# 2013 Freshwater Eionet Workshop - 19/20 September 2013, Copenhagen

## Session 3: WISE reporting process

## Document 3b: GIS reference datasets as basis for WISE and their updating and version history

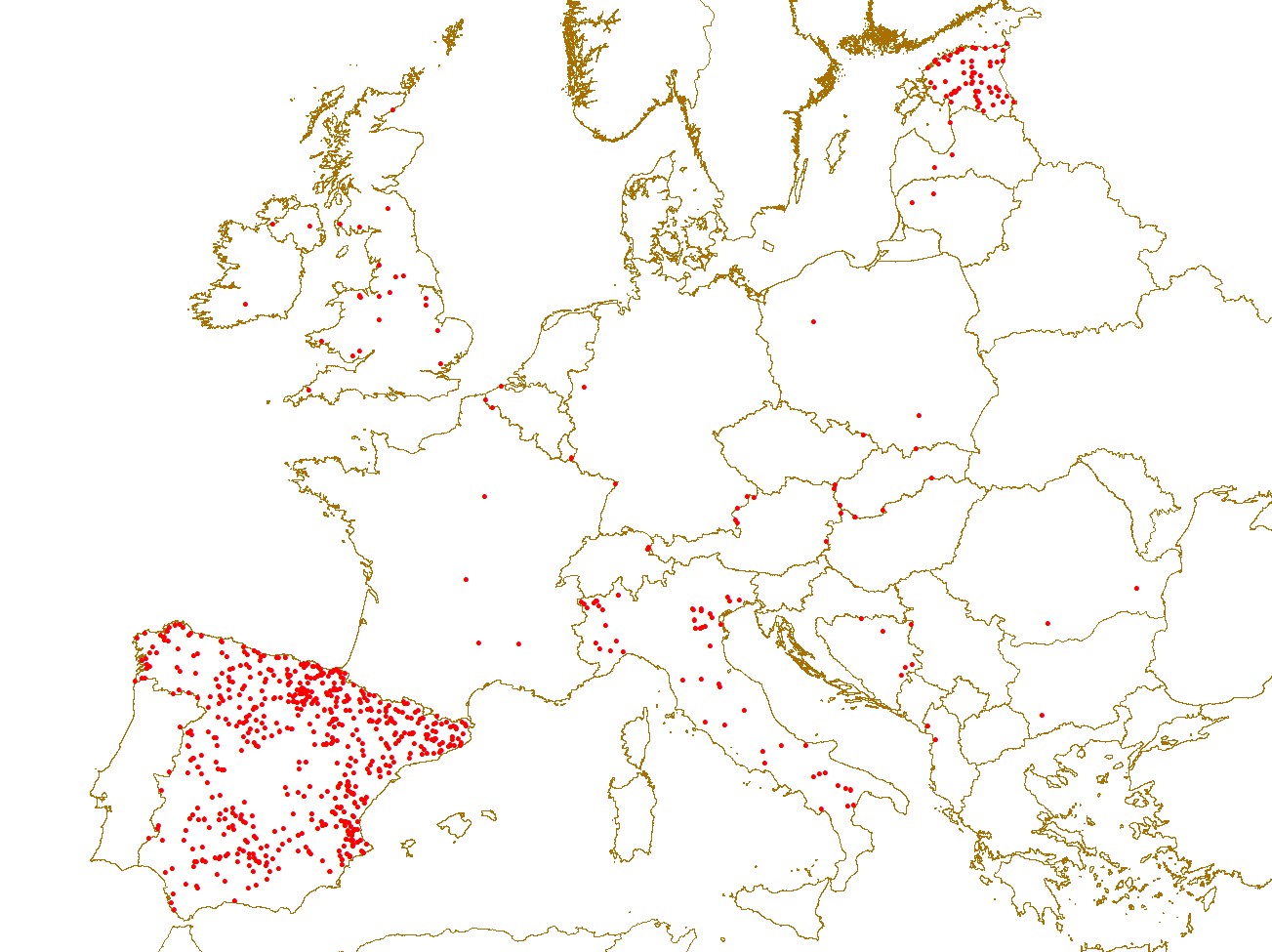
## Background

WISE datasets include two data types, spatial and tabular. Environmental data, disaggregated or aggregated over time or space are reported in tabular format and are related to specific spatial objects (country, river basin district (RBD), location etc). To keep consistency of data sets, the unique IDs of the spatial objects should not change along the time series.

For that purpose, the ETC/ICM develops a data model tracking methodology for the WISE core data structure. The data model will combine and interlink all spatial and tabular data from reference databases. As agreed between EEA and DG ENV, the data model for the WISE core data structure will first focus on the water industry directives: Bathing Water Directive (BWD), Urban Waste Water Treatment Directive (UWWTD) and Drinking Water Directive (DWD). Water Framework Directive (WFD) and Marine Strategy Framework Directive (MSFD) are also considered as well as SoE datasets. WFD and MSFD both present the basic spatial platform for building connectivity (relations) between WISE datasets. Data and information in the WFD and MSFD are linked to structural elements, such as River Basin Districts (RBDs), Water Bodies (WBs), Marine Regions and Subregions (MarRegSubReg). For the basic spatial reference of the data model, the European Catchments and Rivers Network System (ECRINS) database is chosen.

Each of the WISE data sets has one or more spatial objects to which data are reported (chemical, hydrological and biological parameters, water quality aggregates, water quality classes, other descriptive data etc.). These objects have unique ID codes that are present in each data set table to enable cross-table connectivity. Some objects can also be present in other data set (such as water bodies (WBs), RBDs) and can as such be used for cross-database connectivity. But in general, in each WISE data set there is one spatial object that data are related to. When tracking (archiving) object lifetime history, two aspects are important: 1) to track spatial elements of the objects and 2) to track data related to the objects.

### Some suggested data issues to be presented and discussed

* **Quality of reported data**: Some data of monitoring locations might be detected as “suspicious” in the automatic QA/QC procedure (agenda item 2a), but some are only discovered during data analysis (example in figure 1: There are many SoE rivers stations lying within 250 m radius of their nearest neighbour, suggesting that they can be (but not necessarily are) duplicates or can represent same environment conditions; we do visual check or check of names of rivers and their names to found such locations). At this stage highly suspicious values are excluded and reported to the data managers - the data are flagged to be checked with the data providers at a later stage. Errors in the data series can affect the data analysis. The manual QA/QC procedure is time consuming, and there is no guarantee that all possible suspicious values are discovered and corrected. So, it would be a great advantage if such stations and possible errors in locations are identified prior to ****reporting.
* Figure 1: SoE rivers stations within 250 m radius of their nearest neighbour, suggesting that they can be (but not necessarily are) duplicates of each other.
* **Datasets availability through time:** MS are not obliged to report all of the data which are included into WISE datasets. Due to this, there are gaps in significant part of the data time series, which restrict historical statistics calculation.
* **Time series data spatial connectivity:** Time series of different objects’ monitored data are stored within WISE datasets together with unique IDs and geographic coordinates. Each of these should identify one monitoring station, but in a significant number of cases, the same object (e.g. river monitoring station in SoE rivers database) is represented by more unique IDs or spatial locations. An already implemented method for connecting the same object’s datasets stored in different records included spatial joining and aggregating those objects which had complementary time series (i.e. two nearby objects, the first with time series 1992-2006 and the second with time series 2007-2012). However, this method needs further optimization and QA/QC.
* **Spatial datasets needed in future:** Some tabular data reported by countries belong to objects which are not spatially defined (e.g. water supply zones in Drinking Water Directive reporting) or are spatially represented as point objects (agglomerations reported under UWWTD, river water bodies). In the future, reporting obligations might include this kind of spatial objects. This needs further discussion on how to report and store this data (for example, which database should be used for each of new datasets).
* **Reporting of spatial data in a distributed/decentralised system:** It is being planned by DG ENV and EEA to start the distributed/decentralised reporting with the reporting of spatial data (WFD RBDs, waterbodies, monitoring stations etc.) with the aim of having always the newest version of the spatial data available at European level. This requires an agreement with countrieson the procedure using web feature services and INSPIRE rules.

### Coupling GIS reference layers and WISE datasets

Unique IDs between spatial objects of GIS reference layers and WISE datasets are not developed yet. For example, one could not always connect SoE stations and bathing water location with water bodies or ECRINS river basin to SoE river stations through common identifier. It would rather need to use GIS-supported spatial joining to do connections.

### Questions to NRCs

1. **Data quality**: What is the QA/QC procedure in your country before data reporting?
2. **Help with subsequent (data manager) QA/QC:** Is it possible to provide further data on how same spatial objects, represented by different unique IDs, should be connected (providing old and new IDs)? Why and how often are stations’ IDs being changed? What methodology is used to define a station as new (e.g. how far it has to be from nearby monitoring location on the same river)?
3. **Metadata providing:** Which methods are used for data changes tracking in your country?
4. **Distributed/decentralised reporting:** Have you established this in your country already for INSPIRE compliant spatial water data? If yes, how has this been implemented?