EEA/PAN/17/006

European Environment Agency
European Topic Centre on Inland, Coastal and Marine Waters

Support for the implementation of the ENI-SEIS East 2017-2018 regional and country work plans in the thematic area of water

Concept of Implementation of Water Information System (WIS)

Version: 0.9
Date: 30 January 2018
EEA activity:
ETC/ICM task, milestone:

Prepared by / compiled by:
Lidija Globevnik, TC Vode
Gašper Šubelj, TC Vode
Miroslav Fanta, CENIA

Organisation:

EEA project manager: Nihat Zal

Contents

1. Introduction

Among other, the project "Support for the implementation of the ENI-SEIS East 2017-2018 regional and country work plans in the thematic area of water" aims to:

- implement standard data dictionaries of State of Environmental Reporting of the Water Information System of Europe (WISE-SoE) with the aim of harmonizing water quantity and quality data in line with the EU Water Framework Directive (WFD);
- improve the IT capacities for developing the national portal for the water data and information.

The overall objective of the project is to support the SEIS implementation at the country and regional level. The development of the WIS will therefore follow the seven SEIS principles, namely that information should be:

- Managed as close as possible to its source;
- Collected once and shared with others for many purposes;
- Readily available to easily fulfil reporting obligations;
- Easily accessible to all users;
- Accessible to enable comparisons at the appropriate geographical scale and the participation of citizens;
- Fully available to the general public and at national level in the relevant national language(s);
- Supported through common, free, open software standards.

Any rules should serve to reporters, database managers and final users of the data, not to make their work and using of the database more complicated and more difficult.

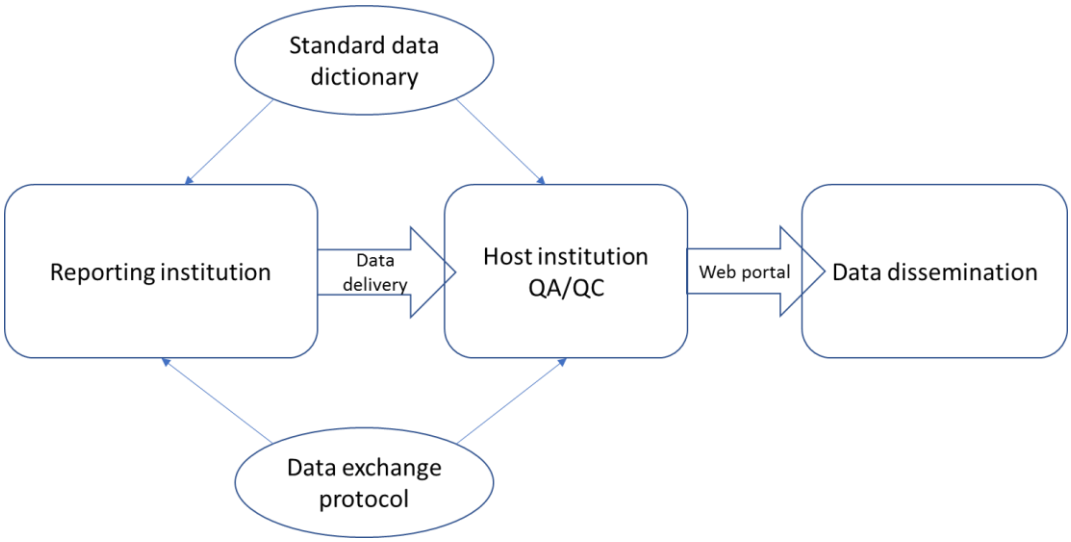
1.1. Glossary of terms

Content management system (CMS)
(Data) harvesting
Front end
Back end
Server

2. Elements of WIS

The following elements of WIS have been identified:
Content: the underlying datasets (raw and derived) as well as the output/dissemination products
Cooperation: work between the partners that makes the data flowing and products available
Infrastructure: technical part of the WIS – hardware and software (computer, database, files, web portal platform)

Figure 2.1 Simplified model of water information system



2.1. Institutions and roles

The following section introduces the roles of the reporting and the host institution. Their successful cooperation leads to data availability and further processing.

2.1.1. Reporting institution

The reporting institution is the entry point in the chain of the data flow. Its main role is to collect the data from the field or corresponding services that collect the field data, compile it to the standard data structure and deliver it to the hosting institution through an agreed transfer technology and in agreed time – typically interval. The reporting institution is also the closest to the source of data and therefore it should be available for correspondence to clarify the data content in later steps of data flow when ambiguities in data content or structure are to be detected.

The reporting institution does not need to have any dedicated IT infrastructure for the purpose of the data flow and involvement in the WIS. Instead, it is under responsibility of the hosting institution to provide straightforward tools of delivery that suit the frequency of reporting and the number of reporting institutions.

Regarding content, the reporting institution should have a thorough knowledge of the data, with a focus on sampling and analysis methodology (e.g. LOQ values used in respective monitoring methods), environmental standards etc. In the field of the data structure, the reporting institution should understand the Data Dictionary used in data reporting.

The reporting institution fulfils the following SEIS principles for information:

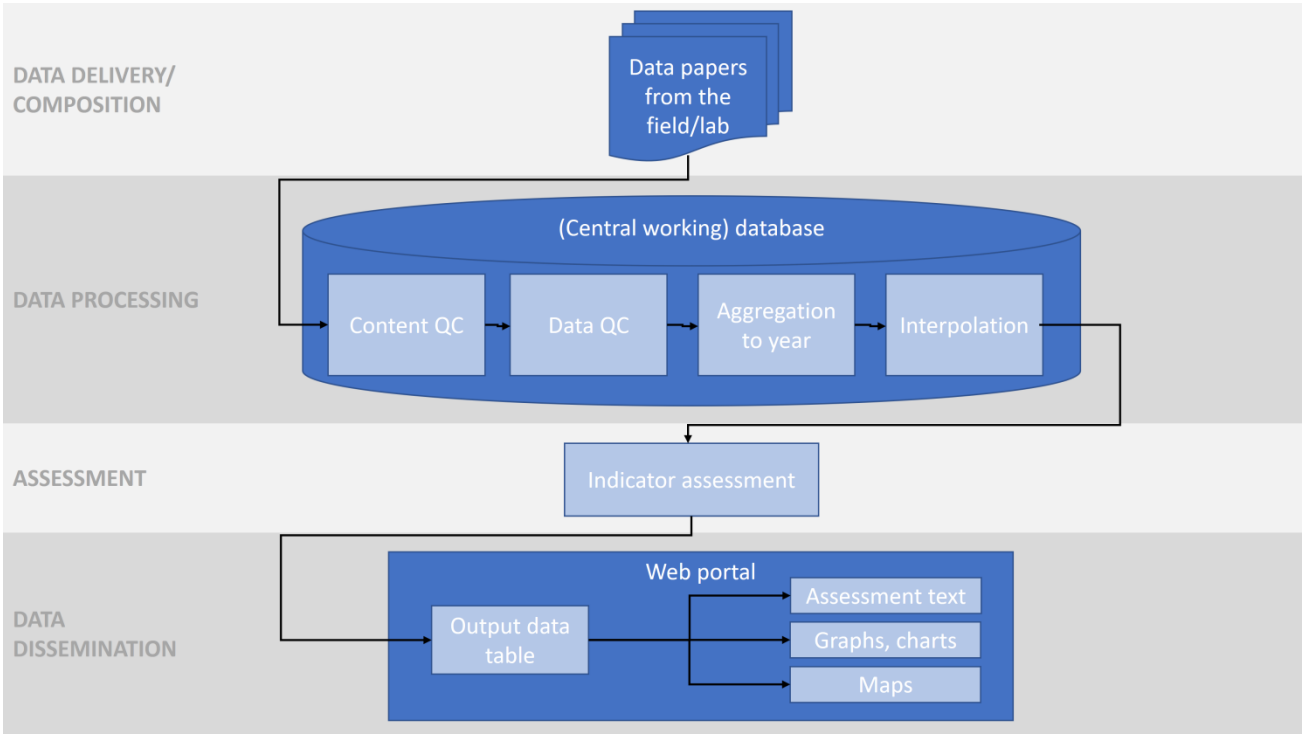
- managed as close as possible to its source;
- readily available to easily fulfil reporting obligations.

2.1.2. Host institution

The host institution collects the data after the delivery point of dataflow and processes it thereafter, being responsible for the majority of the data flow leading to data publication in WIS and dissemination. Its role is both content- and IT-oriented, and it covers the majority of WIS elements' implementation and subsequent maintenance. The host institution should have the expertise specialised in data aggregation and comparison of environmental data (i.e. assessment) as well as the expertise on data centralisation, processing and dissemination.

In producing and publishing the indicator, the reporting and host institutions should cooperate in running the dataflow as presented in Figure 2.2.

Figure 2.2 Dataflow for indicator production



2.2. Data reporting

2.2.1. Data exchange protocol

[This will not be part of the IT group training.]

2.2.2. Standard data dictionary

Data Dictionary is the basic definition and description of requested data. It should contain:

- general and specific instructions;
- data model – table relations and structure;
- data model – table fields;
- enumeration lists (codelist, vocabularies) for predefined values.

General and specific instructions

Specific instructions relevant for given reporting should be provided to reporters:

- Instructions for data analysis (resp. specification of analytical method(s) – e.g. for water quality data)
- Instructions for data sampling (processing and reporting of values below LOQ – if relevant)
- Instructions for data aggregation (important and complex issue especially if data below LOQ should be considered)
- Rules for basic level of data checking and validation before reporting (e.g. specification of low and high threshold values)

Data model – table relations and structure

- Data model (incl. graphic schema), list of tables, relations between tables
- Structure of tables, definition, instructions for filling tables, methodology and format of all fields
- Specification of spatial object to which reported data are related
- Specification of mandatory fields in each table
- Unique combination of fields in each table

Data model – table fields

- Specification of the reported determinands
- Analysed matrix, if relevant (water quality – total / filtered samples / fraction of sediments / biota / ...)
- Reported values and units
- Statistic values relevant for reported record in the case of aggregated data
- Method used for data sampling / analysis / pre-processing (if relevant)
- Validity of the reported record, remarks, ...

Enumeration lists

- Enumeration lists (= codelists, vocabularies) if the data element allowed values are already pre-defined, including additional explanations where appropriate

Reporting template

Template for data in formats suitable and reasonable for data filling and data delivery (xlsx, xml, mdb)

Additional information and recommendations

- Contact to database manager = support for data reporters
- List of significant changes since the last release
- Export of codelists and vocabularies in user friendly tabular format
- Documentation, user's guide
- Analogous data model and data structure should be implemented fo reporting of analogous data
- Parameters and ennumeration lists should be harmonised across categories of reported data, if possible and if reasonable
- Using of international identification of parameters is recommended (e.g. specification of chemical substances – using of international CAS code of chemical substance instead of substance name)
- Specification of rules for identifiers of spatial objects (using of ASCII characters only, implementation of hierarchy of the objects into their identifierscan be considered)
- It is recommended to distinguish reporting of spatial objects (= lists of monitoring sites, water bodies, ...) from reporting of content data (time series of concentrations of chemical substances, of groundwater level, of water flow, ...)

2.2.3. Data delivery

The aim of data reporting is data delivery to the host institution. The prerequisites are the agreed time frame of each reporting and frequency of regular continuous reporting, the reporting infrastructure and systematic support to reporters (experts of the reporting institution).

Data request and reporting deadlines

Data request is the official document of the host institution announcing the start (or defining the period) of the upcoming reporting. In addition, it contains general instructions for delivery of data, policy framework – legal grounds of request, description of data policy and links to relevant documents.

The data request is a written document delivered to the reporting institution in either hardcopy or digital format – via email or the official website of the host institution. In the case of the EEA, the data request and all subsequent data handling is in line with the EEA Data Policy.

An example of the data request is the annually published WISE SoE Data Request.

Reporting infrastructure

The reporting infrastructure is a system of hardware and software that technically enables the delivery of reported data from the reporting to the host institution. A well-structured reporting system should satisfy the following SEIS principles:

- collected once and shared with others for many purposes;
- easily accessible to all users;
- accessible to enable comparisons at the appropriate geographical scale and the participation of citizens.

While data delivery can be technically done using email, the above principles mean that the reporting infrastructure should ideally serve as the repository of raw input data that can be accessed at any time in the future by different authorised parties.

The level of complexity of the reporting infrastructure depends on the needs of the institutions. In the case of the EEA, many reporting institutions are involved in numerous data flows, which requires a complex system of file delivery, preliminary automatic quality assurance, versioning of delivered files and data, administration of user rights of many users on both the reporting and harvesting side. The EEA has implemented these needs with the EIONET Reportnet infrastructure, operational since 2002 and in use today.

For initial reporting infrastructure, the basic recommendation is to have the ability of file upload and storage of uploaded files also after the host institution has harvested them.

Technical requirements:

- hardware: server-capable computer;
- software: online platform with the file upload ability; the majority of Content Management Systems (CMS) have such functionality, however the reporting infrastructure would ideally be integrated into the technical implementation of WIS, which is also based on its own CMS.

Helpdesk support

In order to clear out inevitable ambiguities in reported data, the host institution should provide a straightforward mean of correspondence with reporters – content and data management experts.

While email communication is widely used, the issue tracking system is more advanced way of helpdesk correspondence, serving as a centralised archive of issues and correspondence at the same time. A ticket element, within an issue tracking system, is a running report on a particular problem, its status, and other relevant data. They are commonly created in a helpdesk environment and usually have a unique reference number, which is used to allow the user or help staff to quickly locate, add to or communicate the status of the user's issue or request.

The EEA as the host institution offers helpdesk support via OTRS, a free and open-source ticketing system. On the reporters' side, it merely means sending an email to a central email address through which the OTRS ticket is open. On the host side, it is a web-based portal with lists of tickets, indicating their status, showing all replies in specific correspondence, archiving attachments, automatically or manually prioritising open tickets etc. More helpdesk staff can operate the system as well as the same issue ticket.

Technical requirements:

- hardware: server-capable computer;
- software: issue tracking system platform; more are available for free and as open-source (e.g. ngDesk, OTRS).

2.3. Data processing and storage

2.3.1. Quality Assurance/Quality Check

The Quality assurance/quality control process has to be implemented to ensure data quality. It can be implemented and executed:

1. at the reporting institution (before data delivery);
2. on the data delivery process (automatic quality control after data upload to the server);
3. at the host institution (after data download).

On the reporting institution, at least basic data quality assurance should be implemented to avoid reporting of obviously wrong data.

On the host institution, at least tests containing comparison of newly reported data with data already available in the database (i.e. reported before) should be implemented, to distinguish redelivery of identical records already available in the host institution database; correction of the data reported before; or reporting of new data.

If the control process is not executed by the reporting institution (before data delivery), feedback to this institution should be available in the case the violation of QA/QC rules and the need for correction of the data.

Violation of tests should be classified according to the significance of the error. Examples of classification are as follows:

- for information – data acceptable, possible ambiguities pointed out;
- *warning* – data acceptable, can be corrected in the future;
- *error* – acceptability of the data has to be judged by dataset manager, should be corrected before inserting in the central database;
- *blocker* - data unacceptable.

Figure 2.3 An example of quality issues in datasets

	2008	2009	2010	2011	2012	2013	2014	
Station 115					2.3	2.11	2.67	BROKEN TIME SERIES
Station 115	2.502	2.512	2.411	2.85				
Station 116			1.02	1.05	1.1	1.07	1.21	OVERLAPPING TIME SERIES
Station 116	2.59	2.338	2.244	2.833	2.667			
Station 117	3.12	3.11	87.48	3.05	3.07	2.98	2.83	OUTLIER

Types of QA/QC tests

Formal tests:

- Data format tests
- Data structure tests

Content tests:

- Mandatory values test
- Record uniqueness test
- Data types test (if the format of reported values matches the requested specifications)
- Valid codes test (if reported value is available in appropriate enumeration list)
- Spatial object identifier format test
- Spatial object identifier reference (availability) test (to avoid reporting of content data sampled in "missing" monitoring site etc.)
- Unit of measure test
- Sampling date/Reference year test
- Sampling period test
- Result values - limits test (identification of extremely low and high values out of reasonable thresholds)
- Additional parameters tests (water category, LOQ, sample depth, ...).

Cross–parameter tests

- Conditional dependency of the reported values on the values available in another field

Cross–table tests

- Checking of the relations between tables
- Checking of hierarchical structure of spatial objects
- Conditional dependency of the reported values on the values available in another table

Spatial data checking

- Specification and checking of topological rules for given type of spatial data (points/lines/polygons)

- Location of spatial objects within country bounding box
- Integrity of spatial objects identifiers with reported tabular data

Technical solutions

Description of technical implementation. FME to be mentioned and briefly demonstrated, SQL scripting to be proposed for implementation. Can we share scripts developed by the EEA/ETC?

2.3.2. Composition of central database

Appropriate database engine and database management tool should be used depending on the expected amount of data, management of the database and requested access to the dataset (e.g. MS Access, MS SQL, Oracle).

Database structure should follow general normalisation rules, to avoid storing of duplicated and redundant data. Database should ensure relations between content (time series) data and appropriate spatial objects, as well as storing of hierarchical structure of spatial objects.

It is recommended to store spatial objects in specific tables and ensure the relations to the content data tables and to the spatial objects on the another (lower / upper) level via internal database ID, not by reported spatial object identifier itself (if the spatial object identifier is changed, it has to be updated in one table of spatial objects only).

Database has to ensure storing old versions of reported data. It is recommended to store the archived records in extra tables, separated from tables containing last valid versions of the reported records.

2.3.3. Output data

The input data received from the reporting institution are typically used for assessment. Specifically, the assessment includes selecting representable time series, interpolating and extrapolating, calculating ratios, statistical descriptors and classes. The output data are usually stored in separate group of tables, to be distinguished from input data.

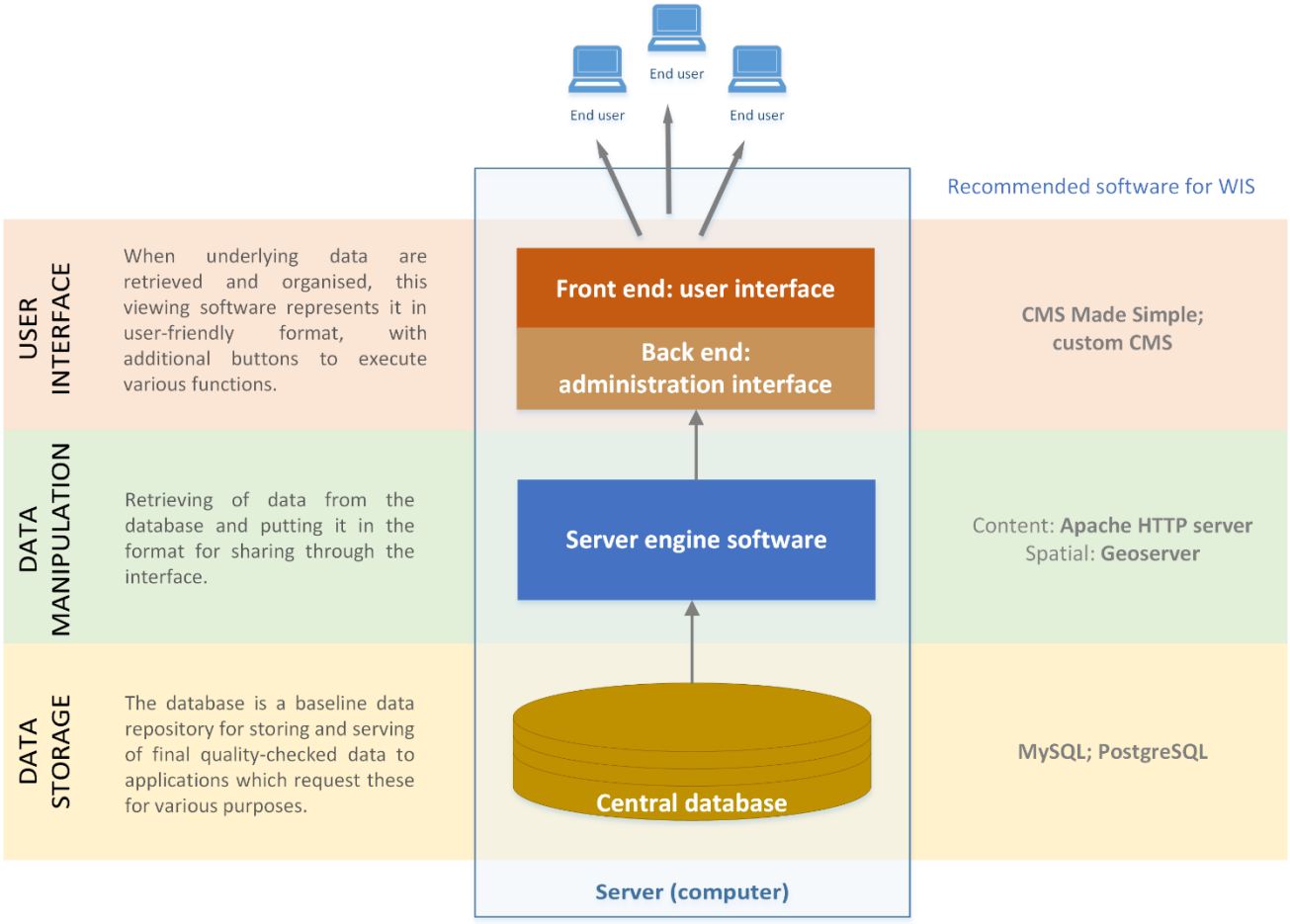
Technical means of composition of the output data depend on the complexity of needed results. The main statistical descriptors, classification and ratios can be produced through SQL querying in the central database, while more complex statistical calculations of trends, outliers and statistical significance may need to be developed in statistical software environment (e.g. SPSS, R).

2.4. Web portal

Web portal is the central software platform, serving as the main tool for storing, processing and presenting the information. The gravity of the web portal is the output data presentation and dissemination, while it can also include other elements of the WIS, from data reporting point to data processing interface and correspondence centre (i.e. helpdesk and issue ticket

system). This section focuses on the web portal that enables data presentation and dissemination.

Figure 2.4 Building elements of the web portal



2.4.1. Platform

The technical elements of the web portal are as follows:

- hardware: server-capable computer connected to the web;
- software:
  - database: the part of the central database serving the data to the users of the web portal, or a separate database, which is dedicated to the web portal;
  - data manipulation engines, e.g. ArcGIS Server, Daviz, Tableau;
  - interface: the connection point between the WIS and the user (administrator, end user etc.), typically in the graphic user interface (GUI) format, as back-end and front-end face;
- data transaction protocols:
  - rules of data transaction between the central database and the web portal;
  - user groups – rules of data serving to users of different clearances or interests.

2.4.2. Back-end – administration interface

The back-end, or the "server-side", is how the site works, updates and changes. This refers to everything the end user cannot see in the browser, for example databases and servers. Back-

end developers and administrators mostly deal with structure, content management and security of the web portal.

Administration interface should allow at least straightforward editing of static elements of the WIS front-end pages: text and images. Depending on level of development, it can also allow direct editing of dynamic components, such as interactive maps or graphs.

Technical requirements:

- CMS with capability of user restriction to site administration section

2.4.3. Front-end – user interface

The front-end is everything involved with what the user sees, including design and some languages like HTML and CSS.

The front-end of the WIS web portal is the interface through which end users access and use the WIS. It contains a system introduction and depending on complexity, menus and links leading to the content: data, assessments and indicators, viewers etc. The proposed elements of user interface are detailed further in chapter 2.5.

Technical requirements:

- CMS with capability to render more static or dynamic web pages
- if no dynamic page generation in place, static HTML files/pages

2.5. Data dissemination

The section discusses the output products of the WIS.

2.5.1. Data tables

Sharing of raw data is the most basic way of following SEIS principles: that data should be easily accessible to all users, and fully available to the general public. The time series used in an indicator assessment should be available directly from the assessment page to allow an interested user to explore and analyse it further. This means that the data should be served in one or more formats that can be directly imported in data storage or analysis software (i.e. Microsoft Excel, databases), most importantly table formats of various file types.

While dynamic export directly from the database is the most practical solution which allows acquisition of the most recent data without the need for administrator's updates of the data tables on web portal, other less dynamic options can also be implemented. A minimum recommended format is the data table that can be copy-pasted into a spreadsheet editor, opposite to image documents (e.g. scanned hard copies) that cannot be directly used for further data processing by interested users.

The level of detail of the data offered to users through the web portal depends on legal requirements, the institution policy and user needs. For this purpose, user group policies can be developed and used in WIS, to tailor the access to different levels of data according to user type (i.e. institution staff, cooperating partners' experts, technical administrators, public).

Which formats are suitable (XLSX spreadsheets, CSV, DB files etc.)?

Technical requirements:

- software:
  - for static data tables: pre-generated tables in any well-known file format (e.g. XLSX spreadsheet, CSV, database export);
  - for dynamically-generated tables: a code that queries the selected data from the database and stores it in the file for download (e.g. in PHP).

2.5.2. Data visualisation

Built upon data tables, data visualisation is the next level of data sharing. It offers initial graphical presentation of trends and other indications that are not directly visible from the data table, and its complexity can range from simple aggregated trend illustrated in a graph to complex compositions of graphs, maps, supplement indicators etc.

Static visualisation

The basic visualisation can be pre-generated in visualisation software (the most well-known may be Microsoft Excel) and uploaded to the web portal as static images, which enriches the sharing of data tables but is less adaptable to user's actions than dynamic/interactive visualisation.

Technical requirements:

- software: data visualisation software (e.g. Microsoft Excel, OpenOffice Calc)

Interactive visualisation

In addition to static images that present data in a way conceived by the publisher, an interactive visualisation allows the user to explore the data in depth: see more detailed information in a pop-up window by hovering each point on the graph, select their own time series of interest, filter the data according to chosen criteria, define their own classes etc.

Implementation of interactive visualisation is naturally more complex and it requires the data manipulation engine between the underlying database and the user interface. Both free-open source and commercial data visualisation software with such engine are available and they need to be integrated to the web portal.

Technical requirements:

- data visualisation software with the ability to publish and serve interactive visualisations (e.g. Chart.js, Tableau)

2.5.3. GIS

Separating presentation of environmental (water) data and statistics to detect spatial patterns (differences), geographic information system (GIS) should be involved in the process. GIS is a computer-based system to analyse and present spatial data. It can be a module integrated to

WIS and serve both data analysis and data presentation in spatially-referenced format. The most typical GIS output is a (static or dynamic) map.

Technical requirements:

- GIS engine connected to the underlying database; this is the software that makes use of spatially-referenced data and puts it in corresponding spatial relationship, the output of which is served through Web Feature Service (WFS) or Web Map Service (WMS) protocols to the front end – user interface in the format of a map (e.g. Geoserver, ArcGIS Server);
- front end:
    - for static presentation: HTML web page with embedded graphical element;
    - for dynamic presentation: script-enabled web page loading the data from WFS/WMS protocols served by the GIS engine.

2.5.4. Static reports

Reports are a composition of different data elements offering a thorough overview of the topic concerned. In the case of WIS, various reports needed for informing decision-makers, public, or other concerned parties can be available for download through the WIS web portal. Like other elements of WIS, depending on the complexity of the system they can either be static (i.e. pre-generated documents in DOCX or PDF format, for example) or dynamically-generated documents – they can be generated on a click and include tailor-made data tables and visualisation, along with the text that is usually prepared in advance by the experts.

3. Implementation of WIS

One of the key objectives of this work plan is to support Azerbaijan in developing water information portal to improve the capacity in meeting the commitments towards the regional/international reporting obligations. The national pilot will focus on the experimental implementation of online data sharing among the institutions responsible either for monitoring of water quality or for managing the water resources, e.g. National Environment Agency, Ministry of Energy and Natural Resources and Ministry of Agriculture etc.

The pilot will use the development of two water quality indicators (C-10 and C-11) as testing the implementation of standard data dictionaries of the State of Environment Reporting of the Water Information System for Europe (WISE-SoE). In addition, it will support developing and operationalising data exchange protocol among the related national institutions for online data collection and processing (Figure 2). EIONET and WISE ( ) experiences on quality check and assurance together with dissemination and visualisation of outputs will also be shared with national institutions and capacity of the national experts will be improved.

In implementing the goals, the following source data for environmental indicators are requested:

- concentrations of pollutants in rivers, lakes and groundwater;
- specification of hydrological or administrative spatial units to which reported data should be related (monitoring sites, water bodies, river basin districts, country), hierarchy of these spatial units and relations between them;
- Specification of temporal resolution (reporting of raw disaggregated data, data aggregated by temporal unit (month, year);
- Reporting period, frequency of sampling and reporting (e.g. monthly, annually).

3.1. Risks

- Lack of expertise or commitment
- Low level of integration of WIS elements