreSQL/PostGIS easily integrates with various leading OS and proprietary GIS, statistical and geovisualisation tools and programming languages, as QGIS, ArcGIS, R and Python. In particular, the QGIS project itself was originally conceived to provide a GIS interface to PostGIS. Spatial data can be easily transferred to other databases, as Oracle, or accessed in read (and, depending upon licensing level, also write) mode from ESRI desktop and server applications.

Cartographic production, spatio-temporal exploratory data analysis and advanced modelling are among the many activities that can be implemented by directly accessing the database, while limiting the risks of data duplication and consistency failure, which arise from bad practices such as the use of poorly structured file formats (**Figure 3**).

Figure 2. Spatio-temporal database: data uploading and accessing


Data collection (monitoring, bibliography)
Data cleaning/tidying
EDD
(Electronic Data Deliverables)
Spatio-temporal database

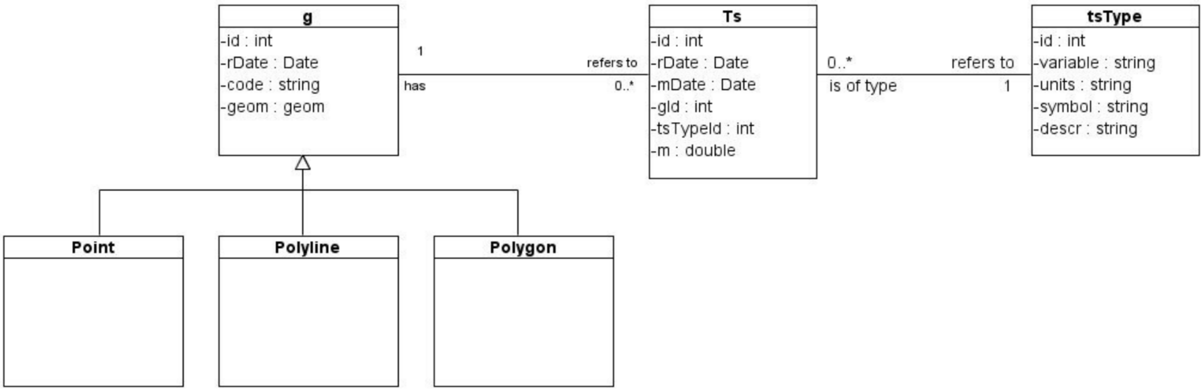
Being an Open Source and multi-user concurrent platform, the PostgreSQL/PostGIS database does not imply any licensing cost and it can be easily installed and configured on a server to provide centralised data management services in the framework of a client-server architecture.

3.2 Implementation

The original Hydro Data Model has been implemented and further extended based on the outcomes of previous research and technical developments (Crestaz, 2014; Crestaz et al., 2015), addressing the need, among others, to flexibly keep track of a full list of (standardised) parameters, laws and EQS (Environmental Quality Standards). Database design focused on traditional relational models, advanced OO (Object Oriented) inheritance features was initially tested for geometry partitioning (**Figure 2**) but then discarded due to PostgreSQL implementation shortcomings (4)

[4] Inheritance is a powerful OO programming (and database) feature that builds, among others, upon the concept of parent and child classes. Let's assume that we expect to store a large number of hydro features in the database; instead of storing all of them in a unique table, we could create an inheritance tree with child tables for specific geometry types (e.g. point, polyline, polygon) and a

Figure 3. UML data model for spatio-temporal data management exploiting OO portioning techniques over the monitoring objects geometry table in PostgreSQL/PostGIS



In line with the original Hydro Data Model specifications, three key tables have been created (**Figure 4**):

- **Monitoring objects** table (named 'p'([5])), storing all hydro features, independently from their geometry type (i.e., points, polylines or polygons). For each record, code and geometry are enforced to be unique and trace is kept of the hydro feature type (e.g. generic monitoring point, well, gauging station) and the data provider, through Foreign Keys (FKs) referencing the related supporting tables. Locations are stored in the database table in geographic WGS84 coordinates (un-projected longitude and latitude; EPSG code 4326). The attempt to upload spatial data provided in any other projection systems (e.g. projected UTM metric coordinates) triggers an automatic database conversion to the WGS84 system, based on the origin EPSG reference code (EPSG, 2022; EPSG.io, 2022). Note that topographic survey devices, relying upon GPS, typically return geographic coordinates, while cartographic production is often based on projected metric coordinates in order to minimise distance and/or areal distortions.
- **Measurement types** table (named '*tstype*'), combining the info on parameter, measurement units and media (e.g. '*Lead (µg/l) in water*') in the form of reference FKs to the related records in supporting tables. Being the details of parameters (code, CAS number, extended reference name, chemical formula), media and measurement unit stored in distinct tables, duplication risks are minimised([6]).
- **Time series data** table (named 'ts'), storing the measurement date, in standard ISO8601 format([7]) ('YYYY-MM-DD [hh:mm:ss]'; ISO, 2022), and the measure, in both text and numeric formats. The

trigger to perform the automatic data reallocation when loading data to the parent table. Unfortunately, PostgreSQL inheritance implementation works fine for tables partitioning, but (differently from Oracle) it falls short in inheriting key attributes as the PKs; this implies that PKs must be redefined in each child table, few of the key benefits of the OO construct being lost.

[5] The table name 'p' implicitly testifies about the PostGIS flexibility, the geometrical attribute within a relational table being able to accommodate different geometries, as (multi)points, (multi)polylines and (multi)polygons. This is a behaviour sensibly different from that of the shape file, a de-facto ESRI standard in geospatial applications, where the shape attribute is constrained to a geometry type only. Data can be partitioned at any time based on the geometry type, in order to be accessed through traditional GIS platforms as QGIS and ArcGIS, where layers are, again, constrained to a specific geometry type. By the way, PostGIS also supports multiple geometric columns within the same table.

[6] The original Hydro Data Model explicitly includes all the details of the measure types in the tstype table (e.g. the extended parameter name, the measurement unit and the media), as detailed in Maidment (2002). This approach has few drawbacks, as the duplications of parameter names particularly when they are complex and difficult to standardise. Other issues attain at the lack of standardisation in media and unit naming conventions, not to forget the parameter synonyms and the translations to be supported in view of the application internationalisation. The adopted approach sensibly differs from the above; the parameters, media and units being reported as FKs pointing to the three related supporting tables.

[7] Technically, the measurement date attribute is set as a 'timestamp without time zone', including both the date and (optionally) the time, to cope with sub-daily frequency measurements (e.g. groundwater levels during a well pumping test). QGIS, having been originally implemented as a front-end for PostgreSQL/PostGIS, has no problems with that. ArcGIS, which since version 10.0 supports direct access to native spatial databases through the 'Query Layer' feature, requires a date only, hence the time extra info must be removed.

measure must be provided as text, as it cannot necessarily be converted to a numeric value; this is the case for qualitative measures (e.g. colour or odour), and for concentrations reported as being below an instrumental detection limit (e.g. '<100'). At database level, the measures are automatically converted to numbers and hence stored also to the numeric format attribute; for the below detection limits data, one half of the value is automatically retained (e.g. '<100' being converted to the numeric value 50), provided that the user can always access the original measure in string format and adopt his/her own conversion strategy (e.g. convert to the detection limit itself or simply discard the data). The record attributes are complemented by the two FKs pointing to the hydro feature (to which the measure refers) and to the measurement type, as detailed for the tables above.

The schema is complemented with a series of supporting tables, whose relational links with the aforementioned ones are graphically depicted in the EAR (Entity-Attribute-Relationship) diagram([8]) (**Figure 4**):

- **Object types** table (named 'Type') - e.g. gauging station, well
- **Data providers** table (named 'Provider') - e.g. data providers references, as it is the case for the analytical laboratories in current case study
- **Parameters** table (named 'Parameter'); a bit more complex table providing information needed to fully characterise each parameter, as its commonly used code (if any), CAS unique number (where applicable), its extended name (and optionally the formula) (e.g. 'Pb', '7439-92-1', 'Lead'). Among the mentioned attributes, the code can provide a good shorthand for referring to the parameter, but unfortunately it is only available for a quite limited subset of parameters (e.g. Pb, Fe, Ca); the CAS number is potentially the unique identifier we would like to have, but of course it is applicable only to chemical compounds and, not less important, is rarely reported when data are transferred; the name is potentially fine, as long as we conform to a standard table using proper naming conventions. We took as a reference the standardised table compiled by the US Environmental Protection Agency (US EPA) as part of the development of an EDD framework (US EPA, 2016); the table, after partial cleaning to remove/correct some inconsistencies and duplications, still retained more than 5800 records, testifying the underlying classification complexity and richness. Using the name to characterise a parameter is anyway challenge, as, apart from the just infinite list, plenty of synonyms and slightly differing names/conventions exist. Another big issue is the need to provide internationalisation support, and hence both the translation of parameter names to other languages and the provision for the synonyms in those other languages. A virtually endless job, for which the 'Alternative names' table hereafter described provides a flexible and powerful solution.
- **Alternative names** table (named 'param_alternate'), storing parameter synonyms and translations in the supported languages. The parameters table described at previous point remains the standardised reference, and hence any alternative name that may be used to report a parameter is expected to be traced back to its official name. From the relational point of view, each record in this table links back through a FK to the related record in the parameters table.
- **Measurement units** table (named 'unit') – e.g. mg/l, gr/Kg, °C
- **Media** table (named 'media') – e.g. Water, Sediment, Fish
- **Languages** table (named 'lang')

Fully integrated, still somehow aside of the main database structure, a set of tables keep track of laws and EQSs provisions as for the maximum admitted contaminants concentrations for the different uses of the resource (**Figure 4**):

- **Laws and EQSs** table (named 'law'), storing national and international laws and Environmental Quality standards and guidelines (e.g. from the EU, WHO, ...).

[8] The label '1 to M' in the EAR diagram clarifies the nature of the relationship between two tables, stating the number of records that we can expect on both sides (e.g. a monitoring point can have many measurements). Actually further details could be expressed in terms of cardinality, stating - in the just mentioned example – that a point could actually have no associated measures (e.g. simply not yet added), one or many.

— **Use** table (named ‘use’), storing information on the different kind of resource uses (e.g. water for drinking purposes, for industrial sites), as officially reported in the law provisions. Depending upon the use, the contaminants concentration limits can vary, also the order(s) of magnitude.

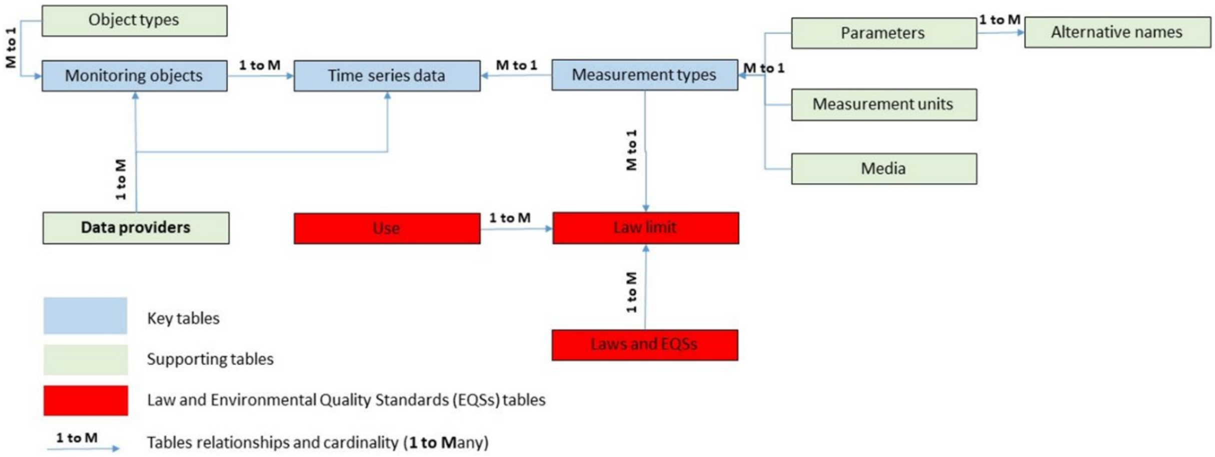
— **Law limit** table (named ‘lawlimit’), linking measurement types (parameter value in given units and media), the resource use and the reference law/EQS. For each record, reference to the specific measurement type (table tstype), laws or EQS, and use is maintained through as distinct FKs.

While data are split in the different database tables, the enforced relational rules guarantee the consistency of the datasets. For example, a monitoring point or a measurement type cannot evidently be deleted if any measure in the time series table already reference both (or any) of them. The same applies to a measurement type for which a specific law limit is already reported in the related table, and so on…

On the other hand, views can be easily built, based on the underlying relational rules, to keep together data otherwise split through different tables. For example, the time series data can be complemented with the monitoring point code and location, the measurement type details (e.g. parameter, measurement unit and media), the concentration law limit as stated in a selected law or EQS; the ratio between the measure and the law limit can be easily derived, discriminating between law compliant and not-compliant data, while reporting also the percentage of exceedance.

Finally, it is worth to highlight that, although the first use case for both the database and accompanying applications (refer to coming chapters for details) is a single user desktop environment (the application being a web one and hence running in the browser), the tool can be easily installed in a more flexible and powerful server, to support a client-server architecture. The database, further to being Open Source, is a multi-user concurrent platform, meaning that different users can connect, query and edit the database at the same time([9]). The database administrator can define new user profiles, granting them specific rights (e.g. limiting tables accessibility, inhibiting records deletion or update) as a function of their roles and work responsibilities. For such a more advanced configuration, the recommendation is to contact your Organisation IT Department. A system administrator will take charge to assess and configure the server and properly install the database.

Figure 4. Spatio-temporal Environmental Monitoring System database EAR schema (Entity-Attributes-Relationships)



[9] Tables temporal locking is transparently managed by the database engine to avoid potential conflicting edits, as distinct users concurrently operate. Versioning is an alternative way to cope with this kind of issues, the database to keep alive parallel versions to be later consolidated at administration level by retaining the modifications and sorting out any potential conflicts. Refer to ESRI (2022) for further discussion, in the context of ArcGIS geospatial application.

4 Environmental Monitoring System application

The Environmental Monitoring System, hereafter referred to as EMS, is a MS Windows application, tested for version 10 or later of the Operative System (OS). EMS was developed in R and Python, largely relying upon the R shiny web framework and the underlying spatio-temporal database, described in the previous chapter.

An installation program has been implemented in C# to setup both the spatio-temporal database schema and the application on the local computer. Full details on the installation tool and process are provided in **Annex 1.**

4.1 Main interface

EMS is characterised by a main GUI (Graphical User Interface; **Figure 5**), giving access to the different application features.

Through the main GUI, the user can access the facilities for setting up and testing the database connection (bottom left; **Figure 6**). Once the connection credentials are provided (server address, database name, user, password and port), the user tests the connection and saves the configuration parameters to disk. Only in case of first connection, the user is expected to initialise the database, the schema detailed in the previous chapter being automatically created and predefined system data (e.g. parameters, units, media, laws, uses and law limits) uploaded. The database can reside on the user computer (localhost), which is supported in the standard installation and for which the user provides his/her own credentials([10]); alternatively, the database can be installed on a remote server by a system administrator, who will be able to provide all the connection details, including the server remote IP address.

Once the database (and connection) is properly setup, the user can access full application features through the related GUI buttons, in order to:

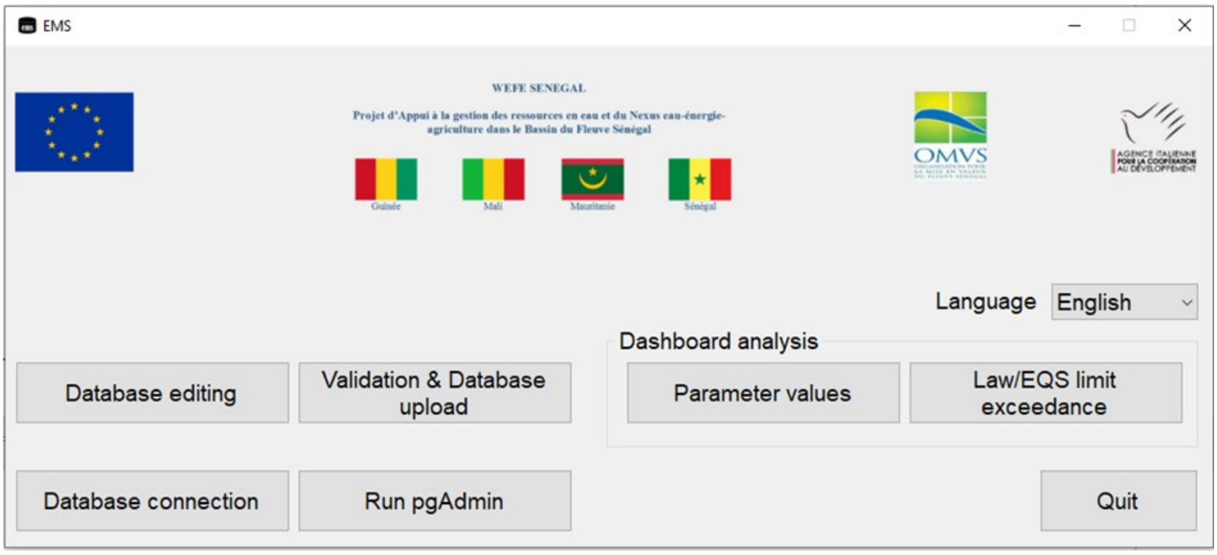
- Edit the database (‘Database editing’ button).
- Validate data files and upload data to the database (‘Validation and database upload’ button).
- Access the dashboard for querying and visualizing data by parameter (‘Parameter values’ button).
- Access the dashboard for querying and visualizing data by multiple parameters and law/EQS limit exceedance (‘Law/EQS exceedance limit’ button).
- Access the pgAdmin IV PostgreSQL-native web administration tool (‘Run pgAdmin’ button).

The last button ‘Quit’, on the bottom right, enables the user to exit the application.

A drop down list supports the language choice for the user interface.

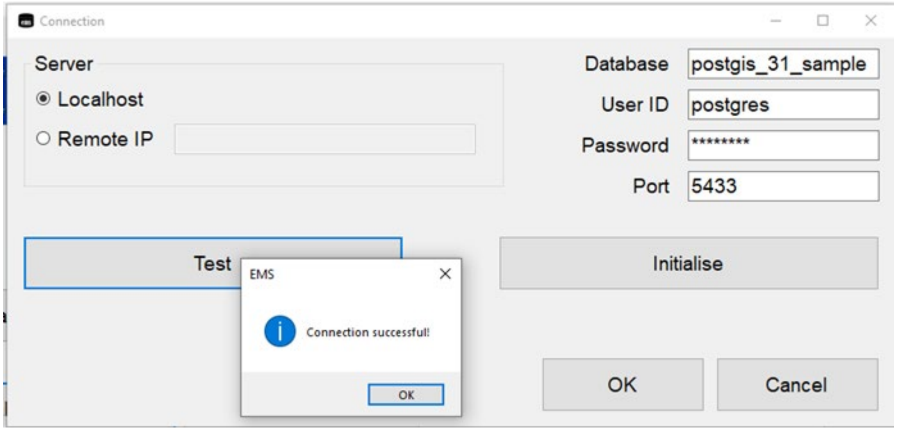
([10]) For local installations, the user is expected to keep note of database access parameters, provided that he will always be able to retrieve them by running the pgAdmin web interface. Suggested password at installation time is ‘postgres’ and port is 5432, except if already reserved (for other application).

Figure 5. EMS: main Graphical User Interface


EMS
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
AGENCE ITALIENNE POUR LA COOPÉRATION AU DÉVELOPPEMENT
Dashboard analysis
Database editing
Validation & Database upload
Parameter values
Law/EQS limit exceedance
Database connection
Run pgAdmin
Quit

Language: English

Figure 6. Database connection testing and initialisation



● Localhost

○ Remote IP

Database: postgis_31_sample

User ID: postgres

Password: ********

Port: 5433

Connection successful!

4.2 Database editing

The component is a CRUD (Create-Read-Update-Delete) web application, enabling simple and fast access and editing of database contents, addressing different tasks as: to visualise database tables; to add new records; to update and to delete existing record(s). The interface (**Figure 7**) characterises for a clean design aimed at improved HCI (Human Computer Interaction). The buttons on the top left provide access to the key features, acting, where relevant (edit, delete), on table selected record(s). Record selection is as easy as left mouse clicking on the table, double clicking on the same record simply resulting in a deselection. Some operations can act on both single and multiple records, as is the case of records deleting. Other operations can act on a single record only, as is the case for editing; if the case occurs that no record or more than one record is selected, a warning message is issued.

Any editing operation is run against the underlying database. In case of failure, a popup window informs the user about the cause, as for example in case of violation of an existing database constraint([11]).

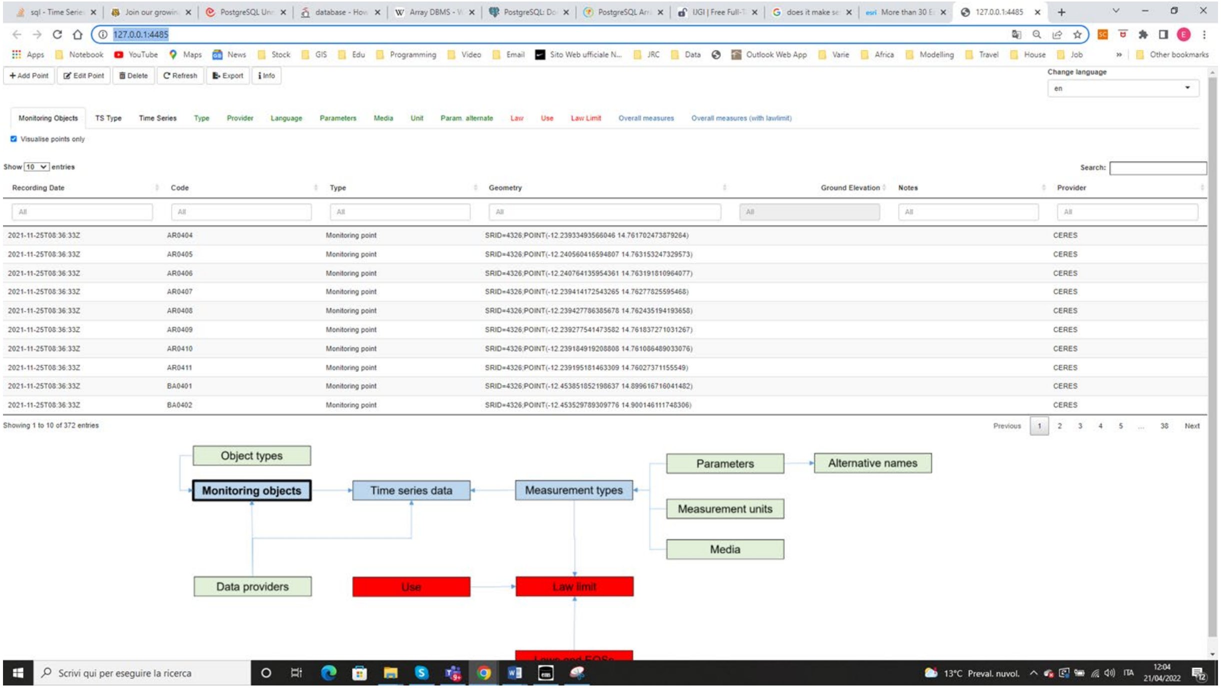
[11] A simple example is the attempt to delete a monitoring point that already has related time series measurements stored in the database; the operation fails because otherwise the time series would remain orphan. In this case, the related time series data must be deleted before attempting to remove the monitoring point. On the same lines, an attempt to enter a new record without inserting

The user can access the different database tables through the top tabs, located immediately below the action buttons (e.g. to add, edit and delete new record(s) - **Figure 7**). The tables are grouped and distinctly coloured to highlight their different role within the database, namely:

- Key tables (in black), hosting application specific data, as locational info and time series, as well as the specific measurement types under analysis.
- Auxiliary tables (in green), hosting supporting system data. These tables are expected to be less prone to changes, as is the case for the 'parameter' table populated with (and cleaned after) US EPA standardised datasets. In case of a client-server architecture with multiple users having access to the database, the optimal use case scenario suggests that a high-level database administrator with advanced knowledge of the water quality domain should take care to update/integrate these tables. Introducing erroneous information in the database tables has always negative aside effects and risks to do so should be minimised; still somehow these tables must be regarded as the database backbone and much attention should be spent to guarantee that their content is consistent.
- Law limits and EQSs tables (in red), hosting data about laws, resource use (e.g. drinking water, industrial use) and related measurement type (combination of parameter, media and measurement unit) vs. law/EQS limits.

A fourth group of tabs (in blue) provides access to the output of queries that join different tables' contents, providing a unified overview of: (i) all time series data; (ii) only the time series for parameters having law or EQS limits reported in the database.

Figure 7. Database editing: web interface accessing tables contents



☑ Visualise points only

Show 10 entries

Search:

Recording Date	Code	Type	Geometry	Ground Elevation	Notes	Provider
All	All	All	All	All	All	All
2021-11-25T08:36:33Z	AR0404	Monitoring point	SRID=4326;POINT(-12.23933493566046 14.761702473879264)			CERES
2021-11-25T08:36:33Z	AR0405	Monitoring point	SRID=4326;POINT(-12.240560416594807 14.763153247329573)			CERES
2021-11-25T08:36:33Z	AR0406	Monitoring point	SRID=4326;POINT(-12.240764135954361 14.763191810964077)			CERES
2021-11-25T08:36:33Z	AR0407	Monitoring point	SRID=4326;POINT(-12.239414172543265 14.76277825595468)			CERES
2021-11-25T08:36:33Z	AR0408	Monitoring point	SRID=4326;POINT(-12.239427786385678 14.762435194193658)			CERES
2021-11-25T08:36:33Z	AR0409	Monitoring point	SRID=4326;POINT(-12.239277541473582 14.761837271031267)			CERES
2021-11-25T08:36:33Z	AR0410	Monitoring point	SRID=4326;POINT(-12.239184919208808 14.761086489033076)			CERES
2021-11-25T08:36:33Z	AR0411	Monitoring point	SRID=4326;POINT(-12.239195181463309 14.76027371155549)			CERES
2021-11-25T08:36:33Z	BA0401	Monitoring point	SRID=4326;POINT(-12.453851852198637 14.899616716041482)			CERES
2021-11-25T08:36:33Z	BA0402	Monitoring point	SRID=4326;POINT(-12.453529789309776 14.900146111748306)			CERES

Showing 1 to 10 of 372 entries

Previous 1 2 3 4 5 … 38 Next

Once a table tab is selected and the table content visualised, the user can:

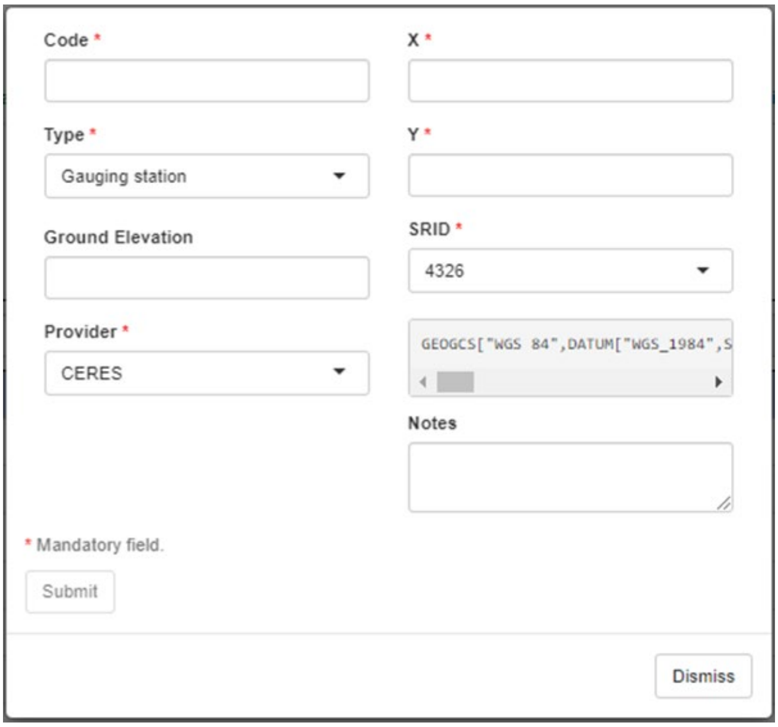
- add a new record (e.g. a new monitoring point in **Figure 8**). The data entry form is customised and dropdown lists provide selection options after database content where relevant, as for the point type, provider and spatial reference system in the specific example.

all the compulsory attribute will fail; examples include exceeding the maximum number of characters for a text string, not respecting an ISO8601 date format, trying to delete a law when referenced law limits are already stored to the database.

— edit an existing selected record (e.g. a law/EQS limit in **Figure 9**).

— delete one or more selected record(s) (given the potential risks, a popup window shows up to ask for user confirmation, **Figure 10**).

Figure 8. Database editing: Add (new measurement) point interface



Code *

X *

Type *
Gauging station

Y *

Ground Elevation

SRID *
4326

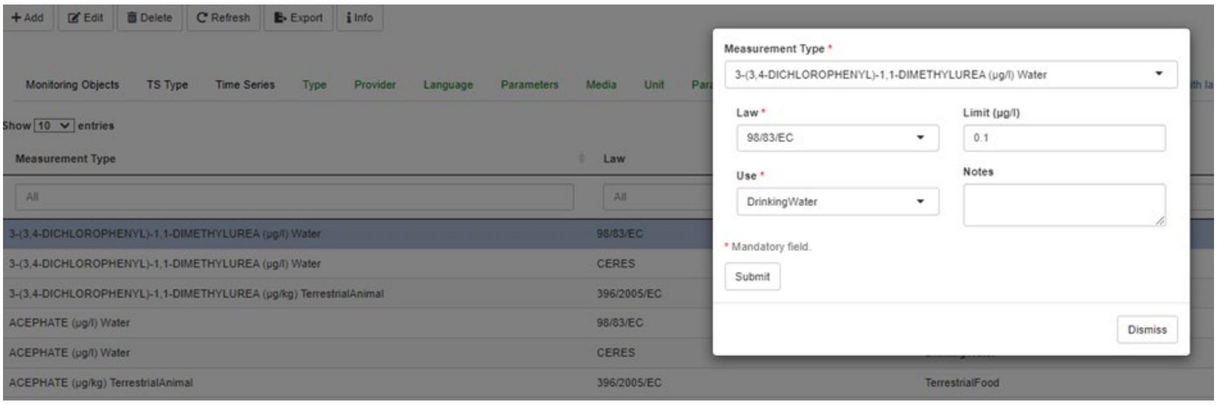
Provider *
CERES

GEOGCS["WGS 84",DATUM["WGS_1984",S

Notes

* Mandatory field.

Figure 9. Database editing: single record editing interface



Show 10 entries

Measurement Type	Law	
All	All	
3-(3,4-DICHLOROPHENYL)-1,1-DIMETHYLUREA (µg/l) Water	98/83/EC	
3-(3,4-DICHLOROPHENYL)-1,1-DIMETHYLUREA (µg/l) Water	CERES	
3-(3,4-DICHLOROPHENYL)-1,1-DIMETHYLUREA (µg/kg) TerrestrialAnimal	396/2005/EC	
ACEPHATE (µg/l) Water	98/83/EC	
ACEPHATE (µg/l) Water	CERES	
ACEPHATE (µg/kg) TerrestrialAnimal	396/2005/EC	TerrestrialFood

Measurement Type *
3-(3,4-DICHLOROPHENYL)-1,1-DIMETHYLUREA (µg/l) Water

Law *
98/83/EC

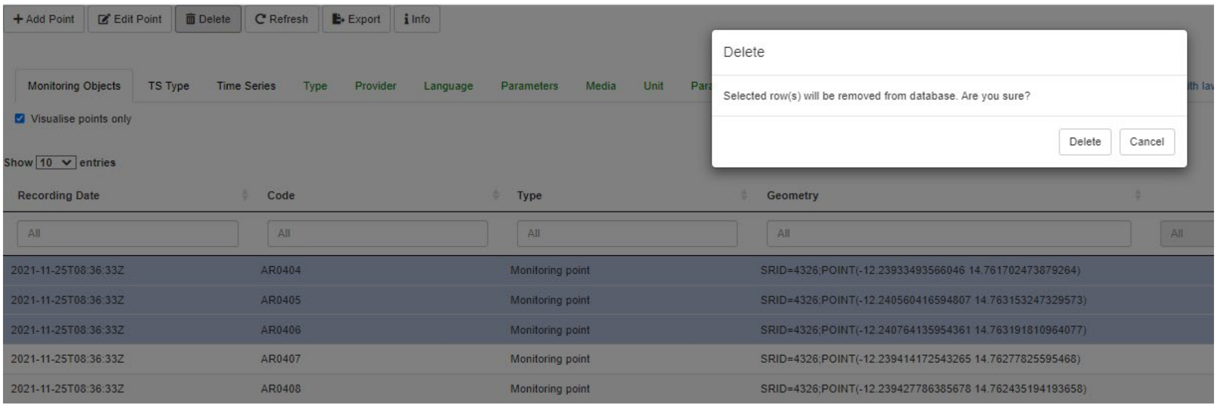
Limit (µg/l)
0.1

Use *
DrinkingWater

Notes

* Mandatory field.

Figure 10. Database editing: record(s) deletion interface



Selected row(s) will be removed from database. Are you sure?

Recording Date	Code	Type	Geometry
All	All	All	All
2021-11-25T08:36:33Z	AR0404	Monitoring point	SRID=4326;POINT(-12.23933493566046 14.761702473879264)
2021-11-25T08:36:33Z	AR0405	Monitoring point	SRID=4326;POINT(-12.240560416594807 14.763153247329573)
2021-11-25T08:36:33Z	AR0406	Monitoring point	SRID=4326;POINT(-12.240764135954361 14.763191810964077)
2021-11-25T08:36:33Z	AR0407	Monitoring point	SRID=4326;POINT(-12.239414172543265 14.76277825595468)
2021-11-25T08:36:33Z	AR0408	Monitoring point	SRID=4326;POINT(-12.239427786385678 14.762435194193658)

Tables or queries visualisation can be refreshed at any time (‘Refresh’ button) and data can be exported to a CSV file (‘Export’ button), typically for use in an external application. Note that many applications have native capabilities to connect, read from and even write to a PostgreSQL/PostGIS database; if this is the case, it is highly recommended to directly access the original data rather than exporting data to file, in order to avoid data duplication that may reasonably lead to inconsistencies.

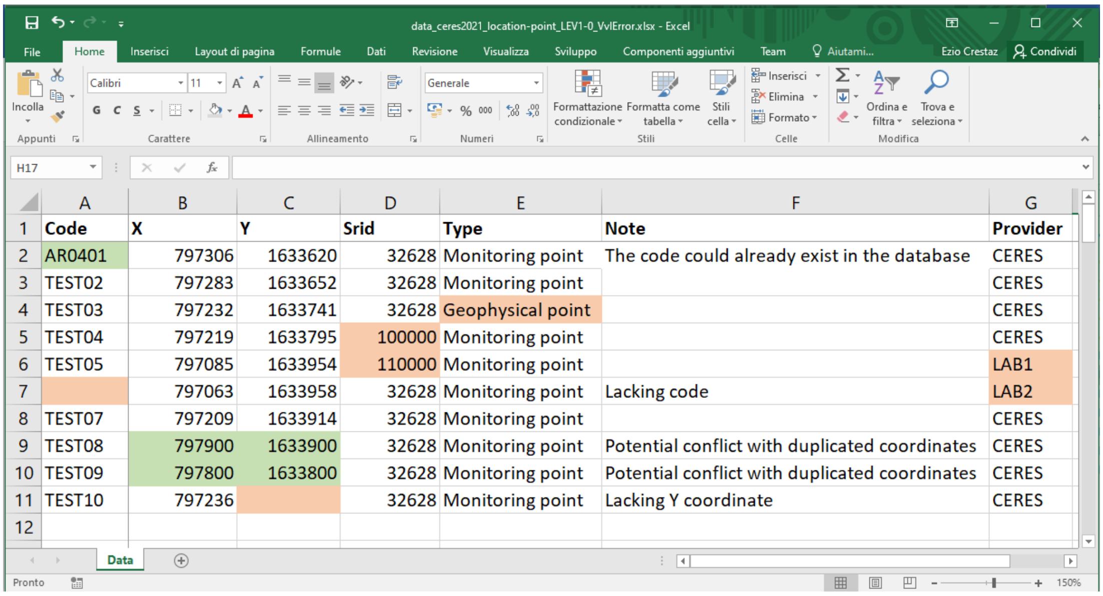
A last option (‘Info’ button) gives access to a detailed EAR database schema (**Figure 4**), reporting also the details of the existing relationships and their cardinality. The schema is slightly simplified and reproduced at the bottom of each table visualisation tab, the specific table been highlighted in bold to keep the user aware of the object (s)he is currently editing.

4.3 Data files validation and massive database uploading

The component enables to validate data files and to upload the data to the database. Populating the database record by record, as detailed in the previous paragraph, is perfectly fine for small datasets and well needed for editing specific records. In most frequent operational scenarios, the use of standard file formats to organise the original raw data (e.g. text, MS Excel) looks by far the most promising option. The data can be already available in digital format (e.g. provided by analytical laboratories) or they can be enough complex to benefit from the use of an advanced spreadsheet as MS Excel. Data files must be checked in order to detect inconsistencies or errors that need to be fixed before any attempt to upload to the database. The current application component makes such a validation task possible, based on constraining rules that are defined through standard templates. Both data files and templates are currently expected to be in MS Excel file format. Once the data files are validated, which means that they contain **locational** or **time series** data compliant with provided templates, the application supports the next step of data uploading to the database.

The MS Excel file of locational data here below (**Figure 11**) provides an example of few of the many different errors that may occur: data lacking for a compulsory attribute (e.g. code, coordinate(s)), reference to non-existent spatial reference ID, monitoring point type or provider. Other errors, as monitoring point code or location already existing in the database (e.g. previously uploaded), can be detected only at the later stage of database data uploading.

Figure 11. Locational data files: MS Excel sample with errors



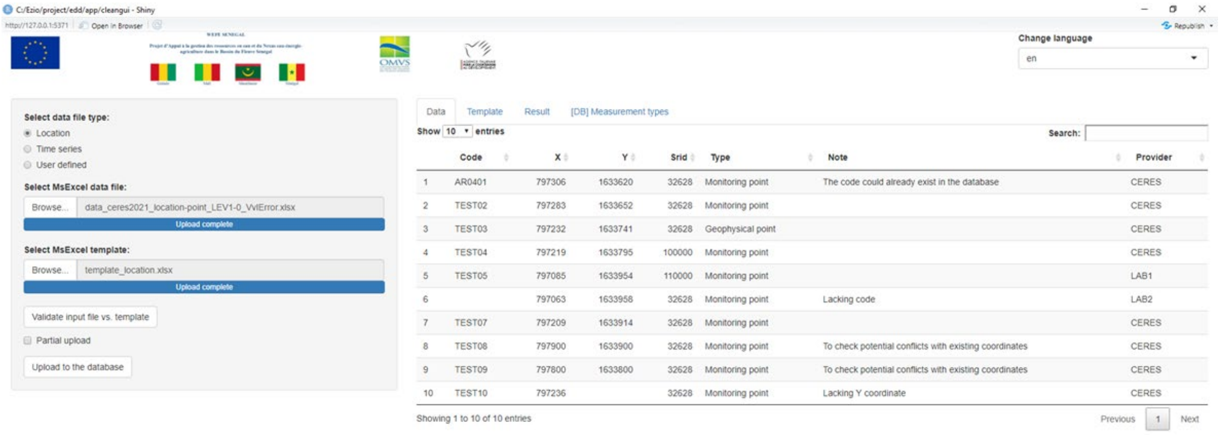
	A	B	C	D	E	F	G
1	Code	X	Y	Srid	Type	Note	Provider
2	AR0401	797306	1633620	32628	Monitoring point	The code could already exist in the database	CERES
3	TEST02	797283	1633652	32628	Monitoring point		CERES
4	TEST03	797232	1633741	32628	Geophysical point		CERES
5	TEST04	797219	1633795	100000	Monitoring point		CERES
6	TEST05	797085	1633954	110000	Monitoring point		LAB1
7		797063	1633958	32628	Monitoring point	Lacking code	LAB2
8	TEST07	797209	1633914	32628	Monitoring point		CERES
9	TEST08	797900	1633900	32628	Monitoring point	Potential conflict with duplicated coordinates	CERES
10	TEST09	797800	1633800	32628	Monitoring point	Potential conflict with duplicated coordinates	CERES
11	TEST10	797236		32628	Monitoring point	Lacking Y coordinate	CERES
12							

The application screenshots below capture the key steps of the validation process for locational data, namely:

- Uploading and visualisation of locational data (**Figure 12**). The user is expected to select the file type ('Location') and browse the directory tree to identify and select the MS Excel containing the locational data. In the example, the data have been imported from the MS Excel file previously discussed.
- Uploading and visualisation of template for locational data (**Figure 13**). The template, provided as a standard reference in MS Excel format, details all the constraints against which the locational data will be tested, namely column ordered by sequence and name, data type and any additional constraint (e.g. maximum number of characters in parentheses for TEXT attributes), if it is compulsory or not, if the content must be validated against a reference list, hereafter referred as VVL (Valid Value List) consistently with the US EPA EDD definition. The user is expected to browse the directory tree to identify and select the MS Excel predefined template.
- Validating the locational data vs. the rules defined in the template (**Figure 14**). Data file content is analysed cell by cell and errors are reported, including the occurrence location (row and column), the current value and a message detailing the error nature.
- Validating the locational data at database uploading stage (**Figure 15**). Data are uploaded to the database record by record, any conflict resulting in an abortion and being reported to the user for further action. In fact, while it is assumed that data are consistent with the template rules at this stage, it is not given for granted that they do not have conflicts with information already stored to the database; a typical conflict arises from the same record already existing in the database, e.g. as resulting from a former uploading. It is worth to highlight that both failing SQL record insertion statement and the error message issued by the database are reported. At first sight, this information can seem a bit weird and too complex, mainly relevant to a database expert who can actually make use of it to properly identify the nature and the location (what record?) of the error. However, a closely look up at the database message reveals that the majority of errors fall in a few main categories, as:
  - A record being already stored in the database or
  - A compulsory attribute constrained to the content of a supporting database table (e.g. provider, language, parameter) not being found in the latter, typically resulting in a NOT NULL constrain violation.

  Database error messages are generally reported in the language of the Operative System (OS) of the computer, but, just in case, PostgreSQL can also be set to use a different language.

Figure 12. Locational data files validation: data uploading



Select data file type:

● Location
○ Time series
○ User defined

Select MsExcel data file:

Browse... data_ceres2021_location-point_LEV1-0_VvlError.xlsx

Upload complete

Select MsExcel template:

Browse... template_location.xlsx

Upload complete

Validate input file vs. template

☐ Partial upload

Upload to the database

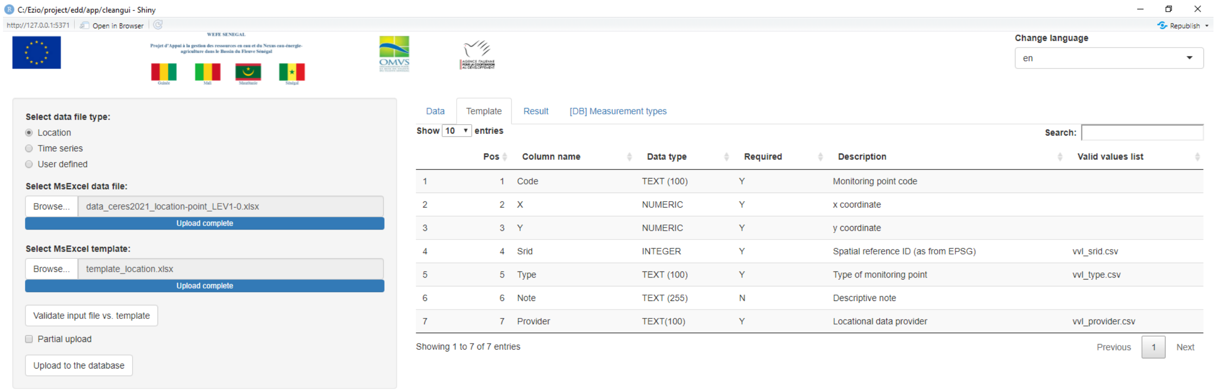
Show 10 entries

Search:

	Code	X	Y	Srid	Type	Note	Provider
1	AR0401	797306	1633620	32628	Monitoring point	The code could already exist in the database	CERES
2	TEST02	797283	1633652	32628	Monitoring point		CERES
3	TEST03	797232	1633741	32628	Geophysical point		CERES
4	TEST04	797219	1633795	100000	Monitoring point		CERES
5	TEST05	797085	1633954	110000	Monitoring point		LAB1
6		797063	1633958	32628	Monitoring point	Lacking code	LAB2
7	TEST07	797209	1633914	32628	Monitoring point		CERES
8	TEST08	797900	1633900	32628	Monitoring point	To check potential conflicts with existing coordinates	CERES
9	TEST09	797800	1633800	32628	Monitoring point	To check potential conflicts with existing coordinates	CERES
10	TEST10	797236		32628	Monitoring point	Lacking Y coordinate	CERES

Showing 1 to 10 of 10 entries

Figure 13. Locational data files validation: reference template



Select data file type:

● Location
○ Time series
○ User defined

Select MsExcel data file:

Browse... data_ceres2021_location-point_LEV1-0.xlsx

Upload complete

Select MsExcel template:

Browse... template_location.xlsx

Upload complete

Validate input file vs. template

☐ Partial upload

Upload to the database

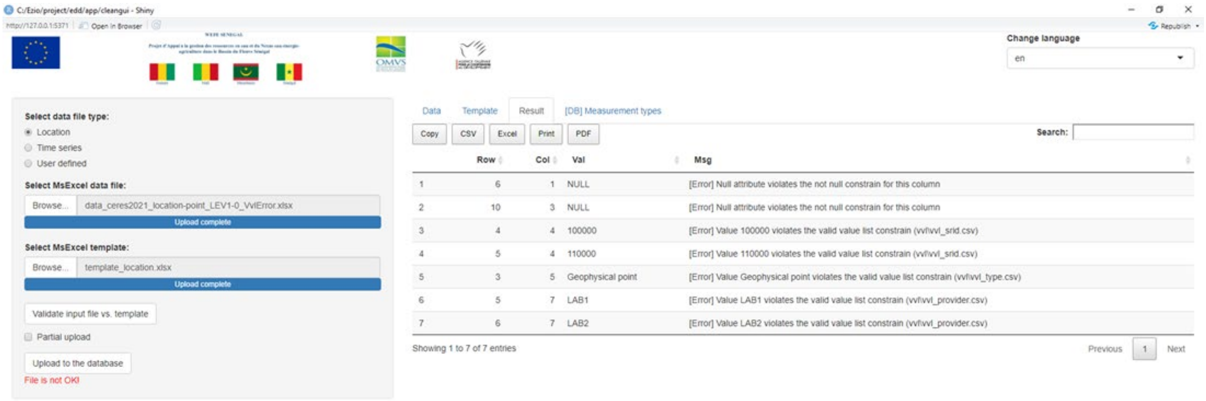
Show 10 entries

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	vvl_provider.csv

Showing 1 to 7 of 7 entries

Figure 14. Locational data files validation: errors reporting



Select data file type:

● Location
○ Time series
○ User defined

Select MsExcel data file:

Browse... data_ceres2021_location-point_LEV1-0_VvlError.xlsx

Upload complete

Select MsExcel template:

Browse... template_location.xlsx

Upload complete

Validate input file vs. template

☐ Partial upload

Upload to the database

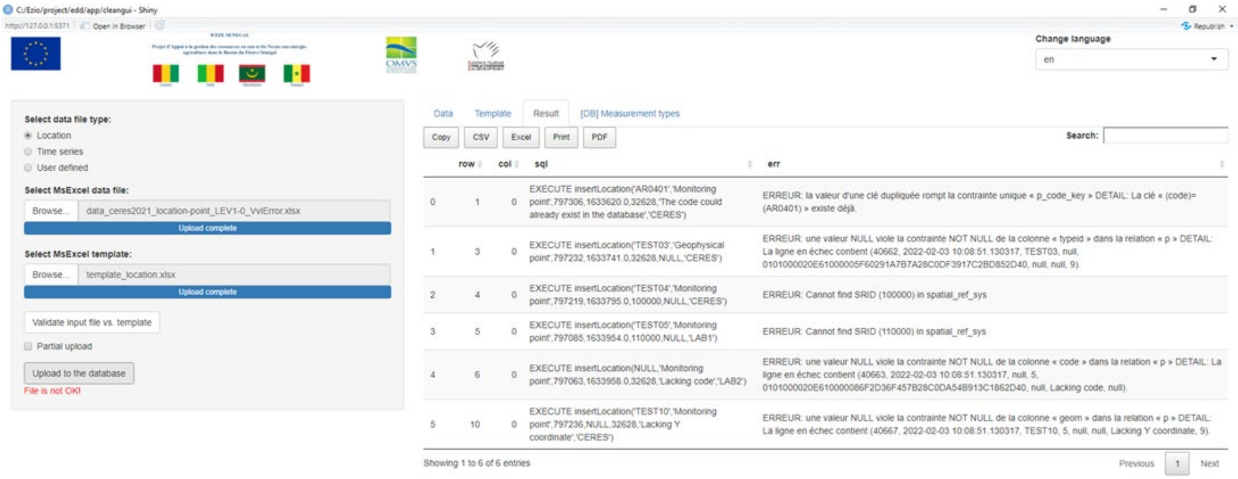
File is not OK!

Search:

	Row	Col	Val	Msg
1	6	1	NULL	[Error] Null attribute violates the not null constrain for this column
2	10	3	NULL	[Error] Null attribute violates the not null constrain for this column
3	4	4	100000	[Error] Value 100000 violates the valid value list constrain (vvl\vvl_srid.csv)
4	5	4	110000	[Error] Value 110000 violates the valid value list constrain (vvl\vvl_srid.csv)
5	3	5	Geophysical point	[Error] Value Geophysical point violates the valid value list constrain (vvl\vvl_type.csv)
6	5	7	LAB1	[Error] Value LAB1 violates the valid value list constrain (vvl\vvl_provider.csv)
7	6	7	LAB2	[Error] Value LAB2 violates the valid value list constrain (vvl\vvl_provider.csv)

Showing 1 to 7 of 7 entries

Figure 15. Locational data files uploading to database: errors reporting



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

Select data file type:

- ● Location
- ○ Time series
- ○ User defined

Select MsExcel data file: Browse... data_ceres2021_location-point_LEV1-0_VvlError.xlsx — Upload complete

Select MsExcel template: Browse... template_location.xlsx — Upload complete

Validate input file vs. template

□ Partial upload

Upload to the database

File is not OK!

Change language: en

Data | Template | Result | [DB] Measurement types

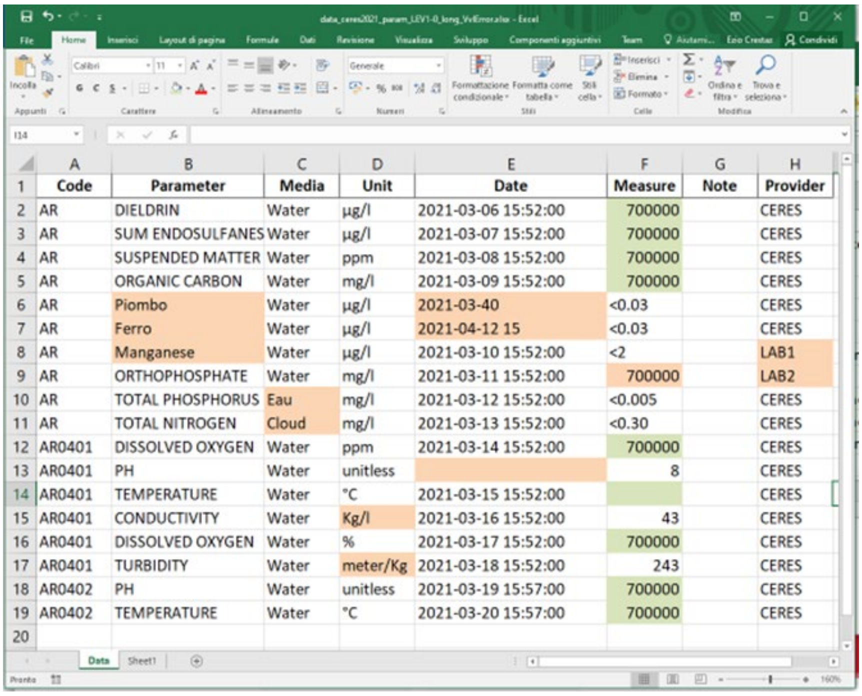
Copy | CSV | Excel | Print | PDF — Search:

	row	col	sql	err
0	1	0	EXECUTE insertLocation('AR0401','Monitoring point',797306,1633620.0,32628,'The code could already exist in the database','CERES')	ERREUR: la valeur d'une clé dupliquée rompt la contrainte unique « p_code_key » DETAIL: La clé « (code)=(AR0401) » existe déjà.
1	3	0	EXECUTE insertLocation('TEST03','Geophysical point',797232,1633741.0,32628,NULL,'CERES')	ERREUR: une valeur NULL viole la contrainte NOT NULL de la colonne « typeid » dans la relation « p » DETAIL: La ligne en échec contient (40662, 2022-02-03 10:08:51.130317, TEST03, null, 0101000020E61000005F60291A7B7A28C0DF3917C2BD852D40, null, null, 9).
2	4	0	EXECUTE insertLocation('TEST04','Monitoring point',797219,1633795.0,100000,NULL,'CERES')	ERREUR: Cannot find SRID (100000) in spatial_ref_sys
3	5	0	EXECUTE insertLocation('TEST05','Monitoring point',797085,1633954.0,110000,NULL,'LAB1')	ERREUR: Cannot find SRID (110000) in spatial_ref_sys
4	6	0	EXECUTE insertLocation(NULL,'Monitoring point',797063,1633958.0,32628,'Lacking code','LAB2')	ERREUR: une valeur NULL viole la contrainte NOT NULL de la colonne « code » dans la relation « p » DETAIL: La ligne en échec contient (40663, 2022-02-03 10:08:51.130317, null, 5, 0101000020E610000086F2D36F457B28C0DA54B913C1862D40, null, Lacking code, null).
5	10	0	EXECUTE insertLocation('TEST10','Monitoring point',797236,NULL,32628,'Lacking Y coordinate','CERES')	ERREUR: une valeur NULL viole la contrainte NOT NULL de la colonne « geom » dans la relation « p » DETAIL: La ligne en échec contient (40667, 2022-02-03 10:08:51.130317, TEST10, 5, null, null, Lacking Y coordinate, 9).

Showing 1 to 6 of 6 entries — Previous 1 Next

Similarly to locational data, the MS Excel file of time series data here below (**Figure 16**) provides an example of a few of the many different errors that may occur: data lacking for a compulsory attribute (e.g. code, parameter, date, measure), reference to non-existent parameter, media, unit or provider in the related database tables. Other errors, as measures already existing in the database (e.g. previously uploaded), can be detected only at the later stage of database data uploading.

Figure 16. Time series data files: MS Excel sample with errors



data_ceres2021_param_LEV1-0_long_VvError.xlsx - Excel

	A	B	C	D	E	F	G	H
1	Code	Parameter	Media	Unit	Date	Measure	Note	Provider
2	AR	DIELDRIN	Water	µg/l	2021-03-06 15:52:00	700000		CERES
3	AR	SUM ENDOSULFANES	Water	µg/l	2021-03-07 15:52:00	700000		CERES
4	AR	SUSPENDED MATTER	Water	ppm	2021-03-08 15:52:00	700000		CERES
5	AR	ORGANIC CARBON	Water	mg/l	2021-03-09 15:52:00	700000		CERES
6	AR	Piombo	Water	µg/l	2021-03-40	<0.03		CERES
7	AR	Ferro	Water	µg/l	2021-04-12 15	<0.03		CERES
8	AR	Manganese	Water	µg/l	2021-03-10 15:52:00	<2		LAB1
9	AR	ORTHOPHOSPHATE	Water	mg/l	2021-03-11 15:52:00	700000		LAB2
10	AR	TOTAL PHOSPHORUS	Eau	mg/l	2021-03-12 15:52:00	<0.005		CERES
11	AR	TOTAL NITROGEN	Cloud	mg/l	2021-03-13 15:52:00	<0.30		CERES
12	AR0401	DISSOLVED OXYGEN	Water	ppm	2021-03-14 15:52:00	700000		CERES
13	AR0401	PH	Water	unitless		8		CERES
14	AR0401	TEMPERATURE	Water	°C	2021-03-15 15:52:00			CERES
15	AR0401	CONDUCTIVITY	Water	Kg/l	2021-03-16 15:52:00	43		CERES
16	AR0401	DISSOLVED OXYGEN	Water	%	2021-03-17 15:52:00	700000		CERES
17	AR0401	TURBIDITY	Water	meter/Kg	2021-03-18 15:52:00	243		CERES
18	AR0402	PH	Water	unitless	2021-03-19 15:57:00	700000		CERES
19	AR0402	TEMPERATURE	Water	°C	2021-03-20 15:57:00	700000		CERES
20								

The application screenshots below capture the key steps of the validation process for time series data, namely:

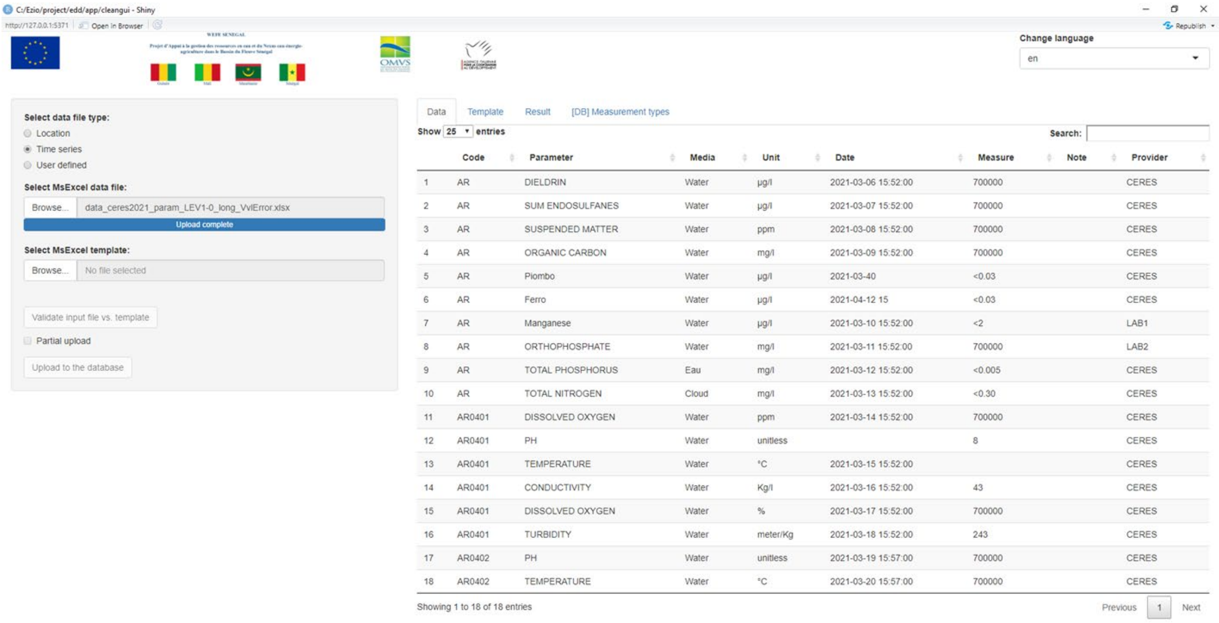
- Uploading and tabular visualisation of time series data (**Figure 17**), the user being expected to select the file type 'Time series'.
- Uploading and tabular visualisation of template for time series data (**Figure 18**).

— Validating the time series data vs. the rules defined in the template (**Figure 19**).

— Validating the time series data at database uploading stage (**Figure 20**).

The validation process, against the reference time series data template and at database uploading stage, follows the same pattern as for locational data, so reference can be made to previous paragraphs for further details.

Figure 17. Time series data files validation: data uploading



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

http://127.0.0.1:5371

Change language: en

Select data file type:
○ Location
● Time series
○ User defined

Select MsExcel data file: data_ceres2021_param_LEV1-0_long_VvlError.xlsx — Upload complete

Select MsExcel template: No file selected

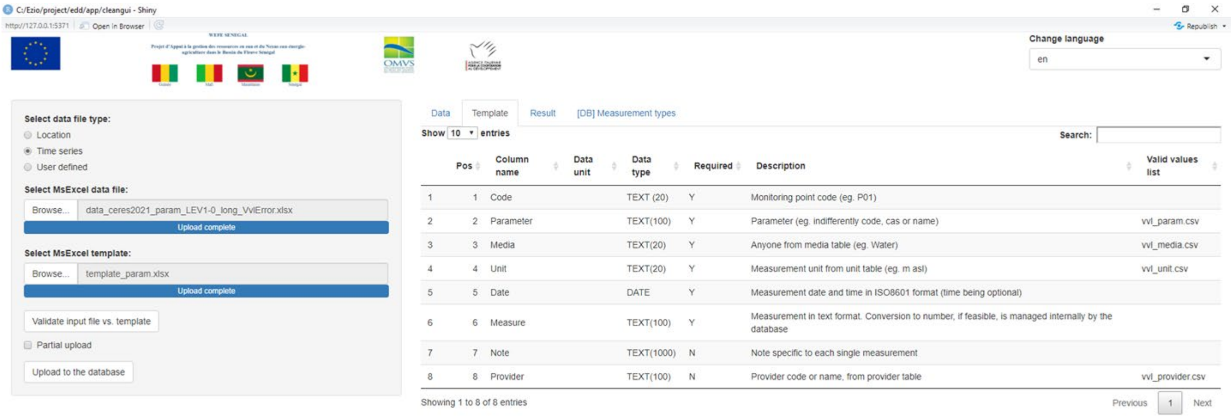
□ Partial upload

Show 25 entries

	Code	Parameter	Media	Unit	Date	Measure	Note	Provider
1	AR	DIELDRIN	Water	µg/l	2021-03-06 15:52:00	700000		CERES
2	AR	SUM ENDOSULFANES	Water	µg/l	2021-03-07 15:52:00	700000		CERES
3	AR	SUSPENDED MATTER	Water	ppm	2021-03-08 15:52:00	700000		CERES
4	AR	ORGANIC CARBON	Water	mg/l	2021-03-09 15:52:00	700000		CERES
5	AR	Piombo	Water	µg/l	2021-03-40	<0.03		CERES
6	AR	Ferro	Water	µg/l	2021-04-12 15	<0.03		CERES
7	AR	Manganese	Water	µg/l	2021-03-10 15:52:00	<2		LAB1
8	AR	ORTHOPHOSPHATE	Water	mg/l	2021-03-11 15:52:00	700000		LAB2
9	AR	TOTAL PHOSPHORUS	Eau	mg/l	2021-03-12 15:52:00	<0.005		CERES
10	AR	TOTAL NITROGEN	Cloud	mg/l	2021-03-13 15:52:00	<0.30		CERES
11	AR0401	DISSOLVED OXYGEN	Water	ppm	2021-03-14 15:52:00	700000		CERES
12	AR0401	PH	Water	unitless		8		CERES
13	AR0401	TEMPERATURE	Water	°C	2021-03-15 15:52:00			CERES
14	AR0401	CONDUCTIVITY	Water	Kg/l	2021-03-16 15:52:00	43		CERES
15	AR0401	DISSOLVED OXYGEN	Water	%	2021-03-17 15:52:00	700000		CERES
16	AR0401	TURBIDITY	Water	meter/Kg	2021-03-18 15:52:00	243		CERES
17	AR0402	PH	Water	unitless	2021-03-19 15:57:00	700000		CERES
18	AR0402	TEMPERATURE	Water	°C	2021-03-20 15:57:00	700000		CERES

Showing 1 to 18 of 18 entries

Figure 18. Time series data files validation: reference template



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

http://127.0.0.1:5371

Change language: en

Select data file type:
○ Location
● Time series
○ User defined

Select MsExcel data file: data_ceres2021_param_LEV1-0_long_VvlError.xlsx — Upload complete

Select MsExcel template: template_param.xlsx — Upload complete

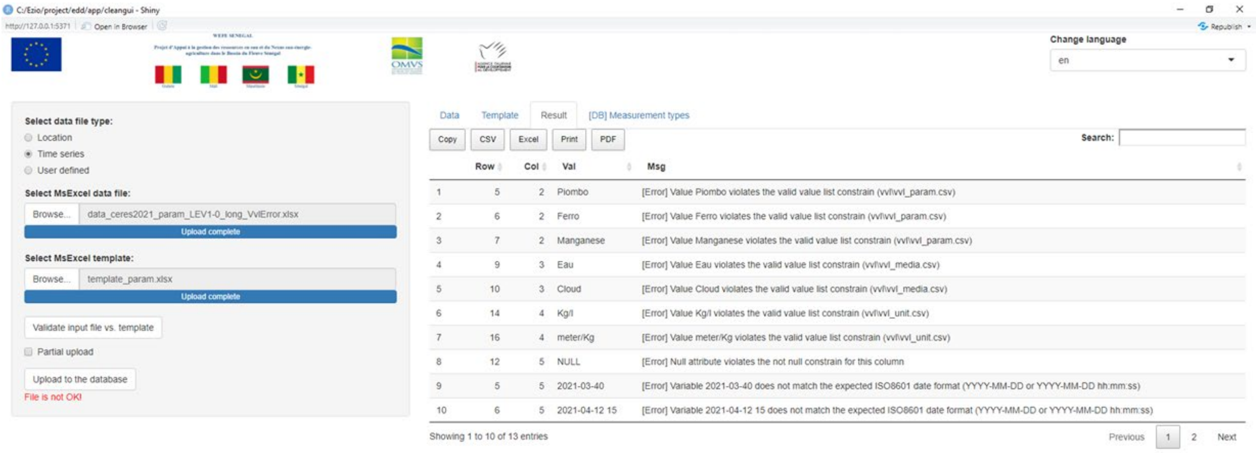
□ Partial upload

Show 10 entries

	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

Figure 19. Time series data files validation: errors reporting (two pages)



Select data file type:
○ Location
● Time series
○ User defined

Select MsExcel data file: data_ceres2021_param_LEV1-0_long_VvlError.xlsx

Select MsExcel template: template_param.xlsx

□ Partial upload

File is not OK!

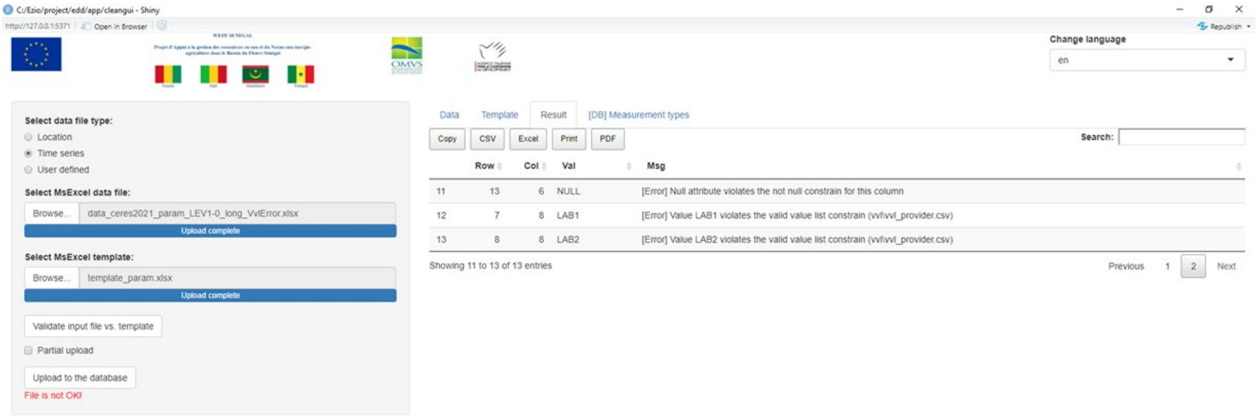
Change language: en

Search:

	Row	Col	Val	Msg
1	5	2	Piombo	[Error] Value Piombo violates the valid value list constrain (vvl\vvl_param.csv)
2	6	2	Ferro	[Error] Value Ferro violates the valid value list constrain (vvl\vvl_param.csv)
3	7	2	Manganese	[Error] Value Manganese violates the valid value list constrain (vvl\vvl_param.csv)
4	9	3	Eau	[Error] Value Eau violates the valid value list constrain (vvl\vvl_media.csv)
5	10	3	Cloud	[Error] Value Cloud violates the valid value list constrain (vvl\vvl_media.csv)
6	14	4	Kg/l	[Error] Value Kg/l violates the valid value list constrain (vvl\vvl_unit.csv)
7	16	4	meter/Kg	[Error] Value meter/Kg violates the valid value list constrain (vvl\vvl_unit.csv)
8	12	5	NULL	[Error] Null attribute violates the not null constrain for this column
9	5	5	2021-03-40	[Error] Variable 2021-03-40 does not match the expected ISO8601 date format (YYYY-MM-DD or YYYY-MM-DD hh:mm:ss)
10	6	5	2021-04-12 15	[Error] Variable 2021-04-12 15 does not match the expected ISO8601 date format (YYYY-MM-DD or YYYY-MM-DD hh:mm:ss)

Showing 1 to 10 of 13 entries

Previous 1 2 Next



Select data file type:
○ Location
● Time series
○ User defined

Select MsExcel data file: data_ceres2021_param_LEV1-0_long_VvlError.xlsx

Select MsExcel template: template_param.xlsx

□ Partial upload

File is not OK!

Change language: en

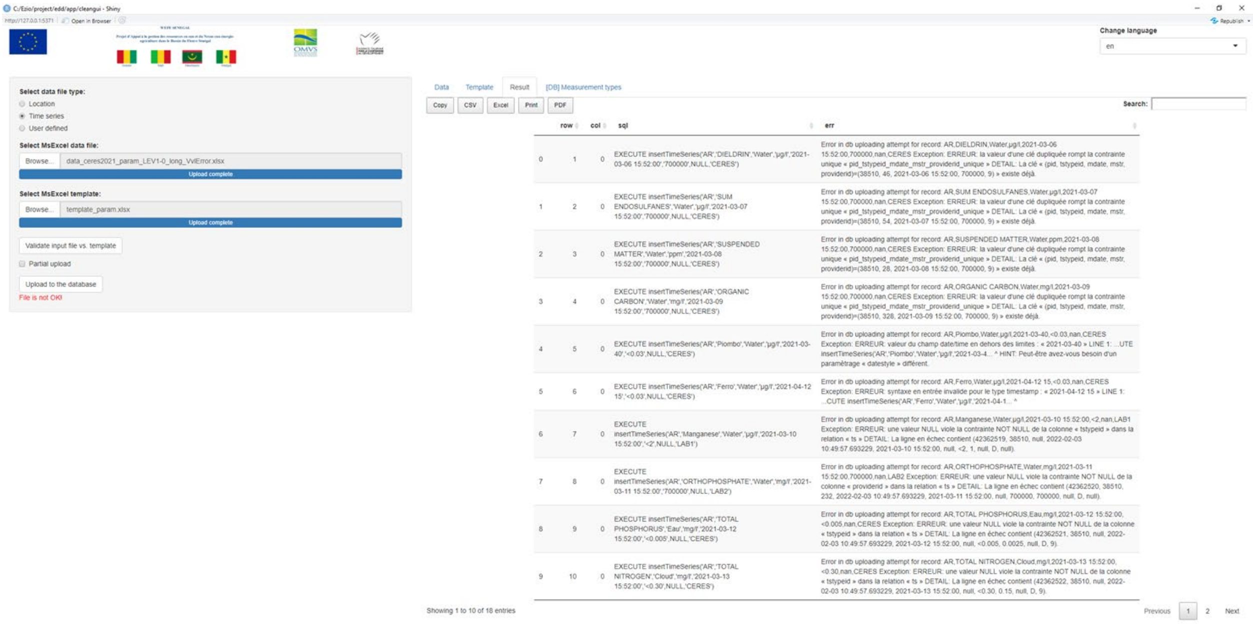
Search:

	Row	Col	Val	Msg
11	13	6	NULL	[Error] Null attribute violates the not null constrain for this column
12	7	8	LAB1	[Error] Value LAB1 violates the valid value list constrain (vvl\vvl_provider.csv)
13	8	8	LAB2	[Error] Value LAB2 violates the valid value list constrain (vvl\vvl_provider.csv)

Showing 11 to 13 of 13 entries

Previous 1 2 Next

Figure 20. Time series data files uploading to database: errors reporting (2 pages)



Select data file type:
○ Location
● Time series
○ User defined

Select MsExcel data file: data_ceres2021_param_LEV1-0_long_VvlError.xlsx

Select MsExcel template: template_param.xlsx

□ Partial upload

File is not OK!

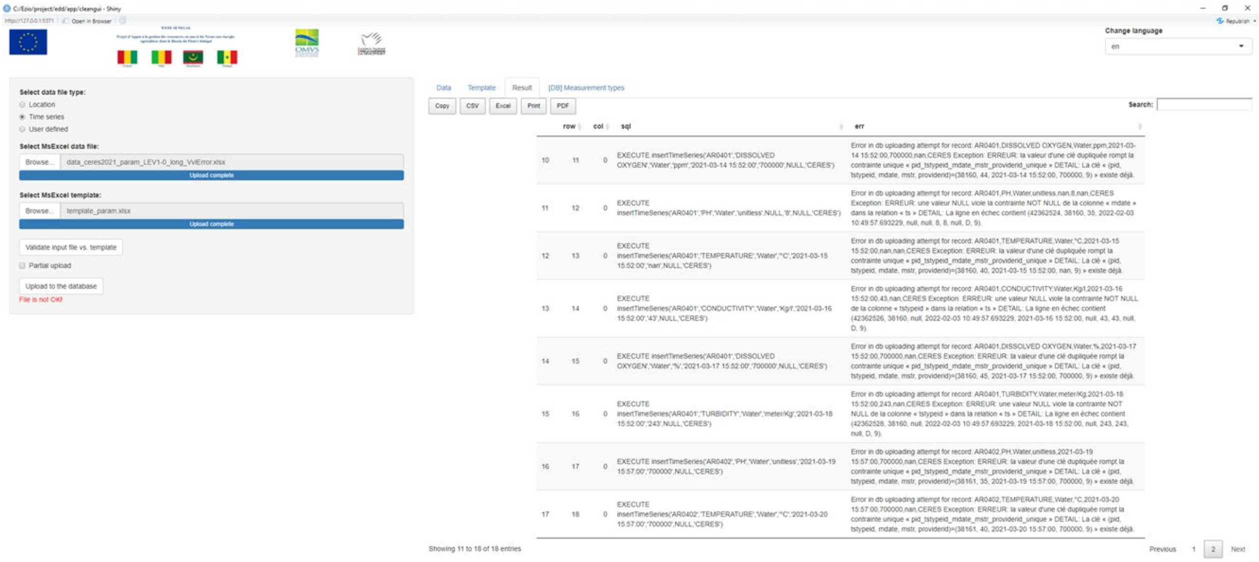
Change language: en

Search:

	row	col	sql	err
0	1	0	EXECUTE insertTimeSeries('AR','DIELDRIN','Water','µg/l','2021-03-06 15:52:00','700000',NULL,'CERES')	Error in db uploading attempt for record: AR,DIELDRIN,Water,µg/l,2021-03-06 15:52:00,700000,nan,CERES Exception: ERREUR: la valeur d'une clé dupliquée rompt la contrainte unique « pid_tstypeid_mdate_mstr_providerid_unique » DETAIL: La clé « (pid, tstypeid, mdate, mstr, providerid)=(38510, 46, 2021-03-06 15:52:00, 700000, 9) » existe déjà.
1	2	0	EXECUTE insertTimeSeries('AR','SUM ENDOSULFANES','Water','µg/l','2021-03-07 15:52:00','700000',NULL,'CERES')	Error in db uploading attempt for record: AR,SUM ENDOSULFANES,Water,µg/l,2021-03-07 15:52:00,700000,nan,CERES Exception: ERREUR: la valeur d'une clé dupliquée rompt la contrainte unique « pid_tstypeid_mdate_mstr_providerid_unique » DETAIL: La clé « (pid, tstypeid, mdate, mstr, providerid)=(38510, 54, 2021-03-07 15:52:00, 700000, 9) » existe déjà.
2	3	0	EXECUTE insertTimeSeries('AR','SUSPENDED MATTER','Water','ppm','2021-03-08 15:52:00','700000',NULL,'CERES')	Error in db uploading attempt for record: AR,SUSPENDED MATTER,Water,ppm,2021-03-08 15:52:00,700000,nan,CERES Exception: ERREUR: la valeur d'une clé dupliquée rompt la contrainte unique « pid_tstypeid_mdate_mstr_providerid_unique » DETAIL: La clé « (pid, tstypeid, mdate, mstr, providerid)=(38510, 28, 2021-03-08 15:52:00, 700000, 9) » existe déjà.
3	4	0	EXECUTE insertTimeSeries('AR','ORGANIC CARBON','Water','mg/l','2021-03-09 15:52:00','700000',NULL,'CERES')	Error in db uploading attempt for record: AR,ORGANIC CARBON,Water,mg/l,2021-03-09 15:52:00,700000,nan,CERES Exception: ERREUR: la valeur d'une clé dupliquée rompt la contrainte unique « pid_tstypeid_mdate_mstr_providerid_unique » DETAIL: La clé « (pid, tstypeid, mdate, mstr, providerid)=(38510, 328, 2021-03-09 15:52:00, 700000, 9) » existe déjà.
4	5	0	EXECUTE insertTimeSeries('AR','Piombo','Water','µg/l','2021-03-40','<0.03',NULL,'CERES')	Error in db uploading attempt for record: AR,Piombo,Water,µg/l,2021-03-40,<0.03,nan,CERES Exception: ERREUR: valeur du champ date/time en dehors des limites : « 2021-03-40 » LINE 1: ...UTE insertTimeSeries('AR','Piombo','Water','µg/l','2021-03-4... ^ HINT: Peut-être avez-vous besoin d'un paramètrage « datestyle » différent.
5	6	0	EXECUTE insertTimeSeries('AR','Ferro','Water','µg/l','2021-04-12 15','<0.03',NULL,'CERES')	Error in db uploading attempt for record: AR,Ferro,Water,µg/l,2021-04-12 15,<0.03,nan,CERES Exception: ERREUR: syntaxe en entrée invalide pour le type timestamp : « 2021-04-12 15 » LINE 1: ...CUTE insertTimeSeries('AR','Ferro','Water','µg/l','2021-04-1... ^
6	7	0	EXECUTE insertTimeSeries('AR','Manganese','Water','µg/l','2021-03-10 15:52:00','<2',NULL,'LAB1')	Error in db uploading attempt for record: AR,Manganese,Water,µg/l,2021-03-10 15:52:00,<2,nan,LAB1 Exception: ERREUR: une valeur NULL viole la contrainte NOT NULL de la colonne « tstypeid » dans la relation « ts » DETAIL: La ligne en échec contient (42362519, 38510, null, 2022-02-03 10:49:57.693229, 2021-03-10 15:52:00, null, <2, 1, null, D, null).
7	8	0	EXECUTE insertTimeSeries('AR','ORTHOPHOSPHATE','Water','mg/l','2021-03-11 15:52:00','700000',NULL,'LAB2')	Error in db uploading attempt for record: AR,ORTHOPHOSPHATE,Water,mg/l,2021-03-11 15:52:00,700000,nan,LAB2 Exception: ERREUR: une valeur NULL viole la contrainte NOT NULL de la colonne « providerid » dans la relation « ts » DETAIL: La ligne en échec contient (42362520, 38510, 232, 2022-02-03 10:49:57.693229, 2021-03-11 15:52:00, null, 700000, 700000, null, D, null).
8	9	0	EXECUTE insertTimeSeries('AR','TOTAL PHOSPHORUS','Eau','mg/l','2021-03-12 15:52:00','<0.005',NULL,'CERES')	Error in db uploading attempt for record: AR,TOTAL PHOSPHORUS,Eau,mg/l,2021-03-12 15:52:00,<0.005,nan,CERES Exception: ERREUR: une valeur NULL viole la contrainte NOT NULL de la colonne « tstypeid » dans la relation « ts » DETAIL: La ligne en échec contient (42362521, 38510, null, 2022-02-03 10:49:57.693229, 2021-03-12 15:52:00, null, <0.005, 0.0025, null, D, 9).
9	10	0	EXECUTE insertTimeSeries('AR','TOTAL NITROGEN','Cloud','mg/l','2021-03-13 15:52:00','<0.30',NULL,'CERES')	Error in db uploading attempt for record: AR,TOTAL NITROGEN,Cloud,mg/l,2021-03-13 15:52:00,<0.30,nan,CERES Exception: ERREUR: une valeur NULL viole la contrainte NOT NULL de la colonne « tstypeid » dans la relation « ts » DETAIL: La ligne en échec contient (42362522, 38510, null, 2022-02-03 10:49:57.693229, 2021-03-13 15:52:00, null, <0.30, 0.15, null, D, 9).

Showing 1 to 10 of 18 entries

Previous 1 2 Next



Select data file type:
○ Location
● Time series
○ User defined

Select MsExcel data file:
Browse... data_ceres2021_param_LEV1-0_long_VvlError.xlsx
Upload complete

Select MsExcel template:
Browse... template_param.xlsx
Upload complete

Validate input file vs. template

□ Partial upload

Upload to the database

File is not OK!

Change language: en

	row	col	sql	err
10	11	0	EXECUTE insertTimeSeries('AR0401','DISSOLVED OXYGEN','Water','ppm','2021-03-14 15:52:00','700000',NULL,'CERES')	Error in db uploading attempt for record: AR0401,DISSOLVED OXYGEN,Water,ppm,2021-03-14 15:52:00,700000,nan,CERES Exception: ERREUR: la valeur d'une clé dupliquée rompt la contrainte unique « pid_tstypeid_mdate_mstr_providerid_unique » DETAIL: La clé « (pid, tstypeid, mdate, mstr, providerid)=(38160, 44, 2021-03-14 15:52:00, 700000, 9) » existe déjà.
11	12	0	EXECUTE insertTimeSeries('AR0401','PH','Water','unitless',NULL,'8',NULL,'CERES')	Error in db uploading attempt for record: AR0401,PH,Water,unitless,nan,8,nan,CERES Exception: ERREUR: une valeur NULL viole la contrainte NOT NULL de la colonne « mdate » dans la relation « ts » DETAIL: La ligne en échec contient (42362524, 38160, 35, 2022-02-03 10:49:57.693229, null, null, 8, 8, null, D, 9).
12	13	0	EXECUTE insertTimeSeries('AR0401','TEMPERATURE','Water','°C','2021-03-15 15:52:00','nan',NULL,'CERES')	Error in db uploading attempt for record: AR0401,TEMPERATURE,Water,°C,2021-03-15 15:52:00,nan,nan,CERES Exception: ERREUR: la valeur d'une clé dupliquée rompt la contrainte unique « pid_tstypeid_mdate_mstr_providerid_unique » DETAIL: La clé « (pid, tstypeid, mdate, mstr, providerid)=(38160, 40, 2021-03-15 15:52:00, nan, 9) » existe déjà.
13	14	0	EXECUTE insertTimeSeries('AR0401','CONDUCTIVITY','Water','Kg/l','2021-03-16 15:52:00','43',NULL,'CERES')	Error in db uploading attempt for record: AR0401,CONDUCTIVITY,Water,Kg/l,2021-03-16 15:52:00,43,nan,CERES Exception: ERREUR: une valeur NULL viole la contrainte NOT NULL de la colonne « tstypeid » dans la relation « ts » DETAIL: La ligne en échec contient (42362526, 38160, null, 2022-02-03 10:49:57.693229, 2021-03-16 15:52:00, null, 43, 43, null, D, 9).
14	15	0	EXECUTE insertTimeSeries('AR0401','DISSOLVED OXYGEN','Water','%','2021-03-17 15:52:00','700000',NULL,'CERES')	Error in db uploading attempt for record: AR0401,DISSOLVED OXYGEN,Water,%,2021-03-17 15:52:00,700000,nan,CERES Exception: ERREUR: la valeur d'une clé dupliquée rompt la contrainte unique « pid_tstypeid_mdate_mstr_providerid_unique » DETAIL: La clé « (pid, tstypeid, mdate, mstr, providerid)=(38160, 45, 2021-03-17 15:52:00, 700000, 9) » existe déjà.
15	16	0	EXECUTE insertTimeSeries('AR0401','TURBIDITY','Water','meter/Kg','2021-03-18 15:52:00','243',NULL,'CERES')	Error in db uploading attempt for record: AR0401,TURBIDITY,Water,meter/Kg,2021-03-18 15:52:00,243,nan,CERES Exception: ERREUR: une valeur NULL viole la contrainte NOT NULL de la colonne « tstypeid » dans la relation « ts » DETAIL: La ligne en échec contient (42362528, 38160, null, 2022-02-03 10:49:57.693229, 2021-03-18 15:52:00, null, 243, 243, null, D, 9).
16	17	0	EXECUTE insertTimeSeries('AR0402','PH','Water','unitless','2021-03-19 15:57:00','700000',NULL,'CERES')	Error in db uploading attempt for record: AR0402,PH,Water,unitless,2021-03-19 15:57:00,700000,nan,CERES Exception: ERREUR: la valeur d'une clé dupliquée rompt la contrainte unique « pid_tstypeid_mdate_mstr_providerid_unique » DETAIL: La clé « (pid, tstypeid, mdate, mstr, providerid)=(38161, 35, 2021-03-19 15:57:00, 700000, 9) » existe déjà.
17	18	0	EXECUTE insertTimeSeries('AR0402','TEMPERATURE','Water','°C','2021-03-20 15:57:00','700000',NULL,'CERES')	Error in db uploading attempt for record: AR0402,TEMPERATURE,Water,°C,2021-03-20 15:57:00,700000,nan,CERES Exception: ERREUR: la valeur d'une clé dupliquée rompt la contrainte unique « pid_tstypeid_mdate_mstr_providerid_unique » DETAIL: La clé « (pid, tstypeid, mdate, mstr, providerid)=(38161, 40, 2021-03-20 15:57:00, 700000, 9) » existe déjà.

Showing 11 to 18 of 18 entries

4.4 Dashboards

The application provides basic spatio-temporal visualisation facilities through two distinct dashboards seamlessly connecting to the underlying database.

The first dashboard (**Figure 21**) is aimed at supporting quick spatio-temporal analysis of single parameters. The user selects the measurement type (combination of parameter, unit and media) through a dropdown list constrained to database time series, sets the symbol size (in map units) and, optionally, filters out locations at which only one measure is available in order to retain time series only. The data are visualised spatially and thematically in the mapping window, where classic zoom in/out, background map (OSM, Toner, Toner Lite) selection and distance/area computation tools are integrated. At the bottom, the data are visualised in both tabular and time-dependent graphs, enabling cross-linked brushing, querying, filtering and sorting.

The second dashboard (**Figure 22**) is aimed at more advanced spatio-temporal analysis of multiple parameters vs. law/EQS limits, based on the following filtering sequence through multi-options dropdown selection lists:

- Selection of provider(s) of interest.
- Selection of media of interest (one or more), all media for which monitored data by the selected provider(s) are available being reported.
- Selection of law(s) or EQS combination with use type (e.g. drinking water).

The user can filter data, retaining only those data exceeding the law/EQS limit, can set the symbol size as normalised to the given limit, and can filter data by date interval, using the related horizontal scroll bars and the check box on the web page.

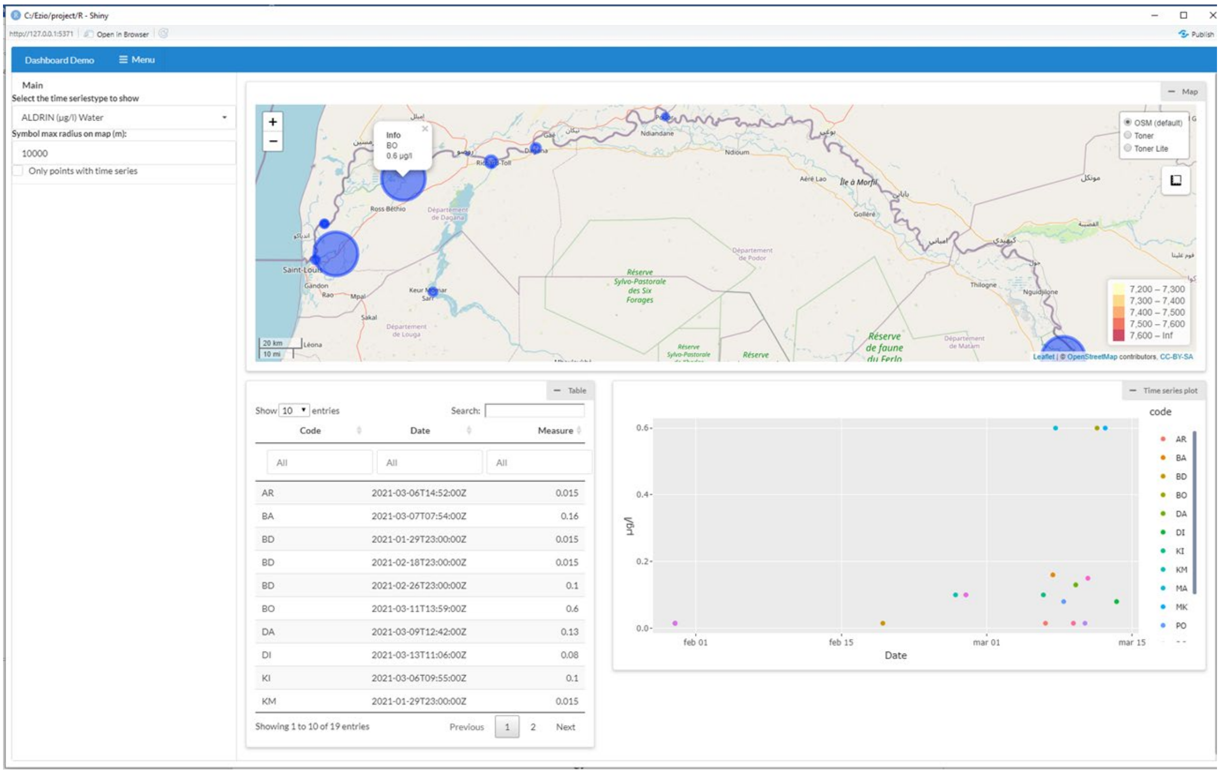
Among the scopes of the dashboard, the following ones can be mentioned:

- to quickly locate law/EQS limit exceedance in space in order to highlight potential spatial relationships, as contamination source, downstream insurgence and so on. A small button of triangular shape, at the right bottom of the date selection bar, supports animation and hence investigation of the occurrence of contamination in time, further than in space.
- to compare the results of different providers as in case of cross laboratories check and validation, and to compare exceedance for different parameters in view of possible relationships. Multiple measures of a given parameter at a monitoring point result in multiple symbols of potentially different size, conveying a first rough idea of the measurement variability over time.

Both dashboards were designed to provide a quick and effective access to the database content, to support user navigation experience through both the spatial and temporal dimensions, not demanding for any previous experience. Being a web application, the dashboards are run in the browser and should hopefully support the user in quickly spotting anomalies or trends that may be worth devoting further analysis. High level executives, senior managers and thematic experts in need of a quick overview of potentially challenging datasets are expected to benefit from such a tool.

It is worth to stress that the database is native, which means that it is not constrained by any specific front-end tool([12]). Hence, as previously stated, the database can be easily accessed from many different state-of-the-art GIS, analysis and modelling tools. This avoids the drawbacks of exporting and duplicating data (forward and backward), limiting the inherent risks of introducing inconsistencies and errors; cuts the development times; contributes to promoting a culture of professional data management; and boosts the chances of analysts to fully focus on their application domains.

Figure 21. Application dashboard for quick spatio-temporal exploratory analysis of single parameters



C:/Ezio/project/R - Shiny

http://127.0.0.1:5371 Open in Browser

Main

Select the time seriestype to show: ALDRIN (µg/l) Water

Symbol max radius on map (m): 10000

□ Only points with time series

Info
BO
0.6 µg/l

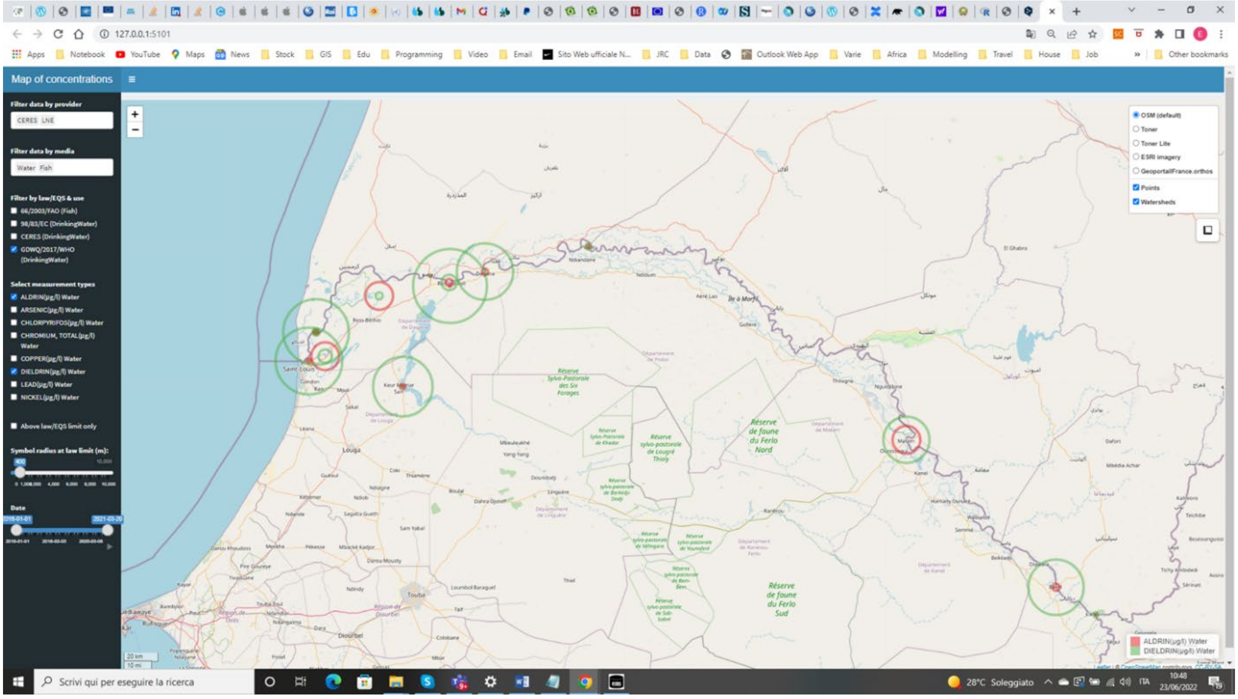
Show 10 entries Search:

Code	Date	Measure
All	All	All
AR	2021-03-06T14:52:00Z	0.015
BA	2021-03-07T07:54:00Z	0.16
BD	2021-01-29T23:00:00Z	0.015
BD	2021-02-18T23:00:00Z	0.015
BD	2021-02-26T23:00:00Z	0.1
BO	2021-03-11T13:59:00Z	0.6
DA	2021-03-09T12:42:00Z	0.13
DI	2021-03-13T11:06:00Z	0.08
KI	2021-03-06T09:55:00Z	0.1
KM	2021-01-29T23:00:00Z	0.015

Showing 1 to 10 of 19 entries Previous 1 2 Next

[12] For example, the ESRI ArcSDE (ESRI, 2004), discontinued since ArcGIS version 10.2, provided an abstraction layer enabling to use different database platforms as the underlying engine. The geographic information was generally stored in ESRI proprietary binary format, which, by the way, was perfectly fine for no spatial database platforms (e.g. MS Access, used to implement the so called personal geodatabase). The use of spatial databases, as PostgreSQL/PostGIS and Oracle, without relying upon the native spatial features, resulted in being 'trapped' within the ArcGIS platform.

Figure 22. Application dashboard for quick spatio-temporal exploratory analysis of multiple parameters vs. law/EQS limits



Map of concentrations

Filter data by provider: CERES LNE

Filter data by media: Water Fish

Filter by law/EQS & use
- ☐ 66/2003/FAO (Fish)
- ☐ 98/83/EC (DrinkingWater)
- ☐ CERES (DrinkingWater)
- ☑ GDWQ/2017/WHO (DrinkingWater)

Select measurement types
- ☑ ALDRIN(µg/l) Water
- ☐ ARSENIC(µg/l) Water
- ☐ CHLORPYRIFOS(µg/l) Water
- ☐ CHROMIUM, TOTAL(µg/l) Water
- ☐ COPPER(µg/l) Water
- ☑ DIELDRIN(µg/l) Water
- ☐ LEAD(µg/l) Water
- ☐ NICKEL(µg/l) Water

☐ Above law/EQS limit only

Symbol radius at law limit (m): 400

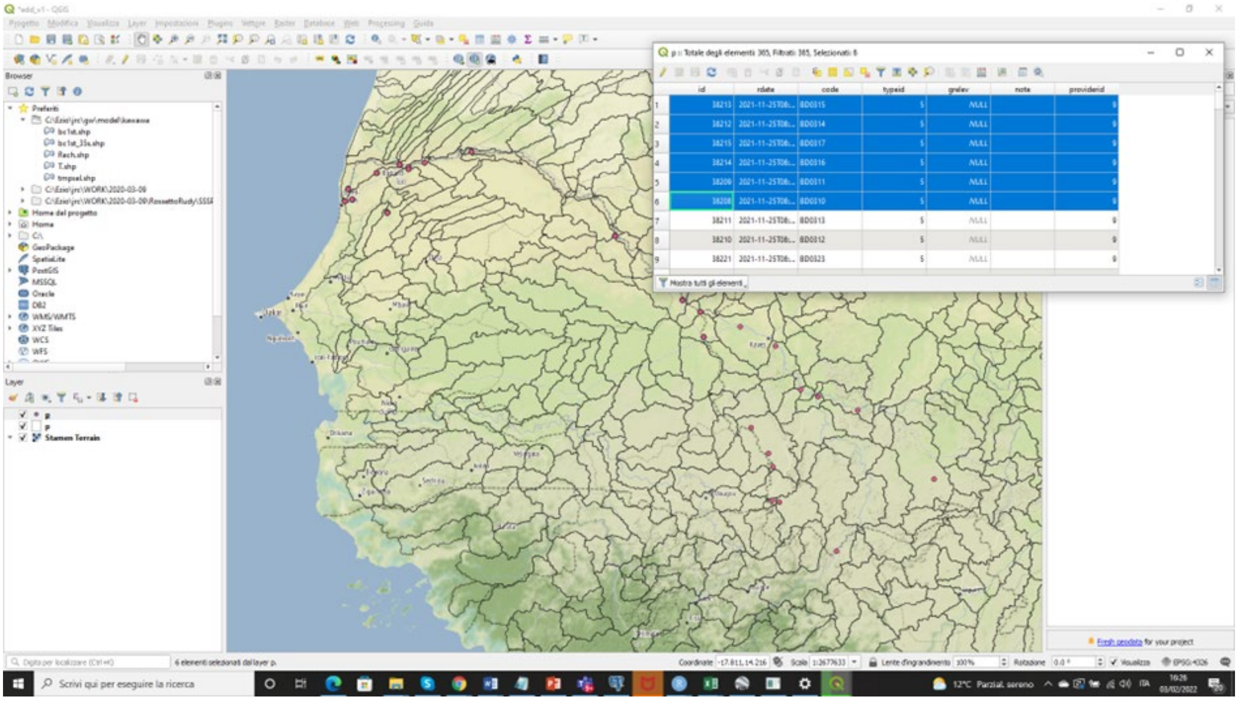
Date: 2018-01-01 – 2021-03-20

- ● OSM (default)
- ○ Toner
- ○ Toner Lite
- ○ ESRI imagery
- ○ GeoportailFrance.orthos
- ☑ Points
- ☑ Watersheds

ALDRIN(µg/l) Water
DIELDRIN(µg/l) Water

The use of the QGIS platform perfectly illustrates the flexibility of the proposed solution. Connecting to the database is as easy as providing the credentials information in the EMS application. Once connected, the geographic data can be visualised, queried and analysed, fully profiting of the available advanced geospatial toolkit (e.g. water quality monitoring points location, with a watershed delineation based on SRTM90 dataset; **Figure 23**).

Figure 23. Database spatial data visualisation and querying from within QGIS



p :: Totale degli elementi 365, Filtrati: 365, Selezionati: 6

	id	rdate	code	typeid	grelev	note	providerid
1	38213	2021-11-25T08:...	BD0315	5	NULL		9
2	38212	2021-11-25T08:...	BD0314	5	NULL		9
3	38215	2021-11-25T08:...	BD0317	5	NULL		9
4	38214	2021-11-25T08:...	BD0316	5	NULL		9
5	38209	2021-11-25T08:...	BD0311	5	NULL		9
6	38208	2021-11-25T08:...	BD0310	5	NULL		9
7	38211	2021-11-25T08:...	BD0313	5	NULL		9
8	38210	2021-11-25T08:...	BD0312	5	NULL		9
9	38221	2021-11-25T08:...	BD0323	5	NULL		9

Mostra tutti gli elementi

6 elementi selezionati dal layer p.

Similarly, the database can be easily accessed using most of programming languages. This has been the case for the EMS tool development, based on R and Python. Both languages provide dedicated libraries, through which the user can connect to the database, easily run SQL queries against the database server, read the data to internal language constructs (e.g. R dataframes, Python (Geo)Pandas), update or delete existing records, as well as creating new records([13]). Once the data are read into memory, the languages can be fully exploited, as is the case for advanced spatio-temporal visualisation and statistical analysis.

4.5 PgAdmin

The application integrates pgAdmin ver. 4.5, a popular web (since version 4) tool providing an intuitive and flexible GUI (**Figure 24**) to administer, develop and query PostgreSQL/PostGIS databases (PgAdmin, 2022).

The pgAdmin interface is organised around two distinct sections:

- On the left, a navigation tree that enables the user to create and access server(s), database(s)([14]) and, for each database, the entire hierarchy of objects, namely schemas, tables, views and functions, just to mention a few ones. Not less important, pgAdmin enables the creation of new users and grant them specific rights (as reading only), depending upon their role and expected tasks.
- On the right, a large window dedicated to the tool output, as database internal working statistics, user queries and related output. For example, the query shown in **Figure 24** returns the time series data with full details on monitoring object properties (e.g. code, geometry), parameter name, measurement unit and media, as organised together from different tables of the underlying database.

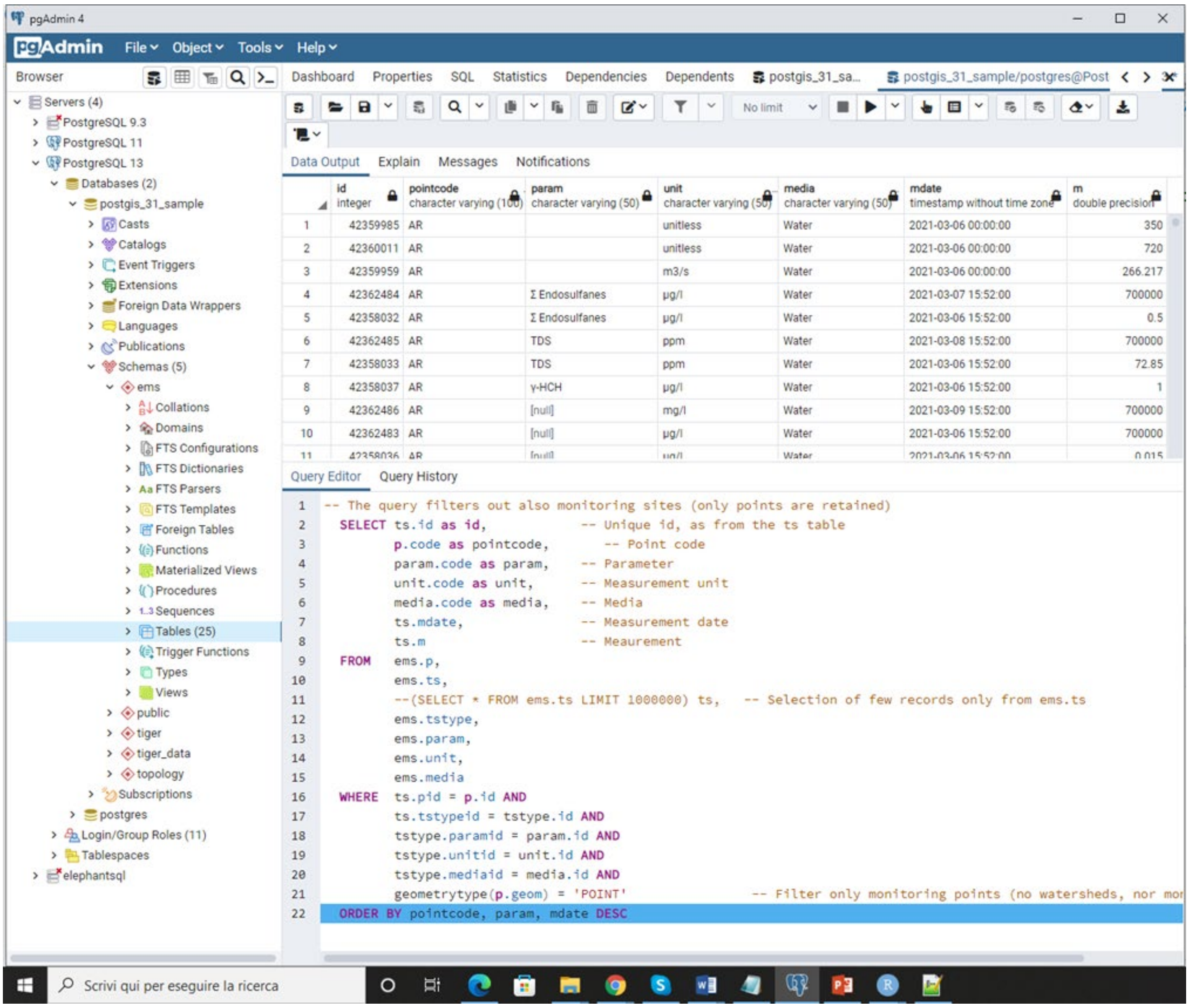
The user is recommended to refer to the many available learning resources for any further details concerning both the tool architecture and functionalities.

[13] RPostgreSQL, one of the R libraries for PostgreSQL connection, is as easy to use as in the example below:

```
drv <- dbDriver("PostgreSQL")       # Load the PostgreSQL driver

con <- dbConnect(drv, dbname = "postgis_25_sample", host = "localhost", port = 5432, user
= "postgres", password = pw)     # Connect to the database

on.exit(dbDisconnect(con))

sql <- " SELECT * FROM ems.p WHERE geometrytype(geom) = 'POINT'” # Define the query

p <- dbGetQuery(con, sql)     # Run the query
```

[14] PgAdmin can support the administration of distinct servers, both locally installed (including different versions of the PostgreSQL engine) and remote ones, as well as different databases. The postgis_31_sample, shown in the tool GUI snapshot, is automatically created on installation of the PostGIS, the PostgreSQL spatial extension.

Figure 24. pgAdmin 4.5: key elements in the interface and database querying



	id integer	pointcode character varying (100)	param character varying (50)	unit character varying (50)	media character varying (50)	mdate timestamp without time zone	m double precision
1	42359985	AR		unitless	Water	2021-03-06 00:00:00	350
2	42360011	AR		unitless	Water	2021-03-06 00:00:00	720
3	42359959	AR		m3/s	Water	2021-03-06 00:00:00	266.217
4	42362484	AR	Σ Endosulfanes	µg/l	Water	2021-03-07 15:52:00	700000
5	42358032	AR	Σ Endosulfanes	µg/l	Water	2021-03-06 15:52:00	0.5
6	42362485	AR	TDS	ppm	Water	2021-03-08 15:52:00	700000
7	42358033	AR	TDS	ppm	Water	2021-03-06 15:52:00	72.85
8	42358037	AR	γ-HCH	µg/l	Water	2021-03-06 15:52:00	1
9	42362486	AR	[null]	mg/l	Water	2021-03-09 15:52:00	700000
10	42362483	AR	[null]	µg/l	Water	2021-03-06 15:52:00	700000
11	42358036	AR	[null]	µg/l	Water	2021-03-06 15:52:00	0.015

```sql
-- The query filters out also monitoring sites (only points are retained)
  SELECT ts.id as id,              -- Unique id, as from the ts table
         p.code as pointcode,        -- Point code
         param.code as param,      -- Parameter
         unit.code as unit,        -- Measurement unit
         media.code as media,      -- Media
         ts.mdate,                 -- Measurement date
         ts.m                      -- Meaurement
  FROM   ems.p,
         ems.ts,
         --(SELECT * FROM ems.ts LIMIT 1000000) ts,   -- Selection of few records only from ems.ts
         ems.tstype,
         ems.param,
         ems.unit,
         ems.media
  WHERE  ts.pid = p.id AND
         ts.tstypeid = tstype.id AND
         tstype.paramid = param.id AND
         tstype.unitid = unit.id AND
         tstype.mediaid = media.id AND
         geometrytype(p.geom) = 'POINT'                -- Filter only monitoring points (no watersheds, nor mor
  ORDER BY pointcode, param, mdate DESC
```

4.6 Data files preparation

Data collected in the framework of ground surveys or repeated monitoring campaigns should ideally be organised in MS Excel spreadsheets, following the rules stated in the locational and time series MS Excel template files (**Annex 3**) From the management point of view, the use of the templates should be agreed upon (if and where possible) and, just in case, expressly required in contracts, in order to promote good and consistent practices, reduce the risks of duplications or the proliferation of ad-hoc and different formats.

While the use of the templates is highly recommended, common practice suggests a certain degree of inertia, most laboratories tending to give preference to their own internal standard procedures and file formats. In these cases, an extra programming effort is required in order to convert the data files to the standard template formats, ready for validation and database uploading in EMS(15).

Quite often, data are organised in wide format, general information about monitoring points (e.g. code, coordinates) being complemented by quality measures organised by parameter in distinct columns. On the

(15) Specialised libraries for programming languages as R and Python are available to support data cleaning and tidying, transposing columns from wide to long formats being a relevant example. Depending upon the quality of the original laboratory data files, the design and implementation of such processing can turn to be quite time consuming and still highly relevant to detect any inconsistency. Experience suggests to keep the original data files and implement the processing tasks so as to be able to refine and re-run the process at any later stage.

other hand, EMS templates rely upon a long format, where each row in a data file uniquely refers to a single measure. The conceptual schema in **Figure 47** should clarify the distinct features and relationships of the two formats.

The conversion from wide to long formats becomes hence a fundamental and quite frequent task to do. Hence a Python source code is provided for reference (**Annex 2**), easy to adapt to specific needs and formats.

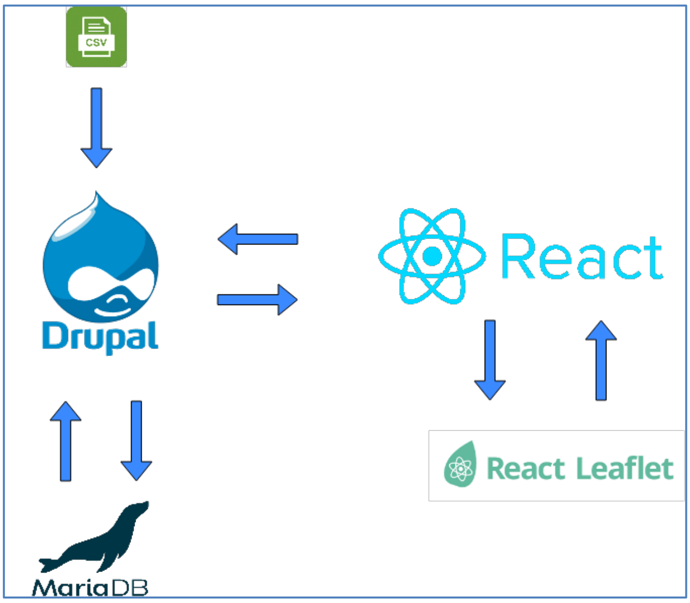
5 Aquaknow: water quality web mapping

Further to the EMS tool detailed above, a web mapping application, hereafter referred to as AquaknowWQM (Aquaknow Water Quality Mapping), has also been developed within the JRC water KMS Aquaknow. The main objective of the tool is to provide a complementary strategy for the dissemination of the environmental and water quality data analysed in the framework of the transboundary Senegal River Basin.

5.1 System architecture

The application, based on React JS/React Leaflet and integrated with Drupal 7 and MariaDB (**Figure 25**), extends key Aquaknow features (as documents sharing and working groups management) with basic mapping, time series visualisation and exploratory analysis capabilities.

Figure 25. AquaknowWQM: system architecture



Above mentioned components are briefly summarised here below:

- 'React' is a JavaScript-based front-end UI development library, first appeared in May 2013 and today one of the most used libraries in the open-source developers community. 'React Leaflet' provides bindings between React and Leaflet. It does not replace Leaflet, one of the most used web mapping libraries, but leverages it to abstract Leaflet layers as React components[16].

[16] React Leaflet can behave differently from how other React components work, notably:

React does not render Leaflet layers to the DOM, this rendering is done by Leaflet itself. React only renders a <div> element when rendering the Map Container component, the contents of UI layers components.

The properties passed to the components are used to create the relevant Leaflet instance when the component is rendered the first time and should be treated as immutable by default. During the first render, all these properties should be supported as they are by Leaflet, however they will not be updated in the UI when they change unless they are explicitly documented as being mutable. Mutable properties changes are compared by reference (unless stated otherwise) and are applied calling the relevant method on the Leaflet element instance.

React Leaflet uses React's context API to make some Leaflet elements instances available to children elements that need it. Each Leaflet map instance has its own React context, created by the Map Container component. Other components and hooks provided by React Leaflet can only be used as descendants of a Map Container.

- 'Drupal' is a Free and Open Source[17] web Content Management System (CMS) written in PHP and distributed under the GNU license[18]. The standard release of Drupal, known as 'Drupal core', contains basic features common to most content-management systems. These include user account registration and maintenance, menu management, RSSfeeds, taxonomy, page layout customisation, and system administration. The Drupal core installation can serve as a simple website, a single- or multi-user blog, an Internet forum, or a community website providing for user-generated content. Drupal also describes itself as a Web application framework[19]. When compared with notable frameworks, Drupal meets most of the generally accepted feature requirements for such web frameworks. Although Drupal offers a sophisticated API[20] for developers, basic Web-site installation and administration of the framework require no programming skills.
- Drupal runs on any computing platform that supports both a web server[21] capable of running PHP and a database to store content and configuration.

'MariaDB' is a community-developed, commercially supported fork[22] of the MySQL relational database management system (RDBMS[23]), intended to remain as a free and open-source software under the GNU license. Operationally, the Aquaknow WQM is fed through an UI with data exported from the EMS database, formatted as CSV files following a standard template. Future development will address the need for enhanced integration through direct and full access to the spatio-temporal EMS database, avoiding data duplication and hence limiting the inherent risks of data integrity failure.

5.2 Operational use

The Aquaknow WQM is characterised by a main web mapping front-end (**Figure 26**), giving access to the different application features. Further to the standard visualisation tools, as full-screen map expansion, base map selection and zoom in/out, the monitoring points are visualised and detailed info is shown on left mouse clicking.

The application provides basic features for exploratory analysis of parameters change in both space and time.

5.2.1 Spatial analysis

The spatial analysis follows the steps below:

- selection of a reference time frame (**Figure 27**), data being dynamically visualised on map depending on sampling date.
- selection of up to three parameters (**Figure 28**) from one media of interest ONLY (Water, Fish, Sediment).
- selection of two up to three maximum monitoring points (**Figure 29**).

[17] Free and open-source software (FOSS) is software that is both free software and open-source software where anyone is freely licensed to use, copy, study, and change the software in any way, and the source code is openly shared so that people are encouraged to voluntarily improve the design of the software.

[18] A GNU license or GNU General Public License (GNU GPL), is a series of widely-used free software licenses that guarantee end users the freedom to run, study, share, and modify the software.

[19] A web framework (WF) or web application framework (WAF) is a software framework that is designed to support the development of web applications including web services, web resources, and web APIs. Web frameworks provide a standard way to build and deploy web applications on the World Wide Web. Web frameworks aim to automate the overhead associated with common activities performed in web development.

[20] An application programming interface (API) is a way for two or more computer programs to communicate with each other. It is a type of software interface, offering a service to other pieces of software

[21] A web server is computer software and underlying hardware that accepts requests via HTTP (the network protocol created to distribute web content) or its secure variant HTTPS. A user agent, commonly a web browser or web crawler, initiates communication by making a request for a web page or other resource using HTTP, and the server responds with the content of that resource or an error message.

[22] In software engineering, a project fork happens when developers take a copy of source code from one software package and start independent development on it, creating a distinct and separate piece of software.

[23] Connolly and Begg define Database Management System (DBMS) as a "software system that enables users to define, create, maintain and control access to the database". An alternative definition for a relational database management system is a database management system (DBMS) based on the relational model. Most databases in widespread use today are based on this model.

— start analysis and build charts on the right, including relevant law limits to highlight concentrations exceedance (**Figure 30**); both data and charts can be exported to standard formats, respectively as PDF/MS Excel and PNG/JPG.

Figure 26. AquaknowWQM: main web mapping interface

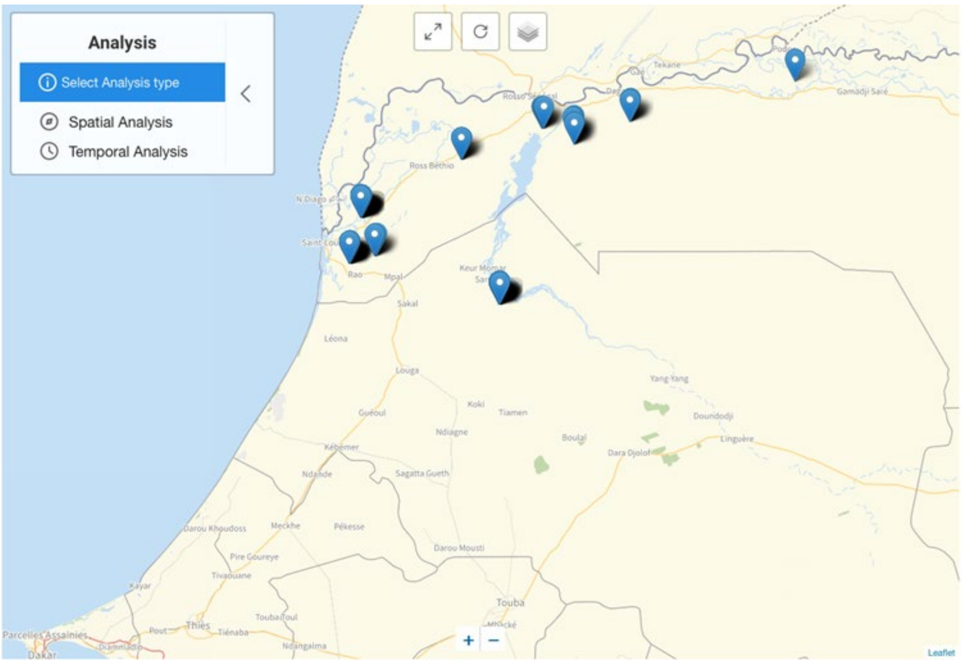


Figure 27. Spatial analysis: Selection of reference time frame

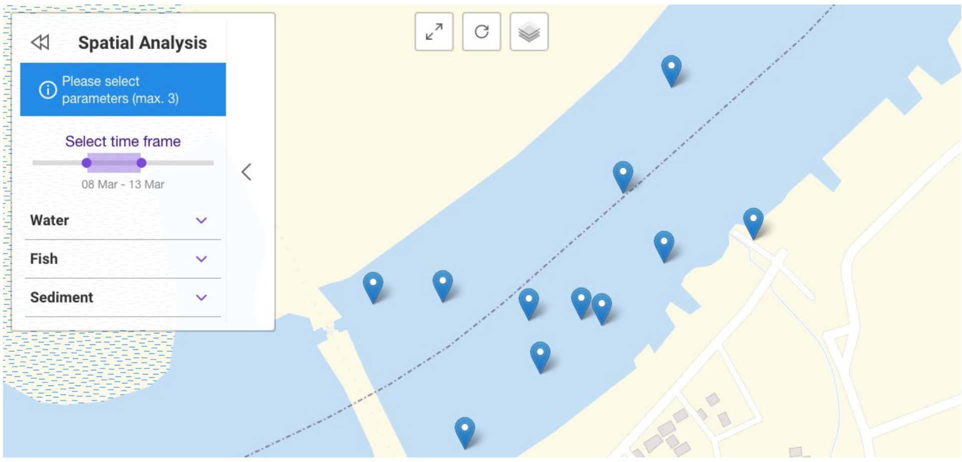
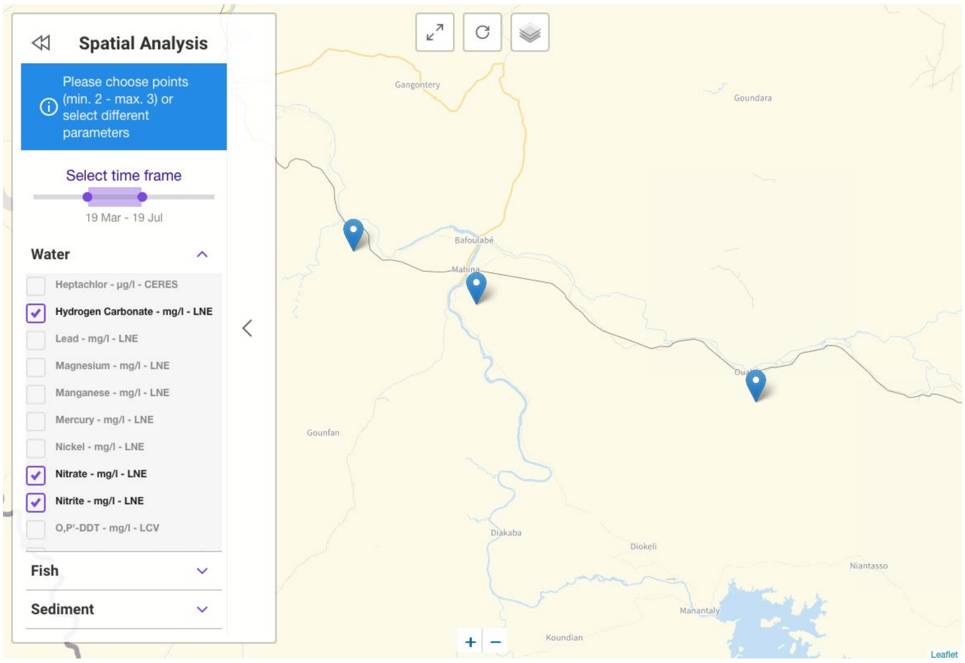


Figure 28. Spatial analysis: Selection of parameters



Spatial Analysis

Please choose points (min. 2 - max. 3) or select different parameters

Select time frame

19 Mar - 19 Jul

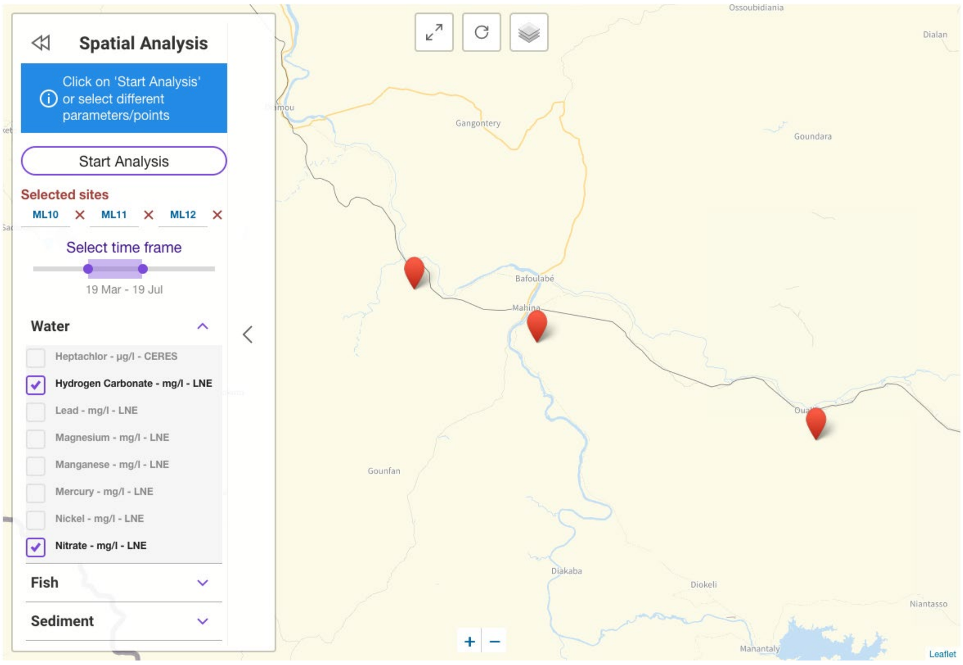
Water

☐ Heptachlor - µg/l - CERES
☑ Hydrogen Carbonate - mg/l - LNE
☐ Lead - mg/l - LNE
☐ Magnesium - mg/l - LNE
☐ Manganese - mg/l - LNE
☐ Mercury - mg/l - LNE
☐ Nickel - mg/l - LNE
☑ Nitrate - mg/l - LNE
☑ Nitrite - mg/l - LNE
☐ O,P'-DDT - mg/l - LCV

Fish

Sediment

Figure 29. Spatial analysis: monitoring points selection



Spatial Analysis

Click on 'Start Analysis' or select different parameters/points

Start Analysis

Selected sites

ML10 ✕ ML11 ✕ ML12 ✕

Select time frame

19 Mar - 19 Jul

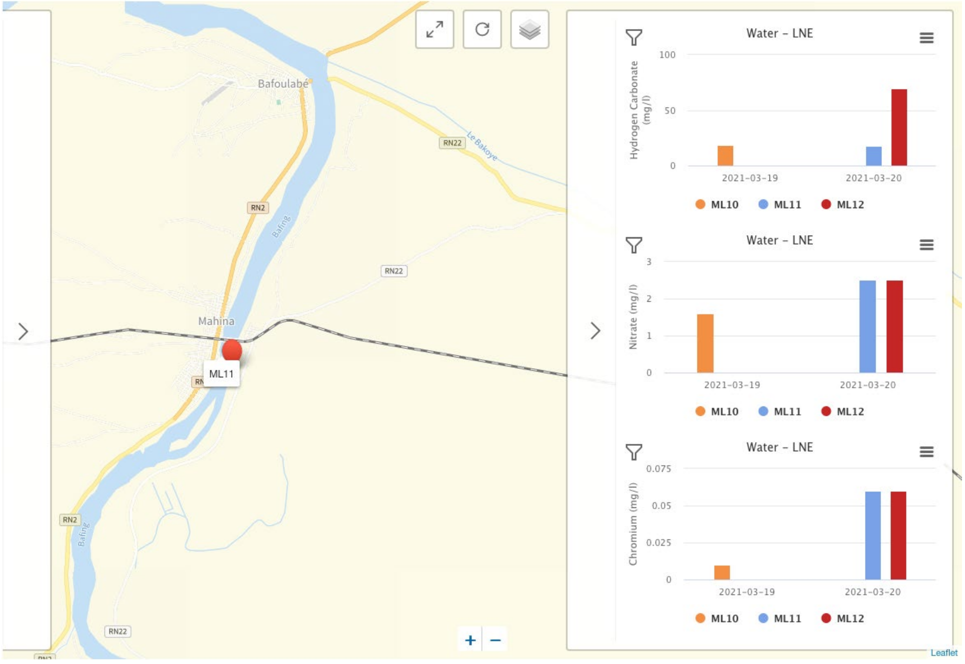
Water

☐ Heptachlor - µg/l - CERES
☑ Hydrogen Carbonate - mg/l - LNE
☐ Lead - mg/l - LNE
☐ Magnesium - mg/l - LNE
☐ Manganese - mg/l - LNE
☐ Mercury - mg/l - LNE
☐ Nickel - mg/l - LNE
☑ Nitrate - mg/l - LNE

Fish

Sediment

Figure 30. Spatial analysis: charts visualisation and download



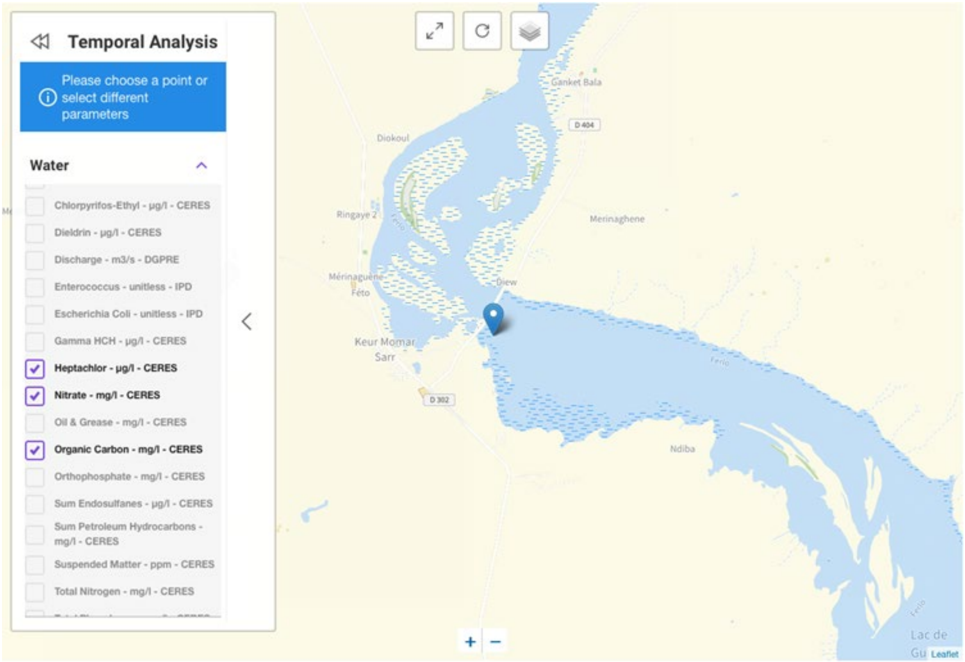
5.2.2 Temporal analysis

Temporal analysis begins with the selection of the parameters. A quick filter on the map is applied immediately, thus excluding points for which no time series (TS) are available.

Temporal analysis, aimed at investigating concentration trends, follows analogous steps to the ones specified for spatial analysis:

- selection of the parameters of interest, a seamless filter being applied to retain time series only, i.e., datasets with at least two distinct measures in time (**Figure 31**).
- selection of the monitoring point(s) (**Figure 32**).
- start analysis and build charts on the right, including relevant law limits to highlight concentrations exceedance (**Figure 33**); both data and charts can be exported to standard formats, respectively as PDF/MS Excel and PNG/JPG.

Figure 31. Temporal analysis: parameters selection



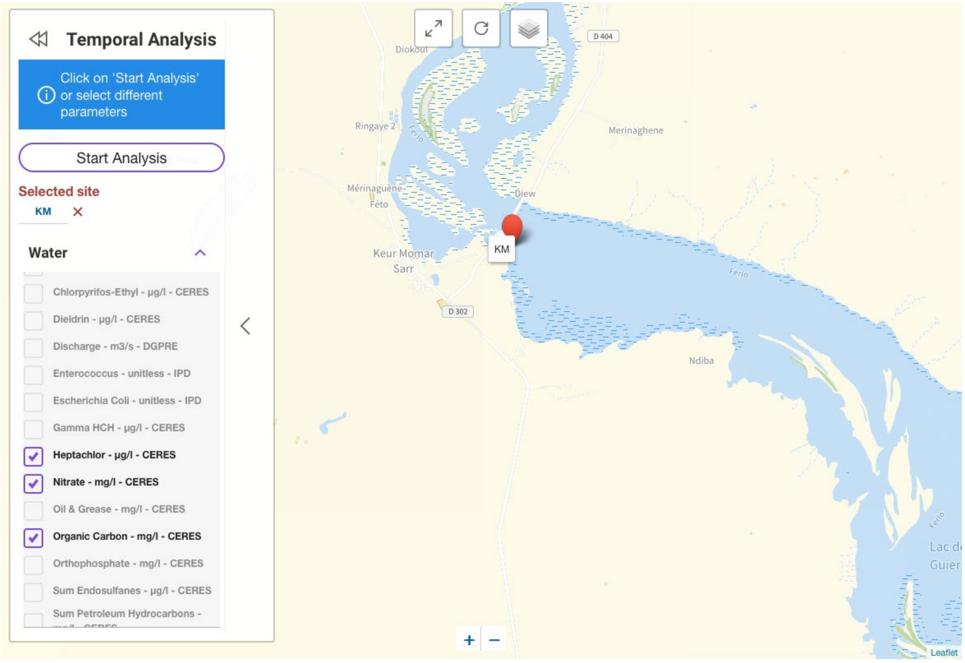
Temporal Analysis

Please choose a point or select different parameters

Water

- □ Chlorpyrifos-Ethyl - µg/l - CERES
- □ Dieldrin - µg/l - CERES
- □ Discharge - m3/s - DGPRE
- □ Enterococcus - unitless - IPD
- □ Escherichia Coli - unitless - IPD
- □ Gamma HCH - µg/l - CERES
- ☑ Heptachlor - µg/l - CERES
- ☑ Nitrate - mg/l - CERES
- □ Oil & Grease - mg/l - CERES
- ☑ Organic Carbon - mg/l - CERES
- □ Orthophosphate - mg/l - CERES
- □ Sum Endosulfanes - µg/l - CERES
- □ Sum Petroleum Hydrocarbons - mg/l - CERES
- □ Suspended Matter - ppm - CERES
- □ Total Nitrogen - mg/l - CERES

Figure 32. Temporal analysis: monitoring point(s) selection



Temporal Analysis

Click on 'Start Analysis' or select different parameters

Start Analysis

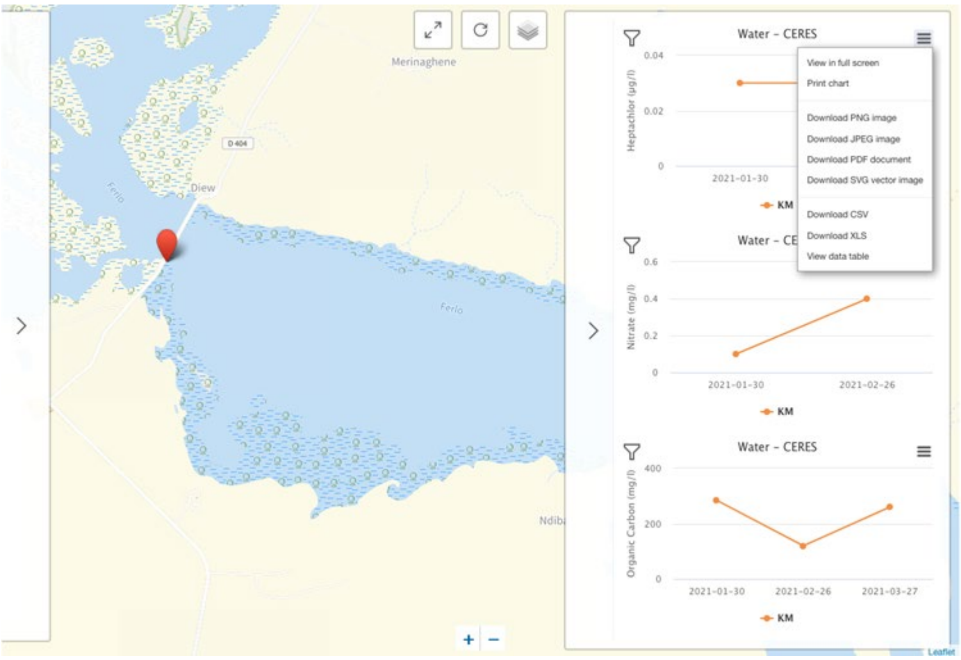
Selected site

KM ✕

Water

- □ Chlorpyrifos-Ethyl - µg/l - CERES
- □ Dieldrin - µg/l - CERES
- □ Discharge - m3/s - DGPRE
- □ Enterococcus - unitless - IPD
- □ Escherichia Coli - unitless - IPD
- □ Gamma HCH - µg/l - CERES
- ☑ Heptachlor - µg/l - CERES
- ☑ Nitrate - mg/l - CERES
- □ Oil & Grease - mg/l - CERES
- ☑ Organic Carbon - mg/l - CERES
- □ Orthophosphate - mg/l - CERES
- □ Sum Endosulfanes - µg/l - CERES
- □ Sum Petroleum Hydrocarbons -

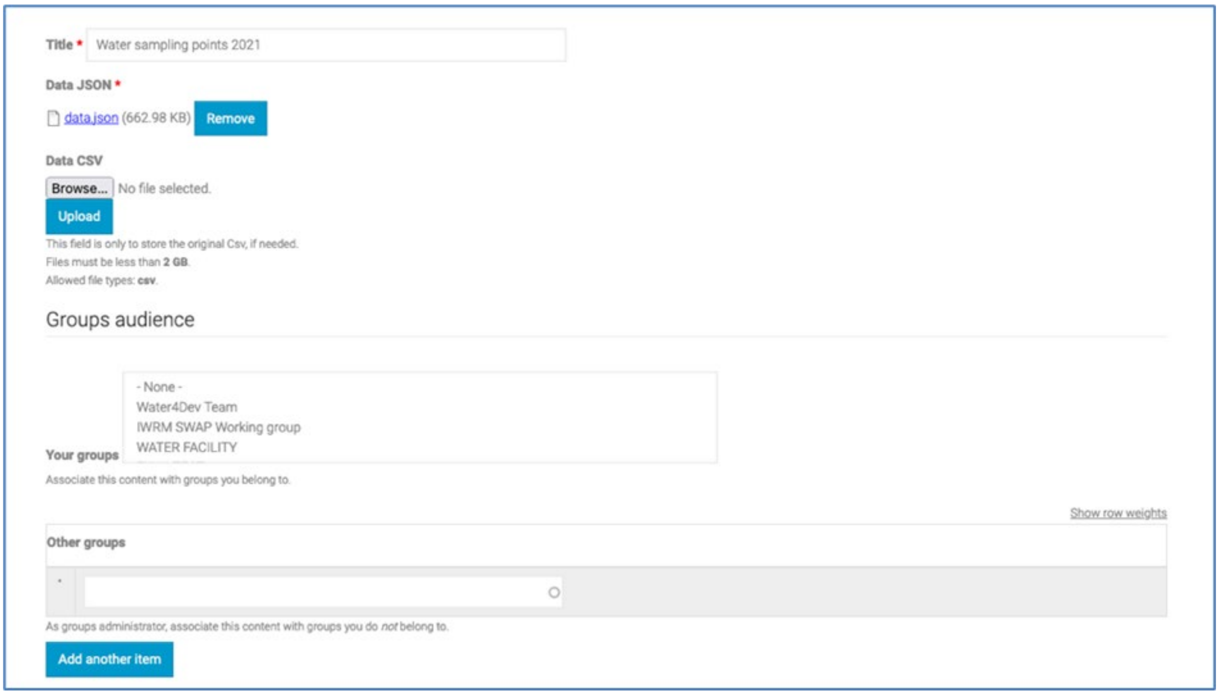
Figure 33. Temporal analysis: charts visualisation and download



5.3 Administrator interface

A dedicated GUI provides support to system administrators for uploading new or editing existing datasets. A new Aquaknow object named 'water station' was created to host information about monitoring points (e.g. code, coordinates). As for other Aquaknow contents (news, documents, events, etc.), the 'water station' objects can be created through a dedicated form (**Figure 34**) and published, both publicly and within private working groups.

Figure 34. Water station entry form



Title * Water sampling points 2021

Data JSON *

data.json (662.98 KB) Remove

Data CSV

Browse... No file selected.

Upload

This field is only to store the original Csv, if needed.
Files must be less than **2 GB**.
Allowed file types: **csv**.

Groups audience

Your groups
- None -
Water4Dev Team
IWRM SWAP Working group
WATER FACILITY

Associate this content with groups you belong to.

Show row weights

Other groups

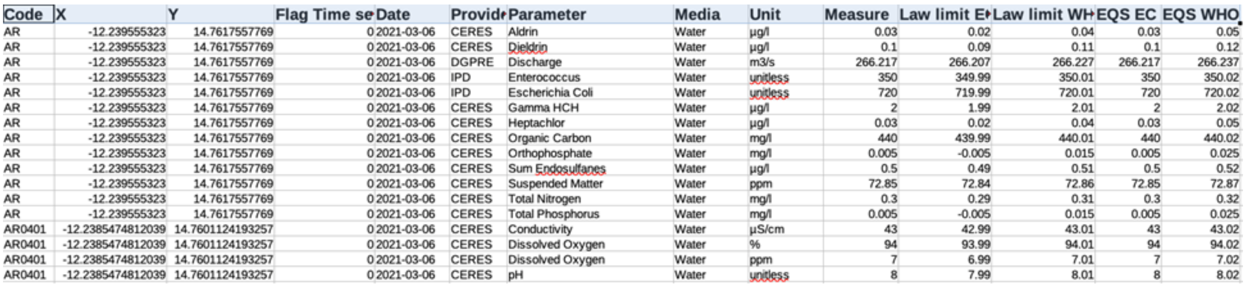
○

As groups administrator, associate this content with groups you do *not* belong to.

Add another item

Datasets are often provided in (or easily converted to) CSV format, hence a standard template has been designed (**Figure 35**) to support data uploading to the Water Point Tool. As a further data conversion to JSON format is needed, also a dedicated Jupyter notebook has been created on a local Docker environment, unfortunately not accessible except for Aquaknow administrators. Future development plans will address this issue, making the tool available to a larger audience.

Figure 35. CSV standard template for data collection



Code	X	Y	Flag Time se	Date	Provid	Parameter	Media	Unit	Measure	Law limit E	Law limit WH	EQS EC	EQS WHO
AR	-12.239555323	14.7617557769	0	2021-03-06	CERES	Aldrin	Water	µg/l	0.03	0.02	0.04	0.03	0.05
AR	-12.239555323	14.7617557769	0	2021-03-06	CERES	Dieldrin	Water	µg/l	0.1	0.09	0.11	0.1	0.12
AR	-12.239555323	14.7617557769	0	2021-03-06	DGPRE	Discharge	Water	m3/s	266.217	266.207	266.227	266.217	266.237
AR	-12.239555323	14.7617557769	0	2021-03-06	IPD	Enterococcus	Water	unitless	350	349.99	350.01	350	350.02
AR	-12.239555323	14.7617557769	0	2021-03-06	IPD	Escherichia Coli	Water	unitless	720	719.99	720.01	720	720.02
AR	-12.239555323	14.7617557769	0	2021-03-06	CERES	Gamma HCH	Water	µg/l	2	1.99	2.01	2	2.02
AR	-12.239555323	14.7617557769	0	2021-03-06	CERES	Heptachlor	Water	µg/l	0.03	0.02	0.04	0.03	0.05
AR	-12.239555323	14.7617557769	0	2021-03-06	CERES	Organic Carbon	Water	mg/l	440	439.99	440.01	440	440.02
AR	-12.239555323	14.7617557769	0	2021-03-06	CERES	Orthophosphate	Water	mg/l	0.005	-0.005	0.015	0.005	0.025
AR	-12.239555323	14.7617557769	0	2021-03-06	CERES	Sum Endosulfanes	Water	µg/l	0.5	0.49	0.51	0.5	0.52
AR	-12.239555323	14.7617557769	0	2021-03-06	CERES	Suspended Matter	Water	ppm	72.85	72.84	72.86	72.85	72.87
AR	-12.239555323	14.7617557769	0	2021-03-06	CERES	Total Nitrogen	Water	mg/l	0.3	0.29	0.31	0.3	0.32
AR	-12.239555323	14.7617557769	0	2021-03-06	CERES	Total Phosphorus	Water	mg/l	0.005	-0.005	0.015	0.005	0.025
AR0401	-12.2385474812039	14.7601124193257	0	2021-03-06	CERES	Conductivity	Water	µS/cm	43	42.99	43.01	43	43.02
AR0401	-12.2385474812039	14.7601124193257	0	2021-03-06	CERES	Dissolved Oxygen	Water	%	94	93.99	94.01	94	94.02
AR0401	-12.2385474812039	14.7601124193257	0	2021-03-06	CERES	Dissolved Oxygen	Water	ppm	7	6.99	7.01	7	7.02
AR0401	-12.2385474812039	14.7601124193257	0	2021-03-06	CERES	pH	Water	unitless	8	7.99	8.01	8	8.02

6 Case study: Senegal environmental quality monitoring network

The pilot project on water quality in the Senegal River Basin (micropollutants and bacteriology) is embedded in the "WEFE SENEGAL PROJECT - Support to the management of water resources and the water-energy-agriculture nexus" and aims at creating a starting position for a future surface water quality monitoring network in the Senegal River Basin. The specific objectives of the project are:

- Identification and prioritisation of chemical and biological pollutants in the Senegal River Basin. It incorporates time series of these data collected in the Delta area during the year 2021, including drinking water aspects in the drinking water reservoirs of St Louis and Dakar. The data from the river in the Delta shall serve as a proxy for the temporal trends of water quality in the whole catchment. In addition, two seasonal spatial pollution monitoring campaigns in different media (water, sediment and fish) are conducted throughout the basin, both during the rainy and dry seasons to investigate spatial trends and origin of pollutants.
- Integration and training of regional experts/institutions in a cooperative and participatory context. Trainings by VUA and the JRC have been intended to support (i) proper laboratory analysis of participating laboratories in Senegal and Mali and (ii) use of the developed EMS tool for data management and exploratory spatio-temporal data-analysis. Due to the COVID-19 health crisis quasi all trainings have been conducted virtually with partner institutions.

6.1 Data sampling and analysis framework

Water quality has been analysed by six laboratories in Senegal (labs DGPRE, CERES, and IPD) and Mali (labs DNH, LCV, LNE) (**Table 1**) and one reference laboratory in the Netherlands (VUA laboratory facilities). 27 monitoring sites have been sampled along the Sénégal, Bafing, Bakoye and Falémé rivers (**Figure 36**). A monitoring site is defined as the outer extent (or its centroid) formed by multiple monitoring points belonging to the same site. Locational information of monitoring points was reported in geographic (WGS84, SRID 4326) or projected (WGS 84 / UTM zone 28N, SRID 32628) coordinate systems. For each monitoring point, basic physico-chemical parameters (e.g. temperature, dissolved oxygen, conductivity, turbidity, pH) were analysed on-site and sampling date and analysis time were indicated. Although these are only one-time measurements, analysis results can support the interpretation of time series and spatio-temporal data.

All river samples (water, sediment) are combined, each consisting of three subsamples, taken from both banks and from the middle of the river, and then combined into one site sample for laboratory analysis. A subset of sediment samples for analysis of heavy metals and pesticides have been taken in parallel. These parallel samples were sent to VUA for reference analysis. Information on transport conditions from the place of sampling to laboratory facilities was provided only in some cases. Fish samples are combined from at least 3 individuals. The fishes were purchased from local fishermen, reporting species and size of the fish and date and time of purchase. Sediment samples were taken from the surface (0-10 cm).

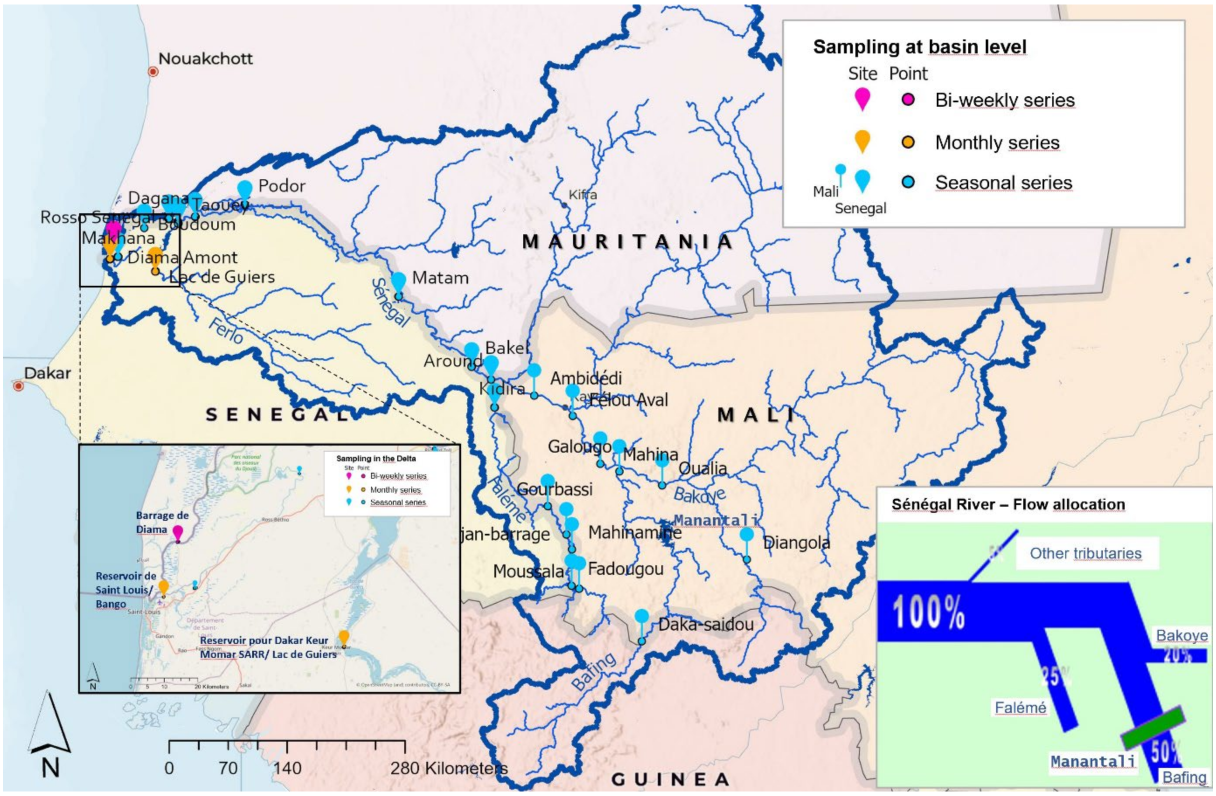
6.1.1 Time series campaign

The time series survey was conducted from January to December 2021 in the Senegal River Delta (Barrage de Diama - BD), including the drinking water reservoirs for St Louis (RS - Bango Reserve) and Dakar (KM - Lac de Guiers at Keur Momar SARR). Water from BD was monitored bi-weekly while the two drinking water reservoirs RS and KM were sampled on a monthly basis. Water was analysed for pesticides (e.g. Aldrin, Bifenthrin, Chlorpyrifos, Dieldrin, Endosulfan, Gamma BHC (Lindane), Heptachlor), nutrients (e.g. Nitrogen (Kjeldahl, total), Nitrogen Nitrate (as N), Phosphorus, Orthophosphate), heavy metals (e.g. Aluminium, Arsenic, Cadmium, Chrome, Copper, Cyanide, Mercury, Nickel, Lead), pathogens (Enterococcus, Escherichia Coli) and physico-chemical parameters (e.g. discharge, total dissolved solids, total organic carbon, total petroleum hydrocarbons, total oil and grease, turbidity). Analyses results for metals were delivered in June 2022.

6.1.2 Spatial/seasonal campaign

The Spatial/seasonal measurements were conducted in the low water (March) and high water (October) season 2021 throughout the entire basin including the Delta, the lower and upper river basin in Senegal and Mali. Here, pesticides and heavy metals were analysed in the three media water, river sediment and fish. River sediment analysis aims to detect long time changes along the river (memory effect) while analysis in biota (here fish) serves to evaluate the influence of pollutants on the food chain. Nutrients, pathogens and physical/chemical parameters were analysed in water only. Results from the wet season (high water series) campaign 2021 were delivered in June 2022 by the Senegalese laboratories and in October 2022 by the Malian laboratories. Reference laboratory values for heavy metals and organochlorine pesticides in the parallel sediment samples were delivered by VUA in June 2022.

Figure 36. Sampling scheme of the pilot study on water quality assessment in the Senegal River Basin



6.1.3 Involved laboratories

The DGPRE coordinated the field campaigns in Senegal and provided discharge data. CERES provided data on sampling (location, sampling point/site name and coordinates, sampling dates), inorganic pollutants (pesticides and nutrients) and physico-chemical parameters. IPD was in charge of microbiological analysis on pathogens. Analysis refers mainly to water, relevant results for sediment and fish were delivered in June 2022 and are still under preparation. For Senegal, 632 monitoring points were reported by CERES. These monitoring points belong to 15 monitoring sites (**Figure 36**).

On the Malian side of the basin (upper Senegal Basin), measurement campaigns were carried out by the DNH, LNE and associated laboratories, as well as LCV. DNH was responsible for field campaign coordination and provision of discharge data. The LNE and associated labs provided information on sampling (location, sampling point/site name and coordinates, sampling dates), inorganic pollutants (pesticides and nutrients) and microbiological analyses (pathogens). LCV took care of analysis on organic pollutants. In Mali, water, sediment and fish samples were investigated at twelve sampling sites in March 2021 (dry season), so far for one date and time only (**Table 1**). Data from the wet season has not yet been delivered by the Mali laboratories. Locational information was provided by DNH. Physico-chemical parameters (e.g. gaga height[24]), maximum and mean river depth) in water were analysed by DNH, nutrients (e.g. Sulphate, Phosphate) and pathogens (E.coli, Faecal Coliform, Total Coliform, Faecal Streptococci) by LNE and pesticides by LCV. All analysis results were assigned to a unique monitoring site only. No information on the exact monitoring point was provided. For instance, samples on different fish types refer all to the same monitoring site location.

[24] 'Gaga height' or 'stage' describes the stream water height above a reference point

Table 1. Key information on water quality campaigns carried out in Senegal and Mali

	Senegal	**Mali**
Involved laboratories	DGPRE, CERES(1), IPD, VUA	LNE, DNE, DNH(1), VUA
Sampling period	January – December 2021	January – December 2021
Analysed media	Water, sediment	Water, sediment, fish
Parameter families	Heavy metals, pesticides, pathogens, nutrients, other parameters	Heavy metals, pesticides, pathogens, nutrients, other parameters
Sampling sites / points	15 / 632	12 / 0
Analysed parameters: 113(2)	56	98
Data entries: 6111(2)	2846	3265
Time series	bi-weekly (BD)(3), monthly (RS, KM)(3) seasonal (remaining sites)	Seasonal (all sites)

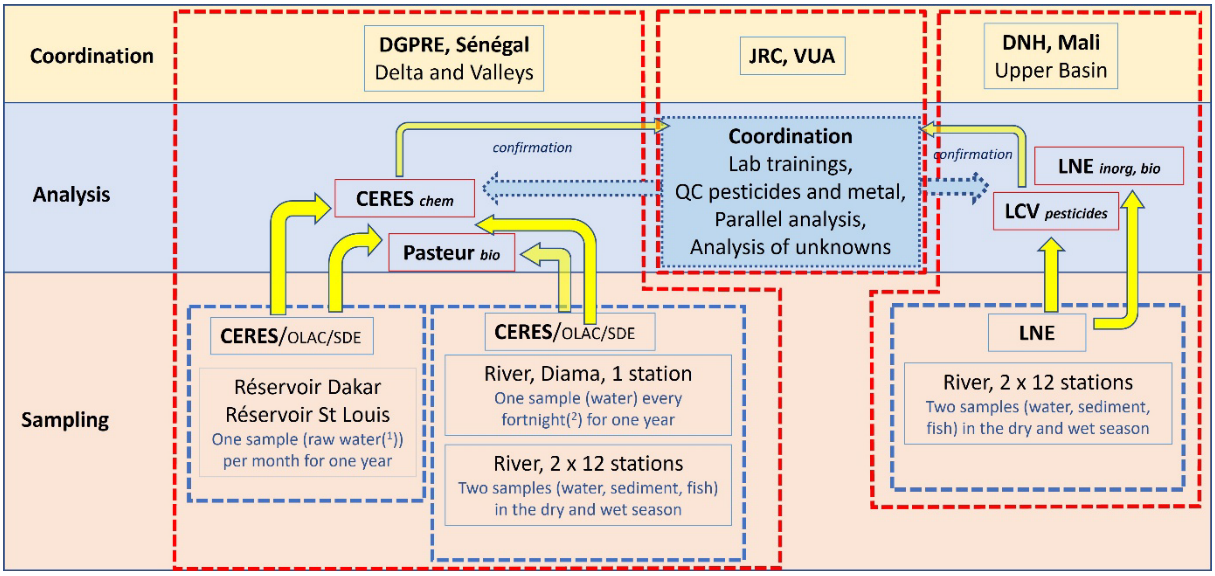
(1) In addition to the parameter data, these laboratories also provided information on the sampling location and date

(2) Only quantitative, consistent and correct data considered; qualitative data is reported under 'Note' in the ts (time series) table of the database

(3) BD: Barrage de Diama, RS: Réserve de Saint-Louis (Réserve de Bango), KM : Keur Momar SARR

Before and during the execution of the field campaigns, VUA external reference laboratory facilities organised online training courses on proper sampling and laboratory analysis of pesticides in water, sediment and biota (fish) to participating laboratories in Senegal and Mali. VUA also conducted a quality control by analysing selected parameters in parallel samples and compared with results from Senegal and Mali laboratories. 20 sediment samples were analysed for organochlorine pesticide (OCP) by VUA and compared with the results obtained in the same samples (10) in Senegal and (10) in Mali. In addition, trace metal concentrations were analysed by Eurofins Omegam, in contract for VUA. Based on the quality control results VUA concluded that "both OCP and trace metal data produced by both laboratories are not reliable". As a consequence, the data from the regional labs were not considered for the assessment of the OCP and heavy metal risk for the aquatic environment of the SRB, since they contain many false positives. Only the OCP data produced by VUA from the parallel samples were considered suitable for this purpose (and hence for the uploading into the DB).

Figure 37. Partner interaction for chemical and microbial sampling and analysis



[1] Analysis of treated water (initially planned, but changed due to COVID-19 crisis).

[2] Not for entire year, only exemplary (initially planned, but changed due to COVID-19 crisis).

6.1.4 Plausibility check

A data plausibility check and interpretation of delivered data is carried out to meet standard data quality targets, as the following:

- Plausibility Check Level 0 (PL0): External Quality Control.
- Plausibility Check Level 1 (PL1): Consistency check between regional laboratories (lab1 vs. lab2).
- Plausibility Check Level 2 (PL2): Consistency check of spatial and temporal data.
- Plausibility Check Level 3 (PL3): Identification of data that deviate significantly from the typical data range.
- Plausibility check Level 4 (PL4): Visualisation and analysis of validated data (via maps, tables, graphs, statistics, etc.)

The objective of **PL0** is a general quality control and the determination of needs for trainings of regional partners (laboratories) to provide support in best practice sampling and analysis. This is ensured by the comparison of results for samples that are analysed in parallel by a reference laboratory. In the frame of this survey, VUA external reference laboratory facilities co-analysed the sediment samples from CERES and LCV for pesticides and heavy metal concentrations. The closer the regional laboratoy values get to the VUA reference values, the better is the analytical performance (**Figure 38**). Data quality is perceived sufficient if analysis results range within a pre-defined lower and upper quality limit around the VUA reference value. Increased analytical performance is usually achieved with more analytical experience and after optimised training sessions. However, in the context of this project the OCP and heavy metal data from the regional labs were not considered eligible.

PL1 targets on the harmonisation of the results by intercomparison exercises between the participating laboratories analysing standardised reference materials. This may include the comparison of results from samples taken near national borders in the same river section but analysed by different laboratories. A deviation in measurement values may indicate systematic inconsistencies between laboratories. The participating labs did not use the reference materials sent.

PL2 serves to clarify whether changes are realistic or related to laboratory procedures (e.g. use of different routines or consumables). This can be indicated by drastic changes in concentrations or outliers in time and space (without an obvious reason), taking also into account the behaviour of other parameters in this context.

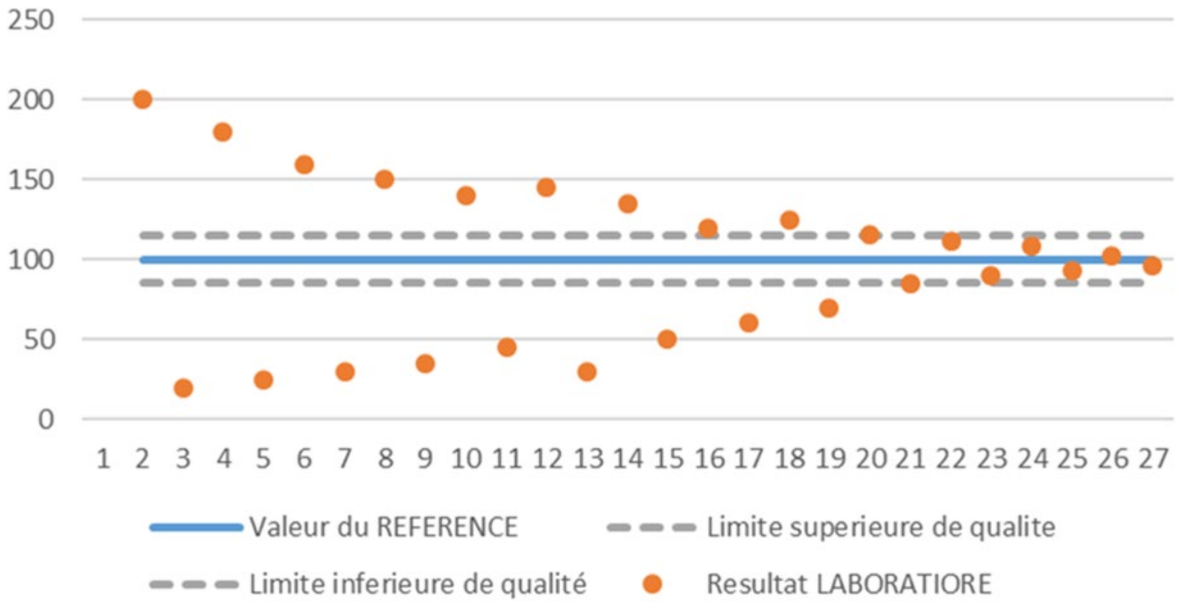
PL3 includes the comparison with existing data in the region. Results that fall outside the typical concentration ranges are discussed and verified with the laboratory. If necessary, sampling and analyses are repeated.

PL4 incorporates a qualitative, quantitative and correlative evaluation of the data. The qualitative assessment is performed through the visualisation of concentrations in relation to ecological quality standards or legal limits, deriving the risk for humans and the environment. In addition, the visualisation of temporal and spatial development of concentrations allows to localise the origin of emissions and understanding of seasonality.

The quantitative assessment (concentration per discharge) describes the balance of total pollutant release. This includes the determination of the quantity of pollutants transported to the sea but also a ranking of the relative importance of the different sections of the river regarding the total pollution of the basin.

The correlative evaluation of all data aims at discovering typical basin consistencies between parameters and the identification of indicators/proxies (parameters) representative of complex mixtures of pollutants. This approach intends to avoid the analysis of the entire set of parameters, which is often too expensive.

Figure 38. Principle of monitoring of the analytical performance by the Senegal and Mali laboratories, compared with reference laboratory values (e.g. VUA)

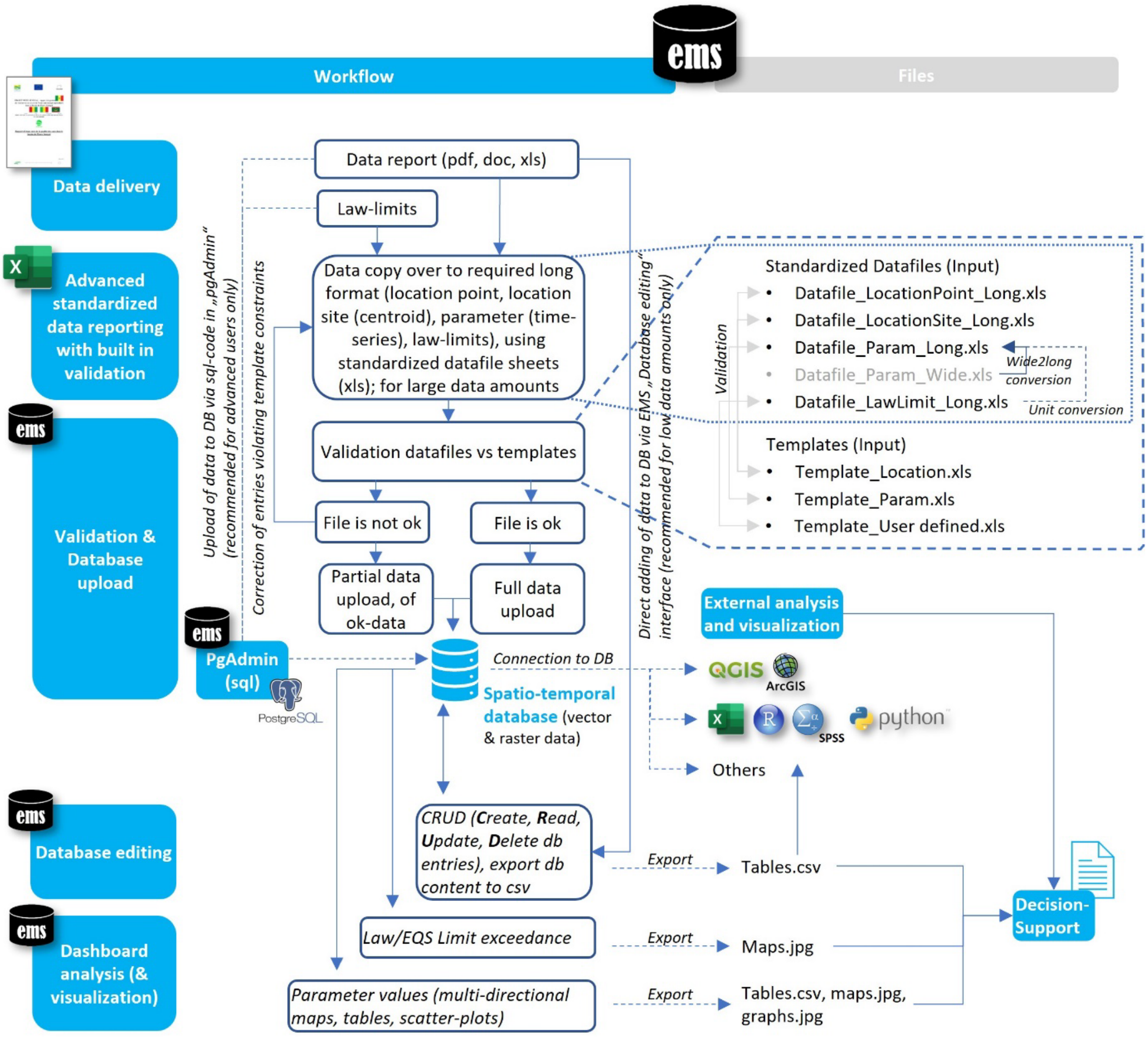


Microbiological data from IPD and sediment data from VUA represent the most reliable data within the survey, while analysis data on heavy metals and pesticides are not so robust in quantity and quality. To ensure high laboratory data quality in the future, the authors recommend setting up a certified African reference laboratory which participates in regular inter-calibration campaigns and uses optimised instruments operated by regularly trained laboratory staff. Regional laboratories could carry out sampling campaigns and send samples to the African reference laboratory for proper analysis. It is key that laboratories share all data in a standardised format using the standardised data reporting sheet in long (or wide) format for both locational and parameter data, made available in this report.

6.2 Validation and visualisation of water quality data using the EMS tool

The stand-alone, multi-lingual EMS tool has been developed in the frame of the water quality survey in the Senegal River Basin to manage, validate, visualise and analyse vector and raster data. The core of EMS is a robust database that can be accessed either via a server or local PC. The latter was explicitly requested by project partners due to the sensitive character of water quality data. Information on sampling location, parameter measurements and law limit/EQS is populated to the database. The workflow of the EMS tool, its modules, input and output files are summarised in **Figure 39**.

Figure 39. Workflow of the management and visualisation of water quality data of the WEFE Senegal survey using the EMS tool and other utilities


ems
Workflow
Files
Data delivery
Advanced standardized data reporting with built in validation
Validation & Database upload
Database editing
Dashboard analysis (& visualization)
Data report (pdf, doc, xls)
Law-limits
Data copy over to required long format (location point, location site (centroid), parameter (time-series), law-limits), using standardized datafile sheets (xls); for large data amounts
Validation datafiles vs templates
File is not ok
File is ok
Partial data upload, of ok-data
Full data upload
Upload of data to DB via sql-code in „pgAdmin" (recommended for advanced users only)
Correction of entries violating template constraints
Direct adding of data to DB via EMS „Database editing" interface (recommended for low data amounts only)
Standardized Datafiles (Input)
Datafile_LocationPoint_Long.xls
Datafile_LocationSite_Long.xls
Datafile_Param_Long.xls
Datafile_Param_Wide.xls
Datafile_LawLimit_Long.xls
Wide2long conversion
Unit conversion
Validation
Templates (Input)
Template_Location.xls
Template_Param.xls
Template_User defined.xls
External analysis and visualization
PgAdmin (sql)
PostgreSQL
Connection to DB
Spatio-temporal database (vector & raster data)
QGIS
ArcGIS
SPSS
python
Others
CRUD (Create, Read, Update, Delete db entries), export db content to csv
Law/EQS Limit exceedance
Parameter values (multi-directional maps, tables, scatter-plots)
Export
Tables.csv
Maps.jpg
Tables.csv, maps.jpg, graphs.jpg
Decision-Support

6.2.1 Data preparation and harmonisation

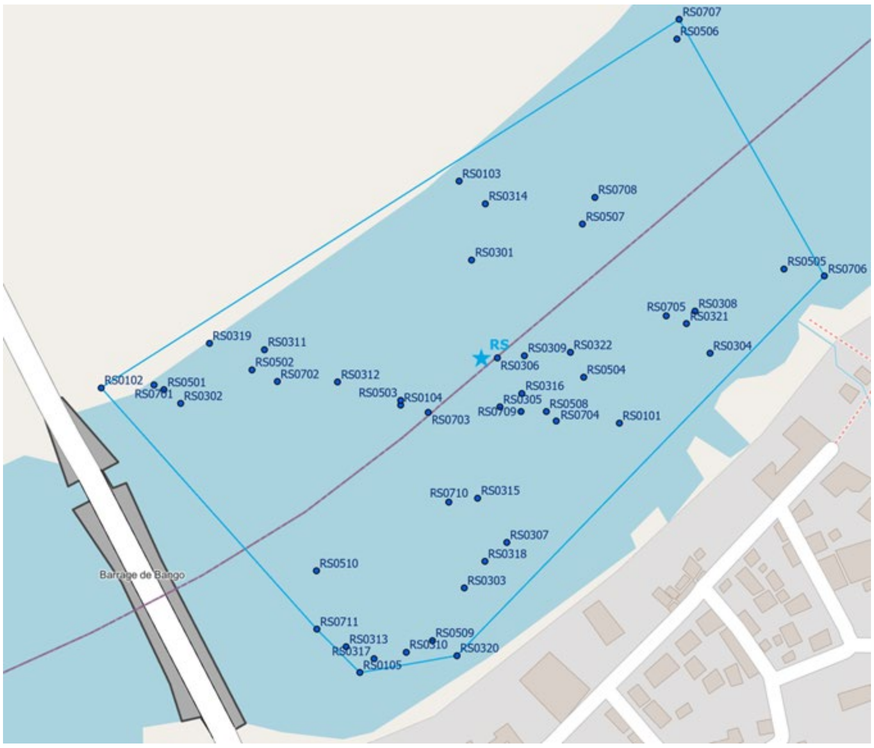
Data provider, here laboratories, delivered sampling (location) and analysis (one-time and time series measurements) data as interim or final summary reports in pdf or MS Word format. To facilitate data copy over, pdf documents were converted to MS Word format using Adobe Acrobat Pro DC (2022) and visually checked for eventual errors which may have occurred during file format conversion. If data tables are provided in non-machine readable format like screenshots, table content needs to be extracted by hand. This procedure is time-consuming and prone to the introduction of additional errors during data compilation process. Laboratories were reminded to provide all data in a computer readable format. All data were transferred to a standardised MS Excel long format file (time series data, locational data, user-defined data) using advanced standardised data reporting sheets with built-in validation, as explained in **Annex 3**. It is strongly recommended that data providers use these pre-formatted data reporting sheets, to avoid additional errors during manual data transfer. These standardised sheets guarantee that data is reported in a proper and unambiguous way, to avoid time consuming and error-prone manual data format checks in multiple non-standardised documents. Moreover, the standardised reporting sheets are already in the correct format for subsequent, automatic data upload to database, facilitating and accelerating the data preparation process remarkably. The standardised data reporting sheets ensure that (i) all sampling locations are unique and correctly assigned, (ii) each measurement can be assigned to a distinct sampling location and a date-time (iii) each contaminant is unique and properly assigned (e.g. ensured by CAS number), (iv) all parameters are provided with a correct measurement unit, (v) all numbers use points as decimal separators. The number of decimals is not restricted but should reflect the proper precision needed. During the data harmonisation process, a continuous dialogue between water quality experts, data editors and laboratory staff is highly recommended to discuss and solve eventual ambiguities together, right when they occur.

6.2.2 Location and time series data provided by the laboratories

Information on sampling locations was compiled from multiple tables in the original document into standardised location datafile in long format (*loc_point_long_LEV1-0.xls). In the case of inconsistent naming of sampling location codes from one report to another, a unique location code is introduced. Information on the original sampling location code as used in the laboratory report is kept in the 'Note' field of the location data file. Senegal and Mali laboratories used different and partly inconsistent naming conventions for sampling points. For Senegal, the name of a monitoring point (e.g. RS0103) contains information on the monitoring site (here RS for Réservoir de Saint Louis / Bango), the sampling mission (RS01) representing the (from-to) date of the mission and the sampling point (RS0103: 3rd sampling point of the mission) (**Figure 40**). The location of the overall monitoring site was not reported by the data provider and, thus, was defined as the centroid of the outer extent of all monitoring points referring to the same site. To ensure that all monitoring points belonging to the same site are captured, a point to point distance threshold of 1000 m was considered, using ArcGIS. For Mali, all measurements were assigned to monitoring sites, named after a nearby place of reference (e.g. Diangola). Information on the exact sampling location (monitoring point) was not shared by data providers from Mali. An artificial location code has been introduced using country code (ML, for Mali), followed by a continuous number for each sampling site (e.g. ML01, site 01 for Diangola). The original location name provided by the laboratory is maintained in the 'Note' field of the location datafile. For the time being, only point locations (monitoring points and centroids of monitoring sites) can be uploaded to the database. The upload of polygonal extents of locations (e.g. monitoring sites) will be enabled soon and stored in WKT format in the database.

Note that due to the sensitive character of water quality data provided by the partner laboratories, the original sampling and analysis data are not shared publicly, neither in the EMS database nor in this report. A dummy file featuring a few fictive locations and associated time series data has been created instead and is described in **Annex 4**.

Figure 40. Concept of monitoring points (dark blue dots) and monitoring site (light blue star as centroid of light blue site polygon), here RS (Réservoir de Saint Louis / Bango)



Information on water quality parameters was copied from original report tables (*report*.doc/pdf) to the standardised time series data file in long format (*ts_long_LEV1-0.xlsx). Major inconsistencies in data reporting experienced in laboratory reports are shown in **Table 2**, contrasted with recommendations for better data documentation. Quantitative information is documented as 'Measure' (number, e.g. 56 or </>56), while qualitative information is kept as text information under 'Note' (e.g. Salmonella: Present, Vibrio Cholerae: Present). To facilitate and accelerate the data documentation of laboratory data, another standardised template in wide format has been created. Here all measurement values for one sampling location are documented in one table row. The conversion from wide to long format is done with the developed python script 'wide2long.py' (see **Annex 2**). A more user-friendly wide2long GUI will be made available soon. To facilitate data comparability and visualisation in the EMS dashboards, measurement units are converted to units commonly used by WHO, EC or US EPA for law limits and EQS (*ts_long_LEV1-1).

Once the data is available in long format, (massive) data can be migrated to the database using the EMS 'Validation and database upload' tool. Here, the datafile is compared against the associated template which defines all constraints and relevant VVL. If the datafile vs. template validation is successful (message: 'File is ok') the data can be uploaded to the database. If any data in the datafile violates the template constraints, detected errors are listed under 'Result. In this case, the user can either (i) do a partial upload of successfully validated data or (ii) correct all erroneous data and retry another full data upload to database afterwards. It must be noted that full data upload is only allowed, if all data in the datafile meets the template constraints. Data not yet included in the VVL (vvl_param.csv, vvl_media.csv, vvl_unit.csv, vvl_provider.csv, vvl_type.csv, vvl_srid.csv, vvl_law.csv, vvl_use.csv) must be added to the associated database table via the 'Database editing' tool. To avoid data duplication, it must be carefully checked via the tool's 'Search' functionality that newly added content is not yet included in the database, e.g. caused by slightly different naming. For instance, in the case of water quality parameters, unicity can be ensured by searching for the CAS number in the 'Parameters' table and considering also synonyms listed in the 'Param.alternate' table. VVL's are automatically updated with newly added or edited database content every time the EMS tool is (re-)loaded. Once all VVL information is updated, new constellations of Measurement Types (parameter-unit-media) can be updated in the TS Type table. For small data amounts, measurement values can manually be added, edited or deleted in the 'Time Series' table by providing mandatory information on Measurement Type (TS Type), Object code (location), Date/Time and Provider. Content of all database tables can be exported in csv format.

Database experts familiar with PostgreSQL can also add data to the database using the pgAdmin open source tool, accessible via the EMS 'Run pgAdmin' feature. However, template constraints need to be met here as well, otherwise data migration will not be permitted. Besides, pgAdmin enables expert users to add new tables to the database, for instance (i) methods applied for field sampling, field analysis and laboratory analysis (e.g. Salmonella (in 2 litres) ISO 19250), (ii) contaminant families (e.g. pesticides, heavy metals, pathogens, nutrients, other parameters or (iii) measurement frequency (hourly, daily, (bi-)weekly, monthly, seasonal (dry, wet), yearly).

All modifications made to the original data (*report_LEV0-0.doc/pdf) by data editor or user are reported as different data levels, explained in **Table 3**. Data levels help to track the number and origin of changes during the entire data editing process and to assess completeness and quality of data deliveries from providers.

Table 2. Major reporting inconsistencies of water quality data observed in this survey and recommendation for proper data reporting

Major data reporting inconsistencies	**Recommendations**
Laboratory measurement values delivered in multiple files over time; newly measured values are updated in tables of past reports. In some cases, measurement values of previous report differ from values in updated data reports	Use standardised data reporting sheets; provide newly measured data in a distinct file (do not update newly measured data in past documents); do not change measurement values which have already been reported in previous data reports
Inconsistent designation of sampling codes (different sampling codes used for same locations from one report to another)	Provide consistent sampling codes across all reports

Sampling periods (from-to) used instead of concise sampling dates	Provide a unique sampling date in UTF-8 format
Ambiguous designation of parameter/contaminant	Provide CAS[1] number for each investigated parameter/contaminant to ensure correct assignment
Sites with incorrect latitude and longitude information and missing SRID[2] information	Provide correct latitude and longitude information and associated SRID for all reported sampling locations
Laboratory results reported for sites without location information	Provide only laboratory results if they can be clearly assigned to a distinct sampling location.
Data provided in formats which cannot be easily copied (e.g. images such as jpg, gif, png, screenshots or pdf documents)	Provide data in machine-readable xls, csv or any other format that allows to copy data in an easy way
Errors in assigning and converting measurement units	Guarantee correct use of units and its conversions (IS[3]). If measurement values differ in order of magnitudes for same contaminants and location, check units
Inconsistent use of decimal separators in numbers: mix of commas and points	Use consistent decimal separators for measurement values, either decimal comma or decimal point (recommended)
Inconsistent documentation of date-time for sampling and analysis	Provide date and time information for both sampling and analysis
Transport conditions from place of sampling (field) to analysis (lab) often not reported	Indicate sample transport condition (e.g. cooling chain, time of transport)
Analysis methods and devices only partially indicated	Assign the used analysis method and analysis devices to each measurement and provide ISO[4] for methods applied

[1] CAS-Number: numerical designation for chemicals, maintained by the Chemical Abstracts Service (CAS) of the American Chemical Society

[2] SRID: Spatial Reference ID

[3] IS: International System of Units

[4] ISO: International Organisation for Standardisation

Table 3. Data (Quality) Levels from raw data (laboratory reports) to tidied data stored into the database, allowing the assessment of data quality and completeness

Data Level	**Description**
*report_LEV0-0 (pdf, doc)	Original data format as received from provider (e.g. laboratory), usually delivered as reports containing parameter and locational data in various tables, text or screenshots throughout the document. The latter impede direct copying of content and, thus, should be avoided. All original reports sent by the laboratories are kept as reference. Desired: Original data provided in an easy to copy format (e.g. xls, csv)
*loc_point_long_LEV1-0 (xls)	Transfer (copy-over) of locational data (location) from report to MS Excel in long format using advanced standardised templates with built-in

*loc_site_long_LEV1-0 (xls)	validation (see **Figure 49**). Monitoring point and monitoring sites are reported in two distinct location MS Excel files. Obvious data errors (e.g. erroneous location name, coordinates format or SRID info) are corrected on the fly and corrections are reported under 'Note'. The harmonised location file ensures that the data is uploaded to the database via the EMS-tool 'Validation & Database upload'. Alternatively, locational data can be directly added one by one to the database, via the EMS tool 'Database editing'. Desired: Error-free locational data in standardised long format datafile
*ts_wide_LEV1-0 (xls)	Transfer (copy-over) of time series data from report to MS Excel in wide format using advanced standardised templates with built-in validation (see **Figure 51**). Obvious data errors (e.g. erroneous naming of parameters/contaminants, units) are corrected on the fly. Desired: Error-free parameter data in standardised wide format datafile; unambiguous naming/assignment of parameters, e.g. by using internationally agreed terminology or providing CAS-number)
*ts_long_LEV1-0 (xls)	Tidied time series data is transposed from wide to long format using wide2long.py script. Time series data in long format ensures the massive data upload to database via the EMS-tool 'Validation & Database upload'. Alternatively, time series data can be (i) directly compiled in long format using datafile template or (ii) directly added one by one to the database, via the EMS tool 'Database editing'. Desired: Error-free parameter data in standardised long format; unambiguous naming convention and assignment of parameters (e.g. standard terminology, CAS-number). Development of wide2long GUI.
*ts_long_LEV1-1 (xls)	Harmonisation of measurement units following the units commonly used for law-limits in international laws (e.g. WHO, EC). This facilitates comparability with parameters, law-limits and EQS. Desired: Time series data with harmonised/converted units, as used in international laws.
*ts_long_LEV2-0([1]) (modifications done in database via EMS tool, based on visual data validation)	Expert corrected time series data after visual data validation in EMS dashboards 'Parameter values' and 'Law/EQS limit exceedance'. Modifications and comments can be done via the EMS module 'Database editing' and need to be reported under 'Note' of the respective database table, stating also the name of the editor. Desired: Add 'flag' option to the EMS 'database editing' interface, to allow the expert to easily indicate modifications done and keep track of changes (LEV2-0 vs LEV1-1 data)
*ts_long_LEV3-0([1]) (modifications done in database via EMS tool, based on data aggregation)	Aggregation of tidied time series data (LEV2-0) for analysis purposes or further assessment of data quality. For example, data from multiple monitoring points referring to the same site and sampling date could be aggregated (e.g. mean, min, max, quantiles standard deviation, variance of temperature). Data aggregation can directly be done in the EMS database. Desired: Add data aggregation functionality as another EMS feature

([1]) *LEV2 and *LEV3 still in progress

6.2.3 Law limit data

To assess water quality, laboratory analysis results are contrasted with law limits and EQS, derived from national or international laws and guidelines (e.g. published by WHO, EC, US EPA, Senegal national laws) or extracted from studies (e.g. MacDonald et al., 2000). A user-defined datafile and associated template have been created to ensure consistent reporting of law limit data (**Figure 51**). The following attributes are compiled in long format, using the law limit datafile reporting sheet with built-in function: code (law code), name (extended law name), type (classification of limit type, e.g. law limit or EQS), parameter (parameter name identifier), law limit (law limit value), unit (measurement unit), media (to which the measurement refers), use (type, e.g. drinking water) and note. To date, the validation of the user-defined datafiles against templates is not yet supported but this feature will be soon integrated in the EMS module 'Validation and database upload'. For the moment being, data compiled in the law limit datafile has to be added one by one using the 'Database editing' module. Here, a new law limit value can be added to the database by clicking on the tab 'Law limit' and then 'Add'. In the pop-up window, measurement type (TS Type as combination of parameter name, measurement unit and media), law and use have to be selected from the scroll down list. The law limit value has to be provided in the same unit as indicated in the measurement type. If some of the mandatory information is not available in the scroll down list, the missing information has to be added in the dedicated table before, e.g. 'law' or 'use'. If a TS Type is not listed, the new measurement type can be created after adding the missing information on parameter, media and unit. Prior to adding new entries, it must be checked that information is not already available in the database.

6.2.4 Visualisation of water quality data

Currently, water quality (time series) data can be visualised and investigated in two dashboards. The dashboard 'Law limit /EQS exceedance' shows sampling locations where measurement values of selected parameters exceed certain law limits or EQS. This helps decision makers to quickly spot areas of water quality conflicts. The second dashboard 'Parameter values' aims at deeper spatio-temporal analysis of selected parameters. By clicking 'Only points with time series', only time series data is shown by filtering out all one-time measurements. Selected data is displayed in inter-linked tables, scatter-plots and maps, making the app responsive and intuitive. Visualisation and analysis features can be extended at any time by further developing the underlined R code, supporting e.g. (i) distance mapping of monitoring points and sites to the river mouth to investigate the change of parameter concentrations along the longitudinal river profile and identifying inputs from tributaries or pollution caused by human activities (cities, agriculture), (ii) correlation diagrams to compare the behaviour of contaminants among each other, originating from same or different media, (iii) a hydrographic model with flow (discharge) diagrams to understand the discharge situation, e.g. spotting river areas with dilution effects and (iv) more web mapping services such as population density and land use, further facilitating water quality data interpretation. These visualisation tools help to analyse the change of contaminant concentrations along the longitudinal river profile and check for data consistency. The content of the dashboards can be exported as figures and tables and integrated in decision support reports. The visualisation and interpretation of validated data can support the preparation of synthesis reports on water quality in the river basin.

All database content about sampling location, parameter concentrations and law limits/EQS can be exported as csv format and serve as input for analysis and visualisation using external open source or proprietary analysis and visualisation tools. Alternatively, external analysis tools (e.g. GIS, spatial statistics and modelling tools) can be directly connected to the database, to benefit from their visualisation, data processing and analysis features.

7 Recommendations

Data harmonisation is generally a challenging task, which can be more complex in the context of a multi-national project focused on water quality data, traditionally lacking common and stringent agreed standards in data collection and organisation.

Even if good basic guidelines were provided since the beginning of the project and national experts made their best to comply with them, still each country dataset turned to be unique on its own. Mostly, they differ in naming conventions, attributes, file formats and overall data organisation strategies (i.e., time series organised in MS Excel sheets or split in a myriad of text files). Moreover,, many minor and/or major relational inconsistencies, as geographic projection issues, no-matching records in spatial coverages and time series, and/or other issues, as undeclared flag data in long time series difficult to spot in advance, strongly impact on harmonisation procedures times and effectiveness.

Based on the above considerations, a major recommendation for more effective data collection, in the future, would be the adoption of more stringent electronic formats standards for data delivery, major inspiration coming from US EPA EDD (Electronic Data Deliverables) and EDP (EDD Data Processing) concepts and software implementations (US EPA, 2016; EarthSoft, 2022). Basic ideas behind EDD/EDP are simple and effective:

- Standardised text formats, organised by data typology (i.e., monitoring points inventory, levels, chemical, discharge data). Standardisation includes attributes and their types, ranges of admitted values and/or related VVLs, Valid Values Lists, relational rules as unicity, not nullity, etc…, much as it would be the case for the rules defined at database schema design phase.
- EDP software enables automatic processing of EDD data, to spot any error/inconsistency far before attempting to migrate data to spatio-temporal database. As an iterative process, validation should ideally be continued until all open issues are properly addressed and fixed.

Ideally EDD file formats should be agreed upon and shared with different stakeholders involved in data collection and harmonisation processes, EDD data validation addressed since the early phases of the project.

The idea of automating data cleaning/tidying processes is well established in common practice (Wickham, 2014), leading to many advantages, as it ensures process documentation and reproducibility. As new issues emerge, data tidying/cleaning processes programming can be refined and re-run, much as the iterative process of EDD validation. R and Python provide ideal platforms with plenty of libraries customised to perform this kind of tasks.

Lessons-learned in the framework of current and similar assignments show that EDD standardisation/validation and automation of data tidying/cleaning processes is the right approach, but also that the latter processes can be extremely time-consuming and a true nightmare if lacking former EDD standardisation. It is much as reinventing the wheel anytime a new dataset is provided.

As data migration to spatio-temporal database is attempted, implemented relational rules contribute to spot any further inconsistency, limiting risks of data duplication and integrity failure, and further promoting iterative data tidying/cleaning processes. Data transfer of unstructured or poorly structured files, i.e., MS Excel and shape files, should be avoided as long as possible, or at least early data validation, including relational consistency checks, should be promoted.

The adoption of OS solutions and native spatial database platforms can add flexibility in combining and integrating data with different OS and proprietary platforms for geospatial analysis (i.e., GIS, spatial statistics and modelling tools), where the concept of native spatial database refers to databases with their own capabilities to load, query, manage (i.e., native spatial data types) and spatial indexing geographic data.

PostgreSQL/PostGIS is a good and well-established state-of-the-art solution, natively integrated (for both reading and writing) with QGIS and, since the introduction of Query Layers, also accessible (unfortunately only in read-mode for entry licenses) from ArcGIS GIS. PostGIS can be easily accessed through (geo)statistical data analysis tools, as R, and even groundwater modelling tools, as DHI-WASY Feflow, since version 6.2.

The above recommendations apply to projects and/or project phases, which would benefit of the stringent integrity controls of a relational database, while lacking direct access to an enterprise SDI (Spatial Data Infrastructure). Data can always be migrated to an enterprise SDI at a later stage, if available and required, as it would be the case for the current project.

While awareness about current state-of-the-art standards in SDI (EC JRC, 2022) design and development is key to improving data flow, organisation and accessibility, it is worth to stress once more that national habits in (ground)water data collection and organisation still suffer of poor standardisation. This is not uncommon in bottom-up data collection efforts, particularly in complex large-scale 3D groundwater hydrology applications.

8 Conclusions

Concise and efficient reporting, management and visualisation of different types (vector and raster) and complexity (high spatial and temporal resolution) of data is a challenge, particularly in an inter-disciplinary and transboundary context. Stand-alone, intuitive, powerful, multi concurrent and open-source data management and visualisation tools are highly needed to ensure reliable data documentation and spatio-temporal exploratory analysis to a wide audience, key for sound decision making. The EMS tool has been developed in the frame of the WEFE Senegal nexus project "Support for water resources management and the Water-Energy-Agriculture Nexus in the Senegal River Basin", requested by the African partners to support the management and analysis of heterogeneous and sensitive water quality data, provided by several laboratories in Senegal, Mali and, regarding reference data, The Netherlands. This pilot study provided a good framework to deal with all the challenges throughout the entire data chain from data acquisition (here sampling), management to analysis and visualisation – resulting in the development of an all-in-one EMS tool. Due to its easy to use modular architecture it contributes to the overall goal of producing and sharing consistent and reliable data in an efficient and standardized manner, even in regions with limited IT infrastructure and expertise. No specific knowledge of the underlying IT and programming architecture is required for the use of EMS modules/dashboards. At the same time, standard geographic data quality (completeness, consistency, spatial and thematic accuracy, timeliness, and usability) and interdependent principles are ensured. The user has full control over his generated and stored data, as he decides whether the tool is used locally or via a web server.

The added values of the developed EMS tool are:

- Open source tool (e.g. no licensing costs, possibility to extend the code)
- Modularised structure
- User-friendly GUI providing quick access to main functionalities, even for less experienced end-users
- Tool that can be installed on single PC (stand-alone) or shared on a web-server
- Standardised data reporting
- A robust database architecture, supporting both raster (e.g. satellite images, aerial photos) and vector data (e.g. monitoring points location, rivers, administrative boundaries)
- Export of database content to text files (csv) or direct connection to database from external tools for spatio-temporal analysis

The three main pillars of the EMS tool are:

- At the core of the system, a concurrent multi-user spatio-temporal database, implemented in PostgreSQL/PostGIS
- An application aimed at supporting data validation (data cleaning/tidying) and data uploading to the database, following the US EPA EDD (Electronic Data Deliverables) concepts
- Dashboards, aimed at supporting exploratory data analysis in space and time, as temporal trends of contaminants concentrations and detection of exceedance of law or EQSs (Environmental Quality Standards) limits

The EMS workflow consists of the following main steps:

- Install, locally or on remote servers, and configure the EMS tool for MsWindows, including database setup (schema creation, system data uploading, connection testing) and web applications
- Edit database (add, update and delete records; export database contents), for both system data (e.g. parameters, laws, law limits) and site-specific monitoring data
- Prepare site-specific data (monitoring location, time series) using standard predefined or new user-defined MS Excel templates
- Validate data and (massive) upload data to the database. The process includes checking for and reporting about inconsistencies/errors in datasets, based on rules defined in the standard templates, and validating data against former database content
- Visualise and analyse data in space and time through dedicated dashboards.

- Spatio-temporal exploratory analysis of parameters distribution and trends, and law/EQS (Environmental Quality Standards) limits exceedance
- Direct access to database content for editing and querying through the web administration tool PgAdminIV, generally dedicated to advanced users and system administrators

A leaflet entitled 'Environmental Monitoring System: Quick Start Guide' was designed to support the user with EMS installation, awareness about system architecture and functionalities (see **Annex 5**).

Among potential future developments:

- Project/contract operational data management, to monitor activities advancement vs. milestones/deadlines
- Keep track of uncertainty and quality of datasets, supporting data cross-validation and laboratories comparison/assessment
- Develop a GUI to facilitate the conversion of reported datasheets from wide to long format
- Further support for automatic validation of user-defined datafiles and templates
- Fine tuning database and dashboards for large datasets performance improvement even in a resource-poor environment
- Expand analysis features, as specialised statistical charts and raster data aggregation over sub-basins, to leverage the full tool potential
- Develop a mobile app version, for use in the field

A public EMS version is accessible through the dedicated page 'EMS-Tool' (https://aquaknow.jrc.ec.europa.eu/document/ems-tool) on Aquaknow KMS. The database contains a dummy dataset (named 3D, as Dummy Desert Dataset) arbitrarily located in the Sahara desert in North Niger. The dataset contains monitoring piezometers, piezometric heads and groundwater quality time series, based on outcomes of a fictitious flow and transport model built using Feflow 6.2. Such a public version addresses the requirement not to infringe confidentiality constrains inherent to Senegal transboundary river basin dataset.

Another version of the tool, the same as above except for the database containing the monitoring data collected and processed in the framework of the Senegal transboundary river basin project, is available through a private project group on Aquaknow KMS (https://aquaknow.jrc.ec.europa.eu/node/18352). The page access is password protected and specific requests for authorisation can be directed to the reference person reported at the beginning of current document.

The EMS tool is complemented by a web mapping application 'Aquaknow GIS TOOL' (video-demonstration: https://www.youtube.com/watch?v=5pzYBx8EPn8) embedded in Aquaknow KMS to quickly inform users about the status of available water quality data in space and time.

The EMS tool setup file (and installation video tutorial), EMS leaflet, the 3D dataset and the Feflow model setup can be downloaded from the dedicated EMS Tool website of Aquaknow KMS (https://aquaknow.jrc.ec.europa.eu/document/ems-tool). The Feflow model setup might be of interest for users who would like to inspect the model and modify it to address different and more challenge scenarios (e.g. multi-aquifer systems, salt water intrusion, surface water and meteo-climate data to inspect statistical relationships).

References

Abramic, A., A. Kotsev, V. Cetl, S. Kephalopoulos, and M. Paviotti, ‘A Spatial Data Infrastructure for Environmental Noise Data in Europe’, International Journal of Environmental Research and Public Health, Vol. 14, No. 7, 2017.

Adobe, ‘Adobe Acrobat Pro’, 2022.

Alam, M.M., L. Torgo, and A. Bifet, ‘A Survey on Spatio-Temporal Data Analytics Systems’, ACM Computing Surveys, Vol. 1, No. 1, 2021, pp. 1–44.

Alamouri, A., M. Hassan, and M. Gerke, ‘Development of a Methodology for Real-Time Retrieving and Viewing of Spatial Data in Emergency Scenarios’, Applied Geomatics, Vol. 13, No. 4, 2021, pp. 747–761.

Bajracharya, B., R.B. Thapa, and M.A. Matin, Earth Observation Science and Applications for Risk Reduction and Enhanced Resilience in Hindu Kush Himalaya Region, Earth Observation Science and Applications for Risk Reduction and Enhanced Resilience in Hindu Kush Himalaya Region, 2021.

Blanes, N., G. Closa, M.J. Ramos, M. Sáinz de la Maza, E. Peris, and D. Lihteneger, Environmental Noise Directive Reporting Guidelines, Kjeller, 2021.

Brinkmann, T., R. Both, B.M. Scalet, S. Roudier, and L. Delgado Sancho, JRC Reference Report on Monitoring of Emissions to Air and Water from IED Installations, Publications Office of the European Union, Publications Office of the European Union, Seville, 2018.

Brovelli, M.A., K.J. Li, and K.S. Eom, ‘Moving toward Open Geospatial Systems: The UN Open GIS Initiative’, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, Vol. 43, No. B4-2021, 2021, pp. 183–188.

Chen, D., S. Shams, C. Carmona-Moreno, and A. Leone, ‘Assessment of Open Source GIS Software for Water Resources Management in Developing Countries’, Journal of Hydro-Environment Research, Vol. 4, No. 3, 2010, pp. 253–264.

Crestaz, E., A. Ancilletta, L. Patata, M. Pellegrini, and F. Tatangelo, ‘Groundwater Flow and Transport Modeling Aimed at an Exploratory Spatial Data Analysis of Ust-Kamenogorsk Aquifer System, East Kazakhstan’, 13th International UFZ-Deltares Conference on Sustainable Use and Management of Soil, Sediment and Water Resources, Copenhagen, 2015, pp. 22–31.

Crestaz, E., N. Habashi, P. Ambrosini, P. Schätzl, and M. Gibin, ‘Advancements in Concurrent Native Spatial Database Technology for Groundwater Monitoring and Modeling Applications. A Case Study Aimed at PostgreSQLPostGIS Coupling with GIS and Feflow’, Proceedings of “MODFLOW and More 2011: Modelling a Complex World”, Denver, 2015.

Crestaz, E., ‘Design and Development of a Prototype Addressing Spatio- Temporal Environmental Vector Data Management, Analysis and Delivery Using Open Source Technology: General Framework and Case Study Focused on Groundwater Management in a Coastal Area’, University of Hertfordshire, 2011.

———, ‘Spatial Data Management in GIS and the Coupling of GIS and Environmental Models’, GIS Based Chemical Fate Modeling, John Wiley & Sons, Ltd, 2014, pp. 217–252.

Criollo, R., V. Velasco, A. Nardi, L. Manuel de Vries, C. Riera, L. Scheiber, A. Jurado, et al., ‘AkvaGIS: An Open Source Tool for Water Quantity and Quality Management’, Computers and Geosciences, Vol. 127, No. December 2017, 2019, pp. 123–132.

CRWR, ‘ArcGIS Hydro Data Model’, 2021. https://www.crwr.utexas.edu.

Diersch, H.-J.G., FEFLOW: Finite Element Modeling of Flow, Mass and Heat Transport in Porous and Fractured Media, FEFLOW, Springer Berlin, Heidelberg, 2014.

Directive, I., Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 Establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)., Official Journal of the European Union, Vol. 108, Vol. 108, 2007.

Earthsoft, ‘EQuIS: World’s Most Widely Used Environmental Data Management Software’, 2022. http://earthsoft.com/.

EC JRC, ‘Inspire Knowledge Base: Infrastructure for Spatial Information in Europe’, 2022. http://inspire.ec.europa.eu/.

EPSG.io, ‘Coordinate Systems Worldwide’, 2022. https://epsg.io/.

EPSG, ‘EPSG Geodetic Parameter Dataset’, 2022. https://epsg.org/home.html.

ESRI, ‘An Overview of Versioning’, 2022. https://desktop.arcgis.com/en/arcmap/10.3/manage-data/geodatabases/an-overview-of-versioning.htm.

———, Understanding ArcSDE, 2004.

De Filippis, G., C. Pouliaris, D. Kahuda, T.A. Vasile, V.A. Manea, F. Zaun, B. Panteleit, et al., ‘Spatial Data Management and Numerical Modelling: Demonstrating the Application of the QGIS-Integrated FREEWAT Platform at 13 Case Studies for Tackling Groundwater Resource Management’, Water (Switzerland), Vol. 12, No. 1, 2020.

Gillwald, A., D. Johnson, L. Cloete, S. Hadzic, S. Kiden, A. Rens, and A.. . Phokeer, African Open Source Digital Infrastructures: Evaluating the Landscape, Cape Town, 2022.

Hipp Wyrick and Company, ‘SQLite’, 2022. https://www.sqlite.org/.

Hřebíček, J., and W. Pillmann, ‘Shared Environmental Information System and Single Information Space in Europe for the Environment: Antipodes or Associates?’, Proceedings of the European Conference “TOWARDS eENVIRONMENT”, 25-27 March 2009, Prague, Brno, 2009.

ISO, ‘ISO 8601: Date and Time Format’, 2022. https://www.iso.org/iso-8601-date-and-time-format.html.

MacDonald, D.D., C.G. Ingersoll, and T.A. Berger, ‘Development and Evaluation of Consensus-Based Sediment Quality Guidelines for Freshwater Ecosystems’, Archives of Environmental Contamination and Toxicology, Vol. 39, No. 1, 2000, pp. 20–31.

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

Obe, R., and L. Hsu, PostGIS in Action, Manning Publications, 2011.

PgAdmin, ‘PgAdmin: PostgreSQL Tools’, 2022. https://www.pgadmin.org/.

Pistocchi, A., GIS Based Chemical Fate Modeling, GIS Based Chemical Fate Modeling: Principles and Applications, Wiley, 2014.

RockWare, ‘A to z All Products’, 2022. https://www.rockware.com/product-category/a-to-z-all-products/.

Rossetto, R., G. De Filippis, I. Borsi, L. Foglia, M. Cannata, R. Criollo, and E. Vázquez-Suñé, ‘Integrating Free and Open Source Tools and Distributed Modelling Codes in GIS Environment for Data-Based Groundwater Management’, Environmental Modelling and Software, Vol. 107, No. May, 2018, pp. 210–230.

Seequent, ‘Leapfrog Products & Software’, 2022. https://www.seequent.com/products-solutions/leapfrog-software/.

Silva, D.G., C. Coutinho, and C.J. Costa, ‘Factors Influencing Free and Open-Source Software Adoption in Developing Countries—an Empirical Study’, Journal of Open Innovation: Technology, Market, and Complexity, Vol. 9, No. 1, 2023.

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

Swain, N.R., K. Latu, S.D. Christensen, N.L. Jones, E.J. Nelson, D.P. Ames, and G.P. Williams, ‘A Review of Open Source Software Solutions for Developing Water Resources Web Applications’, Environmental Modelling and Software, Vol. 67, No. May, 2015, pp. 108–117.

US EPA, Electronic Data Deliverable (EDD) Comprehensive Specification Manual 4.0, 2016.

USGS, ‘MODFLOW and Related Programs’, 2022. https://www.usgs.gov/mission-areas/water-resources/science/modflow-and-related-programs.

Wickham, H., ‘Tidy Data’, Journal of Statistical Software, Vol. 59, No. 10, 2014, pp. 1–23.

List of abbreviations and definitions

ArcSDE	ArcGIS Spatial Data Engine
CAS	Chemical Abstracts Service
CERES	Centre Régional de Recherches en Ecotoxicologie et Sécurité Environnementale, Sénégal
CRUD	Create-Read-Update-Delete
CRWR	Centre for Research in Water Resources
CSV	Comma Separate Values text file
DGPRE	Ministère de l'Eau et de l'Assainissement – Direction de la Gestion et de la Planification des Ressources en Eau, Sénégal
DNH	Direction Nationale de l'Hydraulique, Mali
EAR	Entity Attribute Relationships
EDD	Electronic Data Deliverables
EDP	EDD Data Processing
EEA	European Environmental Agency
EIS	Environmental Information System
EMS	Environmental Monitoring System (also the application acronym)
ENI	Ente Nazionale Idrocarburi (Italian Oil Company)
EPSG	European Petroleum Survey Group
EQS	Environmental Quality Standard
ESRI	Environmental Systems Research Institute
FK	Foreign Key
FREEWAT	FREE and open source software tools for WATer resource management
GPS	Global Positioning System
GUI	Graphical User Interface
HCI	Human Computer Interaction
KMS	Knowledge Management System
ID	Identifier
IP	Internet Protocol address
IPD	Institute Pasteur de Dakar, Sénégal
IS	International System of units
ISO	International Organization for Standardization
IT	Information Technology
LCV	Laboratoire Central Vétérinaire, Mali
LNE	Laboratoire National des Eaux, Mali
OCP	OrganoChlorine Pesticide
OO	Object Oriented
OS	Open Source or Operative System (depending upon context)
PK	Primary Key
QA/QC	Quality Assurance and Quality Control

RAM	Random Access Memory
R&D	Research & Development
SDI	Spatial Data Infrastructure
SQL	Standard Query Language
SRB	Senegal River Basin
SRID	Spatial Reference ID
UCL	University College of London
UI	User Interface
UML	Unified Modelling Language
US EPA	United States Environmental Protection Agency
USGS	United States Geological Survey
UTM	Universal Transverse Mercator projection system
VUA	Vrije Universiteit Amsterdam
VVL	Valid Values List
WEFE	Water Energy Food Ecosystems nexus
WKT	Well Known Text format

List of figures

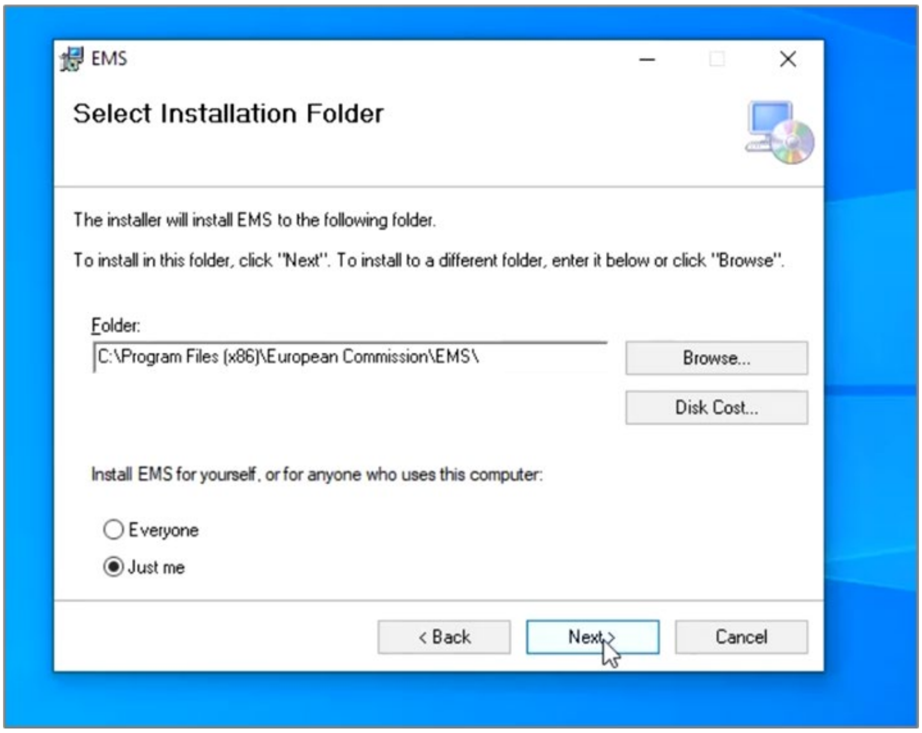
List of tables

Annexes

Annex 1. Installation instructions

The EMS application relies upon multiple languages and environments, including R, Python and the PostgreSQL/PostGIS database. A Windows installation package is provided (**Figure 41**), seamlessly installing all the needed components.

Figure 41. EMS installation menu



EMS

Select Installation Folder

The installer will install EMS to the following folder.

To install in this folder, click "Next". To install to a different folder, enter it below or click "Browse".

Folder:
C:\Program Files (x86)\European Commission\EMS\

Install EMS for yourself, or for anyone who uses this computer:

○ Everyone
● Just me

A basic installation simply demands for the reference path (it is strongly suggested to leave the default Program Files folder) and user profile(s) (either all accounts on device, or just the user one), a shortcut being created on the desktop to run the application (**Figure 42**).

Figure 42. EMS shortcut

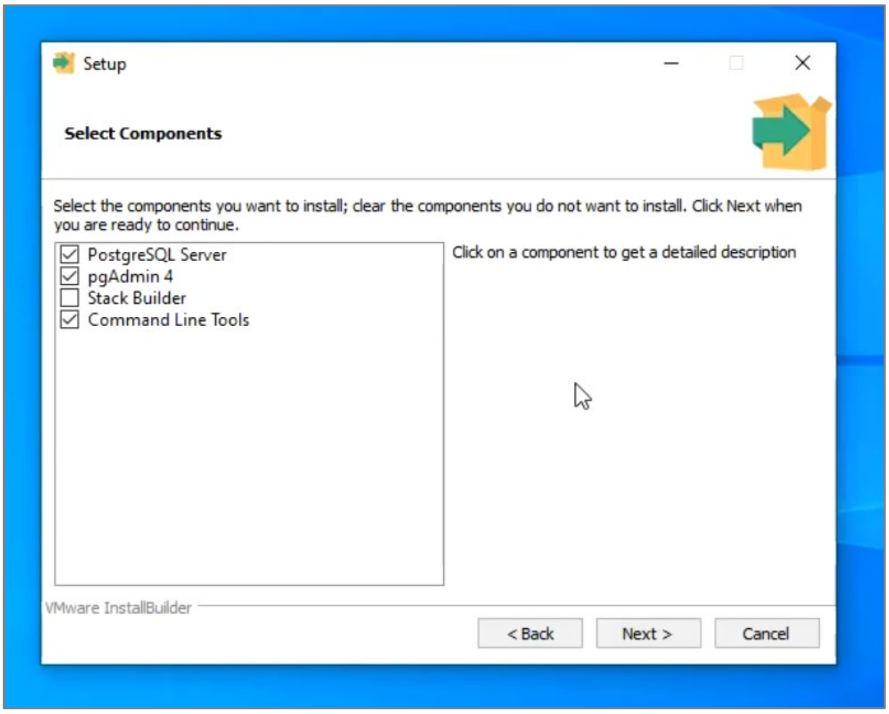
At first run, the application checks for PostgreSQL (13 or later) and PostGIS, and, if needed, it proceeds with the installation following the steps below:

- Installation Directory: leave the default one (should be inside Program Files).
- Select Components: 'Stack Builder' can be unchecked since it is not strictly needed. All other components must be left checked (**Figure 43**).

- Data Directory: leave the default path.
- Password: insert any password. It is important to remember it, since it will used when setting up the PostgreSQL database connection.
- Port: leave the default port number (usually 5432) and memorise it. It will be asked for on database connection too. Occasionally the port 5432 can be already reserved to another installed application, in which case 5433 or other number can be provided.
- Advanced Option: leave the default.

On parameters setup, the installation will start and will be completed in some minutes.

Figure 43. PostgreSQL installation menu



Select Components

Select the components you want to install; clear the components you do not want to install. Click Next when you are ready to continue.

☑ PostgreSQL Server
☑ pgAdmin 4
□ Stack Builder
☑ Command Line Tools

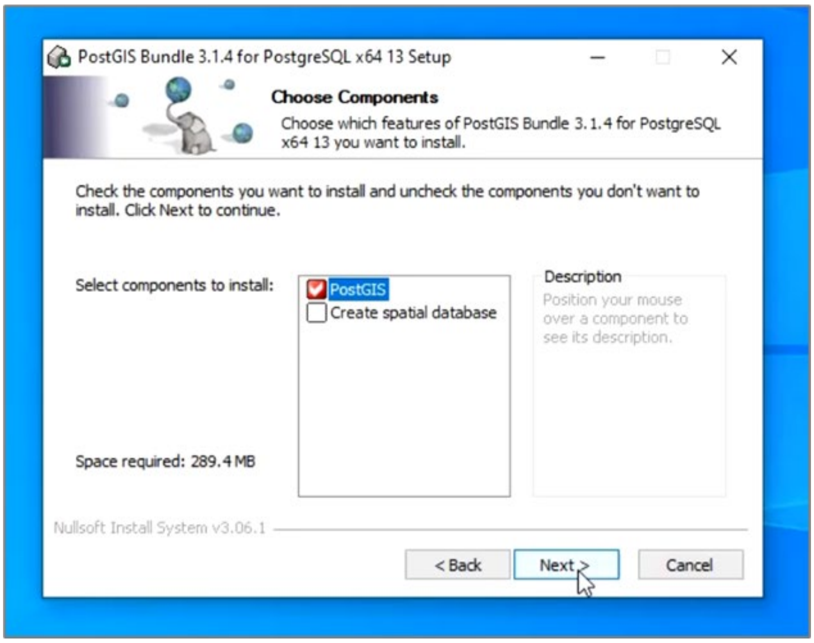
Click on a component to get a detailed description

PostGIS, the geographical extension of PostgreSQL, must then be installed in the same program folder, following the steps below:

- Choose Components: only PostGIS is necessary, the spatial database can be left unchecked (**Figure 44**).
- Choose Install Location: leave the default path, already set to PostgreSQL folder.

Skip further preference requests and the installation will be completed in few minutes.

Figure 44. PostGIS installation menu


PostGIS Bundle 3.1.4 for PostgreSQL x64 13 Setup

Choose Components

Choose which features of PostGIS Bundle 3.1.4 for PostgreSQL x64 13 you want to install.

Check the components you want to install and uncheck the components you don't want to install. Click Next to continue.

Select components to install:

☑ PostGIS
□ Create spatial database

Description

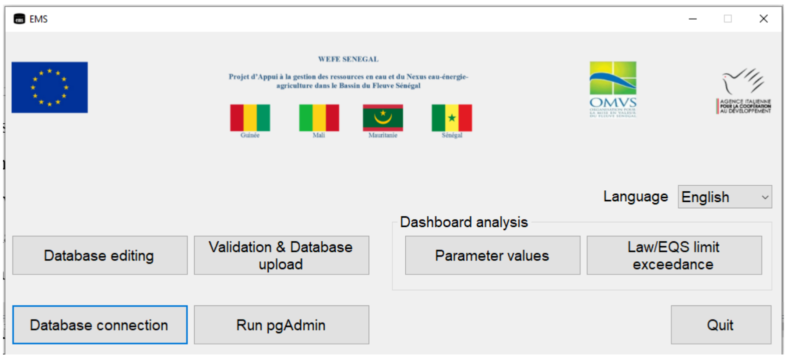
Position your mouse over a component to see its description.

Space required: 289.4 MB

On EMS starting, the main application interface (**Figure 45**) provides access to the different application components, as detailed in this document, further to the selection of the language of preference. The database connection parameters must be entered on first run, or on successive change of reference database, following the steps below (**Figure 46**):

- Server: leave Localhost if data is stored in the local PC (default option). In case the database is accessed through a remote server, ask the system administrator and provide the IP address.
- Database: leave postgres (default).
- User ID: leave postgres (default).
- Password: insert the password selected during PostgreSQL installation.
- Port: insert the port number selected during PostgreSQL installation.

Figure 45. EMS main interface

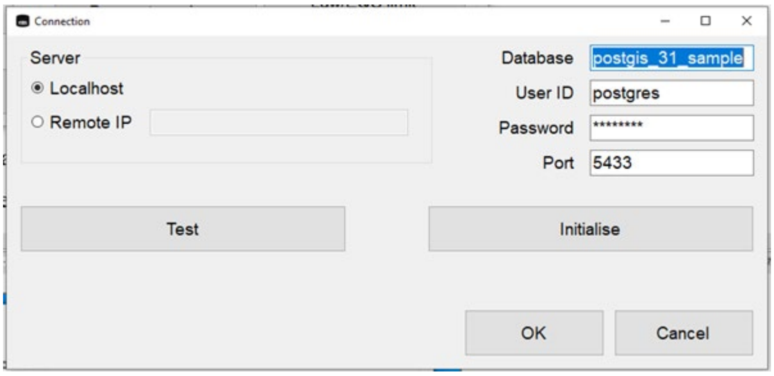

EMS

WEFE SENEGAL

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

Language: English

Figure 46. EMS database connection window



Connection

Server

● Localhost

○ Remote IP

Database postgis_31_sample

User ID postgres

Password ********

Port 5433

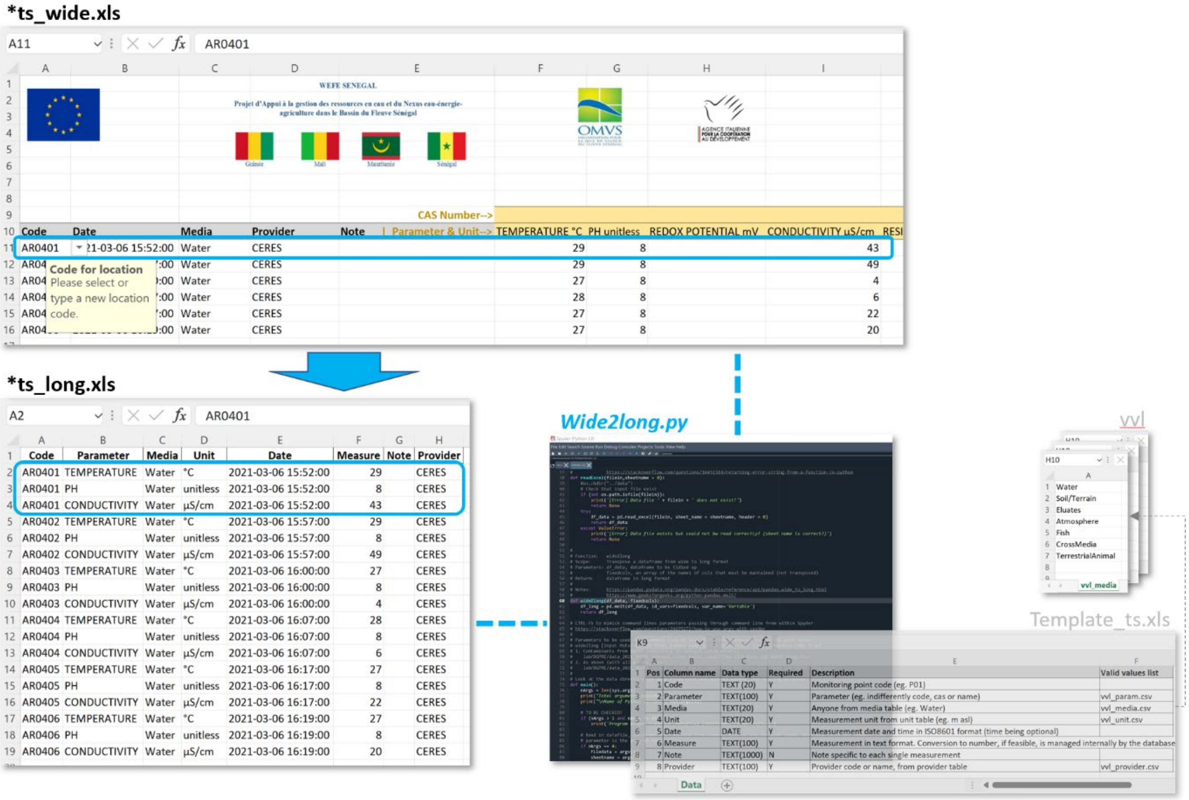
The connection can be tested ('Test' button) and, if successful, configuration can be saved ('Save') to avoid entering the parameters again at next run.

Only the first time the application is run or in case of database reset, the database must be setup ('Initialise' button). The database features are created (e.g. tables, constrains, relationships, views) and the tables populated with the system supporting data (e.g. parameters, laws, law limits) and the Senegal WEFE nexus dataset. The setup file has a total size of about 5 GB, so processing takes quite some time, provided that a command window supports the user in monitoring the processing status advancement. This is the last step, the application being now ready for use.

Annex 2. Wide to long data tidying application

The Wide2Long data tidying utility is provided separately from the EMS application. It aims at converting a standardised MS Excel time series data file from wide to long format, required for massive data uploading to the PostgreSQL/PostGIS EMS database ts (time series) table. Multiple measurement values reported in one row (wide-format) are transposed to one measurement per row (long format), accompanied with all information related to this measurement (**Figure 47**). Cells with no measurement value information in wide format table are not retained in the long format. Long format table attributes are: code (e.g. sampling point), parameter, media, unit, date, measure, note, and provider. The program converts the file to long format compliant with the associated template (e.g. template time series) which defines all data format constraints, including eventual VVL. The Wide2long utility is written in Python (wide2long.py, **Figure 48**) and can be compiled e.g. via Spyder IDE or run from command line providing (i) the MS Excel input data file in wide format (*wide*.xls), (ii) the sheet name of the data file (e.g. 'Data') and (iii) the output file name (*long*.xls). Soon, a stand-alone wide2long utility with user-friendly GUI will be developed to reach also users not familiar with python.

Figure 47. Conversion of MS Excel time series data file from wide to long format, compliant with template file and underlying VVL


*ts_wide.xls
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
AGENCE ITALIENNE POUR LA COOPÉRATION AU DÉVELOPPEMENT
*ts_long.xls
Wide2long.py
vvl
Template_ts.xls
Data
vvl_media

A11 AR0401

	Code	Date	Media	Provider	Note	Parameter & Unit-->	TEMPERATURE °C	PH unitless	REDOX POTENTIAL mV	CONDUCTIVITY uS/cm	RESI
9						CAS Number-->					
11	AR0401	21-03-06 15:52:00	Water	CERES			29	8		43	
12	AR04	:00	Water	CERES			29	8		49	
13	AR04	:00	Water	CERES			27	8		4	
14	AR04	:00	Water	CERES			28	8		6	
15	AR04	:00	Water	CERES			27	8		22	
16	AR04	:00	Water	CERES			27	8		20	

Code for location
Please select or type a new location code.

A2 AR0401

	Code	Parameter	Media	Unit	Date	Measure	Note	Provider
2	AR0401	TEMPERATURE	Water	°C	2021-03-06 15:52:00	29		CERES
3	AR0401	PH	Water	unitless	2021-03-06 15:52:00	8		CERES
4	AR0401	CONDUCTIVITY	Water	µS/cm	2021-03-06 15:52:00	43		CERES
5	AR0402	TEMPERATURE	Water	°C	2021-03-06 15:57:00	29		CERES
6	AR0402	PH	Water	unitless	2021-03-06 15:57:00	8		CERES
7	AR0402	CONDUCTIVITY	Water	µS/cm	2021-03-06 15:57:00	49		CERES
8	AR0403	TEMPERATURE	Water	°C	2021-03-06 16:00:00	27		CERES
9	AR0403	PH	Water	unitless	2021-03-06 16:00:00	8		CERES
10	AR0403	CONDUCTIVITY	Water	µS/cm	2021-03-06 16:00:00	4		CERES
11	AR0404	TEMPERATURE	Water	°C	2021-03-06 16:07:00	28		CERES
12	AR0404	PH	Water	unitless	2021-03-06 16:07:00	8		CERES
13	AR0404	CONDUCTIVITY	Water	µS/cm	2021-03-06 16:07:00	6		CERES
14	AR0405	TEMPERATURE	Water	°C	2021-03-06 16:17:00	27		CERES
15	AR0405	PH	Water	unitless	2021-03-06 16:17:00	8		CERES
16	AR0405	CONDUCTIVITY	Water	µS/cm	2021-03-06 16:17:00	22		CERES
17	AR0406	TEMPERATURE	Water	°C	2021-03-06 16:19:00	27		CERES
18	AR0406	PH	Water	unitless	2021-03-06 16:19:00	8		CERES
19	AR0406	CONDUCTIVITY	Water	µS/cm	2021-03-06 16:19:00	20		CERES

```
#           https://stackoverflow.com/questions/[illegible]/returning-error-string-from-a-function-in-python
def readExcel(filein,sheetname = 0):
    #os.chdir("../data")
    # Check that input file exist
    if (not os.path.isfile(filein)):
        print('[Error] Data file ' + filein + ' does not exist!')
        return None
    try:
        df_data = pd.read_excel(filein, sheet_name = sheetname, header = 0)
        return df_data
    except ValueError:
        print('[Error] Data file exists but could not be read correctly! (sheet name is correct?)')
        return None

#
# Function:   wide2long
# Scope:      Transpose a dataframe from wide to long format
# Parameters: df_data, dataframe to be tidied up
#             fixedcols, an array of the names of cols that must be mantained (not transposed)
# Return:     dataframe in long format
#
# Notes:      https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.wide_to_long.html
#             https://www.geeksforgeeks.org/python-pandas-melt/
def wide2long(df_data, fixedcols):
    df_long = pd.melt(df_data, id_vars=fixedcols, var_name='Variable')
    return df_long

# CTRL-F6 to mimick command lines parameters passing through command line from within Spyder
# https://stackoverflow.com/questions/[illegible]/how-to-use-args-with-spyder
```

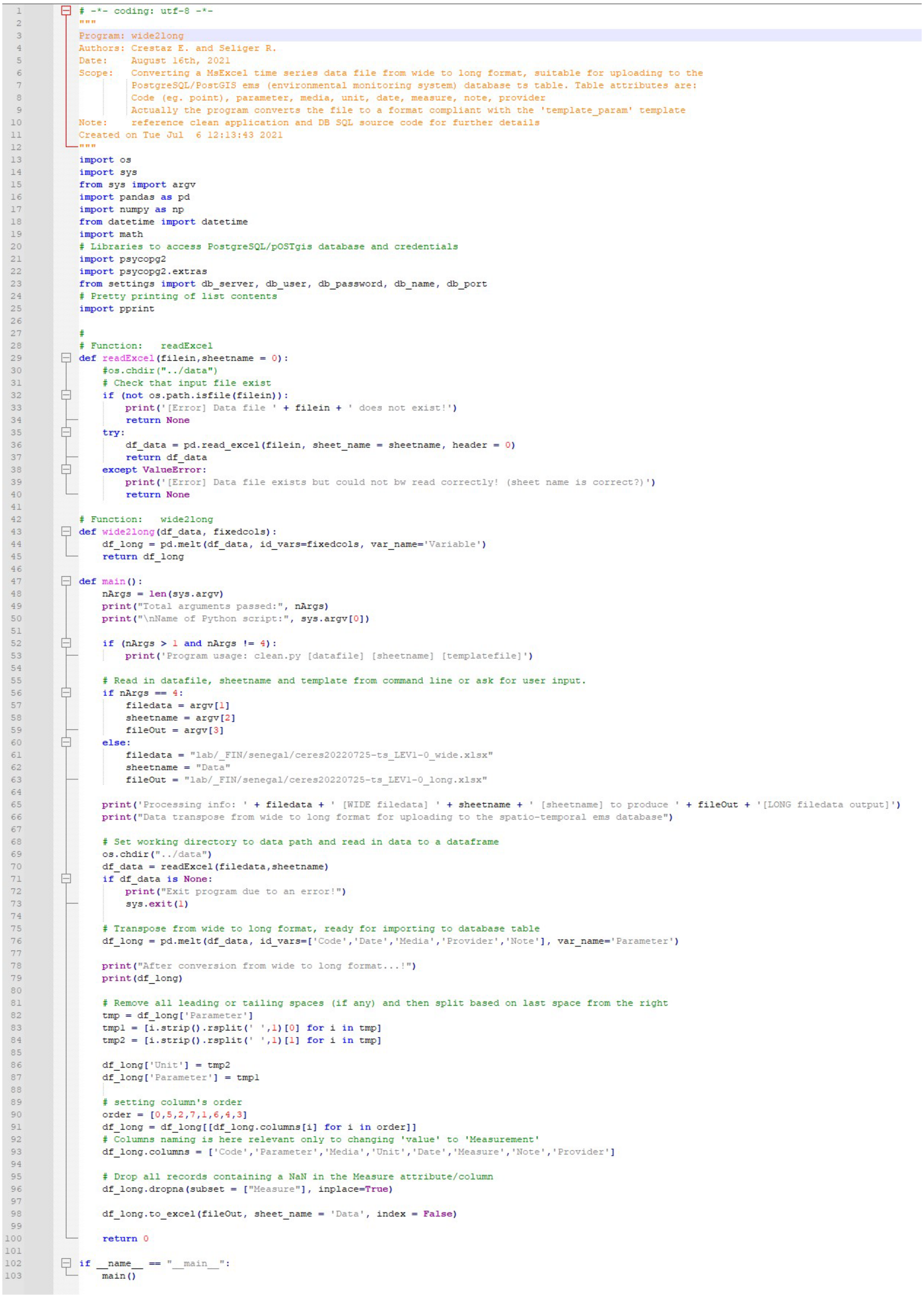
K9

	Pos	Column name	Data type	Required	Description	Valid values list
2	1	Code	TEXT (20)	Y	Monitoring point code (eg. P01)	
3	2	Parameter	TEXT(100)	Y	Parameter (eg. indifferently code, cas or name)	vvl_param.csv
4	3	Media	TEXT(20)	Y	Anyone from media table (eg. Water)	vvl_media.csv
5	4	Unit	TEXT(20)	Y	Measurement unit from unit table (eg. m asl)	vvl_unit.csv
6	5	Date	DATE	Y	Measurement date and time in ISO8601 format (time being optional)	
7	6	Measure	TEXT(100)	Y	Measurement in text format. Conversion to number, if feasible, is managed internally by the database	
8	7	Note	TEXT(1000)	N	Note specific to each single measurement	
9	8	Provider	TEXT(100)	Y	Provider code or name, from provider table	vvl_provider.csv

H10

	A
1	Water
2	Soil/Terrain
3	Eluates
4	Atmosphere
5	Fish
6	CrossMedia
7	TerrestrialAnimal

Figure 48: Wide2long.py code for converting a MS Excel time series data file from wide to long format



```
# -*- coding: utf-8 -*-
"""
Program: wide2long
Authors: Crestaz E. and Seliger R.
Date:    August 16th, 2021
Scope:   Converting a MsExcel time series data file from wide to long format, suitable for uploading to the
         PostgreSQL/PostGIS ems (environmental monitoring system) database ts table. Table attributes are:
         Code (eg. point), parameter, media, unit, date, measure, note, provider
         Actually the program converts the file to a format compliant with the 'template_param' template
Note:    reference clean application and DB SQL source code for further details
Created on Tue Jul  6 12:13:43 2021
"""
import os
import sys
from sys import argv
import pandas as pd
import numpy as np
from datetime import datetime
import math
# Libraries to access PostgreSQL/pOSTgis database and credentials
import psycopg2
import psycopg2.extras
from settings import db_server, db_user, db_password, db_name, db_port
# Pretty printing of list contents
import pprint

#
# Function:   readExcel
def readExcel(filein,sheetname = 0):
    #os.chdir("../data")
    # Check that input file exist
    if (not os.path.isfile(filein)):
        print('[Error] Data file ' + filein + ' does not exist!')
        return None
    try:
        df_data = pd.read_excel(filein, sheet_name = sheetname, header = 0)
        return df_data
    except ValueError:
        print('[Error] Data file exists but could not bw read correctly! (sheet name is correct?)')
        return None

# Function:   wide2long
def wide2long(df_data, fixedcols):
    df_long = pd.melt(df_data, id_vars=fixedcols, var_name='Variable')
    return df_long

def main():
    nArgs = len(sys.argv)
    print("Total arguments passed:", nArgs)
    print("\nName of Python script:", sys.argv[0])

    if (nArgs > 1 and nArgs != 4):
        print('Program usage: clean.py [datafile] [sheetname] [templatefile]')

    # Read in datafile, sheetname and template from command line or ask for user input.
    if nArgs == 4:
        filedata = argv[1]
        sheetname = argv[2]
        fileOut = argv[3]
    else:
        filedata = "lab/_FIN/senegal/ceres20220725-ts_LEV1-0_wide.xlsx"
        sheetname = "Data"
        fileOut = "lab/_FIN/senegal/ceres20220725-ts_LEV1-0_long.xlsx"

    print('Processing info: ' + filedata + ' [WIDE filedata] ' + sheetname + ' [sheetname] to produce ' + fileOut + '[LONG filedata output]')
    print("Data transpose from wide to long format for uploading to the spatio-temporal ems database")

    # Set working directory to data path and read in data to a dataframe
    os.chdir("../data")
    df_data = readExcel(filedata,sheetname)
    if df_data is None:
        print("Exit program due to an error!")
        sys.exit(1)

    # Transpose from wide to long format, ready for importing to database table
    df_long = pd.melt(df_data, id_vars=['Code','Date','Media','Provider','Note'], var_name='Parameter')

    print("After conversion from wide to long format...!")
    print(df_long)

    # Remove all leading or tailing spaces (if any) and then split based on last space from the right
    tmp = df_long['Parameter']
    tmp1 = [i.strip().rsplit(' ',1)[0] for i in tmp]
    tmp2 = [i.strip().rsplit(' ',1)[1] for i in tmp]

    df_long['Unit'] = tmp2
    df_long['Parameter'] = tmp1

    # setting column's order
    order = [0,5,2,7,1,6,4,3]
    df_long = df_long[[df_long.columns[i] for i in order]]
    # Columns naming is here relevant only to changing 'value' to 'Measurement'
    df_long.columns = ['Code','Parameter','Media','Unit','Date','Measure','Note','Provider']

    # Drop all records containing a NaN in the Measure attribute/column
    df_long.dropna(subset = ["Measure"], inplace=True)

    df_long.to_excel(fileOut, sheet_name = 'Data', index = False)

    return 0

if __name__ == "__main__":
    main()
```

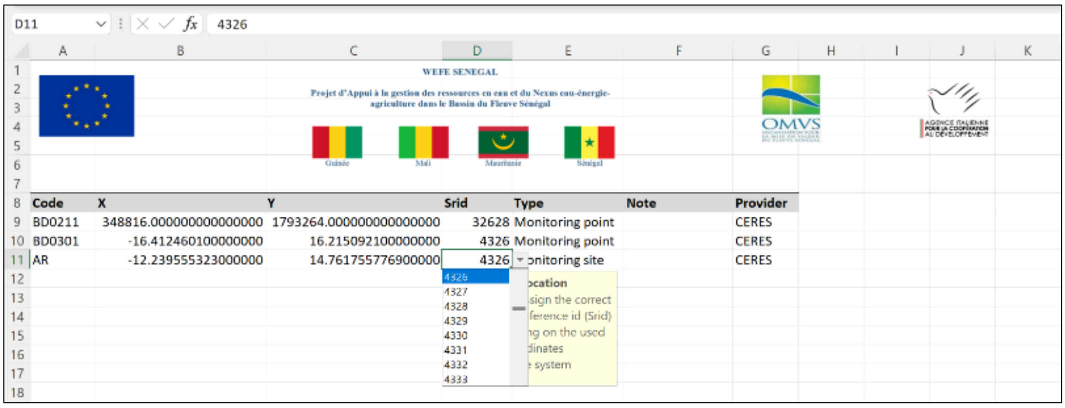
Annex 3. MS Excel file templates

The upload of massive data to the database should be done via the EMS 'Validation and Database Upload' tool, based on advanced standardised datafile templates (MS Excel) with built-in validation function. Standardised data file templates are made available in long format for three file types (i) Location (data_reporting_sheet_loc_long.xlsx), (ii) Time series (data_reporting_sheet_ts_long.xlsx) and (iii) 'User-defined' data (e.g. law limits, data_reporting_sheet_lawlimit_long.xlsx). In addition, another Time Series data file template in wide format is provided, often requested by laboratories for faster data reporting (data_reporting_sheet_ts_wide.xlsx, **Figure 52**). The transformation of wide to long format is automatically done using the python-code wide2long utility (**Figure 47**), described in **Annex 2**. Table attributes of each data file and associated template files are shown in **Figure 49** to **Figure 51**. Built in features in the data file templates provide user guidance during the data compilation process and ensure that data content and format is compiled properly. This includes pop-up windows informing the user on expected column cell input as well as drop-down menus from which the user can select and add correctly formatted content from a list. Cells reserved for date information are already formatted in UTF-8 format. Notes can be made optionally. Measurement values must be reported with decimal points as common decimal separators. The number of used decimals is not limited. 'Less than' and 'greater than' symbols may be used before the measurement values. They will be reported in the database both as text string (e.g. <0.5) and decimal number (e.g. 0.5). In the case of water quality data, the provision of a CAS number for each parameter, e.g. under 'Note', is highly recommended to ensure unambiguous assignment of each contaminant. To facilitate data comparability and visualisation it is recommended to provide/convert measurement units in compliance with units used for law limits/EQS issued by international institutions like WHO, EC or US EPA. All MS Excel sheets, both data files and templates, must be named 'Data' (by default), otherwise an error is reported when validating the datafile vs template file.

The content of each datafile is constraint by an associated template file (Location data: template_location.xls, Time series data: template_timeseries.xls; User-defined data: e.g. template lawlimit.xls). The template file follows always the same column structure, describing the constraints for all data files: (i) the position of the column in the data file ('Pos', e.g. 1), (ii) the name of the column ('Column name', e.g. Code), (iii) the data type ('Data Type', e.g. TEXT(20) or DATE), (iv) if the provision of the info is mandatory ('Required', e.g. Y(Yes) or N(No), (v) the description of the expected column content ('Description', e.g. Monitoring point code (e.g. P01)) and (vi) the VVL, ('Valid values list', e.g. vvl_param.csv). Once a datafile is fully compiled, its content can be validated against the associated template file via EMS 'Validation & Database upload' tool.

It is recommended to entrust a trained data manager with the creation or modification of template and datafile sheets to ensure proper functionality. To guarantee this, all template sheets are password-protected. Data reporters are permitted to compile datafile content within pre-defined unlocked columns/cells.

Figure 49. MS Excel data reporting sheet (top) for standardised documentation of locational data in long format, following template constraints for locational data (bottom)



D11 4326

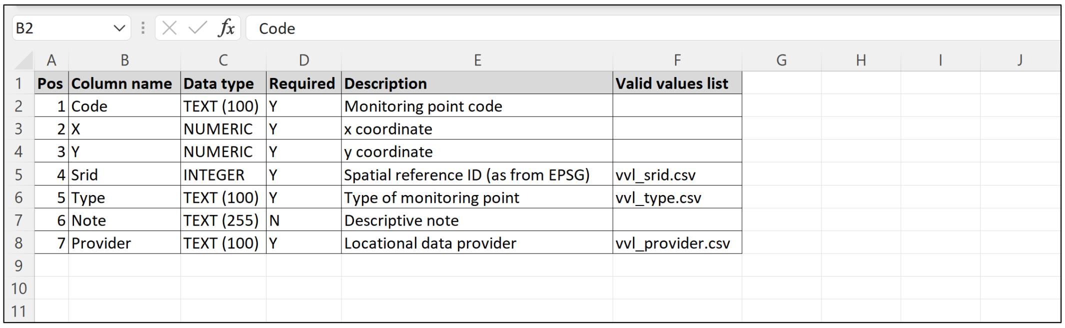
WEFE SENEGAL

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

Code	X	Y	Srid	Type	Note	Provider
BD0211	348816.000000000000000	1793264.000000000000000	32628	Monitoring point		CERES
BD0301	-16.412460100000000	16.215092100000000	4326	Monitoring point		CERES
AR	-12.239555323000000	14.761755776900000	4326	onitoring site		CERES

4326
4327
4328
4329
4330
4331
4332
4333

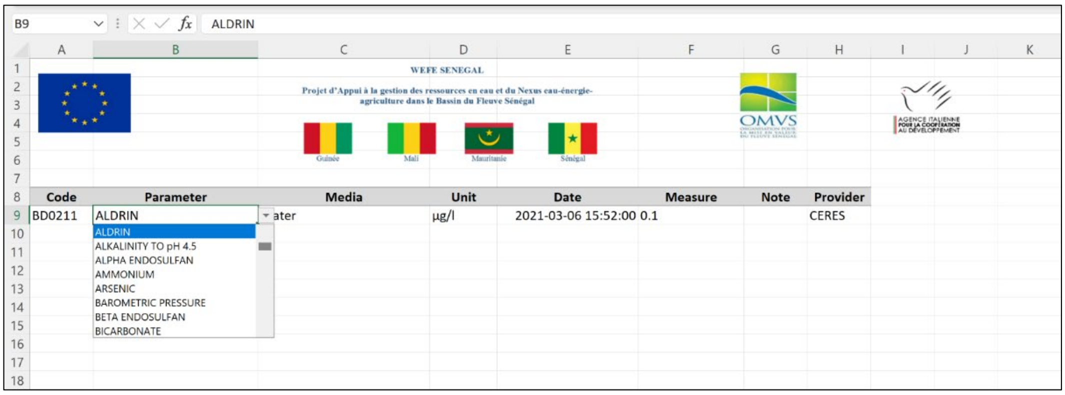
ocation
sign the correct
ference id (Srid)
ng on the used
dinates
system



B2 Code

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

Figure 50. MS Excel data reporting sheet (top) for standardised documentation of time series data (e.g. parameter/contaminants) in long format, following template constraints for time series data (bottom)



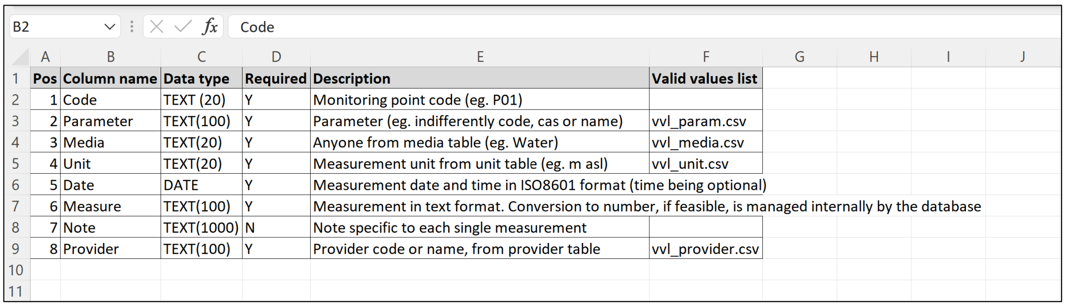
B9 ALDRIN

WEFE SENEGAL

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

Code	Parameter	Media	Unit	Date	Measure	Note	Provider
BD0211	ALDRIN	ater	µg/l	2021-03-06 15:52:00	0.1		CERES

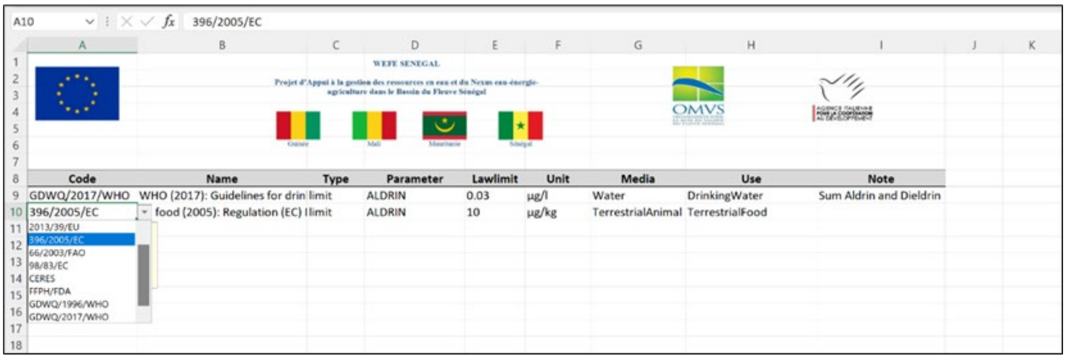
ALDRIN
ALKALINITY TO pH 4.5
ALPHA ENDOSULFAN
AMMONIUM
ARSENIC
BAROMETRIC PRESSURE
BETA ENDOSULFAN
BICARBONATE



B2 Code

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

Figure 51. MS Excel data reporting sheet (top) for standardised documentation of user-defined data (here: law limit data) in long format, following template constraints for user defined data, here law limit data (bottom)



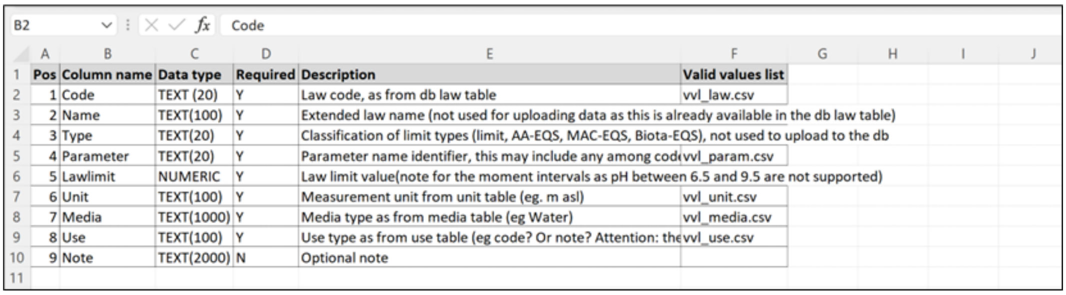
A10 396/2005/EC

WEFE SENEGAL

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

Code	Name	Type	Parameter	Lawlimit	Unit	Media	Use	Note
GDWQ/2017/WHO	WHO (2017): Guidelines for drin	limit	ALDRIN	0.03	µg/l	Water	DrinkingWater	Sum Aldrin and Dieldrin
396/2005/EC	food (2005): Regulation (EC) l	limit	ALDRIN	10	µg/kg	TerrestrialAnimal	TerrestrialFood	

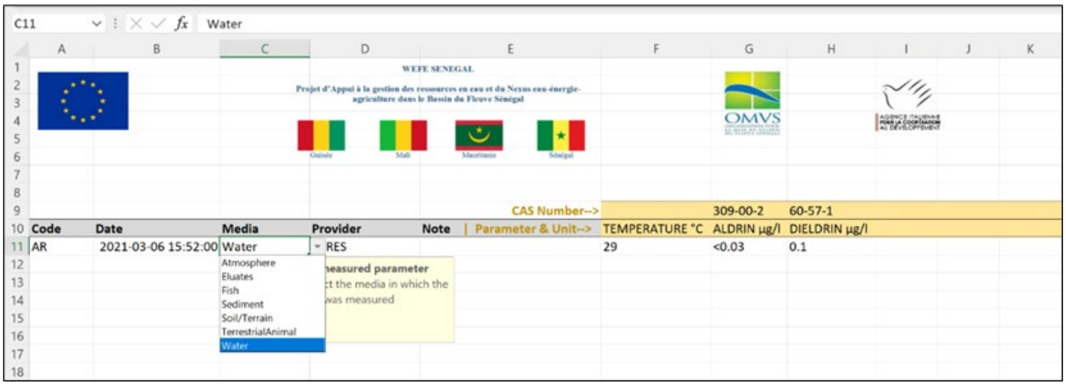
2013/39/EU
396/2005/EC
66/2003/FAO
98/83/EC
CERES
FFPH/FDA
GDWQ/1996/WHO
GDWQ/2017/WHO



B2 | Code

Pos	Column name	Data type	Required	Description	Valid values list
1	Code	TEXT (20)	Y	Law code, as from db law table	vvl_law.csv
2	Name	TEXT(100)	Y	Extended law name (not used for uploading data as this is already available in the db law table)	
3	Type	TEXT(20)	Y	Classification of limit types (limit, AA-EQS, MAC-EQS, Biota-EQS), not used to upload to the db	
4	Parameter	TEXT(20)	Y	Parameter name identifier, this may include any among cod	vvl_param.csv
5	Lawlimit	NUMERIC	Y	Law limit value(note for the moment intervals as pH between 6.5 and 9.5 are not supported)	
6	Unit	TEXT(100)	Y	Measurement unit from unit table (eg. m asl)	vvl_unit.csv
7	Media	TEXT(1000)	Y	Media type as from media table (eg Water)	vvl_media.csv
8	Use	TEXT(100)	Y	Use type as from use table (eg code? Or note? Attention: the	vvl_use.csv
9	Note	TEXT(2000)	N	Optional note	

Figure 52. Data reporting sheet (MS Excel) for standardised documentation of time series data (parameter/contaminants), in wide format (commonly used by laboratories). The transformation from wide to long format is done in Python (code wide2long.py)


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
AGENCE ITALIENNE POUR LA COOPÉRATION AU DÉVELOPPEMENT
Atmosphere
Eluates
Fish
Sediment
Soil/Terrain
TerrestrialAnimal
Water

C11 | Water

Code	Date	Media	Provider	Note	Parameter & Unit-->	TEMPERATURE °C	ALDRIN µg/l	DIELDRIN µg/l
					CAS Number-->		309-00-2	60-57-1
AR	2021-03-06 15:52:00	Water	RES			29	<0.03	0.1

easured parameter
t the media in which the
was measured

Annex 4. Dummy Desert Dataset: a groundwater flow & transport model

The EMS has been conceived to validate, manage and conduct exploratory spatio-temporal analysis of water and environmental quality data delivered in the framework of the Senegal WEFE nexus project. Specific objectives include the operationalisation of a transboundary river basin monitoring network, setup as part of the project commitments, through the involvement of key analytical laboratories from basin countries, as well as the University of Amsterdam for cross-validation purposes. However, it is worth to stress that, well beyond the original mandate (and the environmental sector), the EMS addresses more general concerns attaining at data validation, mature spatio-temporal data management and exploratory analysis. Its application in other domains is highly encouraged.

The EMS setup is freely downloadable from the Aquaknow web site at https://aquaknow.jrc.ec.europa.eu/document/ems-tool. A dummy dataset has been created and used to populate the EMS database. Senegalese quality dataset is not included in the database due to potential privacy concerns, but interested users are invited to keep in touch with contact reference as reported at the beginning of current document.

A fictitious groundwater numerical flow&transport model has been setup using Feflow v. 6.2, a state-of-the-art finite element flow & transport code from DHI-WASY (Diersch, 2014). Based on an oversimplified hydrogeological schema, the model, located in North Niger in the middle of the Sahara desert, has been used to create the datasets at observation points, capturing realistic conditions, although having nothing to do with the local ground truth. Given its origin and location, the dataset is referred hereafter as 3D for 'Dummy Desert Dataset'.

The Feflow model can also be downloaded from https://aquaknow.jrc.ec.europa.eu/document/ems-tool. The model can be used to further refine and extend the simulation scenarios in view of extending or building more complex dummy datasets (e.g. reactive and radioactive contaminants, 3D and multi-aquifer conditions, density-dependent salt water intrusion).

Hydrogeological framework

The hydrogeological system is assumed to be located in a remote arid area in North Niger (**Figure 53**), to further stress that the model (and the simulated dataset) has nothing to do with the real conditions on the ground. The hydrogeological system, elongated in E-W direction for 100 km and a width of 50 m, characterises for (**Figure 54**):

- All limits along groundwater divides, that's no flow exchange with the external groundwater bodies, except for the eastern limit where a lake is assumed to exist at constant elevation of 450 m asl.
- Phreatic porous aquifer system, characterised by very high hydraulic conductivity of 1*10-3 m/s, typical of coarse sand lithology.
- Basement at constant elevation of 400 m asl.
- Initial mean effective yearly recharge over the entire domain set to the value of 100mm/yr, unevenly distributed in 3 months (July-September), respectively of 20, 60 and 20 mm/month; for the other months, null effective recharge is assumed.
- River, elongated in W-E direction, with highest elevation of 500m asl at the Western boundary, a steeper initial section degrading up to 460m asl, at about one half of the model, and a lower section where river meandering develops.

Figure 53. Hydrogeological system: location (dark blue dot) in Western Africa.

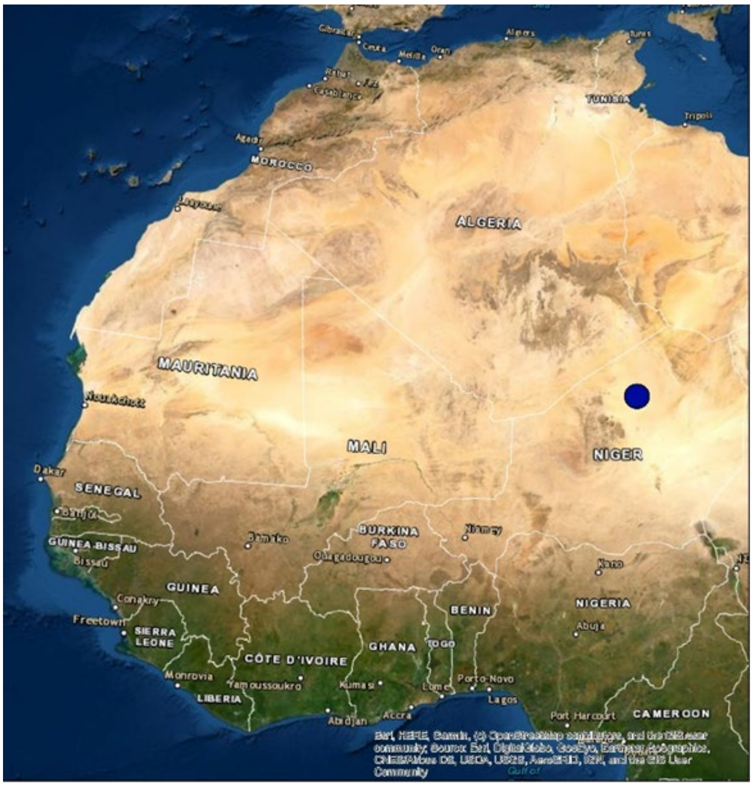
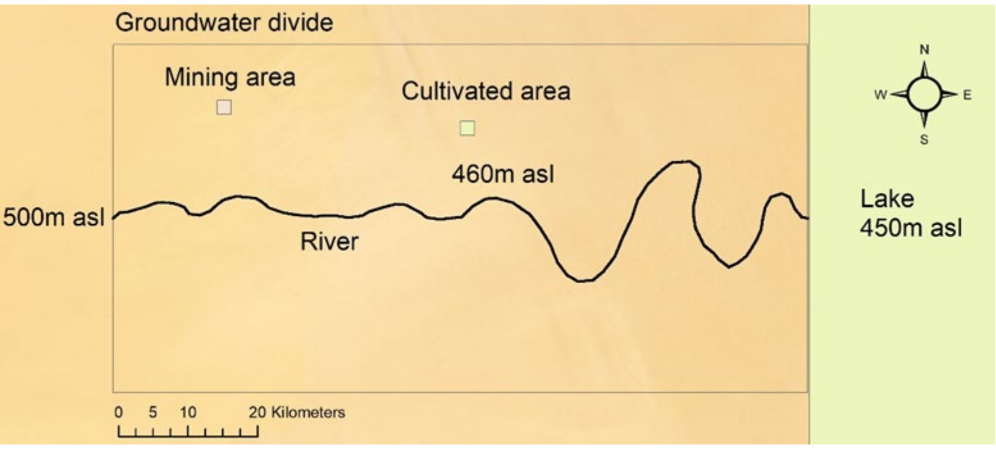


Figure 54. Aquifer system conceptual schema



Mesh discretisation design is key to minimise numerical oscillations and improve efficiency of numerical convergence process, particularly for the simulation of contaminants migration in aquifer systems.

Groundwater flow model setup

Based on the hydrogeological schema above, the numerical flow model has been setup as follows (**Figure 55**):

- A finite elements discretisation mesh has been created with a high refinement along the main stream, in the proximity of potential sites of interest (mining and cultivated areas) and downstream of contamination sources to be set at a later stage; impermeable basement elevation, aquifer hydraulic conductivity and yearly effective recharge time series have been assigned consistently with previously detailed conceptualisation.
- The eastern boundary along the lake has been set to 1st type Dirichlet BCs (Boundary Conditions) with constant head equal to the lake elevation (450 m asl).
- The river has been set to 1st type Dirichlet BCs , with elevation lowering from W to E, consistently with the presented profile.

Based on an initial constant piezometric head of 450 m asl over the entire modelling domain, a long term simulation has been run to produce a pseudo steady-state reference piezometry (**Figure 56**).

Figure 55. Feflow flow model setup

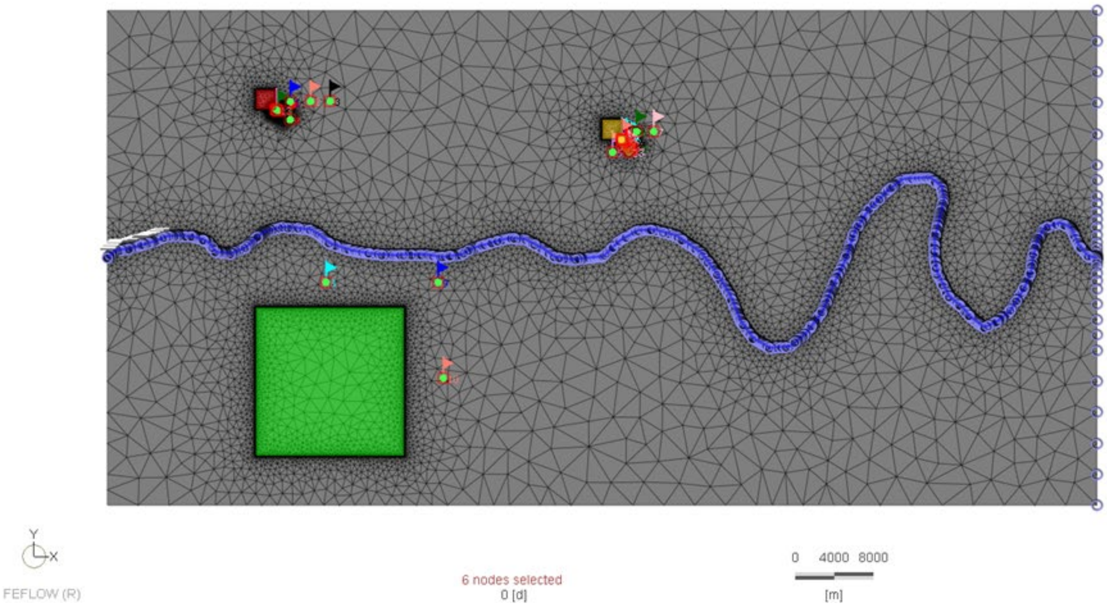
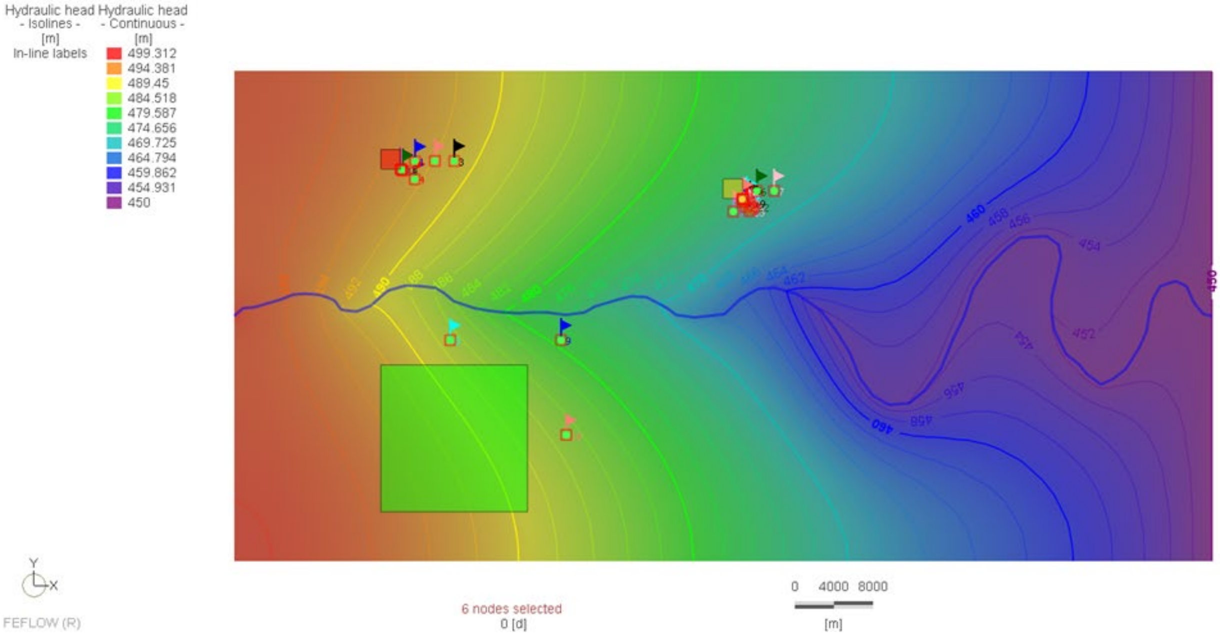


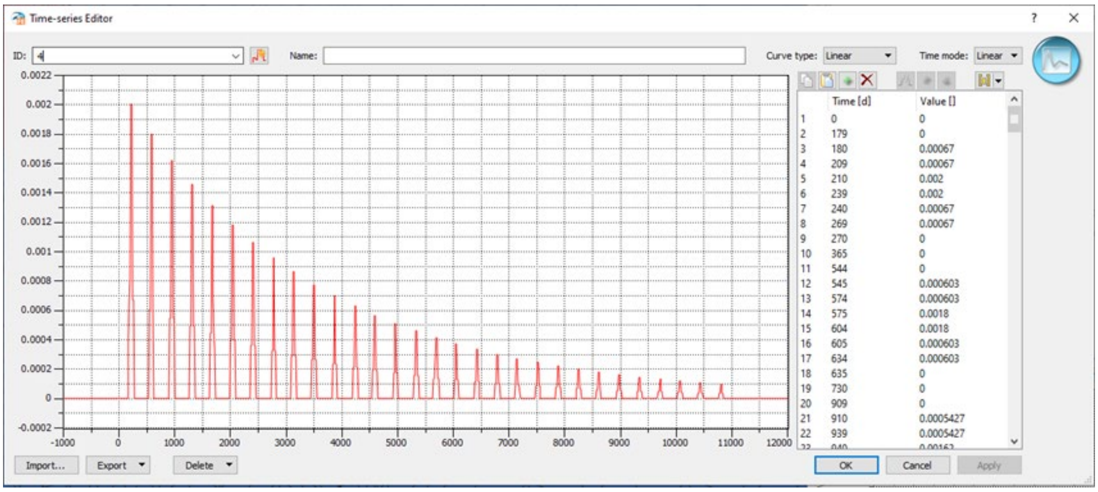
Figure 56. Steady state simulation: piezometric head distribution



Groundwater transport model setup

Based on the initial reference pseudo steady-state flow conditions, a 30 years transient flow & transport simulation has been set, assuming a contamination taking place at the SE limits of the mining and the agriculture sites. A uniformly distributed recharge over the modelling domain is set, concentrated in the three months July-September as previously detailed, gradually decreasing of 10% each year to simulate increasingly dry conditions (**Figure 57**). The effect of declining recharge is evident in the piezometric heads lowering trends simulated over the entire modelling domain (**Figure 58**), although different depending upon the distance from the main river, acting as a major draining feature. Effects of pumping start are also clear in the mid of the simulation (15 years), when pumping at wells is activated.

Figure 57. Effective daily recharge (m/day) over the 30years



ID: 4

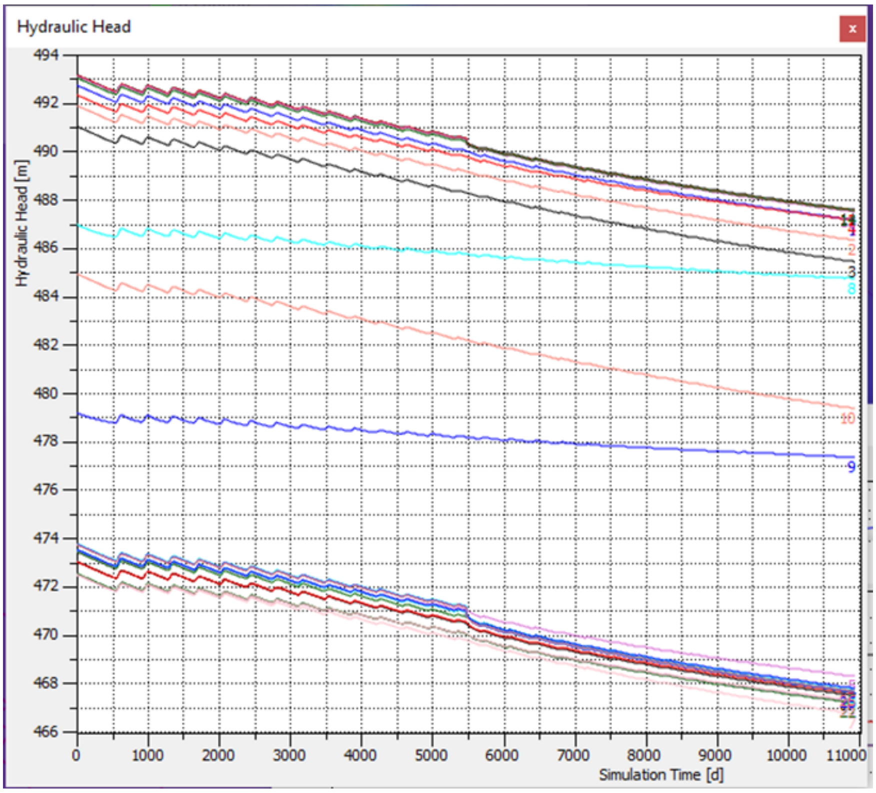
Name:

Curve type: Linear

Time mode: Linear

	Time [d]	Value []
1	0	0
2	179	0
3	180	0.00067
4	209	0.00067
5	210	0.002
6	239	0.002
7	240	0.00067
8	269	0.00067
9	270	0
10	365	0
11	544	0
12	545	0.000603
13	574	0.000603
14	575	0.0018
15	604	0.0018
16	605	0.000603
17	634	0.000603
18	635	0
19	730	0
20	909	0
21	910	0.0005427
22	939	0.0005427
23	940	0.00162

Figure 58. Simulated piezometric trends over time



The model downstream of the mining site, in the NW of the study area, characterises for:

- A highly refined discretisation mesh, to reduce numerical oscillations and improve convergence process, around and downstream of the contaminant source in the mining site (**Figure 59**).
- Contamination from Selenium and Cadmium (Se and Cd) occurring at the SE boundary of the mining site, the two contaminants sensibly differing for their sorption coefficients, resulting in quicker

migration of Se compared to Cd; Se and Cd concentrations are set to 400 mg/l and 10 mg/l respectively, constant over the entire simulation assuming no contaminant source removal; initial concentration for both contaminants over the entire domain is 0 mg/l.

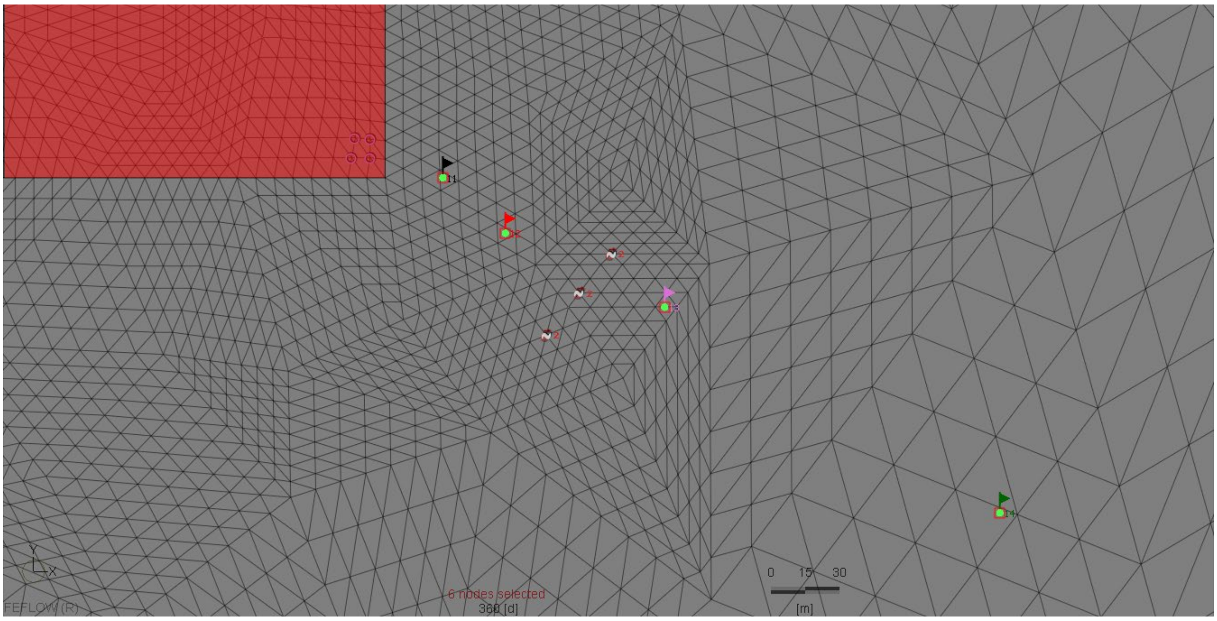
- Three wells pumping barrier, located at about 100m downstream of the contamination source and transversal to the contamination plume; aimed at capturing the contaminated flow downstream of the source of contamination, the barrier is activated after 15 years, with a withdraw of 1000 m^3/day/well (equivalent to about 11.6 l/s/well).

The simulation conceptually captures a scenario where a long time and still ongoing contamination gave rise to a multi-contaminant plume and a pumping barrier is activated to capture the contaminants. The concentration maps of Se (**Figure 60**) and Cd (**Figure 61**) approximately at 1 (360 days), 10 (3660 days) and 30 years, clearly highlight:

- The higher mobility of Se compared to Cd, resulting in a more elongated contamination plume of the former.
- The capture effect of the pumping barrier, in the last simulation step, particularly for Se where residual contamination downstream of the barrier continues to migrate towards the SE.

The piezometric low at wells is quite evident and the extended upstream capture zone can be intuitively inferred (**Figure 62**).

Figure 59. Discretisation mesh refinement downstream of mining site contamination source (three pumping wells barrier and four observation points)



6 nodes selected

Figure 60. Se contamination at about 1, 10 and 30 years with pumping wells barrier activated at 15 years

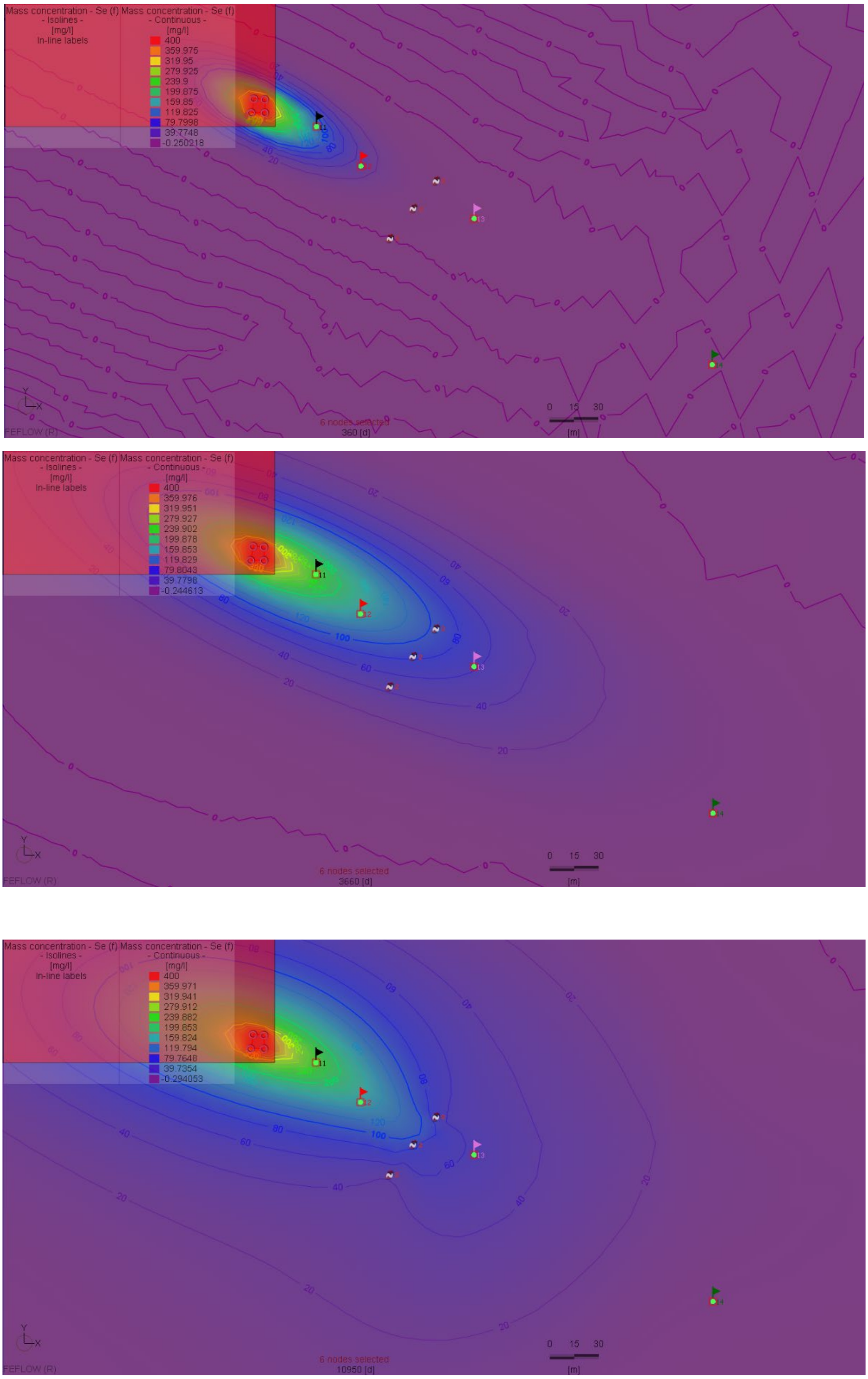


Figure 61. Cd contamination at about 1, 10 and 30 years with pumping wells barrier activated at 15 years

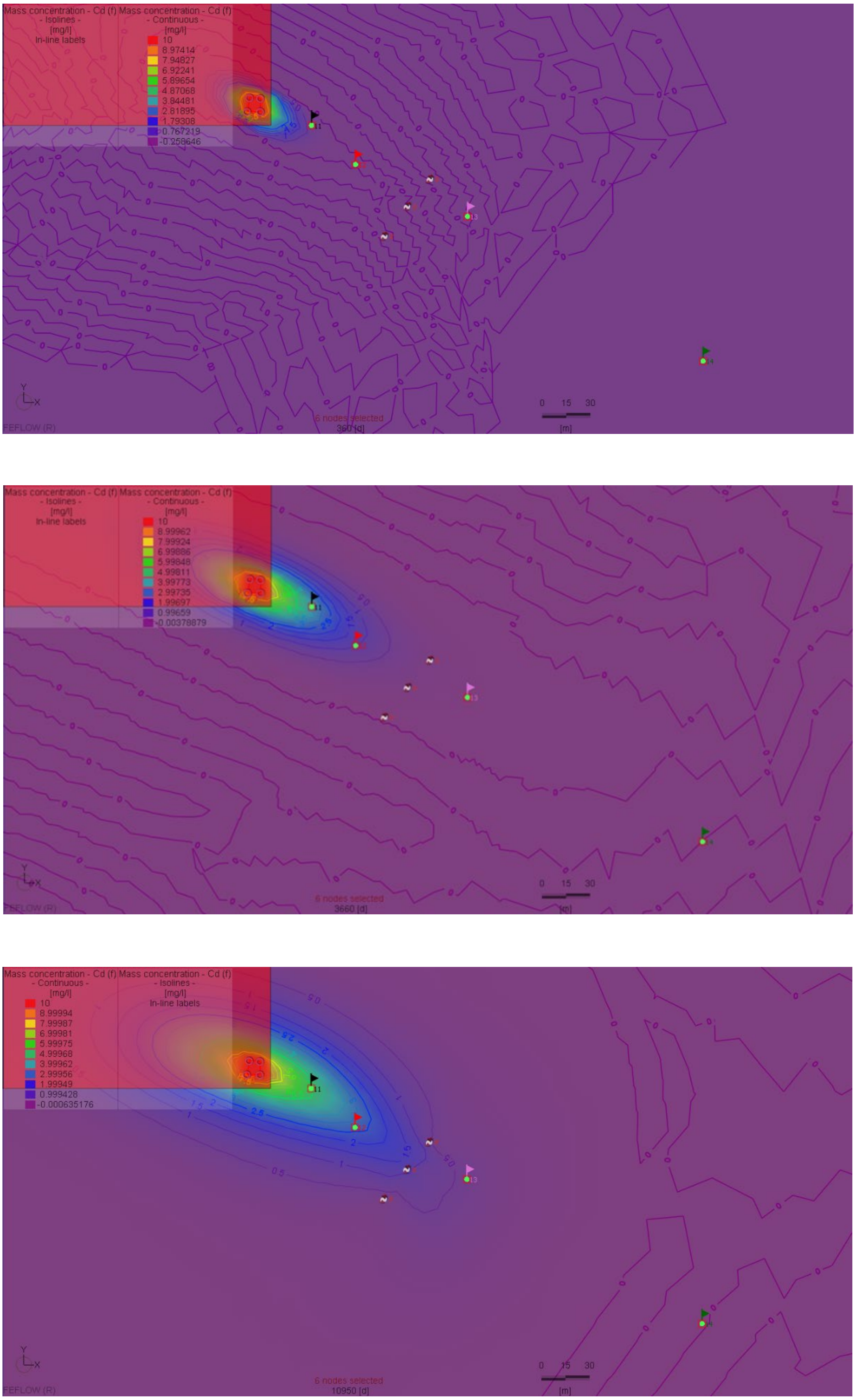
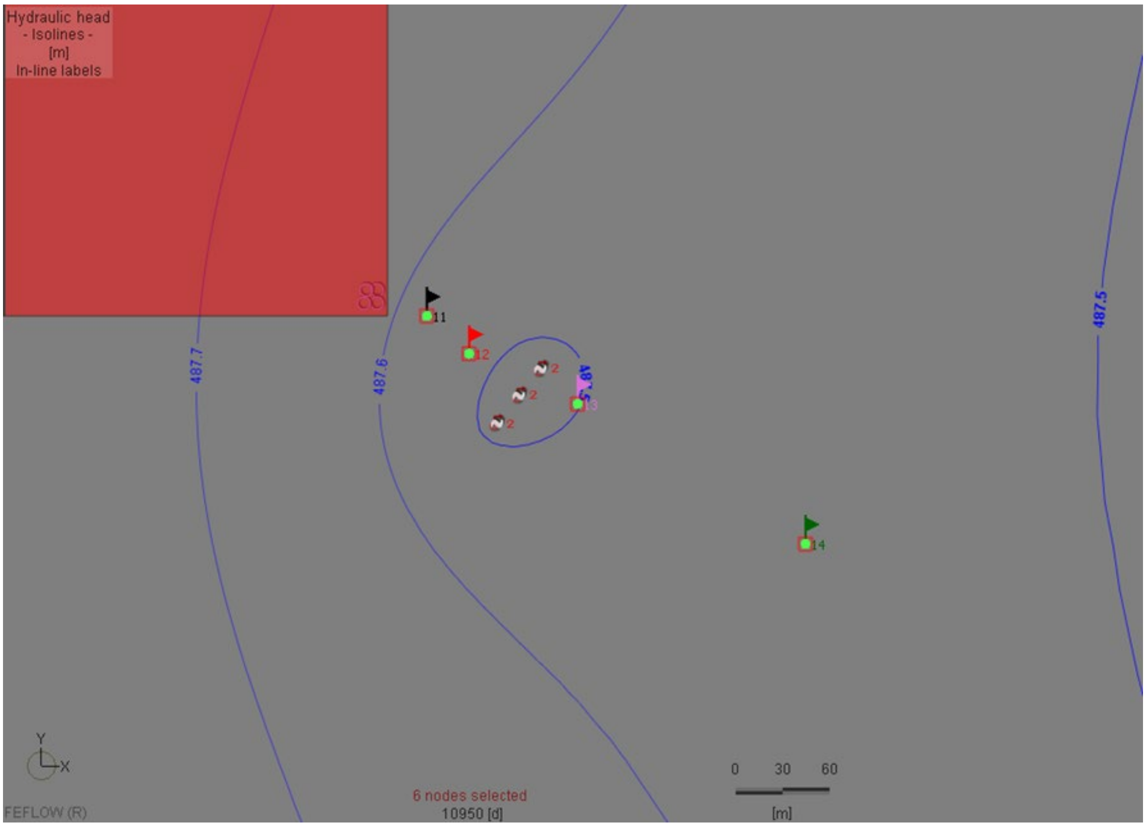


Figure 62. Piezometric head at 30 years, at the end of the simulation downstream of the mining site



The model downstream of the agriculture site, in the central northern part of the study area, characterises for:

- A highly refined discretisation mesh, to reduce numerical oscillations and improve convergence process, around and downstream of the contaminant source (**Figure 63**).
- Contamination from Aldrin occurring at the SE boundary of the agriculture site, concentration being set to 1 mg/l, constant over the entire simulation assuming no contaminant source removal; initial concentration for Aldrin over the entire domain is set to 0 mg/l.
- Six wells pumping site, located SSW of the contamination source and aside of the expected plume; the barrier is activated after 15 years, with a withdraw of 1000 m^3/day/well (equivalent to about 11.6 l/s/well).

The simulation conceptually captures a scenario where a long time and still ongoing contamination gave rise to a plume, while the activation of pumping at an aside well field (e.g. for water supply) gives rise to an undesired effect of impacting on contaminants migration. The concentration maps of Aldrin (**Figure 64**) approximately at about 1 (360 days), 10 (3660 days) and 30 years, clearly highlight:

- The long contaminant plume arising from the source, clearly evident after 10 years.
- The (undesired) capture effect of the pumping at well field, in the last simulation step, as the piezometry clearly highlights (**Figure 65**); residual contamination plume clearly extends towards the SE, being not captured by the wells.

Figure 63. Discretisation mesh refinement downstream of agriculture site contamination source (six pumping wells for water supply aside of the expected contaminant plume and observation points)

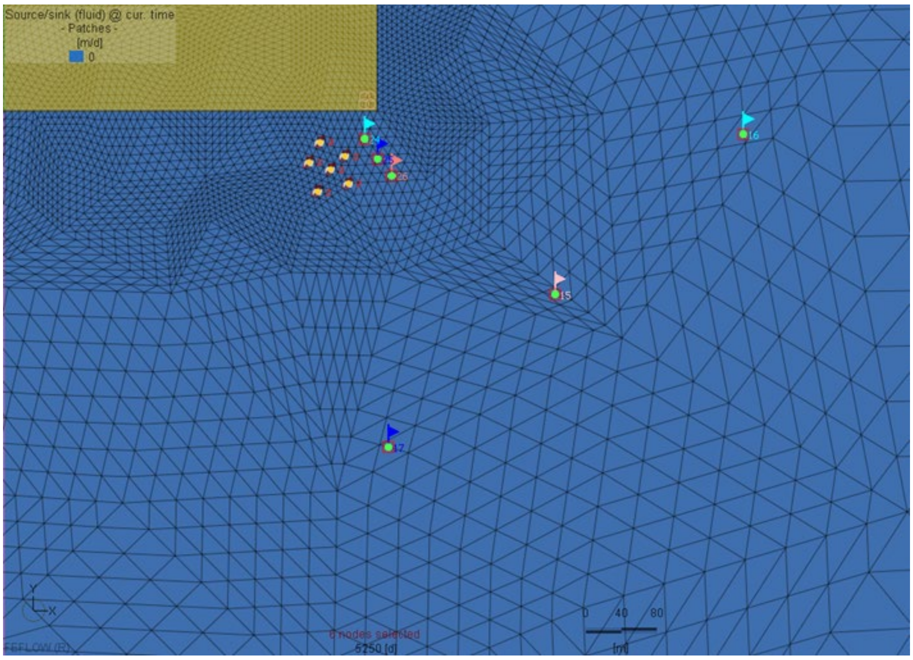
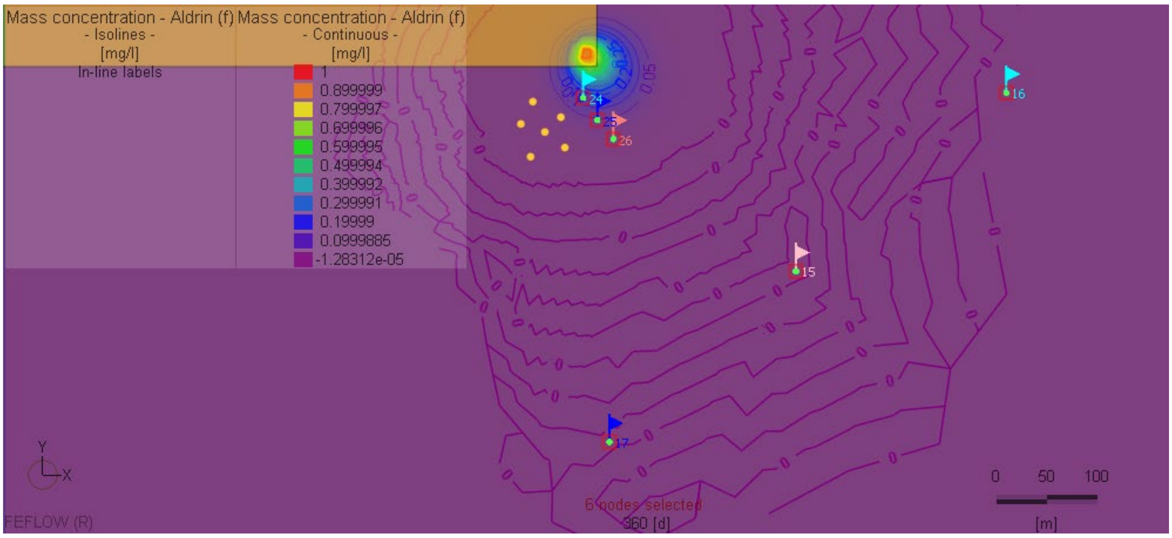
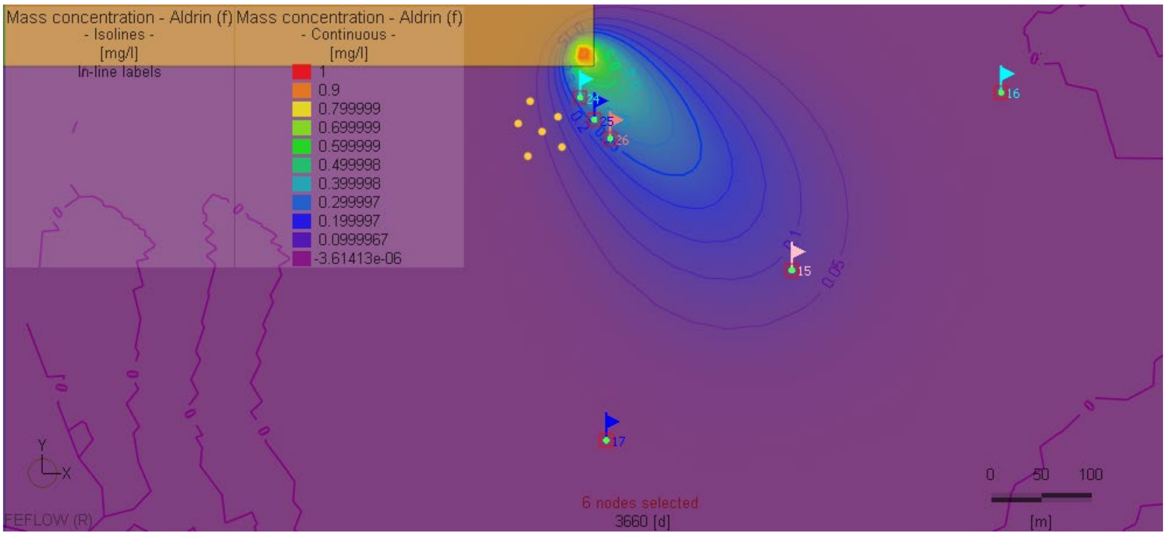


Figure 64. Aldrin contamination at about 1, 10 and 30 years with pumping wells barrier activated at 15 years




Mass concentration - Aldrin (f)
- Isolines -
[mg/l]
In-line labels
Mass concentration - Aldrin (f)
- Continuous -
[mg/l]
🟥 1
🟧 0.9
🟨 0.799999
🟩 0.699999
🟩 0.599999
🟩 0.499998
🟦 0.399998
🟦 0.299997
🟦 0.199997
🟪 0.0999967
🟪 -3.61413e-06
24
25
26
16
15
17
0.1
0.05
0
Y
X
FEFLOW (R)
6 nodes selected
3660 [d]
0 50 100
[m]

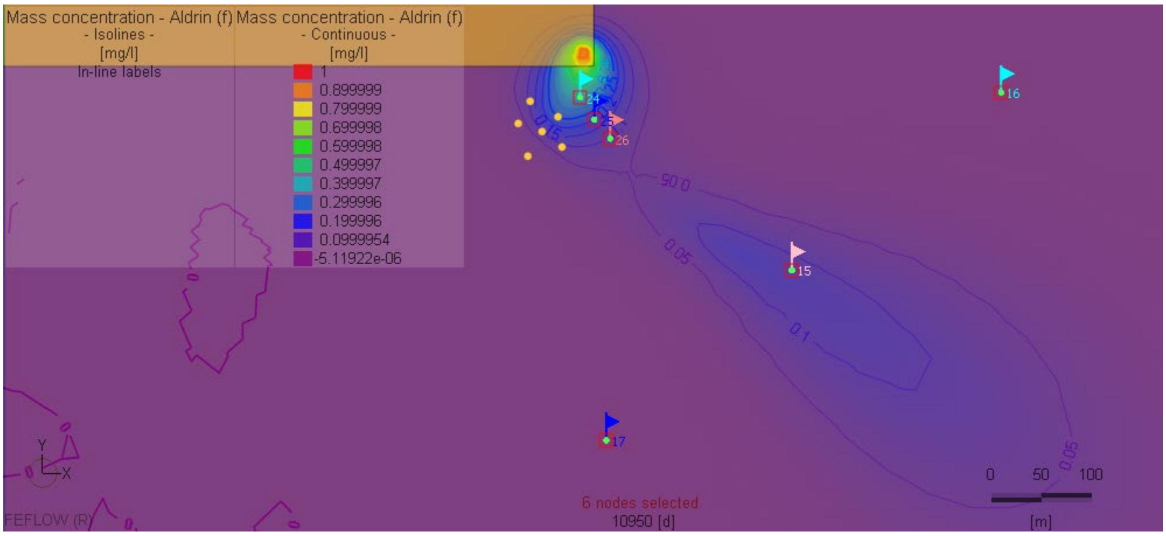
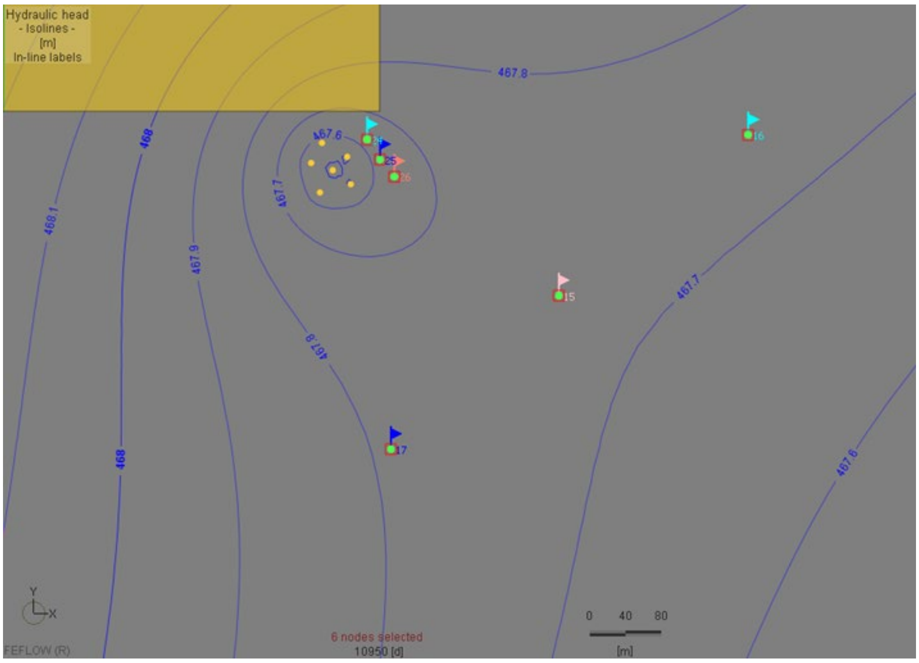

Mass concentration - Aldrin (f)
- Isolines -
[mg/l]
In-line labels
Mass concentration - Aldrin (f)
- Continuous -
[mg/l]
🟥 1
🟧 0.899999
🟨 0.799999
🟩 0.699998
🟩 0.599998
🟩 0.499997
🟦 0.399997
🟦 0.299996
🟦 0.199996
🟪 0.0999954
🟪 -5.11922e-06
24
25
26
16
15
17
0.05
0.1
0
Y
X
FEFLOW (R)
6 nodes selected
10950 [d]
0 50 100
[m]

Figure 65. Piezometric head at 30 years, at the end of the simulation downstream of agriculture site



Dummy Desert Dataset quick description

Based on the above transient flow and transport simulation, relevant data at observation points have been exported to MS Excel files, as if they would be observed ground data. Computed contaminant concentrations below a reasonable detection limit has been substituted with the classical below detection limit (e.g. '< 0.01'). Data have then been formatted for validation and import to EMS tool, following the standard template formats (**Annex 3**).

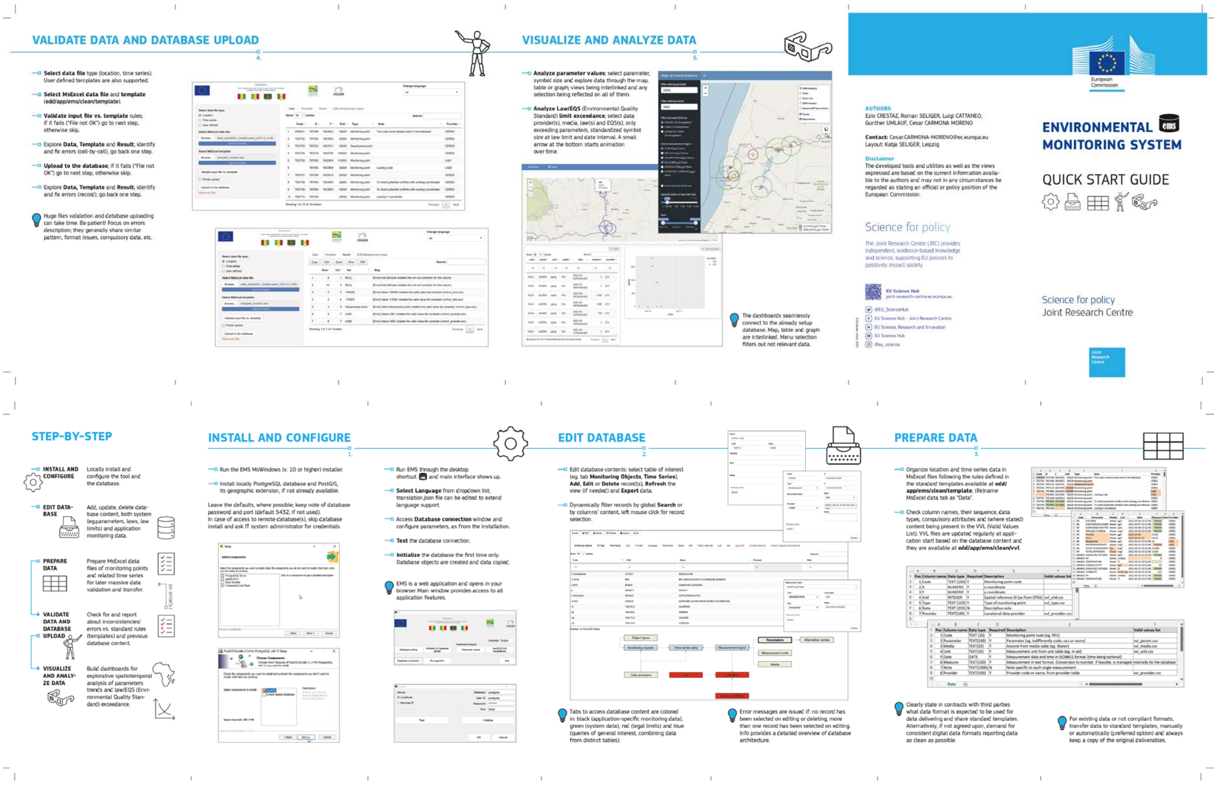
The Dummy Desert Dataset, in line with previous detailed descriptions, include:

- Piezometric heads all over the modelling domain, capturing the seasonal variation of effective recharge as concentrated in July-September, the long term depressurisation due to the effective recharge reduction, and, last but not least, the activation of pumping at 15 years after the simulation start. The entity of piezometric depressurisation also changes over space, depending upon distance from pumping wells and the draining river that acts as a major reference elevation constrain. Piezometric heads south of the river are of course not impacted by pumping.
- Se (Selenium) and Cd (Cadmium) concentrations over time, downstream of the mining site, capturing the evolution of the plumes as reflecting different contaminants sorption coefficients, and the effects of the wells pumping barrier. The latter, located about 100 m downstream of the contamination source and activated 15 years from the contamination start, can be regarded as a barrier aimed at capturing flow for groundwater quality treatment. Part of the Se plume, the most mobile of the two contaminants, is not captured downstream of the wells barrier, falling outside of the capture zone as evident from the piezometry.
- Aldrin concentration over time, downstream of the agriculture site. Similarly to the above, contaminant plume extends downstream and effects of pumping, activated 15 years after the contamination source start, result in a strong local distortion of the groundwater flow field. The contamination plume is largely deviated towards the wells, provided that a large plume continues to migrate downstream. The simulation refers here to a pretty different conceptual scheme, where a water supply well field, located aside of an existing contamination plume, is unexpectedly impacted by the contamination.

Annex 5. Leaflet ‘Environmental Monitoring System: Quick Start Guide’

A leaflet has been designed as a quick start guide to demonstrate the use and main features of the EMS tool. The leaflet file (pdf) of 600dpi resolution can be downloaded through the dedicated page ‘EMS-Tool’ on Aquaknow KMS (https://aquaknow.jrc.ec.europa.eu/document/ems-tool). If printed, the use of the following print parameters is recommended: Format: Flyer DIN long (closed format: 105 x 210mm, open format: 705 x 210mm), Extent: 14 pages, Print: 4/4-colour scale, Paper: 135g/m² picture print matt, Processing: wrap folding.

Figure 66. Leaflet ‘Environmental Monitoring System: Quick Start Guide’, unfolded: Front view (top) and rear view (bottom)



VALIDATE DATA AND DATABASE UPLOAD

4.

- **Select data file** type (location, time series). User defined templates are also supported.
- **Select MsExcel data file** and **template** (edd/app/ems/clean/template).
- **Validate input file vs. template** rules; if it fails ("File not OK") go to next step, otherwise skip.
- Explore **Data, Template** and **Result**; identify and fix errors (cell-by-cell), go back one step.
- **Upload to the database**; if it fails ("File not OK") go to next step, otherwise skip.
- Explore **Data, Template** and **Result**; identify and fix errors (record), go back one step.

Huge files validation and database uploading can take time. Be patient! Focus on errors description; they generally share similar pattern, format issues, compulsory data, etc.

VISUALIZE AND ANALYZE DATA

5.

- **Analyze parameter values**; select parameter, symbol size and explore data through the map, table or graph; views being interlinked and any selection being reflected on all of them.
- **Analyze Law/EQS** (Environmental Quality Standard) **limit exceedance**; select data provider(s), media, law(s) and EQS(s), only exceeding parameters, standardized symbol size at law limit and date interval. A small arrow at the bottom starts animation over time.

The dashboards seamlessly connect to the already setup database. Map, table and graph are interlinked. Menu selection filters out not relevant data.

AUTHORS
Ezio CRESTAZ, Roman SELIGER, Luigi CATTANEO, Gunther UMLAUF, Cesar CARMONA MORENO

Contact: Cesar.CARMONA-MORENO@ec.europa.eu
Layout: Katja SELIGER, Leipzig

Disclaimer
The developed tools and utilities as well as the views expressed are based on the current information available to the authors and may not in any circumstances be regarded as stating an official or policy position of the European Commission.

Science for policy

The Joint Research Centre (JRC) provides independent, evidence-based knowledge and science, supporting EU policies to positively impact society

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

@EU_ScienceHub
EU Science Hub – Joint Research Centre
EU Science, Research and Innovation
EU Science Hub
@eu_science

© European Union, 2023

European Commission

ENVIRONMENTAL MONITORING SYSTEM

QUICK START GUIDE

Science for policy
Joint Research Centre

Joint Research Centre

STEP-BY-STEP

- **INSTALL AND CONFIGURE** Locally install and configure the tool and the database.
- **EDIT DATABASE** Add, update, delete database content, both system (eg.parameters, laws, law limits) and application monitoring data.
- **PREPARE DATA** Prepare MsExcel data files of monitoring points and related time series for later massive data validation and transfer.
- **VALIDATE DATA AND DATABASE UPLOAD** Check for and report about inconsistencies/errors vs. standard rules (templates) and previous database content.
- **VISUALIZE AND ANALYZE DATA** Build dashboards for explorative spatiotemporal analysis of parameters trends and law/EQS (Environmental Quality Standard) exceedance.

on error(s)

INSTALL AND CONFIGURE

1.

- Run the EMS MsWindows (v. 10 or higher) installer.
- Install locally PostgreSQL database and PostGIS, its geographic extension, if not already available.

Leave the defaults, where possible; keep note of database password and port (default 5432, if not used).
In case of access to remote database(s), skip database install and ask IT system administrator for credentials.

- Run EMS through the desktop shortcut and main interface shows up.
- **Select Language** from dropdown list; translation.json file can be edited to extend language support.
- Access **Database connection** window and configure parameters, as from the installation.
- **Test** the database connection.
- **Initialize** the database the first time only. Database objects are created and data copied.

EMS is a web application and opens in your browser. Main window provides access to all application features.

EDIT DATABASE

2.

- Edit database contents: select table of interest (eg. tab **Monitoring Objects**, **Time Series**), **Add**, **Edit** or **Delete** record(s), **Refresh** the view (if needed) and **Export** data.
- Dynamically filter records by global **Search** or by columns' content, left mouse click for record selection.

Tabs to access database content are colored in: black (application-specific monitoring data), green (system data), red (legal limits) and blue (queries of general interest, combining data from distinct tables).

Error messages are issued if: no record has been selected on editing or deleting; more than one record has been selected on editing. Info provides a detailed overview of database architecture.

PREPARE DATA

3.

- Organize location and time series data in MsExcel files following the rules defined in the standard templates available at **edd/app/ems/clean/template**; (Re)name MsExcel data tab as "Data".
- Check column names, their sequence, data types, compulsory attributes and (where stated) content being present in the VVL (Valid Values List). VVL files are updated regularly at application start based on the database content and they are available at **edd/app/ems/clean/vvl**.

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

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

Clearly state in contracts with third parties what data format is expected to be used for data delivering and share standard templates. Alternatively, if not agreed upon, demand for consistent digital data formats reporting data as clean as possible.

For existing data or not compliant formats, transfer data to standard templates, manually or automatically (preferred option) and always keep a copy of the original deliverables.

Science for policy

The Joint Research Centre (JRC) provides independent, evidence-based knowledge and science, supporting EU policies to positively impact society

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

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

@eu_science