EEA Technical report | No 7/2012

EEA Catchments and Rivers Network System ECRINS v1.1

Rationales, building and improving for widening uses to Water Accounts and WISE applications

ISSN 1725-2237

European Environment Agency

EEA Catchments and Rivers Network System ECRINS v1.1

Rationales, building and improving for widening uses to Water Accounts and WISE applications

European Environment Agency

Design: EEA
Layout: Rosendahl-Schultz grafisk/EEA
Cover photo: EEA

Luxembourg: Publications Office of the European Union, 2012

ISBN 978-92-9213-320-7
ISSN 1725-2237
doi:10.2800/51667

EMAS
Verified environmental management
REG.NO. DK-000244

European Environment Agency
Kongens Nytorv 6
1050 Copenhagen K
Denmark
Tel.: +45 33 36 71 00
Fax: +45 33 36 71 99
Web: eea.europa.eu
Enquiries: eea.europa.eu/enquiries

Contents

Acronyms

AWB	artificial water body
CCM	Catchment Characterisation and Modelling
CDR	central data repository
CLC	Corine Land Cover
COBOL	Common Business-Oriented Language
DEM	data elevation model
ECRINS	European Catchments and Rivers Network System
EEA	European Environment Agency
ERICA	European Rivers and Catchments
FRBD	Functional River Basin District
GIS	geographic information system
GISCO	Geographical Information System at the Commission
GMES	Global Monitoring for Environment and Security
HMWB	heavily modified waterbody
IHO	International Hydrographic Organization
IRWS	International recommendations for Water Statistics
JRC	Joint Research Centre
LEAC	Land and Ecosystem Accounting
MSFD	Marine Strategy Framework Directive
NGO	non-governmental organisation
NUTS	Nomenclature of Territorial Units for Statistics
PEBLDS	Pan-European Biological and Landscape Diversity Strategy
RBD	river basin district
SEBI	Streamlining European Biodiversity Indicator

SEEA	System of Economic Environmental Accounts
SEEAW	System of Environmental-Economic Accounting for Water
SWTP	sewage water treatment plant
WFD	Water Framework Directive
WISE	Water Information System for Europe

Acknowledgements

This report was prepared by the European Environment Agency (EEA), and Pöyry with some support from EEA topic centre on Land use and spatial information The lead author was Philippe Crouzet. Contributors: Guillaume le Gall (Pöyry), Oscar Gomez (EEA). The EEA project manager was Philippe Crouzet.

This report presents the most recent update of the European Catchments and Rivers Network System (ECRINS) developed by the European Environment Agency (EEA) based on previous work carried out by the Joint Research Centre (JRC) Catchment Characterisation and Modelling (CCM) and the EEA (European Lakes, Dams and Reservoirs Database (Eldred2), European Rivers and Catchments (ERICA)).

This product was intended to become a reference system, at medium resolution (~ 1/250K), for all applications requiring high modelling capability that can be fulfilled with average geometrical accuracy, until a second version is developed in future.

It is not intended to substitute any geographical supports. From the EEA's perspective, it was primarily meant to serve as a reference system for all wide-scale water- and hydrosystems-related assessments: water accounts, river fragmentation, representative statistics, vulnerabilities, etc.

As reference system, its main use is the integration of Member State reporting and deliveries. To this end, much care has been taken in the ancillary data sets to maintain the equivalence tables between nationally managed objects and their European mirroring. One of the aims of setting up ECRINS is to facilitate integration of national data at European level, without imposing an added reporting burden on Member States.

As a fully topological system, its main use is data integration, and to allow the development, testing and production of the different environmental asset accounts relating to water issues.

A system that can be freely disseminated, ECRINS supports the involvement of scientists, students and non-governmental organisations (NGOs) in environmental assessment, which is often jeopardised by the absence of data and lack of reference systems. ECRINS aims to bridge this gap by providing, free of charge, a comprehensive hydrosystem that is accompanied by information making its use possible with limited computing capacity. It completes the Land and Ecosystem Accounting (LEAC) continental data set, based on a systematic kilometric gridding of the EEA area, also a source for populating the ECRINS features.

This disseminated version has been prepared as carefully as possible. However, its authors are fully aware of certain gaps, errors and mistakes that could not be corrected simply. The next scheduled versions will possibly address some of these; they may complete the naming of rivers and add ancillary data sets for those data not mature enough to be disseminated, despite being utilisable for specific purposes.

Executive summary

The European Catchments and Rivers Network System (ECRINS) is the hydrographical system currently in use at the European Environment Agency (EEA) as well as widely serving as the reference system for the Water Information System for Europe (WISE).

The first version of ECRINS (version 0 or β) was released in December 2008, and was completed throughout 2009. This version has been used to carry out assessments and simulations for the EEA. It is based on Catchment Characterisation and Modelling (CCM)[19], with resolution equivalent to that of a 1:250K map. This followed former attempts to develop the European Rivers and Catchments (ERICA) database (European Environment Agency, Collins Bartholomew, 1998) in 1998, that was restricted in its use. The later development of the ERC (Bredhal and Sousa, 2006), based on a combination of CCM and the EuroGlobal map, produced maps, but no calculable system.

During the late autumn of 2009, it was agreed that ECRINS would be the most suited host for integrating the Member States' reporting on 'main rivers and main lakes' under Art. 13 of the Water Framework Directive (WFD) (Directive 2000/60/EC). This decision made it necessary to improve the current version by inserting river names, a key driver for accurately defining the 'main drains' that are the candidate main rivers.

The version was upgraded by populating ECRINS β with river names, removing errors resulting from CCM data sets and bugs in the application developed to make it, and updating other components required to build ECRINS from CCM. Altogether, changes took longer than expected, and development of the new version was also delayed due to other important tasks, for example building the water accounts and producing results with ECRINS β.

The new version, named ECRINS v1.x, whose rationales, data model and contents are described in this report, is very similar to the former one. First and foremost, the data models are almost identical, making all applications tested with ECRINS β fully utilisable with the new one.

The life expectancy of ECRINS v1.x is in the range of three to five years. At the end of 2012, the Global Monitoring for Environment and Security (GMES) Reference Data Access Service for Europe should have provided an updated data elevation model (DEM) and rivers/catchment system that can be used to build ECRINS v2. The GMES reference data service is designed to have a resolution equivalent to that of a 1:100K map. However, considering the resources required to build an ECRINS layer, it is possible that the construction of v2.x will be carried out stepwise.

ECRINS v1.x: objectives and methods

Objectives

Version 1 of ECRINS was developed under stepwise improvements that will be released as upgraded versions. The objectives of building a generic ECRINS v1.x are as follows.

- Provide a stable conceptual model, taking stock of past experience. Adjusting, where necessary, the original data model to take into account experience from applications carried out (water accounts, stratification, river fragmentation, WISE references, etc.), and incorporating the 'main rivers and main lakes' from Member States' reporting so as to facilitate uses.
- Insert systematically an operational system of versioning that takes into account the many types of kernel and ancillary layers and their differential update rates.
- Insert dummy/real river unique identification (nicknamed 'routing' and 'naming' respectively). Naming is based on the most complete reference maps that are free of charge to use, so that the main drains should be based on named rivers (i.e. the usual concept of rivers). In parallel, use stable route information, that is equivalent to

having dummy rivers fully operational when naming is not required or achievable.
- Correct the improper allocation of elementary catchments (some hundreds of units over 2 million) resulting from the processing of CCM source data sets.
- Correct the inaccurate construction of islands catchments resulting from incomplete populating of island codes in CCM; correct the few errors in coastal catchments (islands and continent) resulting from the incomplete populating of islands and the inadequacy of the processing application when aggregating very small island objects.
- Reshuffle completely the seas delineation proposed by the CCM, and align it with both the preparatory work for the EU Marine Strategy Framework Directive (MSFD) (Directive 2008/56/EC) and the international agreements made under the International Hydrographic Organization (IHO) (IHO, 1953 and 2002); in parallel, make it possible to insert a consistent shoreline in a flexible way to allow transition waters and continental waters to be accurately placed.
- Adjust, where possible, topological errors recorded, as well as those resulting from the source processing for the CCM.
- Update and complete documents to facilitate the updating of ECRINS v1.0 with data sets that may change rapidly (dams, monitoring stations, etc.), so that applications using ECRINS are as accurate as possible.
- Create, based on the geographical elementary catchments, different layers of aggregation catchments, fitted to the legal entities (river basin districts (RBDs), regions, etc.), to allow the computation of water asset accounts.
- Integrate, as far as possible, static water bodies (lakes, reservoirs) so that corresponding layers can be disseminated and populated with important hydrographical information (e.g. volume), and their elements related to the drainage systems and catchments.

Methods

The experience gained during the development of ECRINS β, i.e. the WERC ([1]) application, has been used. The application is systematically tested and corrected or simplified where necessary and the ArcGIS® procedures needed to create the intermediate geographic information system (GIS) data sets tested and upgraded to scripts (instead of models) to make them more transferable.

On the documentation side, the 'trials and errors' that were reported in the development notices [4] have been omitted from this document; it only presents the hard findings. However, the means of producing the intermediate and final layers is presented, so this report can also serve as a manual for the next generation of ECRINS, v2.0.

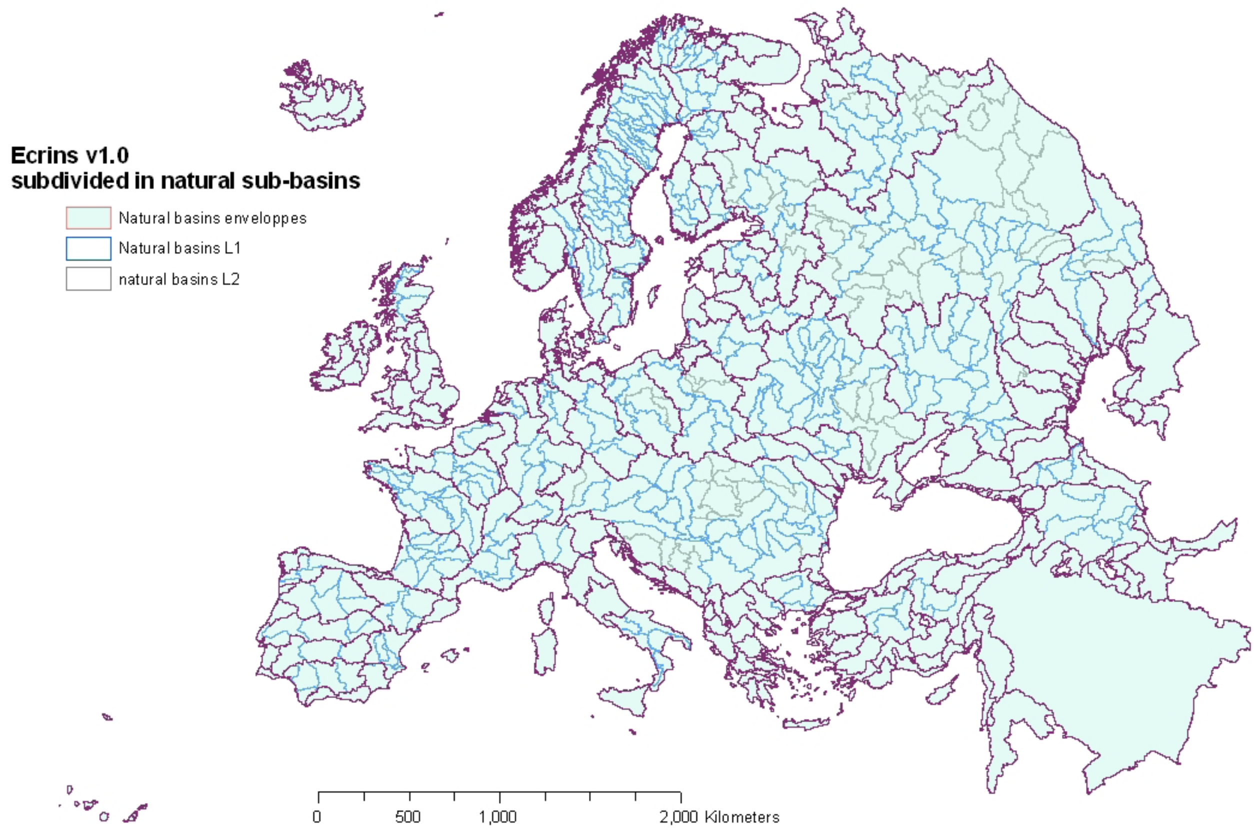
Scheduled versions

A new version of ECRINS is imperative; not all scheduled changes could be carried out to meet urgent needs for the scheduled uses. The following ECRINS versions are planned.

1. ECRINS v1.0, third quarter of 2011. New sea delineations have been taken into account, and where rivers are named according to the most commonly available source, so have first Member State deliveries under Art. 13 and EuroRegionalMap (ERM) v2, thereby creating new main drains. This version is designed to host the main rivers and main lakes from the WFD reporting. It is fitted only with the current CCM derived shoreline. It includes, as an ancillary provision, the water bodies, attached to main drains and dynamically routed.
2. ECRINS v1.01, early 2012. This version replaces the lakes layer from different sources by a layer uniquely taken from the old CCM, Corine Land Cover (CLC) and Art. 13 deliveries.
3. ECRINS v1.1, early 2012. This version is scheduled to take stock of the topological errors identified thanks to the naming procedure, and considers comments from Member States. Addressing these errors will call for improved connectivity and recalculations of the topology and Strahler levels accordingly. This is quite a hefty operation and its duration will depend on the number of errors needing correction. Using this opportunity, the shoreline derived from CCM shall be replaced by another shoreline, possibly extracted from the SeaVoX data set. ECRINS v1.1 will present the flood plains polygons next to main rivers.

([1]) Compression of Watersheds Rivers Catchments.

Map ES.1 Geographical extension of ECRINS, displayed with natural watersheds



Note: Basins envelope's stick to RBDs when existing and subdivide the watersheds into level 1 sub-basins, possibly apportioned again if the watershed is too large.

Source: EEA processing of ECRINS v1.0.

4. ECRINS v1.2, not before the end of 2012. This version should not differ in essence from 1.1; it will be completed with important objects and completed with canals and derivations. Where available, river attributes such as width will be added.

ECRINS v2.x is not due to be analysed before the end of 2013, it and could be built from the current GMES production process. Version 2 is planned to be produced at 1:100K resolution from RDA deliveries, of which the new DEM is at 30 m resolution.

Geographical coverage

ECRINS encompasses ~ 10 million km^2, including the European part of Russia and countries on the east coast of the Black Sea, as well as all watersheds with springs in Turkey, i.e. the Tigris and the Euphrates. For practical reasons, it has not been possible to break elements down into sub-basins; this will be undertaken in another update.

Licensing issues

With exception of lakes, all ECRINS layers were primarily derived from data sources that were entirely free of charge. Catchments are aggregates of CCM Strahler 1 catchments, and rivers segments are taken from the CCM as well.

By contrast, the lakes layer inserted in ECRINS and limited to internal EEA use was a complex composite of many sources: CCM lakes (free of charge), EuroRegionalMap (ERM) v2 lakes (may be licensed), country sources and WFD

Art. 13 deliveries (by definition, free of charge). Licensing issues are being considered; the terms of ERM licences are still unclear.

To avoid any risk of misuse of the licensed data sets, all were removed from tshe ECRINS version described in this report: lake polygons are derived from the latest CLC, just checking if the CLC water masses are identified as lakes or not. Complementary information is taken from different sources such as Art. 13 deliveries and Wikipedia.

The RDA mentioned above is expected to resolve pending licensing issues concerning certain attributes that must be derived from geographical sources.

Introduction

Purposes of rivers and catchment reference system

The need for operational watershed layers has long been recognised by the EEA. It came about with the release of ECRINS β, that was driven by the need to implement the different methodologies required to meet the medium and longer term objectives of the EEA in assessments. Examples are the production of water accounts (resources and quality), the sound stratification of catchment-based divers to representatively analyse quality and quantity data, the production of migratory fish Streamlining European Biodiversity Indicators (SEBIs), and inputs to the sea (mass discharges). All the above depend entirely on a reference layer. In the EU area, homogenous contextual information on water bodies (under the WFD) is a prerequisite for analysing the necessary elements of policy effectiveness assessments.

River catchments are modelled objects: there is no means of observing them since they do not constitute a geographical reality. By contrast, rivers are geographical objects that can be observed, albeit requiring some level of modelling to become part of a GIS. The complex web mimicking a braided river must be represented as a valid centre-line network for calculation purposes with affordable tools.

Rivers and catchments are highly interconnected systems that must be completed by many other features: lakes, dams, monitoring points, abstraction points, flood plains, sewage treatment plants, etc. All these features must be identified and related to the elements that conventionally represent the rivers and the catchments. These connexions in turn open up the possibility for these objects to be populated with information stemming from different sources: 'How many people live upstream of this lake/water body?', 'What is the flow discharge at this point?', 'How many cubic metres of water are abstracted to irrigate fields?', 'What is the annual water balance/indicator for this basin?', 'What is the length of accessible rivers below that dam?'. The potential number of questions and possible responses is enormous, but addressing such issues calls for a flexible and soundly connected system of features connected using the appropriate geometry.

There are few essential specifications of a catchment and rivers reference system suited to the EEA and to European needs.

- It should provide comprehensive coverage of the EEA area (i.e. encompassing areas beyond the EU).
- It should have a reasonable resolution (in the approximate range of 1:500K to 1:100K) that is simultaneously compact enough to be manipulated and sufficiently detailed to address local differences that matter at European level.
- It should comprise elementary objects (catchments, river segments, etc.) having both a relevant and a narrow range of sizes to allow for it to be accurately populated with the required information and to be statistically representative. Additionally, the elementary catchments sizing should be appropriate for making non-hydrographical aggregates with small differences (e.g. for the Nomenclature of Territorial Units for Statistics (NUTS) equivalent of catchments).
- It should be fully connected and routed to allow any type of populating and assessment, and sufficiently gazetted to allow data from external sources (such as reporting under the WFD and from other important EU legislation) to be attached to it.
- It should be disseminated free of charge, and be accessible and available to the general public.

When the development of ECRINS β started, the CCM catchments and drains were considered adequate candidates to form the building bricks of this new reference layer. However, the CCM is not the only data set used for building ECRINS: the gazetting is poor, the seas do not have international sea delineations, it lacks an appropriate lakes layer, it lacks dams, and last but not least, its elementary catchments layer is 'rough and ready' — CCM source data needed a lot of post-processing and complements to meet criteria listed above and to make it an 'end-user' product meeting EEA requirements. During the creation of ECRINS, many data errors were found that sometimes delayed the production of ECRINS as well.

The CCM developed and produced by the JRC has the additional advantage of being free of copyright and can therefore be disseminated.

Definitions

ECRINS development required determining precise meanings for certain words. The selected words are not necessarily the best for all circumstances; the criterion for selecting a word is to avoid possible confusion with other candidate words.

ECRINS is being built in a complex administrative environment in which many hydrological terms have been used with specific meanings. Some are clearly contextual; others may be ambiguous.

Table I.1 Definition of key terms used in ECRINS

Term	Definition
Basin	In ECRINS v1.x, this is used when all the envelopes of elementary catchments have a single outlet in a terminal recipient. A basin is defined by CCM source data sets and is identified by a unique code. The CCM basins are a kernel feature.
Cardinality	In any GIS processing, objects seldom pair exactly. The term 'cardinality' refers to different relationships that can be one to one (A pairs exactly with I), one to many (A pairs with I, J, K, etc.), many to one (A, B, C matches with I), many to many (A, B, C, etc. match with I, J, K, etc.). Notations are 1 – 1, 1 – M, etc. No pairing is noted as 1 – 0 and 0 – 1.
Catchment	This term is used systematically by CCM authors to refer to all the elementary polygons, of any Strahler level.
Coastal basin Non-costal basin	In ECRINS, the distinction between costal and non-coastal basins is quite arbitrary: a non-coastal basin has a size larger that a certain threshold (e.g. 200 km^2) and is defined as such. Hence, coastal basins are the non-coastal ones.
Code Identifier	A 'code' conveys semantic information about the object, for example, the Pfafstetter coding system (Furnans and Olivera, 2000) provides information on the branching and subcatchment, among others. By contrast, the purpose of an 'identifier' is limited to identifying the object, and possibly locating the dataset from which the object having this identifier originates. Identifiers are nicknamed as IDs when referring to a field in a table. A code might have the advantage of implicitly containing topological information, for example the Pfafstetter code, but in case of an update of the catchment structure or shoreline, this codification may become obsolete.
Drain Main drain Secondary drain	A drain is the representation of the water channel in which water flows across and out from a catchment. Drains are represented by segments and nodes. The main drain is defined in several ways: i) the river stem that enters and exits the catchment for standard FECs; ii) the set of river stems having the same name or the longer extension in the most upstream FEC; and iii) the set of river stems having a name and the longest extension in the largest of the catchments aggregated into a coastal FEC.
Feature	Used in modern GIS applications, this term designates any object having a geometry: catchment, basin, point, line, etc. A feature is stored as a row in feature class table (e.g. catchments table).
FEC or Functional Elementary Catchment	Stands as the central element of ECRINS. FEC refers to the smallest catchment identified as an ECRINS elementary catchment. A FEC is built via aggregating elementary CCM catchments. It could be either a 'continental FEC' when built by aggregating elementary CCM catchments from a non-coastal basin, or a 'coastal FEC' when elementary CCM catchments belong to a coastal basin.
Island	The 'island' concept has been simplified. In the source CCM, not all islands were coded as such (e.g. Great Britain was not an island). In ECRINS v1.x, all land not conterminous to continental Europe and Asia is an island. All codification has been revised to address this issue, making it possible to use different delineations.

Term	Definition
River (main)	The WFD uses 'main rivers' to refer to certain rivers under the reporting obligation. The term 'rivers' hence becomes ambiguous. In ECRINS, it is replaced by the term 'drain' when designating a feature, until it is named or routed, hence making it either a 'true' or a 'dummy' river. The term 'river' is retained when referring to a river in layperson terms. The Common Implementation Strategy GIS Informal *Guidance Document on compliance reporting using the Water Information System for Europe (WISE)* (WFD CIS, 2005) defines main rivers as those rivers that drain at least 500 km^2 of watershed; these are required to be reported under Annex I (ii) of the WFD. According to the guidance document, a unique ID and the name of these rivers should be reported. This source could constitute a major information source for ECRINS, thanks to the fact that the FECs are substantially smaller than the 500 km^2 threshold mentioned above.
River	Is defined under the layperson understanding as a series of stretches having the same identification (which may be a name, regardless of translation issues) from a starting point (usually its spring) to an ending point (the terminal recipient/confluence). A stretch is a polyline between two nodes; hence a river has or shares n+1 nodes if it has n segments. In the CCM, rivers do not exist, although a 'named rivers' data set is provided. By contrast, the CCM provides segments and nodes of drains. In the absence of information indicating otherwise, a 'true river' may be defined as the series of segments having the same name between a starting and an ending node. In most cases, a river has a single 'name'; this name may, however, have different spellings related to different languages (e.g. the English Oder, and the Czech and Polish Odra), that can be coped using a single reference language (e.g. English when a translation exists). By contrast, different rivers may have identical spellings in English despite referring to different objects, for example Don/Don river/(le) Don may designate a river in Russia, in the United Kingdom and in France respectively).
River (dummy) Route	In CCM, rivers are not all clearly identified; only the segments and nodes are. In the absence of information indicating otherwise, a dummy river is defined as the series of polylines having the longest length from the final recipient outlet or confluence to the most upstream node. Hence a dummy river has no river_id but has a cgenelin (river CODE).
River basin	Has a specific definition in the Article 2 of the WFD: a river basin is defined as 'the area of land from which all surface run-off flows through a sequence of streams, rivers and, possibly, lakes into the sea at a single river mouth, estuary or delta'. This definition is fully acceptable and is just generalised as 'streams, rivers and, possibly, lakes into a terminal recipient at a single river mouth, estuary or delta'.
River rank Route rank	The Strahler rank indicates a degree of branching that starts at a spring. By contrast, river and route ranks start at a mouth. A river emptying at sea has river rank 1, a river emptying to river rank 1 is ranked 2, and so on. Routes are ranked in the same way. However, river ranks and route ranks may have different values for the same set of segments, depending on the way rivers were named and the naming of the upstream segments.
Strahler order	The river branching has been defined by Strahler (Strahler, 1957) [14] from the former Horton classification. The most upstream stretch has the order 1, two stretches of the same order s make a stretch of the order s+1 and two stretches of different order result in a stretch keeping the largest order. The Strahler order depends on branching structure: small catchments of wide and impervious bedrock have a high Strahler order, narrow and pervious catchments have a smaller order ('fishbone' catchments, for example). The Strahler order is a very powerful tool for analysing river system arborescence. By extension, a catchment having no drain is ordered as 0.
Sub-basin	The Common Implementation Strategy GIS Guidance Document No 9 (EC, 2003b) defines a sub-basin as 'the area of land from which all surface run-off flows through a series of streams, rivers and, possibly, lakes to a particular point in a water course (normally a lake or a river confluence).' This term is used to define natural sub-basins that are aggregates of elementary catchments within a given size range.
Watershed	Generically designates (in this report) any level and type of polygon perceived as a drainage basin. This is done to avoid ambiguities, since it has not been used in a specifically accepted sense either by the WFD or the CCM authors. The naming of the different types and levels of watersheds is described in the data model.
Segment Stretch	Is a polyline (of vertices) between two nodes. The term 'segment' is used when dealing with the GIS object, and 'stretch' when considering the river piece it mimics. A drain is made up of segments.

1 ECRINS v1.0: delivered data sets

1.1 Structure of libraries and tables for public data sets

1.1.1 Rationales

The purpose of ECRINS v1.x is to provide EEA staff, European Topic Centres (ETCs), and the different EU institutions working in the environmental field with a ready-to-use, completed, and gazetted catchment and river system, free of charge. This can be achieved by taking stock of the research carried out under EU funding by the JRC, that outcome with CCM.

However, while using ECRINS β, it became evident that many other users could also use it for support in their activities, provided that the delivered data sets could be manipulated and computed using different software on the one hand, and adequately documented, on the other. The obvious solution to the former requirement was the delivery of the different layers and ancillary tables in the form of MS Access® personal geodatabases (PGDBs). The PGDBs, albeit in Microsoft® proprietary format, can be operated with both MS Access® ([2]) and ArcGIS® ([3]), the software used to build ECRINS v1.x that can also be used with most open source GIS and database managers.

The PGDB also offers the important advantage of hosting both feature class tables (tables with geometric information and attributes) and flat tables (containing only attribute data). The negative aspect is that the maximum size of individual PGDBs is 2 GB, the size eventually required by large data sets. This size can be also be reached during processing, a system inconvenience in MS Access® being the tendency of any database to become 'bloated' due to the presence of temporary objects, even in their absence. A safe size of 1.4 GB has been set as a target. Hence the design and the processing of ECRINS databases are carried out to mitigate this inconvenience. The solutions are presented in the sections below. The process in WERC analyses the target database size and, when necessary, compacts it during computations.

1.1.2 Naming conventions

Naming conventions for final ECRINS databases and their tables is essential. First of all, the databases are apportioned in logical groups defined to meet the size criterion indicated above. The final databases are named EcrXxx where Xxx indicates the contents.

The naming of tables inside the PGDBs follows NOPOLU Système 2 ([4]) requirements, the application used to carry out water accounting. The other developed applications (river fragmentation, small hydropower, stratified assessment of water quality, etc.) were carried out using this platform, which in turn helped improve the design of ECRINS. Applying this naming convention sped up the development of these applications, required less funding that would other naming means, and presents no inconvenience for users.

The conventions are that all tables describing a characteristic are named C_xxx where xxx names the characteristic, and tables relating to values of these characteristics are v_xxx. Table names are listed in the sections above: the principle is to facilitate an understanding of relationships between tables.

1.1.3 Working data sets

ECRINS is built with a central MS Access® application (WERC) that uses PGDBs and is based on ArcGIS® models and scripts that either use file geodatabases (ESRI proprietary format), shapefiles or PGDBs. The delivered data sets are final copies of the feature classes and flat tables, whose organisation is far more complex and numbers are

([2]) First built with MS Access® 2003 and completed under MS Access® 2010, maintaining compatibility.
([3]) Version 9.3.
([4]) Developed by Pöyry consultants.

Table 1.1 Names and contents of final delivery data sets of ECRINS v1.x

PGDB name	Contents
EcrFEC	All ECRINS elementary watersheds (FECs) layer as feature class in table C_Zhyd
EcrRiv	All drains (as segments) table C_Tr and nodes C_Nod. River names are in other DB
EcrLak	All lakes in feature class C_lak Ancillary tables are: 1. V_LakInOut: IDs of the inlet and outlet segments attached to the lake and main FEC ID; 2. V_LakperFec apportionment of lake per FEC (for water accounting purposes); 3. LakCtroid feature class of lakes centroids (true or forced centroid); 4 V_LakperCtry is the relation lake to country; 5. V_LakperNUTs is the relation with standard regions. The EcrLak database contains only unlicensed objects
EcrFAy	Ancillary features than cannot be in the same database as a source feature C_Zhyd_Centro: FEC centroids
EcrAgg	Aggregation basins/areas. All FC exist under a double geometry: the original geometry and the one adjusted to the FEC layer. Tables are: 1. countries Ctry (source and C_Ctry (adjusted to FECs); 2. regions NUTsX and C_NUTS (adjusted to FECs); 3. shore watersheds in relation with shores C_B; 4. river basin districts (RBDs) C_FRBD and C_FRBD_BC (adjusted to FECs); 5. sub-basins Strahler C_Ss (by definition adjusted to functional river basin districts (FRBDs) and FECs); 6. natural sub-basins C_Sb (by definition adjusted to FRBDs and FECs)
EcrGaz	Gazetting of ECRINS. In v1.0, only rivers are gazetted, with two related tables: 1. RivNaming, with river ID and main name; 2. RivAlias with the aliases of names and duplicate IDs where relevant
EcrAnc	Ancillary tables. The most important one is river IDs and names. This is a flat database. It contains a miscellaneous set of tables: 1. FECvsW1: relationships between FECs and source level 1 CCM catchments,
EcrPFt	Point features. It contains all point features of relevance: 1. dams and obstacles C_Dam; 2. gauging stations, C_Quan; 3. water quality stations C_Qual

much larger than the delivered ones. These files exist for ECRINS primary administrator purposes, and are not described in this report. When necessary for understanding a point at ECRINS user level, further explanation will be provided in the relevant section.

1.1.4 Relationships between data sets

No built-in relationship has been created; such relationships are unnecessary and would jeopardise further uses. Using referential integrity could have compromised the use of ECRINS data sets with non-Microsoft applications.

The relationships are ensured by appropriate IDs placed at the most effective place. When the creation of the relationship is time consuming, the relationship stable within a version, and the use of the relationship frequent, it has been computed and is set in the table. For example, the ID of catchments upstream of a catchment can be computed from the ID of the catchment downstream; however, this is a long computation and the need for such upstream catchments frequent. Hence it has been computed in advance.

1.2 Contents of the different data sets

1.2.1 Database of elementary catchments (EcrFEC)

The FECs database EcrFEC.mdb contains feature class C_Zhyd, which is the most important data set in ECRINS. The structure of C_Zhyd is reported in Annex 1. This table sets out the FEC IDs (field ZHYD) and all the required IDs of the useful data sets: aggregation watersheds and reference watersheds, the connection between FECs and sources of information.

Table 1.2 Key statistical information of FEC number and sizes (v1.0)

Threshold area (km^2)	No of FECs larger or equal than threshold	Cumulated area of category	% in number in the category	% in area in the category	% in number in category and below	% in area in category and below
0	181 071	12 097 491	100.0 %	100.0 %	0.0 %	0.0 %
5	164 997	12 064 345	91.1 %	99.7 %	8.9 %	0.3 %
10	153 202	11 976 383	84.6 %	99.0 %	15.4 %	1.0 %
50	77 566	98 48 760	42.8 %	81.4 %	57.2 %	18.6 %
100	35 472	6 850 228	19.6 %	56.6 %	80.4 %	43.4 %
200	9 125	3 248 108	5.0 %	26.8 %	95.0 %	73.2 %
500	955	977 235.4	0.5 %	8.1 %	99.5 %	91.9 %

The key statistics on FECs are set out in Table 1.2.

The average area is 66.8 km^2, significantly smaller than the 92 km^2 average area in ECRINS β. The most significant improvement is in the better computation of coastal catchments and increased use of Strahler 2 (instead of Strahler 3) aggregates in flat areas. Almost 100 % in number, and close to 92 % in cumulated area of FECs are smaller than 500 km^2; the larger FECs are mostly situated in the eastern part, in desert areas. By contrast, there is a certain proportion in number (almost 9 %), cumulating in only 0.3 % in areas of FECs smaller than 5 km^2 related to the source data set. This cannot be corrected for two reasons: i) the complexity of the task outweighs the benefits, and ii) the need to distinguish connecting small FECs, FECs whose size is driven by geography (e.g. small islands), and useless ones.

FECs systematically belong to a basin, with ID Bas0_ID.

The FECs IDs are stable under a calculation, within the ECRINS version. This stability cannot be permanent since the FECs' composition may change if their building targets change. Consequently, there is no way to maintain the IDs, since the referenced objects may change.

1.2.2 Aggregation and reference catchments (EcrAgg)

This data set comprises three categories of objects:

- aggregates from C_Zhyd, that can be reconstructed by breaking down the FECs into the appropriate fields;
- source data sets used to build some categories: seas, RBDs, etc.;
- source aggregates used for building the FECs.

1.2.3 Drains, nodes and rivers (EcrRiv)

This data set comprises two important features classes: C_Tr, and C_Node, i.e. the river segments and river nodes, respectively.

Each segment is located between two nodes (FNode: upstream; TNode: downstream) and is, in CCM, the drain (when existing) of a Strahler level 1 elementary catchment. In CCM, level 1 elementary catchments without identified drains are qualified as 'Strahler 0' by convention.

All segments are allocatsed to a FEC, through the ZHYD field that connects C_TR to C_Zhyd. The ancillary table FECvsW1, placed in the EcrAncl flat database, contains the connections between the permanent layer of segments ([5]) and the potentially changing layer of FECs.

The segments table is permanent until topological corrections are inserted. In this case, the segment ID (TR) of the suppressed segments is definitively lost and new segments have new IDs. Permanency in segment ID does not guarantee that other attributes are also permanent: some topological changes could make a segment shift to another basin, for example.

River names are deliberately placed in another database, since the gazetting is likely to expand with time. The connection between name and river is the standard CGENELIN ID that is set in the C_Tr, the river being named (true CGENELIN) or forged.

([5]) This layer is permanent until topological errors are corrected in v1.1.

The first delivery takes a new and systematic approach to 'main drains' that are fully documented, as provisional in the absence of full naming.

Compared to ECRINS β, the new version proposes full revision of the drains, as follows.

- Unique data sets comprising all drains, instead of a set of main drains and no information on the secondary ones. Since ECRINS v1.x, all drains are considered equally. Only their attributions indicate the category depending on FEC versions (a main drain where FECs are small may become secondary if a larger FEC system is implemented).
- Systematic information on main drains, considering that 'trunk main drains' that relate to all the FECs not upstream is a stable tree defined only by the delineation of FECs. The drains in upstream FECs (that include coastal FECs) are the only subject to change with the naming of the river that continues the main drain in its downstream recipient.
- All segment and node information, incomplete in CCM, is fully calculated. CCM had computed the category of segment and nodes only for ~ 55 % of objects. In ECRINS v1.x, all nodes are categorised, the value of length to mouth is computed for all segments, and some topological errors have been corrected.
- Full routing of segments, the route ID being set in the CGNELIN field, whereas the name, if attributed, is set in the River_ID field.
- Spurious segments have been 'shaved' and their upstream node removed (about 160 000).

Nodes are points connecting two segments. In principle, CCM nodes connect three segments (one segment and its two upstream segments) except the final node of the outlet segment and the starting node of the spring segment. However, ECRINS processing identified nodes having up to five connected segments, which could not be all be corrected.

As segments, node IDs are now unique and disconnected from the ECRINS version. The node table carries very important information — the altitude at the node. Hence, the slope of any segment can be computed from node altitude and segment length. In ECRINS, since many spurious segments have been removed, a node may have a single upstream. The number of upstream, set in C_TR, is hence 0, 1 or 2. Source CCM comprises 7 090 nodes having more than 2 upstream; these cases have been cleaned and the suspect segments systematically removed, resulting, however, in orphan sets of segments.

Of more than 3.65 million km of rivers contained in the CCM source dataset, and distributed over 1.35 million segments (2.7 km length on average), 98.7 % could be allocated and computed. Some are located in areas where the drain could not be properly set; others were discarded during the process (for any reason). The total length actually computed is 3.61 million km, of which 2.1 million km (57.7 %) are main drains. Table xx indicates the apportionment of main drains amongst the three categories of FECs. Assuming that the intermediate segments are the most stable, 1.65 million km of drains (79.3 % of total length of main drains) are main drains for certain, and the remaining 0.431 million km are subject to revision. This will not substantially change the total length, since upstream catchment drains rejected as main drains will be replaced by other drains in the same catchment.

After cleaning and shaving, the database still comprises 510 000 springs (segments having no

Table 1.3 Key summary information on river segments

Source data	Allocated to coastal FECs	Allocated to source FECs (non-coastal)	Allocated to intermediate FECs	Total	SumOfNinter
TS: 1 348 163	9 485	66 1020	660 433	TA: 1 330 938	TA/TS: 98.7 %
Of which main	5 953	87 822	529 289	TM: 623 064	TM/TA: 46.8 %
Cumulated length of segments in km Same as above, for Cumulated length.					
CS: 3 646 582	29 523.98	1 663 848	1 912 417	CA: 3 605 789	CA/CS: 98.9 %
Of which main	2,2779.1	408 344.4	1 647 658	CM: 2 078 781	CM/CA: 57.7 %

Note: Statistics computed before cleaning and shaving.

TS: Total number of segments; TA: total number allocated; TM: total number in main drains.

upstream) that create as many routes, since a route is measured from the spring to the confluence having another route or terminal recipient.

1.2.4 Lakes (EcrLak)

This database comprises the C_Lak feature class, with polygon lakes and ancillary information, and two categories of flat tables.

- Structural tables forming the relationships between C_TR and C_Zhyd. Table V_LakInOut marks all the inlets and the unique outlet of each lake for which such drains were found, along with the FEC of lake centroids, plus ancillary information (altitude, etc.)
- Ad hoc table required for the water accounts procedure, that apportions lakes under the FECs to which they belong: some lakes extend over several FECs, since they are defined by catchments before lake (in the case of artificial lakes, for example) or by connecting drains within the lake in the other cases. This table is named C_Lak_I_Zhyd.

The lakes feature class has been widely revised in v1.01 that replaces v1.0: the lakes have been systematically freed of any licensed polygon. Structure, by contrast, has been retained.

1.2.5 Ancillary data sets (EcrAnc)

The number of possible ancillary data sets is unlimited. This document addresses only those ancillary data sets that relate to the ECRINS structure and that are necessary for its operation.

The most important ancillary data set is the RivNaming table. This table contains, for all defined rivers, their ID (information source, if any, and ECRINS unique ID), with possible duplicates in the case of shared rivers. For example, France and Germany do not have the same national river ID for the border reaches of the Rhine. This table is related to the RiverAlias table that contains the different aliases for a river name.

In ECRINS v1.x, there are ~ 22 000 rivers named. The number is deliberately kept imprecise because it is certain that there are some naming errors, albeit unidentified.

Both tables are new for ECRINS; they replace the CCM table of NamedRivers that was a feature class of dissolved segments having the same name under CCM and comprised 1 480 named rivers. This table was discarded in ECRINS since keeping it would result in a duplicate linear feature class.

1.2.6 Aggregation catchments (EcrAgg)

This data set was inserted in case the ancillary data sets become too large. It comprises all the catchment aggregates and their ancillary data. Details are in Table 3.3, page 37.

- C_B: basins built from shores and islands. There is a basin for each shore and island, plus one for the continent pouring out to the same shore. The ID structure allows processing individually or aggregating all catchments for the shore or all catchments for islands and continents in the same shore.
- C_FRBD: aggregates of FRBDs.
- C_FRBD_BC: are the closest set of FRCS that match the RBDs (international RBD or aggregated).
- C_SS: aggregates based on Strahler levels 6 or 7 (sub-basins).
- C_SB: aggregates based on natural sub-basins; these aggregates comprise different levels. C_SB is the most detailed one, having the smallest heterogeneity in size distribution.
- C_Ctry: countries aggregates; this is the set of FECs best suited to the countries.
- C_NUTS: NUTS aggregate (pseudo-NUTS 2), this is the set of FECs best suited to the NUTS within the countries.

The C_ZG level is still incompletely defined; it shall be added during the upgrade of the 'water accounting' application in the months to come, and is likely to be completed in v1.2.

The source 'seed' data sets are copied as well, as they are the source of basin aggregation or supplementary data sets. They are:

- CTRY: source of countries delineations;
- NUTS: compilation of NUTS level x having been used for the C_NUTS;
- RBD_INT_CLP: merging of the RBD as international RBDs making or mimicking a basin (for example, the Belgian RBDs have been merged so that basins result from the merge);
- C_SB_HIERARCHICAL_CLASSIFICATION is the table that indicates the structure of C_SB. This is the only non-feature class table in the data set.

1.2.7 Point features database (EcrPFt)

Point features do not have large size requirements; it is therefore possible to combine several point features into a single data set. The other logical reason for this is that point features are updated quite often and are therefore not suited for placement in the same data set containing those major features to which they belong: lakes for dams, rivers for monitoring points, etc.

This database comprises the C_Dam feature; the dams are extracted from Eldred2 database and snapped to the segment and to the lake to which they belong.

Other point features are all those important features (e.g. monitoring stations C_Quan and C_Qual) reflecting time-related data. There is some necessary lack of synchronisation between the monitoring point features classes and the attached time-dependent values. In the case of NOPOLU Système 2, for example, time-dependent values are so numerous that they are necessarily hosted on a SQLServer database alongside the tables of monitoring points (by the end of 2011, there were over 68 million daily river discharge data, encompassing only ~ 70 % of the EU). The ideal solution of having only one such table cannot be achieved in practice.

The point features database is not part of the ECRINS delivery; however, when ready, this database shall be placed in the EEA data service.

1.2.8 Documentation

Documentation takes the form of this EEA technical report. Documentation of the WERC application is not an EEA report, but a working paper.

2 Structure of ECRINS

2.1 Conceptual model

2.1.1 Rationales

The ECRINS conceptual model has to deal with two radically different concepts that are nevertheless both imposed by EU legislation.

1. The system must be hydrologically compliant; this is built in with the development of the system.
2. The system must be WFD-compatible and must not breach INSPIRE requirements. This imposes some special conditions and precautions. However, the development of ECRINS was started before INSPIRE requirements apply; consequently, it has not been developed under the INSPIRE format. The verification on possible compliance breaches, carried out by the ETC/LUSI in 2010, did not find any particular problems (Ansorge, 2010). By contrast, the development of ECRINS v2, from the RDA process carried out under the aegis of GMES and the Directorate-General for Enterprise and Industry, will strictly follow INSPIRE recommendations.
3. The system must meet sea limits; a basin empties into a certain sea and a coastal FEC pours out into a single shore part of a terminal recipient (most often a sea, thought sometimes an endorheic or karstic system).
4. The system should be INSPIRE-compliant; this is the case in its features and capabilities, albeit not systematically so in its metadata.

The following sections describe the conceptual model representing the way the system is built, based on these constraints; details are set out in the data model in further section below.

2.1.2 Hydrological design constraints

The main orientations of ECRINS are the average area and range of size of the elementary building blocks, the FECs. The largest FECs are floored by the size of the largest elementary catchments in the CCM source. The intermediate FECs are defined by the ECRINS aggregation rules: target area and hydrological structure.

The smallest FECs mostly result from the necessity of having 'junction' watersheds to meet the branching constraint ([6]). The analysis carried out for the development of ECRINS β resulted in targeting 100 km^2 on the average that yielded ~ 92 km^2. This target size drives the type of final system and is therefore a key feature of ECRINS.

The second important technical constraint is that any watershed may have 0, 1 or 2 upstream and 0 or 1 downstream watersheds. In the event of three rivers (each having a watershed) converging at the same point (which is never the case in the real word because rivers have a width), a dummy junction catchment must be created to meet the rule above. This rule applies to the elementary CCM catchments and the process of building ECRINS strictly respects that rule. Consequently defluences are forbidden and hence deltas are not represented as such.

ECRINS is constructed to allow calculations (at catchment and drain levels). This calls for complete topological information, that can be either explicit or implicit; explicit topology was considered better for meeting the last specification set out in the Introduction, that of keeping the system free of charge, and available and accessible for the general public. The data sets are made independent of the GIS software used, a technical constraint reflected in

([6]) Nevertheless, a couple of triple branching instances were found in CCM source data sets, that were not corrected in v1.0.

the conceptual and data models, since river segment watersheds are explicitly related with appropriate identifiers. Explicit topology nevertheless allows for the use of implicit topology as 'routes' under the GIS manager chosen by the operator.

The last technical constraint is related to the choice of **identifiers** instead of **codes** in the information attached to objects. This choice is at the verge between conceptual and data model and is no longer discussed.

2.1.3 Legal constraints

The legal constraints that apply to the conceptual model derive from the WFD. The WFD has binding requirements on the definitions by Member States of RBDs and water bodies that are respectively the jurisdiction area to implement measures and the smallest reporting unit on which assessments and measures are implemented. RBDs range from quasi- hydrographical (albeit not real) systems to purely administrative delineations. Water bodies are, from a data modelling perspective, quite fuzzy partitions of the river and catchment system as it can be operationally modelled and implemented. The lake water bodies are even more complex in their relationships with the hydrographical objects, and present a full range of possible cardinalities.

The matching to a greater or lesser extent between a river basin and a RBD depends directly upon the concerns of stakeholders. For example, the Seine watershed has small upstream basin in Belgium, and an international RBD was created. By contrast, the upper Garonne river catchment (~ 500 km^2) is in Spain, but for practical reasons France and Spain decided not to create an international district; consequently, the upper Garonne is attached to the Ebro RBD (emptying in the Mediterranean), whereas the Garonne reaches the Atlantic through the Gironde estuary.

To address this issue, the RBDs have first been aggregated as 'great districts' (the generalisation applied to any RBD of the 'International RBDs'), and the concept of 'functional river basin districts (FRBDs)' has been developed. A FRBD is defined as the hydrological area, made of hydrologically consistent FECs best mirroring the 'great district', limited to the shoreline should the district encompass coastal seas. Since RBDs may flow out in different seas, there is no constraint related to the sea domain in the FRBD delineation.

INSPIRE is a general set of legal constraints that were not followed when not called for. For example, INSPIRE demands that the direction of flow be set in the layers containing segments. This is relevant when the segments are digitised from maps, but useless when each segment is related to its upstream and downstream segments by explicit information.

2.1.4 Technical constraints

The technical constraints are self-imposed commitments to make the different layers of ECRINS compact enough to be produced as a set of 'personal geodatabases (PGDBs)' (MS Access® workspaces) that can be each manipulated either by ArcGIS® and MS Access® (not processing geographical objects of course) and by some 'open source' GIS manipulating software.

In principle, this meant that the number of objects must help meet this constraint. In practice, the other specifications (e.g. FEC target size) were set independently, and the result was that the technical constraints were met.

2.1.5 Resulting conceptual model

The ECRINS v1.0 conceptual model is slightly different from that of ECRINS β: it is more explicit and includes the rivers and lakes as components. It is presented in different graphics to make it clearer.

The proposed 'Unitary catchment' in ECRINS β aiming to make an intermediate entity between the sub-catchments/sub-basins has been abandoned, because it could never be made univocally from the existing data sets. The FEC is the central source of exchange between the hydrographical entities on the one hand and the legal objects on the other, through their avatars, respectively the functional subunit for the subunits and the functional RBD for the 'great districts'. Subunits not yet fully defined at the end of 2011 are not included as feature in ECRINS v1.

The conceptualisation for rivers follows the same rationale as does the catchment.

This data model indicates that the main drains result from a selection of river segments, sorted as 'main drains' or 'secondary drains', secondary drains being all those not marked as main drains. Implicitly, the WFD 'main rivers' should all be a subset of the main drains, since they normally drain over 500 km^2 and the average FEC is 62 km^2. However, this assumption has not been verified to date; drainage areas larger than the 500 km^2 threshold should be encountered only exceptionally in western Europe.

Figure 2.1 ECRINS v1.0 conceptual model for reference watershed layers

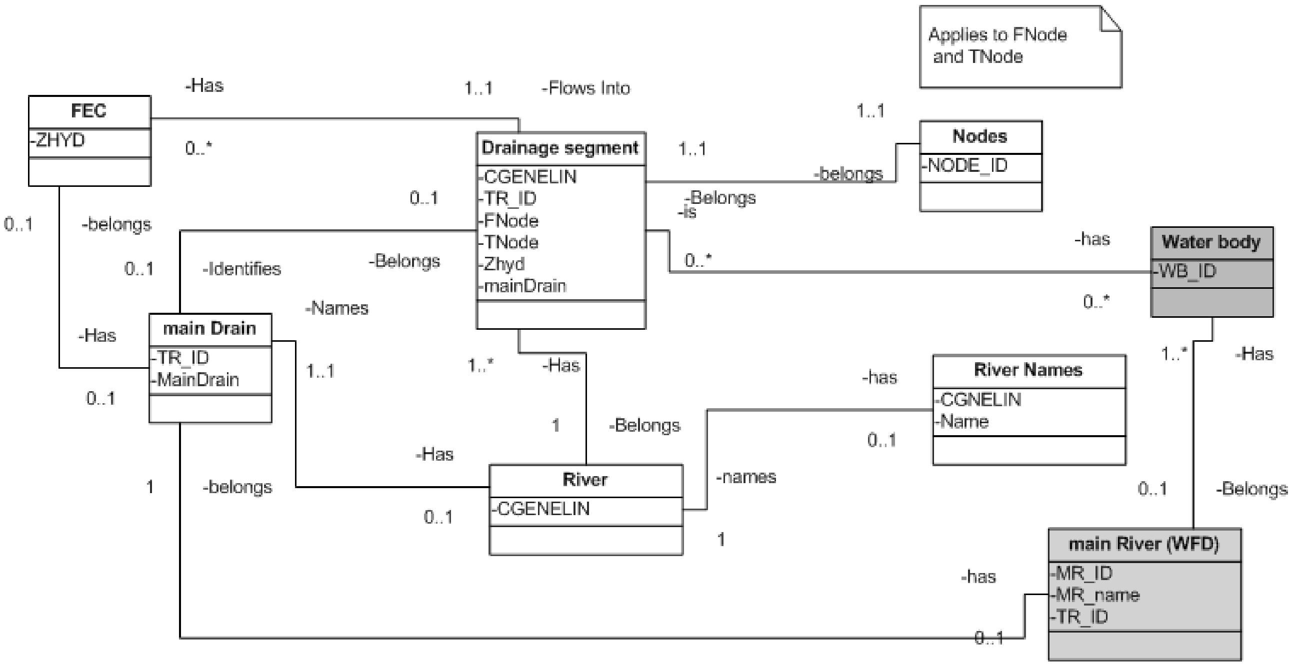


Figure 2.2 ECRINS v1.0 conceptual model for drains and rivers layers

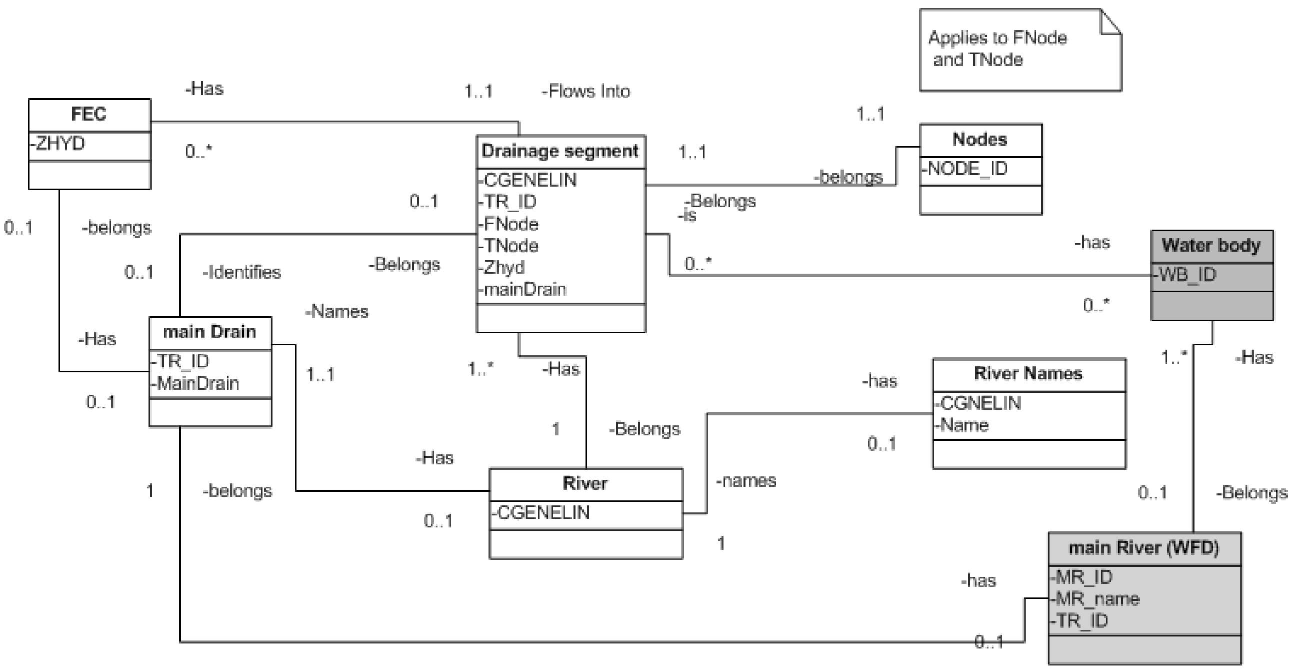
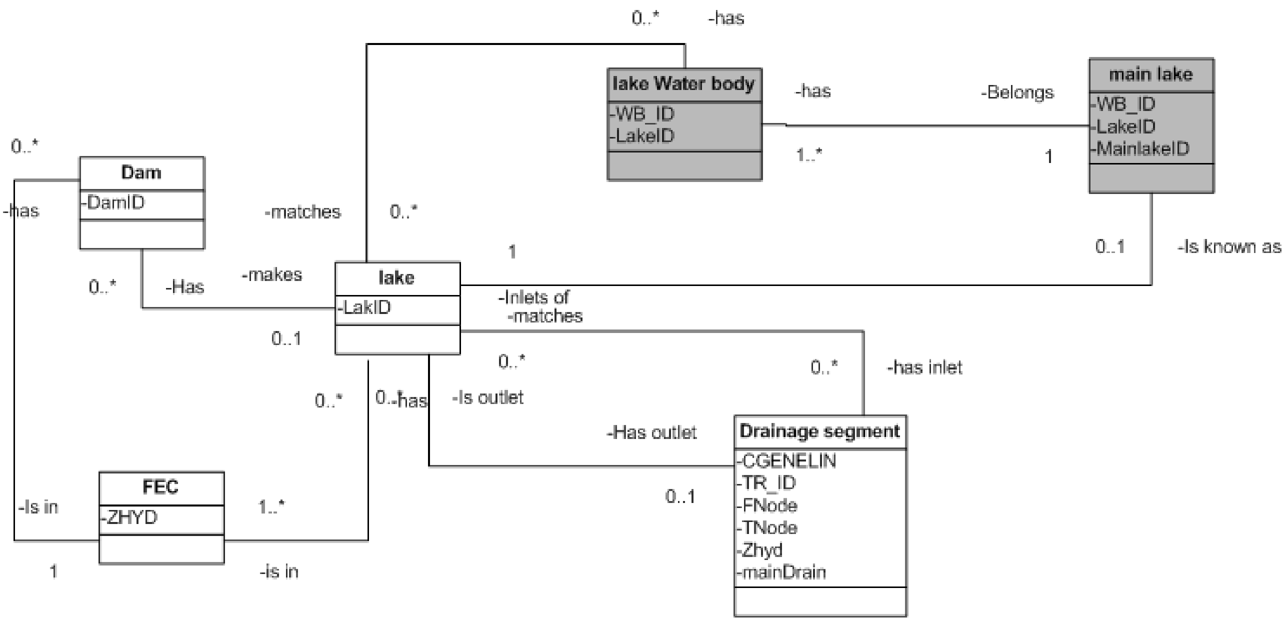


Figure 2.3 ECRINS v1.0 conceptual model for lakes and dams



The data structure indicates the following.

- Drainage segments exist in strict association with FECs; since drainage segments are produced at the most elementary level of CCM, there can be many within a FEC. By contrast, a main drain is unique within a FEC, which, however, may have no river at all.
- Dummy rivers, as a strict series of drainage segments sharing the same route ID (CEGENELIN), may or may not have a name. This is expressed by the separation between river routes and rivers. A route ('dummy river') is created by procedure and is not subject to change when the name becomes available.
- Water bodies are still quite fuzzy in definition vs hydrographical features. It is assumed that any water body may overlap one to many drainage segments, and reciprocally that they might be small enough to be many inside a single drainage segment.

It became clear during the incorporation of Art. 13 deliveries that no simple relationship could be established between segments and water bodies. In the first step, water bodies were matched to segments (with some remaining orphans if their target segment had been related to a larger water body). In the second step, dynamic segmentation was analysed and tested to make the relationship of water bodies (possibly changing) to segments one of events to a stable reference. Lack of resources prohibited developing the computation procedures.

The allocation of a drainage segment/river to a FRBD is ensured through the FEC/FRBD relationships.

The situation for lakes and dams is similar to the one described above. The conceptual model illustrates the following rules.

1. A lake may extend over several FECs, whereas a certain FEC may have no or several lakes.
2. A dam belongs at most to a unique lake (it is possible to find dams having no lake as well as having dams with no matching lake, which is a different issue); by contrast, a lake may have several dams (in Eldred2, the concept of main dams (that make) and secondary dams (that contribute) is processed).
3. A lake may have several inlets, but has a maximum single outlet (lakes next to the sea have no river outlet). Some lakes have neither outlets nor inlets in ECRINS v1.0 because the river is not represented, notwithstanding with the possible endorheicity of the lake.

A lake waterbody normally belongs to an ECRINS v1.0 lake; otherwise it flags a possible error. This is why this option is not in the conceptual model, as with the 'main lakes' from WFD reporting. The huge difference between minimum size for including lakes in ECRINS v1.0 and the main lakes makes it necessary to consider the mismatch between a declared main lake and an ECRINS v1.0 lake object as an error.

The conceptual model has been updated after processing the WFD Art. 13 deliveries; it transpired that many lake waterbodies are involved in a many-to-many cardinality with lakes.

It was also observed that in a couple of cases, artificial lakes had several outlets because the valley of several rivers had been dammed. Such special instances would be outstandingly complicated to model, and have not been addressed: the largest river has been considered a single outlet.

2.2 Data sources

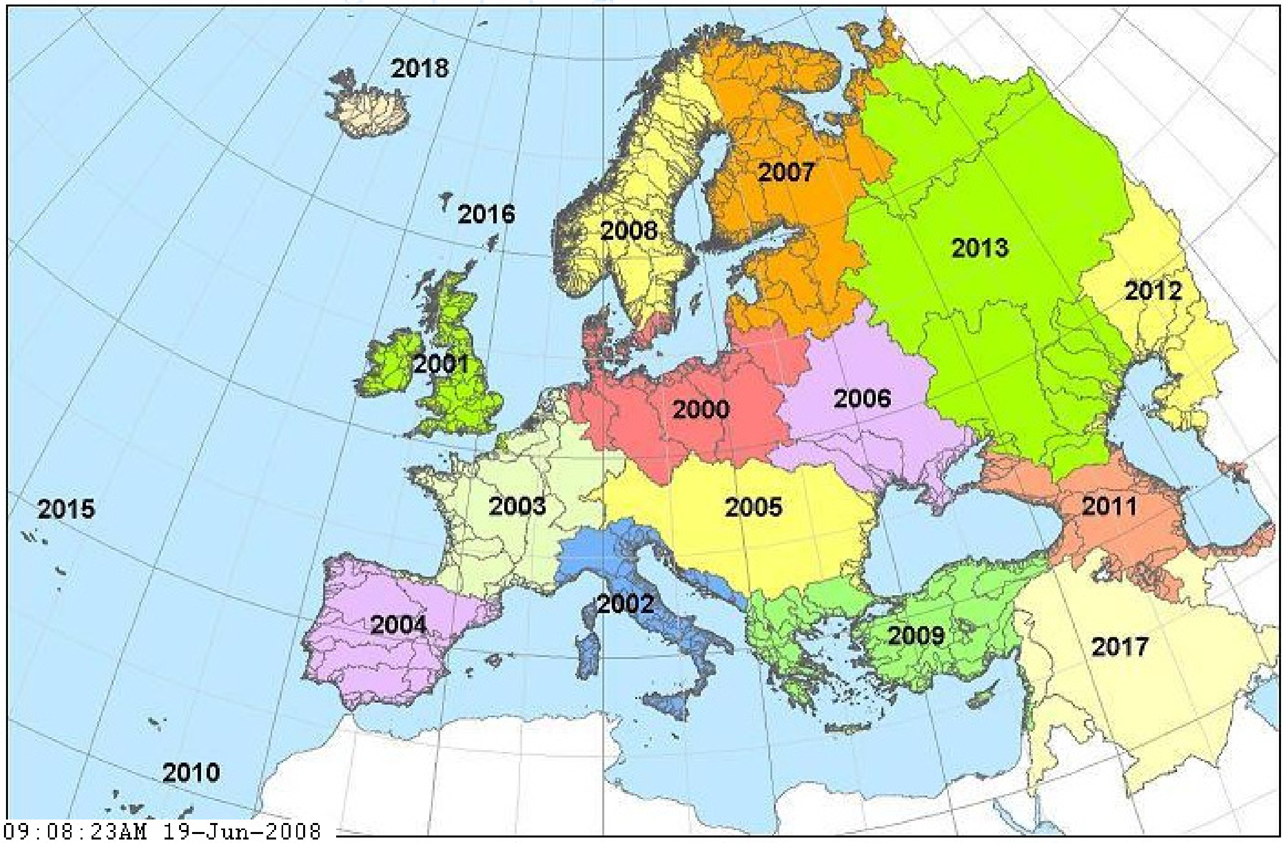
2.2.1 *Catchment Characterisation and Modelling (CCM)*

The primary data source used to carry out the production of ECRINS v1.0 is CCM v2.1 (JRC, 2008; Vogt et al., 2007). This data set was released in mid June 2008 and has been processed since to produce usable layers. The name of the data set indicates clearly that it consists of a **model** developed to **define** catchments, in this case from the DEM and the hydrology model.

The area covered by the data sets is geographical Europe plus the western Caspian catchments and the complete rivers originating in Turkey.

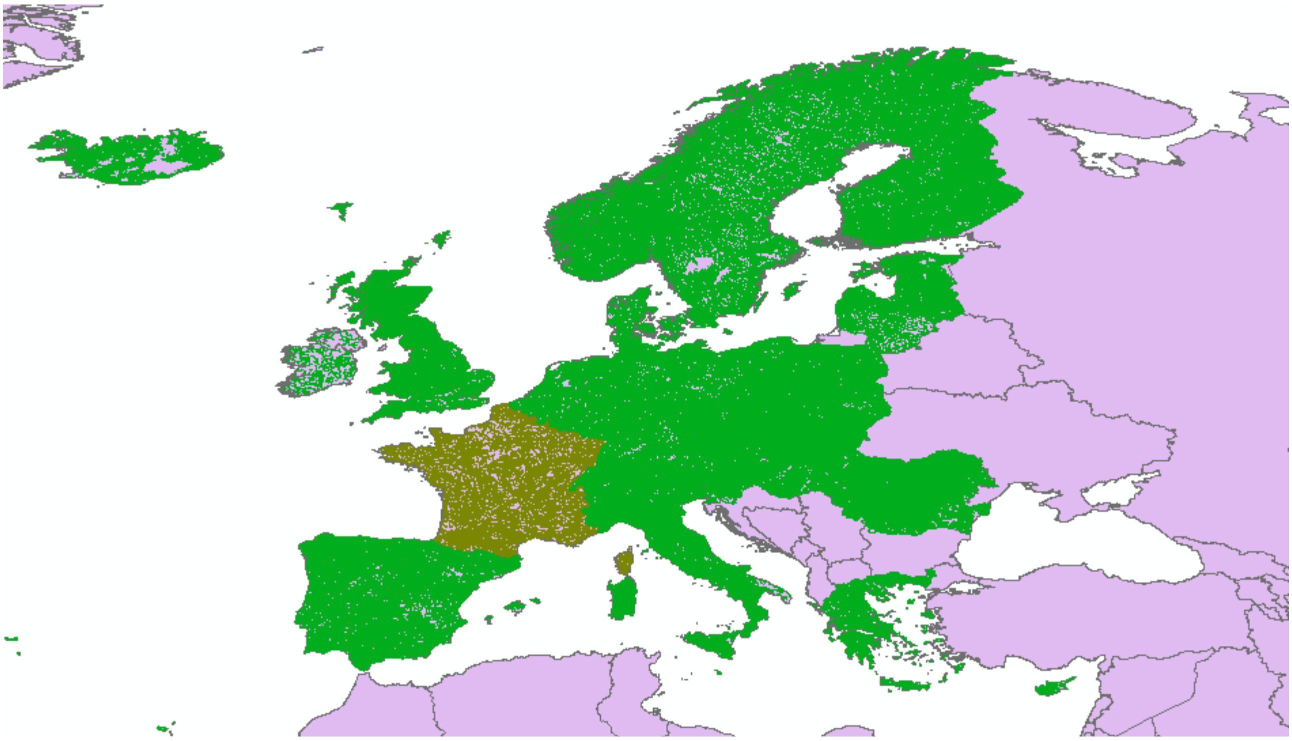
CCM is the outcome of a research programme not primarily aiming to produce a directly usable hydrological system. Its main components are the elementary watersheds, understood to be the smallest drainage area delineated by the computing under the modelling constraints imposed by the authors. These elementary watersheds are primarily defined by the accumulation flow defining a likely drain. Being the outcome of a model based on the DEM, the catchment polygons are aggregates of 1 ha (100m × 100m) polygons. Since river nodes are placed at the centre of the pixel, the downstream node of a segment is often outside the CCM catchment to which it belongs.

Map 2.1 Area covered by CCM data sets



Source: CCM report. map_21_WindowsAsFromJRC.png.

Map 2.2 Area covered by ERM data sets



Note: Map displays delivery 2.2 in green and delivery 2.0 in brown-green, where 2.2 is not present. EU and EEA coverage are incomplete compared to CCM in Map 2.1.

Source: EEA processing of ERM deliveries. map_22_ERMCoverage.png.

The data sets are provided as 'windows' that clip the complete coverage into 18 data sets of catchments and as many rivers and nodes, plus some overarching data sets that cover the whole area. The area covered and the indicative 'windows' are illustrated in Map 2.1. When considering the details, the contents of windows are locally questionable, and allow for further processing. A reallocation of items between windows is hence a prerequisite for FEC delineation; otherwise, the FECs could not be constructed simply ([7]).

Following Strahler's logic, the CCM elementary catchments are clustered into their Strahler levels 2, 3 and beyond, up to 11, the maximum envisaged in the data sets, with 10 being the actual deeper branching level. All elementary catchments converging to a single outlet are grouped into basins, which are given a 'sea outlet' code. The rules backing this clustering are analysed in Annex 2, since they are key to topological corrections in following versions and not described in the CCM report.

The sea outlets are ranked along the coastline, indicating their relative vicinity in order of appearance. The islands case is partly addressed: not all islands are declared as islands (e.g. Ireland and Great Britain are not islands, and nor is Sicily) and some islands lack a complete sorting order: some of their basins are just not populated.

The data sets comprise ancillary information that is used in data models.

2.2.2 EuroRegionalMap (ERM)

The ERM data set is a geographical map of Europe, made by assembly of national deliveries to EuroGeographics and hence limited to those countries that have delivered data. The ERM is licensed to the EEA, and the EEA had envisaged using this data source to build its hydrographical reference layer before developing ECRINS with CCM as a primary source.

([7]) It is not possible to group all elementary catchments in a single table, because this would outreach the size of PGDB and it is not possible to group in a file geodatabase since it cannot be read with MS Access®.

The data structure is described in the technical guide (EuroGeographics, 2005). The key characteristics of the ERM are that there is no topology in the maps and the features are poorly identified; as a rule, there is no unique identifier. Two distinct deliveries, covering different areas, have been used in this production to assess some quality issues and to double-check with other sources. The quality of data and the degree of populating attributes varies widely from country to country, making the usability of this data set quite erratic. The ERM has no catchment, being a geographical map and hence not addressing the modelled objects.

The main issue with the ERM data sets is the fact that data layers are licensed. The decision to not include any ERM-sourced feature was eventually taken to avoid any risk of conflict once the ECRINS layers have been disseminated. By contrast, under the licensing terms, it can be helpful for internal use, for example to cross-check data from other sources.

2.2.3 *Corine Land Cover (CLC)*

CLC is the major land occupation data set, it was updated in 1990, 2000 and 2006. The data sets comprise 3 and possibly 4 categories that may contribute towards identifying lakes: 511 (Water courses), 512 (Water bodies) and 521 (Coastal lagoons). Category 411 (Inland marshes) can be identified as lakes in many cases as well.

Systematic identification of water masses and their assessment as 'lakes' was carried out, comparing the obtained polygons from CLC2006 with reference sources, resulting in the identification of 34 333 CLC lakes have a one-to-one relationship with ERM lakes (Out of the 36 732 ERM lakes bigger than 25 ha, 35 784 are matched with one CLC lake (97.4 %). Of the remaining CLC lakes, 4 417 have a many-to-one relationship with additional 11 138 ERM lakes. The same process needs to be repeated for Greece (with CLC2000) since Greece has not delivered CLC 2006.

2.2.4 *Wikipedia*

Wikipedia is a reliable, albeit oddly populated, source of factual information. For the largest lakes, it is important to get comparative information related to their depth and volume. This information can be extracted from this source and related to the lake polygons thanks to the centroid coordinates from Wikipedia.

2.2.5 *European Lakes, Dams and Reservoirs Database (Eldred2)*

Eldred2 is the second version, following the first release issued in 1998 of a European dams database (Leonard and Crouzet, 1998). It has been developed stepwise and populated since 2004 in the EEA. The second version has not yet been fully published, despite being used in many works and provided to several organisations. The key improvements of Eldred2 are a systematic placement of dams and the possibility to refine its coordinates, thanks to the DamPos web application and the appending of modelling capacities making it possible to use Eldred2 as source for modelling migratory fish passing capacities. In contrast with the two sources described above, Eldred2 content is rapidly evolving: the implementation of dams in ECRINS v1.0 may change every time new dams are inserted and positioned in Eldred2.

2.2.6 *International Hydrographic Organization (IHO)*

The IHO (IHO, 2012) is an intergovernmental consultative and technical organisation that was established in 1921 to support safety in navigation and the protection of the marine environment. One of the activities of the IHO is the delineation of seas (IHO, 2002) that is used in many contexts.

ECRINS v1.0 required an updated sea delineation, since the coastal FECs are aggregated taking into account their bordering sea. An attempt has been made to use the draft documents issued in 2002 that were downloaded from the IHO site. Unfortunately, these later issues were withdrawn from the site on 13 January 2010. Following correspondence with the responsible IHO official, the use of IHO delineation was dismissed.

2.2.7 *SeaVoX data sets*

SeaVoX (BODC, 2012) has the responsibility, under the aegis of the British Oceanographic Data Centre, for the content management of all controlled vocabularies for the SeaDataNet and the IOC ISO19139 profile, except specific platform identifiers. Maritime organisations use code lists and controlled vocabularies to regulate the population of metadata. The appropriate denomination of water bodies from which samples are taken are of key importance for these communities.

The SeaVoX team delivered to the EEA its layer of seas called 'Object geometries for the SeaVoX salt and fresh water body gazetteer' (SeaVoX, 2009), hereafter referred to as 'SeaVoX seas' in this report.

The delineation of seas in 'SeaVoX seas' follows the same overall logic of nesting as does the draft IHO. This source is hence primarily used to carry out, where relevant, adjustments and sub-delineations to meet ECRINS v1.0 and WISE requirements.

2.2.8 GISCO data sets

Countries and regions limits are provided by the Geographical Information System at the Commission (GISCO) (Eurostat, 2012). This is a permanent service of Eurostat that, among others, manages and disseminates the geographical reference database of the European Commission. The national borders and regional limits are therefore best taken from this source.

2.2.9 Deliveries under Arts 5, 8 and 13 of the Water Framework Directive

The WFD demands reporting of water bodies' characteristics (Article 5), water status (Article 8) and basin management plans (Article 13). These different deliveries have provided the delineations of RBDs, some subunits and in 2010, lakes, rivers and groundwater waterbodies.

These later sources of information are extremely important since they address the smaller components of ECRINS: lake waterbodies and river segments. Unfortunately, not all countries had reported their 'waterbodies' at the end of 2010 and no more data are available in March 2011. However, this source of information proved to be of paramount interest for completing the lake data set and naming the rivers. Being compliance reporting, these deliveries are submitted to a complex traceability process implemented under ReportNet and operated jointly by the EEA and the Directorate-General for the Environment. Complements delivered in late 2011 have been incorporated.

The central data repository (CDR) (Eionet, 2012) stores and provides access to reported data for authorised users.

2.2.10 Other sources used

To research ambiguity of objects, their placement or relationships, two major web services are used.

- GoogleEarth®, which has the invaluable advantage of having both a very accurate coordinate adjustment (more accurate than needed, at 1:250K) and displaying aerial photography. This is very helpful, for example, if close, albeit distinct, objects from different sources apparently overlay because of differential accuracy.
- Recent developments in ArcGIS® and open source Microsoft ® products, namely ArcBruTile ([Author], 2012), allow superimposing GIS layers with maps or satellite images. This is unfortunately not possible with GoogleEarth® images, due to licensing constraints.

These sources were mostly used to better understand specific characteristics of the objects in CCM and their geographical meaning before developing the appropriate computation algorithms in WERC.

2.3 Requirements

2.3.1 Water Information System for Europe (WISE)

WISE is the reference for all elements that relate to the implementation of the WFD. Its implementation explicitly demands the reporting of 'main rivers' and 'main lakes' by Member States (EC, 2003c) within each RBD, possibly apportioned into sub-basins. The definition of main rivers is non-univocal (it has been agreed that all rivers draining at least 500 km^2 should be reported as such, even though the possibility of placing the threshold at 1 000 km^2 for 'very large basins' is offered). The 'main lakes' are those lakes whose mirror area is larger than 100 km^2. This value is downscaled to 10 km^2 in the WISE viewer.

ECRINS v1.0 had to be made in such a way that any main river or main lake meeting these criteria could be matched with a ECRINS v1.0 feature. In parallel, the RBDs and their apportionment into sub-basins should find their matching equivalents.

The case of rivers and lakes on the one hand, and of RBDs and sub-basins on the other, are not equivalent: the former are true hydrographical objects, whereas the latter are pseudo-hydrographical objects according to WISE and modelled hydrographical objects according to ECRINS v1.0. This difficulty has been addressed in the conceptual model.

The issue of waterbodies is somewhat more complex, since they are not precisely defined. The waterbody is the elementary bit of the hydrosystem that is subject to measures and assessment. The 'waterbody' has no immediately applicable GIS definition; it is determined by each Member State.

According to Art. 2 of the WFD, a water body is defined in the following terms: '"Body of surface water" means a discrete and significant element of surface water such as a lake, a reservoir, a stream, river or canal, part of a stream, river or canal, a transitional water or a stretch of coastal water'.

The Guidance Document no 2 (EC, 2003a) (in the introduction to Section 5.1), notes that Article 5 of the WFD requires Member States to identify surface waterbodies that will be used for assessing progress with, and achievement of, the WFD's environmental objectives. In addition, under certain conditions, Article 4(3) of the WFD permits Member States to identify and designate artificial water bodies (AWBs) and heavily modified water bodies (HMWBs). HMWBs and AWBs are required to achieve Good Ecological Potential by 2015.

Identifying the size of water bodies was an important parameter that had implications on the design of the monitoring programmes and on the development of appropriate programmes of measures. A stepwise process for the identification of HMWBs resulted in a provisional identification by 2004.

Article 5 of the WFD also requires Member States to analyse the characteristics of surface waterbodies and provide a summary report on surface water characterisation including general information on their typology. Article 6 of the WFD requires that a register of the water-related protected areas lying within each RBD be established. This will help to ensure that the management of the relevant water bodies also ensures the objectives of these protected areas are achieved. Annex IV of the WFD specifies what types of protected areas should be included in the register as well as what the summary of the register, which should be part of the RBMP, should include.

The above definition and recommendations for defining waterbodies do not immediately match any target features, something that is addressed in more detail in further sections.

2.3.2 *SEBI and ecological assessments*

The SEBI 2010 process was started in 2005 to provide a streamlined set of biodiversity indicators for Europe. It tries to ensure consistency between biodiversity indicator sets at national and international levels without creating new monitoring or reporting obligations. The SEBI 2010 process brings together national administrations, NGOs and international organisations; they have compiled a first set of 26 indicators.

SEBI 2010 relies on the contribution of more than 120 experts across the pan-European region and from international intergovernmental organisations and NGOs. Its institutional partners are the EEA (and its European Topic Centre on Biological Diversity), the European Centre for Nature Conservation, the United Nations Environment Programme's World Conservation Monitoring Centre, the European Commission, the Joint Secretariat of the Pan-European Biological and Landscape Diversity Strategy (PEBLDS), and the Czech Republic (as lead country for the Kiev Resolution action plan on biodiversity indicators) (EEA, 2012).

One important indicator addresses the fragmentation of rivers that impacts migratory fish journeys. Requirements for its production are the existence of topologically connected rivers, effective placement and relationships between dams, lakes and river stretches, for computing possible vs optimum routes followed by different fish species at a certain moment in time in relation to the presence of obstacles. The establishing of fish-optimum routes calls for scientific expertise (Which areas were explored in the past? Which are the preferences of fish?) that is eventually stored as a certain segment of river or a certain upstream watershed where certain species had been observed.

This SEBI-related requirement concerning fish migration is fully supported by the data model developed for ECRINS β and is retained in ECRINS v1.0.

Along the same lines, albeit not set out as such in the WISE requirements, is the assessment of potential changes in river type: damming (large or small objects) practically turns a free-flowing river (having a slope) into a still body of water (no slope, reduced velocity, but water passes through).

Fulfilling this potential requirement calls for information on river drain slopes, at least the elevation at nodes and the distance between nodes (slope= $\delta Z/D$).

2.3.3 *Water accounts*

The water accounts process is the representation of the water cycle under the accounting framework defined by the System of Economic Environmental Accounts (SEEA). The water accounts methodology is beyond the scope of this report and forms part of the overall SEEA (UNSD, 2003). The methodology of implementing the water resource accounts (quality and quantity) has been analysed and presented in a working document (UNSD, 2007). The later document has been in turn supplemented by the *International recommendations for Water Statistics* (IRWS) concerning data collection (UNSD, 2010).

The implementation of the Water asset resource accounts by the EEA has several targets that are more ambitious than those of the UNSD. The EEA wishes to derive indicators on water usage and possible vulnerabilities, carry out soundly based statistical assessments at the sub-basin level, and also expose seasonality issues. To this end, the definition of the 'statistical units' as envisaged in the IRWS is more rigorous.

Without going into details of the development of the different exercises, the Water asset resource accounts fundamentally require three classes on ingredients: a calculable reference system, data on natural assets (rainfall, run-off, etc.) and data on human uses (agriculture, urban abstractions, etc.). These requirements, from the ECRINS perspective, are analysed as the need for:

- a system of nested and routed elementary catchments (home for the elementary statistical units as catchments and rivers), used both as recipient for meteorological events and for any surface-related water use/return, of which the soil water compartment which is common to many other accounts and ecosystem assessments;
- an interlinked layer of connected underground aquifers ([8]);
- an interlinked system of main drains inside these catchments to convey run-off water from catchment to catchment and from catchment to the sea;
- an interlinked layer of lakes and reservoirs that are the saving tanks of water;
- a related layer of monitoring points (attached to the feature) to which observations are attached;
- a related layer of usages and restitutions (abstraction points, return point through sewage water treatment plants (SWTPs), etc.).

In the event that the exportable energy from the river (the exergy) (Antonio Valero et al., 2007) is addressed, the elements above and the slope mentioned in Section 2.3.2 suffice.

2.3.4 *Vulnerabilities*

The term 'vulnerability' covers many different approaches that can be to a large extent represented as 'the degree to which a system is likely to experience harm due to exposure to a hazard' (Füssel, 2007).

The development of ECRINS aimed at supporting the analysis of vulnerabilities as far as possible. What a topological system can provide is the possibility of relating targets of vulnerabilities which even the events making the harm are.

Potential vulnerabilities in relation to hydrosystems are set out below.

- Those related to floods that inundate the flood plain and may affect some targets (e.g. natural sites in the flood plain), if the source of harm exists upstream and may be inundated as well. The issue of hazards to humans is not in the mandate of the EEA; this is dealt with by the competent organisations. By contrast, the role of floods in nature conservation or in pollution risks fall under the integrated environmental assessment realm. The attachment of flood plains (in their wide acceptation of lit majeur, and not only from the observed floods perspective) is scheduled in the next version of ECRINS.
- Those related to scarcity and droughts are analysed with the water balances; ECRINS is designed to this facilitate this.

([8]) This was not part of Ecrins β and has been partly incorporated into the 2011–12 water accounts exercise, albeit not assimilated as an ECRINS feature class.

3 ECRINS v1.x data model building

3.1 Versioning

3.1.1 Rationales and core ECRINS layers

ECRINS is an evolving data set. All its components must be accurately connected to function properly; if this is not the case, the topology system will be breached. Accurate connections are ensured by the production method. By contrast, the production method complicates versioning: some data sets may be upgraded, computed again, etc. Hence, versioning is prepared during the process and inserted at the appropriate step of the production process.

From the versioning perspective, the components of ECRINS fall under one to four groups.

- Definition of the corrections imposed on the relationships of the elementary building bricks; this imposes the export version (export from most elementary data sets).
- Definition of terminal recipients that is overarching for any other ECRINS calculation; this defines the shore version (what set of 'shores' the basins belong to).
- Core ECRINS data sets, some of which are totally stable under an export version: rivers, river segments. This defines the ECRINS versions; catchment versioning is shore dependent. The combination of export + shore make up an ECRINS version.
- Complementary or ancillary data sets: names, lakes, dams, canals, big abstractions. All can be updated and assimilated under a certain ECRINS version, despite these having their own versioning.

Fields ExportVers and ShoreVers (source data and terminal recipient version, as integer), in tables LogExportvers and LogShoreVers capture respectively the export (allocation to windows and corrections) and the shore version. A true ECRINS version is a compound of ExportVers and EcrVers (recording the settings and thresholds), in each of the tables of the produced data sets pointing to an information table, copied in each database. The final version nickname, expressed, for example, by 'v1.0', is allocated by the administrator.

The reference tables LogExportVers, LogShoreVers, LogEcrVers are managed by the WERC application. The versioning data model is shown in Figure 3.1.

This featuring is implemented from ECRINS v1.0 onwards. Table LogShoreVers is updated when processing the basins table; table LogEcrVers is updated once at the beginning of the ECRINS development process.

The export version is required to ensure that the statistical table used to score the Strahler levels included in continental FECs is in line with the currently linked export.

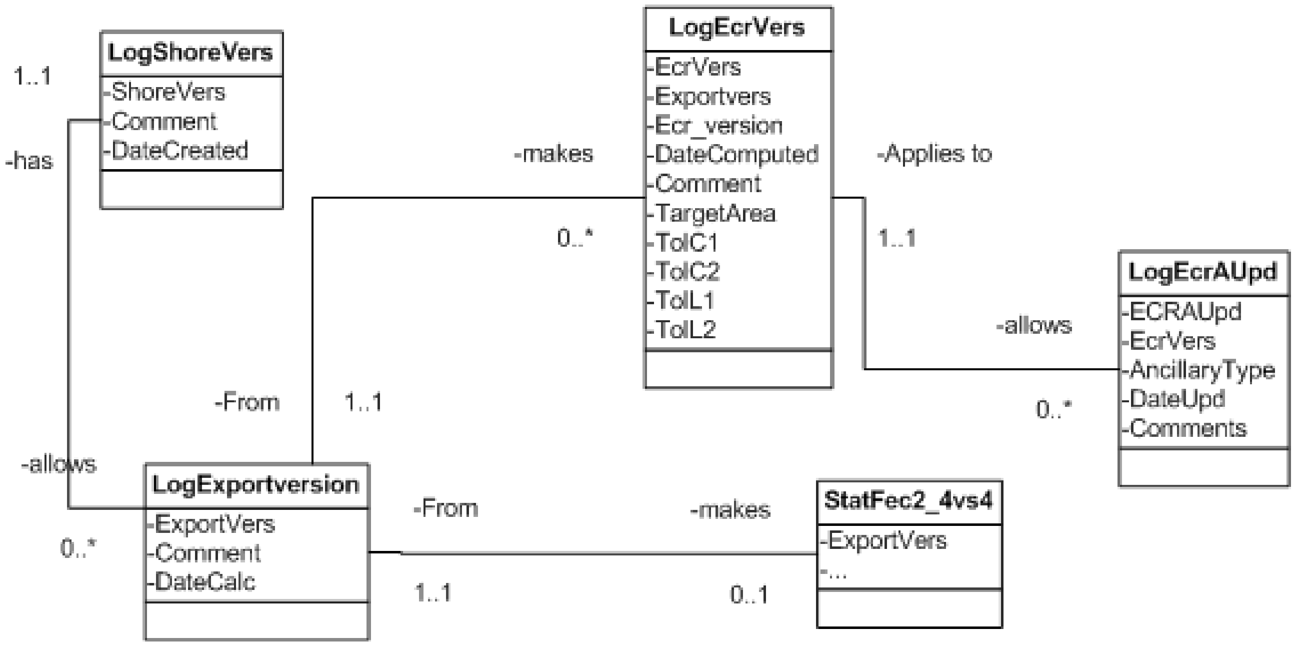
3.1.2 Ancillary data sets

Versioning the ancillary data sets is a complex issue since the semantics behind versioning are not as straightforward as they are for the core layers. This is managed thanks to an update in a central table, using keywords.

Depending on type, the logic behind versioning is as follows.

- Names: the naming level is highly dependent on the accuracy and update of the source of names. The keyword is 'NAME'.
- Lakes: the lakes layer is independent of catchments. A new lake version occurs when lakes are added or removed from the layer. By contrast, the ECRINS version attached to the lake is the current cores data set version, updated when the lakes-to-segments/FECs relationship has been established. Once entered, a lake keeps its lake version, thus allowing the history of lake insertions to be traced (however, deleted lakes are lost). The keyword is 'LAKE'.
- Marine basins: they are updated if the modification, not changing the shores, is inserted. The keyword is MASH (MArine SHores).
- Countries: Countries' limits are used to identify the FECs mostly in a certain country. Country

Figure 3.1 ECRINS v1.0 data model for versioning



delineations may change, impacting the allocation of a certain FEC to this or that country. The keyword is 'CTRY'.
- NUTS: The same stands as for countries. The keyword is 'NUTS'.
- Districts: RBD delineations are used to create the FRBDs; as they may change, their version should be mentioned. The keyword is 'RBDT'.
- Dams: Dams are regularly updated: i) new dams are located from the Eldred2 database; new dams not in Eldred2 are entered. The keyword is 'DAMS'.
- Canals: Keyword is 'CANL'.

Table LogEcrAUpd has been created and is updated every time a new ancillary feature is processed. This table has the following structure. According to the nature of the update, the update of the version is set in the appropriate table. For example, ECRINS versions are in the baseline ECRINS data set (C_Zhyd, C_TR). By contrast, these tables don't have the lake version which is found along with the ECRINS version in the Lakes dataset.

The general principle is that any update in such layers creates a temporary version number for this layer. This number is reused as many times as the insertion data set is redeveloped. Once this insertion data set is actually added to the main data set for updating, the number is frozen and the data set updated, if required, with the ECRINS version.

For example, the lakes layer inherited from ECRINS β has the ECRINS version #1, and the lakes have version #1. When preparing the first set of Art. 13 lakes to append, this data set is named ECRINS version #2 (because ECRINS v1 replaces ECRINS β) and lake version #2, until the data set is to be added to the source lakes layer.

Once appended, and after the different ancillary tables (LakInOut, etc.) are finalised, the resulting lakes layer is updated to ECRINS version #2, because all lakes are attached to segments and FECs as in ECRINS version #2, but the originally present lakes keep their lake version #1, whereas the appended lakes are marked as #2. If a supplementary set of lakes is appended, it will be named lakes version #3, with the ECRINS version keeping the latest update of #2, until a new ECRINS version is prepared.

This naming tells the user that the lake-to-river segments relationships have been computed with ECRINS version #2, and that hence relationships with former ECRINS versions are certainly false. It also flags that that the lake polygons are inherited from a previous computation.

Figure 3.2 Structure of ancillary layers update log

LogEcrAUpd

Field Name	Data Type	
EcrVers	Number	Ecrins version
AncillaryType	Text	'LAKE' Or 'DAMS' Or 'CTRY' Or 'CANL' Or 'NUTS' or ...
EcrAUpd	Number	is the version under the key
DateUpd	Date/Time	Date of the last movemùent
Comments	Text	comments
is_finalised	Yes/No	TRUE if the update has fully be integrated (see report)
Sou_tot	Number	source number of elements
Input#	Number	change (+/-) in elements
Fin_tot	Text	total number of elements= sou_tot + input

3.2 Elements of design for hydrological layers data model

3.2.1 Production-related rationales

The production of the ECRINS v1.0 data sets and the use of sources have a profound impact on the data model. The main prerequisites are the traceability of objects to ensure the production is maintained and errors corrected, and the appropriateness of the derived objects and their interlinkages. Consequently, the final data model has some redundancies and correspondence tables between identifiers.

The production and constraints for foreseen uses by the target public also impact the capabilities and known limitations of the software used for the production. For example, certain variable (field) names cause errors in ArcGIS® ([9]) and had to be avoided. The model used to carry out the water accounts and most derived applications (NOPOLU Système 2) systematically uses certain variable names derived from the French language and the French catchment GIS (BD Carthage) on which it was first developed in the early 1990s. These were retained, since the production of results would have been more costly otherwise; it seemed preferable to adjust ECRINS field names over making useless developments.

Last but not least, the final data sets had to meet the MS Access® limitations of database size (2 GB), as previously mentioned. The choice of the MS Access®–compatible PGDB has been explained in a former section. The size limit of the PGDB also acted as a trigger for aggregating the CCM-derived features most effectively.

3.2.2 Common structure of identifiers

ECRINS v1.0 and ECRINS β handle dozens of features and hundreds of attributes. The identification of the important features can be made with a code or with an identifier (ID). 'Code' and 'Identifier' are used with different meanings that are precisely defined in the definitions section in the Introduction. The ID only provides **neutral information** concerning the object, possibly including its type and location; hence it is less subject to change if the reference of the object changes. For example, if the FEC delineation changes, the FEC ID will probably change as well. By contrast, the river segments contained in it should keep their IDs.

([9]) If a feature class contains a field named 'level', this prevents the displaying of the attributes table under ArcGis® v9.3.

The use of IDs requires that the topological relationships are made explicit, a requirement which is translated into the data model to indicate the structure of such tables. The use of IDs must simultaneously meet different specifications driving the identifier syntax, that must achieve the following.

1. Be capable of handling all features and allow the appending of new ones if required,
2. Facilitate development and maintenance (and hence be intelligible for the developer and the user (in debugging, is this object a lake or a catchment?), while retaining some contextual flexibility.
3. Be short enough to take up minimum space in databases, if identifying an item with many instances of attached data. For example, each monitoring station ID may be repeated thousands of times in data sets containing the measured values, and river_ID is repeated in each drainage segment belonging to a river. Hence, each supplementary character increases the required storage space in the table proper and in the index tables as well, slowing down the data processing. This constraint concerns the development of current database managers which have practically abandoned the structures and tables of structures used in the 1980s with Programming Language One (PL/1) or Common Business-Oriented Language (COBOL), for example.

The analysis of the possible number of features (ETC/LUSI and EEA, 2008) suggested that a **10-character identifier**, made with an alphabetic prefix and any (preferably numeric) body would meet the above specifications and allow extensions. A unique length for all identifiers simplifies programming when the processing of empty records ('Null' and empty being differently or equally processed depending on software) is processed differently by the applications. The use of unique identifier lengths also led to the use of systematic jokers to fill a variable in a record when keying was required ([10]).

The selection of the prefix and its codification depends on the context; it is to some extent a code because it contains information in relation to itself. The identification aims at marking two sets of objects:

1. the source CCM objects, for which the ECRINS ID aims to inform about the nature and origin of the object, and to provide tracing information as well;
2. the new objects defined by ECRINS, for which the identification of the object is the single goal.

The general format of all ECRINS identifiers is Pssssssss where P is a prefix (always a letter) and sssssssss a nine-character body that may comprise a suffix and a leading chain.

The syntax of the prefix and the body are slightly different when considering the identification of source CCM objects turned into ECRINS objects or newly created objects. Moreover, some practical constraints may necessitate the use of some alternate methods for populating the nine 'free' digits of the ID.

([10]) A null value cannot be part of a key, and an empty one should be avoided because they create unforeseen duplicates.

Table 3.1 Source CCM objects identifier syntax

Type of object	Left part		Body	Comment
	Prefix	Suffix		
Source CCM watersheds, to be kept	'A' to 'S' ('O' not used): source window coding (Julius Caesar's), after reallocation	'01' to '11', indicating the Strahler order	Rank in the column	In source, is rank of the watershed inside window, for the Strahler order, from 1 (first of the highest Strahler order in window 1) to N (last of Strahler 1 in last window) in the set (*)
Source CCM watersheds, to be discarded	'A' to 'S'	'S1'	Inserted in the rank, as above	Spurious catchments are systematically Strahler 1
Source CCM Basins	A' to 'S' ('O' not used): source window, after reallocation	'SO'	CCM source numeric ID, padded left with zeroes	SO for 'sea outlet', plus arbitrary number computer form the minimum ID in the window and incremented. Field is WXSOID. Relationship with source is through the source WSO_ID
Islands	'X'	'I'	Source CCM island ID, padded left with zeroes. There are 37 990 islands	I included as suffix to make it possible to use X as prefix for other purposes, if required
River segment	'Z' Taken because of visual analogy	'A' to 'S' ('O' not used): source window coding (Julius Caesar's), after reallocation	Source CCM ID, padded left with zeroes (8 positions). Source window could be abandoned. In current situation, a 70-fold increase is still possible	The total number of nodes is 1 410 647 (7 digits), hence leaving space for increase in the perspective of ECRINS 2 at 1:100K
River node	'Y' Taken because of visual analogy	none	Source CCM ID, padded left with zeroes (9 positions). There is a 730-fold possible increase, allowing node creation	The total number of nodes is 1 369 641 (7 digits), hence leaving space for an increase for ECRINS 2 at 1:100K

Note: (*) See Annex 2 for information on the way the Strahler order is imposed on catchments.

The source CCM objects (including those partly taken from CCM) are systematically identified as ABByyyyyyy. 'A' is the prefix, 'BB' the suffix and 'yyyyyyy' the body. The choice and meaning of prefix and suffix are reported in Table 3.1. The body string is padded to the left with zeroes or underscores, so that no space appears in the string.

The syntax of the objects newly created in ECRINS can be a bit different, although they always have a prefix letter. If, for any reason, the prefix letter has had to be used previously for CCM objects, the suffix is such that no confusion can be ensue. In principle, and if no errors occur, IDs within ECRINS are all unique. The syntax for these objects is reported in

Aggregation entities are either catchment type (having hydrological consistency) or administrative entities. All the aggregation entities are stored as follows: i) field populated in C_ZHYD, and ii) as dissolved feature class in the EcrAgg PGDB.

Table 3.2 ECRINS objects identifiers syntax: basic objects

Type of the object	Left part		Body	Comment
	Prefix	Suffix		
Primary basins	W	'som', where: s is the Strahler level of cutting/ aggregate om are ocean and sea system, as below	Sequential number, created in the order of making	All basins are Wsomxxxxxx, being the container of the FECs. Coastal catchments are both container and FEC, continental catchments are fully inserted into a primary basin. Field is Bas0_ID. W is for W(atershed)
Ecrins FEC	If the FEC is coastal, is 'W' If FEC is continental, is 'A' to 'S' ('O') excluded	If the FEC is coastal, is System + sea code (1 char) If FEC is continental, is the Strahler order of the source (in practice '02' or '03')	Sequential number, in the order of catchment sorting. The type of catchment (aggregate, singleton or continental) is indicated by the absolute value. See Section 5.3.5, page 47 for complements	This is the only circumstance where an ID may be duplicated between categories; this was implemented for development and maintenance reasons. This ID syntax ensures full traceability with source data sets
River code (= points to names)	'Z'	River pre-existing in CCM, '_C' River created from naming, '_E'	CCM pre-existing rivers are 1 480, their body is 0000001 to 0001480. Added rivers are per basin groups, same as for CGENELIN	The number of allocated rivers/routes is computed from the maximum between number of springs and of Strahler level changes. The resulting number is sorted by classes 10, 100, 1 000, etc. Final allocated range is computed by Bas0_ID, and allocated to allow enough supplementary space
River connecting code: CGENELIN	'Z'	'_G'	Body is numeric 'xxxxxxx', as truncated 10000000+rank per basin	
Lakes Lakes do not exist as such in CCM. There is a layer of polygons, not CCM related	'O' Taken because of visual analogy	'OC0' is source is CCM water/ lakes layer, 'OE0' is source is EEA and 'OCC' (CC= ISO code of the country) if resulting from country source	From the source object ID, by simple incrementing	Letter 'O' is not used as window and mimics a pond. Latest update provided. In future versions, all lakes should be either 'OE0' or 'OEcc' prefixed
Dam	'D'	'cc', as the ISO country code	Eldred2 code plus a '00' inserted right of the left part, to match ECRINS syntax	D is kept to stick to Eldred2 syntax which is based on 8 character Dccxxxxx code, that is expanded to Dcc00xxxxx, giving a 100-fold extension possibility and capability of managing all weirs

For historical reasons, and owing to many developments carried out with NOPOLU Système 2, the field names populated by the aggregates are inherited from this application. They are set out in Table 3.3 below.

Table 3.3 Syntax of nomenclature of aggregation entities set in ECRINS

Definition	Field name	ECRINS ID structure	Comments
Natural large basins pouring out to the same shore. A shore is defined as the contact between coastal FECs and a sea, on the same continental mass	B	WBcsxxxxxx where CS are respectively the ocean system and the shortened sea code. xxxxxx is numeric and contains the island id (rightmost part of the island ID in ECRINS) or '_____C' for continents	Comes from French BD Carthage 'main basins'. Joker values are 'WB_#Joker#'. Corresponding feature class is C_B
Natural sub-basins, part of FRBDs and based on Strahler level aggregates. Mnemonic is 'Subcatchment Strahler'	SS	WSSxxxxxxx where xxxxxxx is based on either the Strahler level aggregates (where relevant) or on sum of areas where Strahler level is not relevant	Comes from French BD Carthage sous-secteur. Corresponding feature class is C_SS
Natural sub-basins, based on geographical terms, where possible being sub-basins of FRBDs or FRBD aggregates	SB	WSBxxxxxxx, TBD	Computed by watershed of large affluent to rivers. Corresponding feature class is C_SB
n/a	ZG	WZGxxxxxxx	Was used in ECRINS beta version. Corresponding feature class should be C_ZG
n/a	BV	WBVxxxxxxx	Was used in previous versions for the French accounting catchments 'BV RNDE'. Corresponding feature class should be C_BV
Functional subunits. A subunit is a partition of RBDs, as defined by the MS. A functional subunit is the mirroring of the subunit, having the same hydrological consistency as the FRBDs	FSU	WFUxxxxxxx, TBD	Not yet implemented, owing to no available source of data. No related feature class
International RBDs. They are made from the breaking down of the national RBDs into a consistent international RBD where relevant, whose outer limit is taken from shoreline	FRBD_BC	WRBbbbbbbb	Field name to recall that they are basis for FRBD, _BC stands for 'base catchment'. Corresponding feature class is C_SRBD
Functional RBDs. They are the transcription of the FECs belonging to a FRBD_BC, plus correction to ensure hydrological consistency using the natural basins, as defined by CCM	FRBD	WFBbbbbbbb	This construction is primary to other constructions. Despite slight geometric differences, the trailing 'b's are the same in FRBD and _BC source because they designate the same entity with different accuracy. Corresponding feature class is C_FRBD

Table 3.3 Syntax of nomenclature of aggregation entities set in ECRINS (cont.)

Definition	Field name	ECRINS ID structure	Comments
Natural large basins pouring out to the same shore. A shore is defined as the contact between coastal FECs and a sea, on the same continental mass	B	WBcsxxxxxx where CS are respectively the ocean system and the shortened sea code. xxxxxx is numeric and contains the island id (rightmost part of the island ID in ECRINS) or '_____C' for continents	Comes from French BD Carthage 'main basins'. Joker values are 'WB_#Joker#'. Corresponding feature class is C_B
Country, as represented by all the FECs (hydrologically connected or not) whose centroid falls under the country	Ctry	WCYxxxxxxx	New feature of ECRINS. Chain xxxxxxx is #Joker# in case no match is found. Corresponding feature class is C_Ctry
Region, based on covered area criterion, so that the region that apportions a country is in the average range of the French/Spanish NUTS 2 taken as reference	Nuts2	WNTxxxxxxx	New feature of ECRINS. Chain xxxxxxx is #Joker# in case no match is found. Corresponding feature class is C_Reg
Field used on request	ZU1, ZU2	Should be WZX and WZY	To be used on demand, for local applications, if any. Inserted for compatibility with NOPOLU; no envisaged systematic feature classes

3.3.1 Taking data sources into account in the data model

Since ECRINS is constructed from various data sources each having a strong specific structure, they deeply influence the data model of the resulting product. The most constraining features of CCM are the Strahler order of catchments and the definition of basins. These structures are the key drivers for the building of ECRINS.

The other important structuring feature is related to prioritising topology over all other considerations. Most river systems are tree-like; in these cases, from upstream to downstream, there are only confluences. In the real world, there are defluences. River deltas are the best known example. As things stand with processing systems at the moment, there is no affordable way to take such features under consideration. Defluences **cannot exist in ECRINS** and hence are not modelled as rivers; a special conceptual and data model have to be defined to include these features in the topology.

The plan is to handle defluences practically, and model them as small rivers and a couple of outflow/inflow items, which makes it difficult to keep all information, except discharge. This implementation is scheduled from v1.1 onwards.

3.3.2 Elementary catchments key featuring

An 'elementary catchment' is the smallest areal entity participating in surface drainage. Their structure has been described previously.

The method by which the Strahler level is given to an elementary catchment aggregate is not documented in the source publication (Vogt et al., 2007); the analysis carried out at the EEA resulted in the rationales and rules that are detailed and exemplified in Annex 2.

This clarification is essential for planning the topological corrections that will form the keystone of version 1.1 production. The elementary catchments are the single source of topological information and are the target for adjustments in the case of correction of connecting errors that will be applied to river segments. Any change in their connexions must be reflected in the composition of the upper Strahler level aggregates and must be done automatically.

In source CCM the IDs are numeric and carry no information about the object category. This has been changed in ECRINS to make processing more maintainable. The CCM elementary catchments are hence in a 0-1 to 1-1 respective cardinality, elementary catchment to drainage segment and drainage segment to catchment. The table of ID conversion is hence a key working table of ECRINS.

3.3.3 Aggregation entities

As stated as a constant baseline for ECRINS, all aggregates are based on FECs. Hence the final resolution of the aggregate and its possible difference with the reference layer depends on the size of the FECs and on the greater or lesser relevance of the aggregate vs the catchment delineations.

All aggregates are hence either defined from a calculation carried out on FEC features (their basin, the seashore, etc.) or derived from the previous overlay between the FEC layer and the source information providing the delineation of target aggregate system. Depending on issues, the FEC-to-layer comparison is carried out by intersecting or by seeking out the overlay between target layer and FEC centroids. To achieve this, the FEC layer is systematically converted to a centroid layer provided as an ancillary data set, taking into consideration the possible misplacements of the centroid from ArcGIS®. In the most complex cases, the drainage system is also used to finalise the aggregation.

The main difficulty in the definition of aggregates is the quasi-legal obligation to make the aggregates match with the WFD where it applies, while covering the whole ECRINS area with aggregates.

3.3.4 Drain, nodes and rivers

In the source CCM data set, the elementary catchment-to-drainage segment cardinality is 0-1 or 1-1. The cardinality has dramatically changed between the two latest versions of CCM, the former having thousands of 'spurious catchments' that were eventually aggregated with true catchments in the latter version.

Since FECs are aggregates of CCM elementary catchments, each one potentially contains several drainage segments. This made it necessary to develop the concept of 'main drains' composed of those drainage segments that relate upstream FECs to their downstream peers. By contrast, the non-main drainage segments are secondary drains. Segments start and end with nodes. The cardinality between segments and nodes is unchanged from CCM to ECRINS.

In source CCM, a node is by definition the junction of two segments, except the starting node (spring) and outlet. Only springs have a 1-1 relationship with segments; some outlets are 'V' shaped and hence have two upstream. During the naming and routing process, it was observed that a huge number of segments were spurious as well, and consequently 'shaved'. As a consequence, the node receiving these segments now has a single upstream. Of course, no spring segment has been 'shaved' to preserve Strahler levels.

3.3.5 Lakes

The layperson understanding of a 'lake' is simple; when addressing it from a GIS perspective it becomes extremely complex: rivers and lakes are both geometric objects and cultural objects. In ECRINS, a lake is a polygon representing a water mass and having an ID.

The different sources of information relating to lakes provided as many possible cardinalities as was feared, and the inclusion of WFD 'lake waterbodies' introduced still more complication; some waterbodies are parts of lakes and even parts of lakes or collections of lakes.

From a data modelling perspective, a lake (when identified as single polygon) has inlets and a single outlet, and is inside, or shared, between FECs. It may have inlets and no outlet if the lake is coastal, in which case its outlet is a sea.

A pending issue is the case of endorheic lakes that are not part of the catchment area in CCM. This makes the watershed of endorheic lakes inaccurate. This is not significant for systems having large watersheds, but significant in the case of Trasimeno lake (Italy), for example.

4 ECRINS data sets preparation

4.1 Rationales

Building a version of ECRINS requires accurate data preparation so that no error or omission of watersheds occurs during the calculation process. The calculation process is something of a craftsman's production assisted by a computer, and calls for a sequence of MS Access® processing under WERC and feature-building with ArcGIS® processes: the limited number of times the process has to be carried out does not deserve full the kind of automation that would probably be impossible to develop. Since this handling of procedures under different software may induce errors, care has been taken to limit the steps to the minimum and to keep track of the calculations.

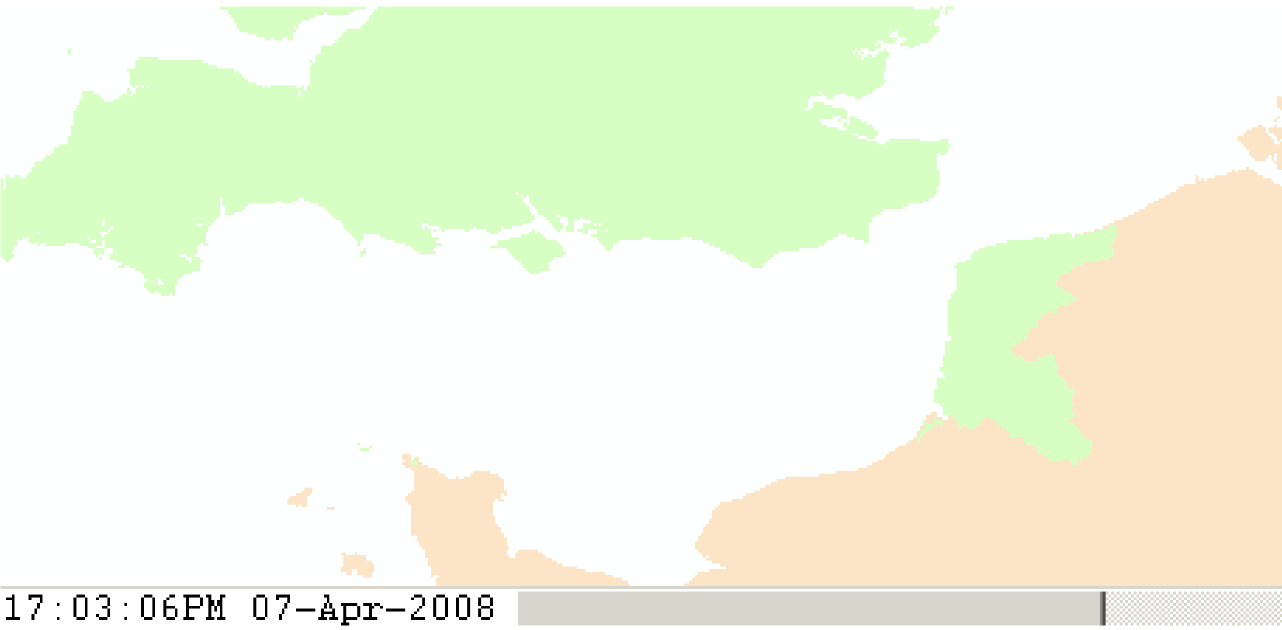
4.2 Reallocating CCM windows

The source delivery has been performed as 18 'processing windows', having limited hydrological meaning since they result from clipping the whole CCM area into rectangles of handleable area. This resulted locally in erratic distribution of catchments between windows, exemplified in Map 4.1. The source of this misallocation is precisely the clipping process that takes full basins. This misallocation jeopardises the process of computing coastal FECs that require processing all catchments belonging to a certain shore. The reallocation from 'source' to 'data' windows is corrected as a preliminary step. The selection of the basins to be moved is done by hand and stored in a data set to allow further updates and corrections and new computations. Hence the 'source' data sets are not processed after this.

A second objective of reallocation is to balance the size of the window tables; the largest (north Caspian, _2012) is limited to the Volga basin, the other catchments being set to _2011 (western Caspian).

The reallocation of windows consists in creating intermediate feature classes that contain the basins to be moved, then moving them (appending to target and removing from source) with an

Map 4.1 Misallocation of regions (from _2001: British Isles to _2003: France and Rhine)



automated script under ArcGIS®. It was made necessary by the fact that the data sets are either kept in ArcGIS® file geodatabases (that cannot be processed by database managing programmes) or consolidated so that actual coastal FECs could be produced.

4.3 Seas and shores: terminal recipient outlets codes

4.3.1 Rationales

To make accurate FECs, the elementary catchments and basins must be connected to the correct shore. A shore is the contact line between sea and the continent. Since the coastal FECs are made by aggregating together basins of any size, the shore must be continuous; otherwise a resulting aggregate might be split into two disjoined polygons.

The CCM sea outlets implicitly define shores following the definitions accepted at the time of CCM processing. The delineation of sea outlets is more precisely a pseudo-codification of terminal recipients, since it considers both seas proper (any body of water directly connected to the interconnected body of salt water often referred to as the 'World Ocean') and endorheic lakes (e.g. the Caspian Sea). Following limnologist definitions, any still body of water not connected to the global ocean is a lake (SeaVoX, 2009, p.1). Any lake or terminal recipient not connected to the river system is an endorheic system that acts as a terminal recipient, analogous to the sea.

Since ECRINS is a system including topology, the information on the terminal recipient is essential. CCM had proposed and used a codification of terminal recipients, named 'seas' for practical reasons. What ECRINS v1.x needs are 'shores' that can be defined only if the corresponding seas are defined first.

The final ID of any 'sea' in CCM is System + Sea (e.g. the Baltic Sea is A6). This coding, appropriate for CCM, is no longer suited to ECRINS v1.x since it is not in line with the MSFD definition of seas and it renders it impossible to update the endorheic recipients, if required. Endorheic systems are coded as distinct systems: over time, a lack of letters may present, since endorheic systems are by definition all distinct and hence potentially extremely numerous.

The identification of terminal recipients is attached to the basins (the outlet of the system of elementary catchments) and had to be revised prior to further calculations.

Map 4.2 ECRINS v1.x terminal recipients used to allocate basin outlets

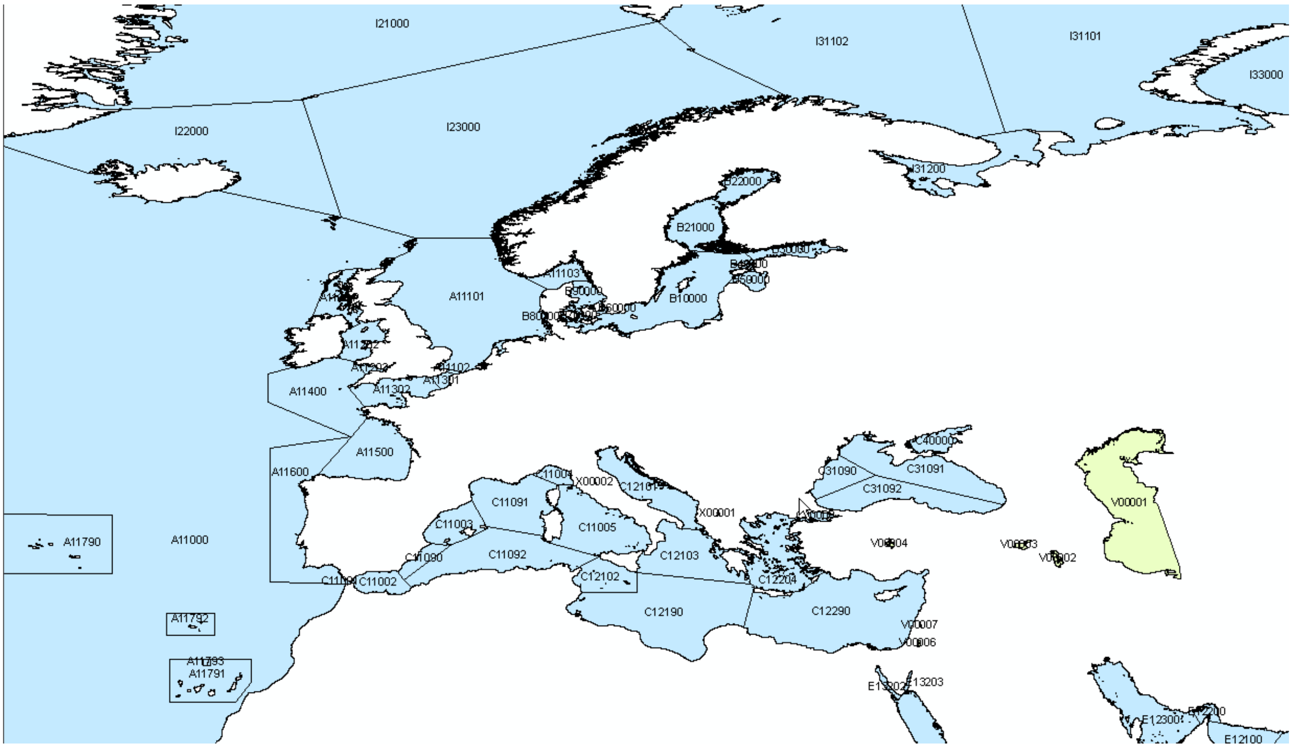
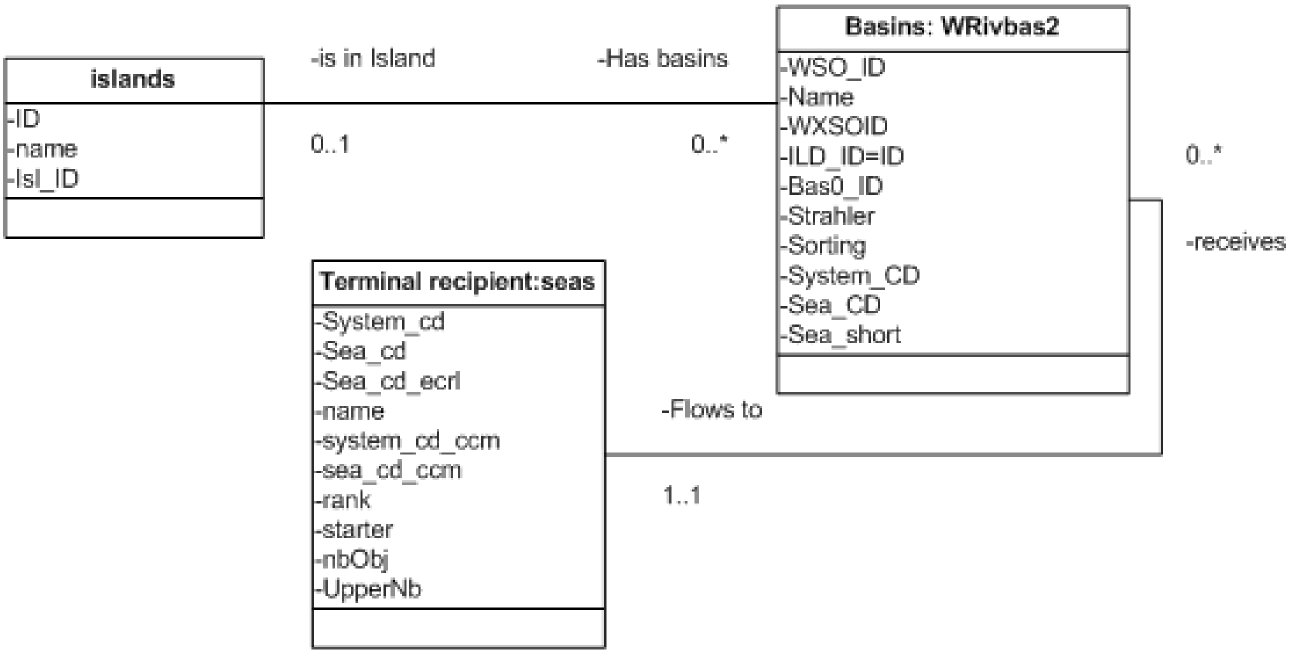


Figure 4.1 Local data model of basins



The new codification of seas for ECRINS v1.x is based on two simple principles.

- A 'system' is the largest recipient. From the sea perspective, there are 8 to 10 major oceanic systems. Logically, the continental endorheic systems are coded at the continent level, mirroring the oceans and requiring seven letters only (one per continent).
- The seas (connected to the World Ocean) within each system are coded from the hierarchy proposed by SeaVoX that suggests three levels plus a sub-sea. This codification serves as the basis for the definition of shores.
- The endorheic systems are coded by a systematic number inside a system since there is no hierarchy between them.

At the end of the process, the original ID of seas in CCM is radically changed, which in turns makes it necessary to adjust the former island coding to match the new definition of islands (Table I.1, page 13) that now is identical to that of a layperson's understanding. The sea codification follows the nesting principles defined by SeaVoX but adjusts them to match ECRINS processing application requirements for shores.

The final codification for structural reasons is a 5-digit integer, with an observed maximum of 19 shores in a single System_CD. A sub-codification for internal needs is therefore implemented as a single character 0-9, A-X allowing FEC identification to be kept unchanged. Since it is unlikely that more than 36 endorheic systems would be considered within an continent, this trick is acceptable.

The equivalence is given in Annex 2, Table A3.2, page 95.

Compared to ECRINS v0, the sea definition variables are slightly different: System_CD and Sea_CD contain the new codification of terminal recipients. The added fields are the source CCM system and sea codes, plus working fields that make the shore delineation more flexible and preserve the FEC-building requirement of considering which basin is next to another.

4.4 Completing island codification and sorting

4.4.1 Basin- and island-related data model

Basins, seas and islands are tightly interconnected.

The data model in Figure 4.1 above shows that an island may have several basins, and that a basin belongs to a single shore (as identified by a terminal recipient code), hence making it possible for any island to be surrounded by many seas. This is a substantial change compared to ECRINS v0. This change offers more flexibility in updating the shores that are now related to a single entity, the basin.

4.4.2 Issues and rationales related to islands

As defined, an island is composed of basins not belonging to the main continental mass. The new definition (see definitions in Table I.1, page 13) allows changing the seashores. This operation should be considered carefully because the knock-on effect is that the entire ECRINS may need to be recomputed.

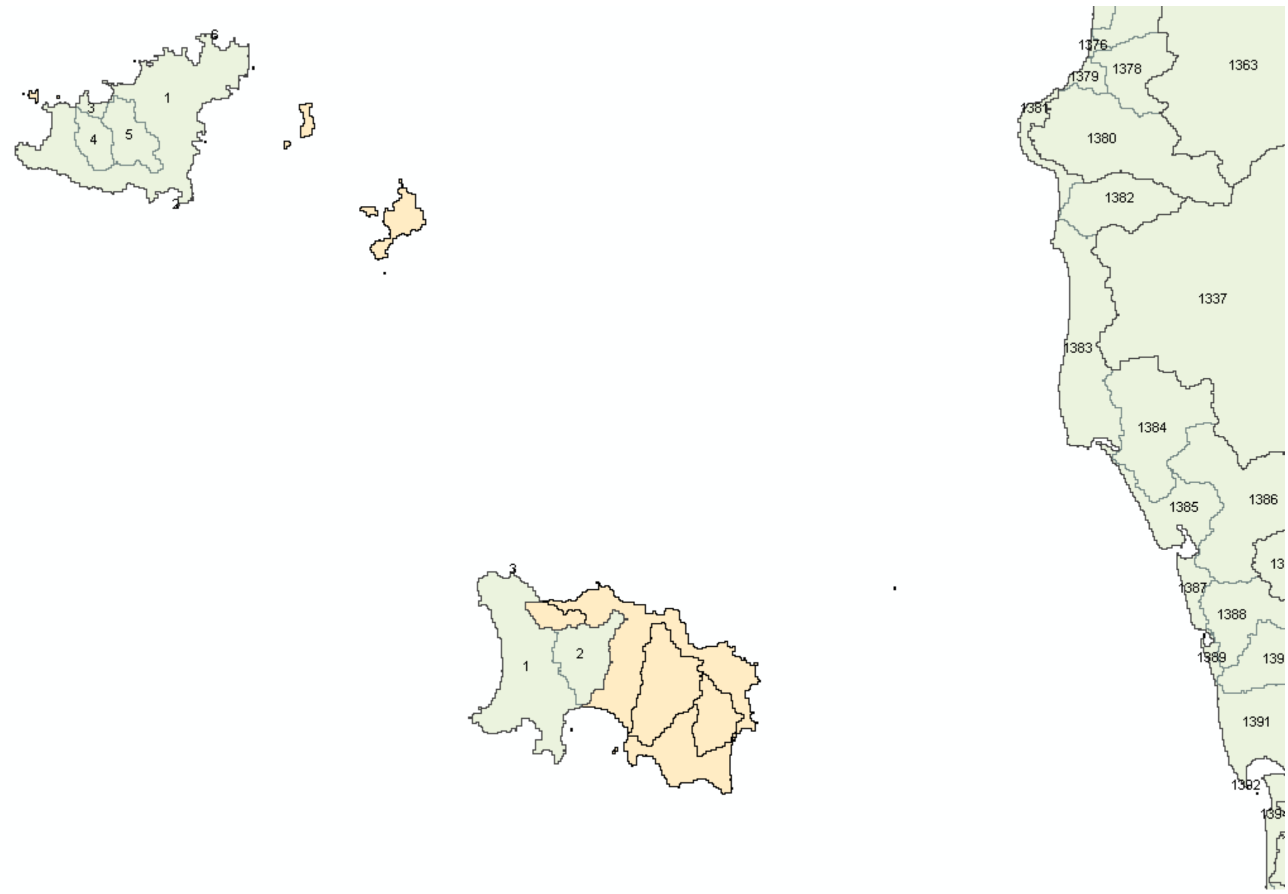
The source CCM considers neither main Ireland nor main Great Britain nor Sicily as islands. This has been changed, and the CCM-like IDs of 37900 and 37990 respectively were given to Ireland and Great Britain, and 37950 to Sicily. Then, the islands features were created and added to table islands (making it pass from 36988 to 36991 objects).

The changes inserted as seas make it possible (as in ECRINS v0) for an island to be surrounded by several seas. This renders obsolete the CCM sea coding in the islands table. In the new version, islands are also given an ID in the ECRINS syntax. Islands are prefixed 'X' and followed by their ID in the source data set.

The basin sorting issue is a key element of the building of coastal FECs. It is an arbitrary number that numbers the basins and marks their relative position (the basin immediately to the right of the preceding one). Ordering is complete for continents (and the largest islands). CCM carried out the sorting per shore, as defined by CCM, which makes for obvious duplicates and risks of errors in the new shore definition.

To correct this, a function has been added in WERC so that the basins are counted and a new unique ranking ignoring the shores is used instead of the

Map 4.3 Disappearing part of island> incomplete sorting values



Note: The numbers indicate the sorting values. Several basins in Jersey have no value populated; the corresponding catchments will be lost in further steps. Guernsey island is complete, and Sark island is entirely not populated.

Table 4.1 Distribution of basins per shore in source and corrected CCM

SYSTEM_CD	SEA_CD	Number of items	Number of items, after correction
		603	0
A	1	6 361	6 361
A	2	8 057	8 059
A	4	2 677	2 677
A	5	6 937	6 950
A	6	30 779	31 009
C	1	1 668	1 715
H		4	0
H	9	189	306
I	2	1	1
J	9	1	1
M	2	3 828	3 832
M	4	11 815	11 824
M	5	2 890	2 888
N	7	315	316
N	8	1 207	1 239
N	9	2 021	2 025
Q		2	0
Q	9	81	89
S	1	863	863
T	9	16	21
V	9	135	149
Z		2	0
Z	9	97	224
		1196	1346

CCM one. These modifications allow any further sea delineation to be considered and operated without error.

By contrast, some errors are still present; they relate to the absence of SORTING for some basins. In the absence of sorting (SORTING = 0 or SORTING = null), the basin is discarded because it becomes impossible to aggregate catchments. This may have unexpected results. For example, half of the island of Jersey has vanished owing to this lack of SORTING in the eastern part.

As a result, the FECs (light green) may miss a part of an island (remaining beige background).

The changes in island attachments are reflected in the updated data model: terminal recipients are information at the basin level; the cardinalities make it possible for any island to have several shores, depending exclusively on its feature in basins. The data model is common with basins, as reported in Figure 4.1, page 42, making the basins in islands fully equivalent to basins in any terrestrial area.

4.4.3 Completing data for sea outlets

The 'basins' table has been updated after having selected back all the basins within the envelopes to take into account the insertion of new islands (Great Britain, etc.). This operation helped discover that in CCM, many basins have no proper sea outlet coding. This has been corrected, as displayed in the correction statistics reported in Table 4.1.

After correction, the number of basins posing problem drops to seven; this relates to areas declared as basins and that are situated in 'the middle of nowhere'. These basins are all spurious and are made up of a couple of pixels.

During this correction process, it came obvious that many geographical islands are not set as islands in source CCM. This is all the more the case in the eastern basins not in the EEA area. In ECRINS v1.0, this is not fully corrected since the errors were detected at an advanced stage of ECRINS development.

4.5 Summary of preparatory processes

The source data sets are processed in two distinct stages.

- Correcting major data inconsistencies and errors locally, to create new calculable data sets (this is done once only and is only reported in the preparatory report).
- Computing the ECRINS identifiers and building the additional fields, data sets, etc. This operation should preferably be performed anew for each new version. This is carried out thanks to systematic and calculable data set (databases and tables) nomenclature. This second stage comprises many steps that are organised from the least (e.g. creating ECRINS IDs for the source CCM features) to the most frequent (e.g. adding new dams to the lake layer).

Drawing up preparatory data sets follows a procedure that is articulated between WERC and ArcGIS®. The analysis of original IDs only has the historical function of counting the original Ids, and was used to define the syntax of the ECRINS ID.

The definition of shores is an important step that governs all others; once completed, the original CCM basins must be populated with the new shore information. The database names in the figures are explained in the next section. Defining new shores is not identical to substituting shorelines.

5 Building ECRINS

5.1 Programmed steps

The key steps, once source data has been secured and pre-processed are as follows.

1. The creation and populating of continental FEC envelopes is made by selecting the appropriate Strahler level, that can be 2 or 3. This is based on a scoring method, which requires prior assessment of size distribution that is supported by the application as well. The produced table analyses for each aggregate level 4 (3+1) the distribution of catchments on the 2 lower levels, and at the same time, the preparatory information for retaining the links (what flows to what?).
2. Populating the continental FECsflat table with the appropriate grouping of Strahler 1 level catchments. This step involves flagging each CCM elementary catchment with its computed FEC. Again, this step requires a decision from the operator on target size and tolerance around target size.
3. Under ArcGIS®, making the FEC candidate feature class from the source feature classes containing the appropriate Strahler groupings. This step consists in dissolving the continental FECs from the CCM elementary catchments based on flags set by WERC, and appending the coastal FECs, ending with a complete feature class of FECs. This feature class is incomplete at this stage and should be updated in the next steps.
4. Post-process the FEC tables with appropriate relationships, append to them the important featuring required by further applications (upstream and downstream FECs, arborescence code defined by NOPOLU Système 2, ID of the aggregation catchments, summed area upstream). This is done using the facilities launched by special button in WERC.
5. Create from internal sources or from external overlays the aggregation catchments, and populate the C_Zhyd with the FECs to aggregate relationships.
6. Create drains and compute, in relation with their hosting FECs, the distance to the sea of any segment (this is partly done by CCM, but it is very incomplete), complete with the main drains.
7. Link the lakes to the drains and the FECs. This step can be carried out at any time after step 6.
8. Create the routes, that are independent of names and that will become the mask for naming in next naming versions.
9. Create rivers by transferring names from external sources to segments; build true and dummy rivers.
10. Carry out the needed ancillary production.
11. Copy the finished feature classes as final C_XXX tables in the Ecryyy PGDB.

5.2 Supporting application

5.2.1 Justification of need

All the steps needed to create an ECRINS release must be carried out in the appropriate order, and the identification of objects performed accurately; the possibility of correcting the data sets must be offered.

To this end, a stepwise-improved application is being developed and used to carry out ECRINS. Named WERC, is allows swapping between MS Access® and ArcGIS®. In order to minimise the number of swaps, that are time consuming and a source of errors, several functions to handle feature classes (e.g. copying a feature class from an Access database to another and restructuring the attributes) have been developed. These improvements dramatically cut the number of swaps compared to the previous version development and allow for production of releasable databases clean of temporary tables.

The WERC application has also been designed to allow quick and accurate linking of working databases so that it can be run from any configuration (e.g. desktop or laptop).

5.2.2 *Versioning management*

When building ECRINS β, versioning had not been implemented. It is now a fully operational feature that seeks and controls the articulation of data sets so that consistency of coding is ensured across the different layers.

5.2.3 *Exporting produced layers to the data service*

Once the working layers and data sets are produced, specific functions in WERC allow selection of the appropriate data sets, renaming them if necessary and placing them in the final deliverable databases.

5.3 Building the Functional Elementary Catchments

5.3.1 *Definitions and size criteria*

The FEC layer is defined as the homogeneous layer of 'elementary' watersheds covering the whole area. By definition, the FEC is made of all watersheds belonging to basins, as defined above, and all coastal watersheds. The coastal watersheds are made from the individual watersheds aggregated together, between two basins. This definition is clarified by the example in Map 5.1.

The data availability constraints inherited from CCM result in the following processing definition: 'the FEC groups all the Strahler level X catchments belonging to basins having Strahler order X+1 or more and all clusters of the other catchments (Strahler order X, X-1, etc., 0) placed between the aforesaid basins'.

The definition seems complicated, but an illustration can help clarify it. In Map 5.1, the X Strahler level has been set to 2, which happened to be the most practical. The different clusters (in dark yellow) shall become FECs populating the coastal strip. These small watersheds are just a series of headwaters whose common trunks are now submerged beneath sea level. The other FECs shall result from the selection of all the Strahler 2 catchments inside the true basins (true because they empty to the sea through a single river). This selection is performed by setting apart the basins having Strahler orders equal or larger than 3. Hence, all coastal basins having Strahler orders equal or smaller than 2 are candidates for aggregation as costal FECs.

The procedure has been further elaborated to deal with the cumulated area as a secondary criterion: if a single basin exceeds the size within the tolerance range (target –x %; target +y %), it is considered as a 'singleton' and not further aggregated, even if its

Map 5.1 FEC-building rationales

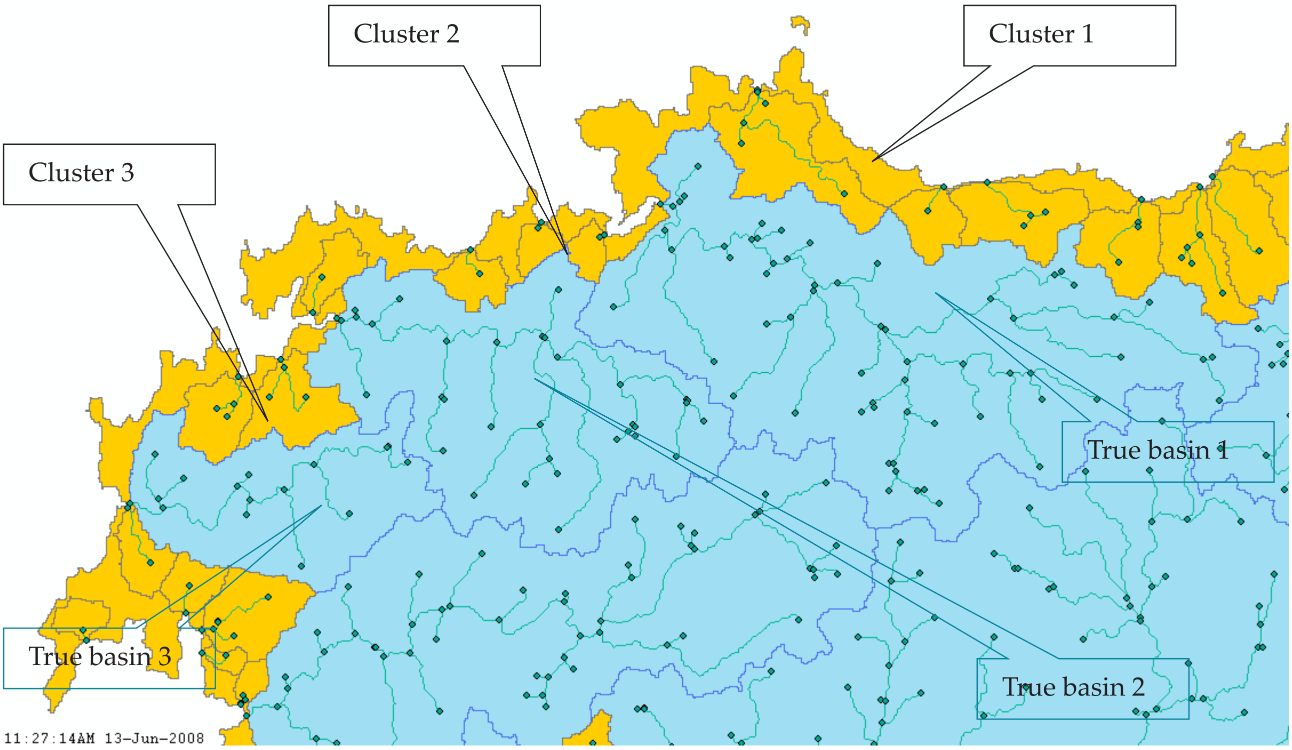


Table 5.1 Elements for selecting the pivot Strahler level for building FECs

Considered elements	Strahler level 2 (basins at level 3)	Strahler level 3 (basins at level 4)
Reference area (km^2) (all watersheds, spurious excluded)	12 131 416	12 131 416
Area covered by catchments (basins with Strahler level >= 2 or 3)	11 768 741	11 529 855
Average watershed area	39	171

rank makes it a candidate for aggregation. Hence coastal FECs can be either aggregates of basins or basins. This distinction exists for production purposes only and has no relevance elsewhere.

The size criterion is set with lower and upper tolerance so that the coastal FECs between two larger basins would be of comparable size. Costal FECs can be smaller than the lower threshold boundary in islands, depending on the island size.

The preliminary analysis suggested that FECs must result from hybridising watersheds from different Strahler orders aiming at partitioning the total area into hydrological watersheds having a narrow range of sizes. There is no indisputable method to define this range of sizes. However, experience shows that if the majority of watersheds lays between 25 km^2 and 250 km^2, the layer is very compatible with stratification purposes. Moreover, each FEC can contain enough information from raster (CLC, population density, etc.) or be reasonably populated from other sources of areal information, for example the climatic data, whose mesh ranges from 25 km^2 to 625 km^2, eventually disaggregated at the squared kilometre into the LEAC grid.

The practical value of the target area is hence in a range from 40 km^2 to 100 km^2; the application allows selecting the target area, the Strahler level target and the tolerance. Such tuning is key in processing an ECRINS release.

5.3.2 Rationales for defining FEC envelopes

The practical solution to the issue of constructing FECs is to build FEC envelopes that define the outer borders of the two categories of FECs:

- coastal FECs that are aggregates of small (possibly single pixel) basins; the envelope is the outer limit of these basins aggregated together;
- continental FECs, contained in the natural envelope of a basin whose size defines it as such.

The FEC production procedure therefore requires the following:

- define the coastal FECs envelope and list their contents based on a definition and aggregation algorithm that produces the desired range of sizes;
- populate the continental FEC envelopes with catchments of appropriate size based on simple rules.

The simplest rule is based on the Strahler order; a simple assessment suggests that the appropriate Strahler order should be between 2 and 4, with possible mixing. Selecting a Strahler order for the FECs implicitly defines the basin as being of this order +1. The question is addressed by considering the requirements and the statistical distribution of basin sizes.

In conclusion, the FEC layer is the union of:

- all the coastal FECs envelopes (in this case one envelope = one FEC);
- the apportionment of continental envelopes into FECs, carried out by filling each envelope with appropriate aggregated catchments, based on the Strahler order, to reach homogeneous size distribution.

Summarising the building of FECs is the outcome of aggregating elementary CCM catchments as coastal FECs and disaggregating CCM basins into continental FECs. To this end, a 'pivot Strahler' is defined, which is the most likely Strahler order aggregate to be tested as source for populating the continental envelopes; this selection is made according to the statistics in Table 5.1 which reports essential information on sizes. The selection of Strahler level 3 to identify basins makes the FEC level Strahler 2 (3-1), providing ~ 300 000 entities. However the total area of land covered by a certain Strahler level decreases with the Strahler level selected: respectively 97 % and 95 % of total area is covered by Strahler levels 2 and 3 because coastal catchments are decreasingly encompassed by the growing Strahler levels.

The smallest unitary catchment suggested by the WFD is 10 km^2. The overall aim of the project, however, is to centre the watersheds between 50 km^2 and 100 km^2, since such a small area is not compatible with production constraints and the source data available. The average area of possible FEC building from basins at Strahler 3 level is 39 km^2, which is compatible with both this WFD threshold and specific requirements; using basins at level 4 would not allow small enough FECs.

For these reasons, the final FEC layer is a blend of different Strahler aggregates, since this is the only way to make the final layer homogeneous and relevant. Therefore the production first identifies the envelopes of continental FECs by selecting those with a Strahler order larger than or equal to 3. The application allows the use of other thresholds; this function was developed for testing but could produce a more aggregated system as well. The development of coastal envelopes was described earlier.

5.3.3 *FEC local data model*

The FECs are the heart of ECRINS: the data model in Figure 5.1 shows that envelopes are made of basins, not all being taken (those considered spurious or too small being discarded); the envelopes are set in a very important intermediate feature class.

The envelopes define homogeneous natural basins, which in turn are used to create the FRBDs as well as the sub-basins. The model is identical for any aggregate; the difference is located in the source of information used to build the sub-basins and FRBDs.

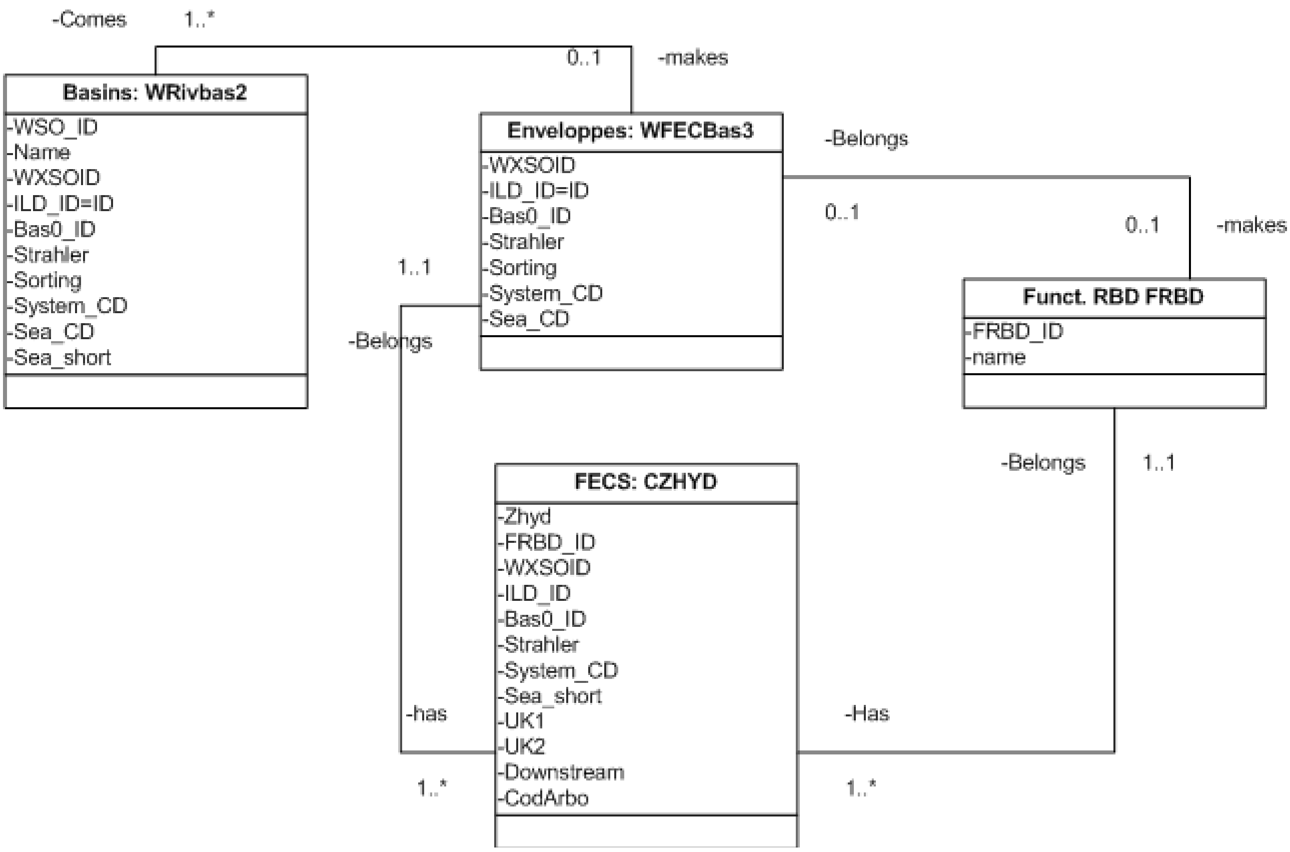
5.3.4 *Algorithm for building coastal FEC envelopes*

Selecting basins is obvious, based on the Strahler level. By contrast, aggregating the watersheds from the list of the non-basin watersheds is more complex.

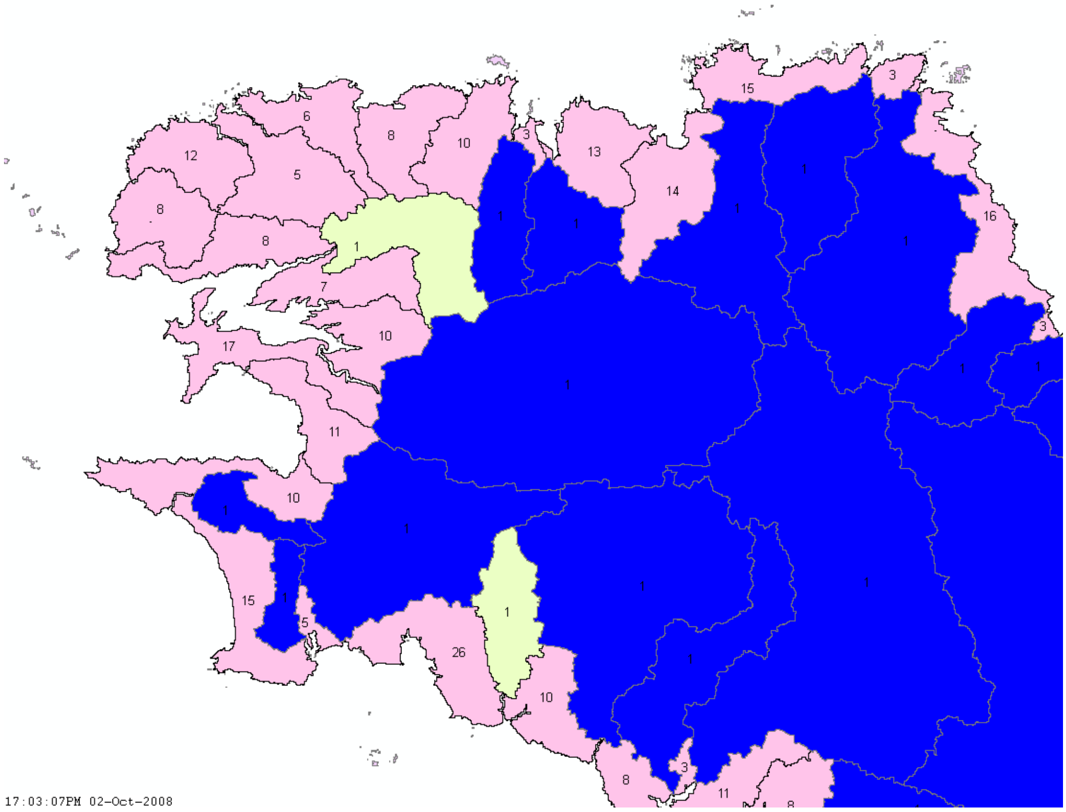
The problems to be addressed are threefold:

- the aggregation method of coastal FECs must be reproducible and provide FECs of comparable size;

Figure 5.1 FECs local data model



Map 5.2 Built aggregates for coastal catchments: Brittany



Note: Continental basins are in blue, singletons in pale yellow, and final sub-aggregates in pink. The number catchments (here from Strahler order 0 to 2) in these aggregates is reported. Singletons are aggregates of a unique CCM basin.

- since continental envelopes are defined once only by their Strahler level, their ID must be independent of the corrections on the coastal aggregates;
- islands may have no continental envelopes and should be nevertheless broken down into reasonable FECs.

The development of the algorithm resulted in elaborating the simple principles exemplified in Map 5.1, page 47. The principle of calculation is based on the use of the improved SORTING field that provides the relative position of basins, in a clockwise order along the coastline. This correction generalises the computation and allows any new shore delineation to be set quite simply, in case of political developments.

Even though continents and islands are sorted in a slightly different way, the same coastal catchment identification applies. It is therefore described only once.

In ECRINS v1.x, the field SORTING_cor is populated for all terrestrial basins and islands. A later version will have all islands appropriately sorted and identified (this is not yet the case).

Parameters are Strahler 3 and area threshold 200 km^2 ±15 % in this case.

The processing comprises the following steps ('sea shore = System_CD + Sea_CD within an island or for the continent).

- List (sorted by order of SORTING) all catchments on a seashore.
- Give an ID to each continental basin (Strahler order >=threshold value (set to 3 here)).
- Identify blocks of catchments (all those between continental basins), if no single block exists.
- Inside each block, identify the 'singleton' catchments, whose individual area is > (areal threshold tolerance), hence creating sub-blocks.

- Inside each sub-block, compute the number of aggregates to build, so that each aggregate meets the cumulated area threshold range. The computation of the number of aggregates is performed as n= round(average (sum of area/(threshold-tolerance), sum of area/(threshold + tolerance)),0). This method minimises the number of items, for example if the sum of areas is just a bit larger than (threshold + tolerance), e.g. 501 instead of 500.
- Assign an ID to each aggregate/singleton. Within the number 'n' of sub-aggregates, the cumulated area is computed, following which (threshold–tolerance) the sub-aggregate is created and its ID set.
- Eliminate from the coastal set all aggregates made of a single catchment and whose area is below elimination threshold. This last condition originates from analysis of the results of first runs that suggested that many very small (few hectares) coastal catchments were present. This is when a small basin could not be aggregated because it was situated between two adjacent continental catchments or between a continental and a singleton catchment. Setting the threshold to 10 ha resulted in suppressing 1 067 basins, reaching a total of 22.3 km^2 for the whole area, which is negligible with respect to the precision of CCM2.

All thresholds are parameters in the WERC application. At the end, all basins are populated with appropriate values of fields from the envelope ID and flag Is_2delete to make the subset of river basins. Those not selected have to be aggregated in pseudo-river basins, to complete the previous set.

5.3.5 *Envelope ID syntax*

The coastal catchment ID creation principles are defined following the general ID syntax of ECRINS.

- Prefix is Wsyo where W is a fixed letter (for watershed), 's' is the Strahler level (should be 3 in current production), 'y' the System_CD letter ,and 'o' the compacted sea_CD number. The sample prefix is therefore W3C1 or W3ZZ, following the internal Sea_CD coding ([11]).

The rest of the ID is numeric and carries information on the type of basin:

- 000100 to 099900 (by hundred), giving 999 distinct possibilities, are the true basins inside continents (islands excluded);
- 100100 to 999900 (by hundred), giving 8 999 possibilities, are the continental basins in islands, bearing in mind that no duplicate can be found between two islands within the same sea shore;
- 99 possibilities of costal basins are provided between two continental basins. For example, the values are 000101 or 0023510 for continental costal basins, and 100001 or 102308 for island basins.

With these conventions, any basin whose last digits are '00' and leftmost digit is '0' are true continental basins. Any basin whose two last digits range between '01' and '99' and leftmost digit is '0' are coastal basins. Any basin whose leftmost digit is '0' is continental; otherwise it is on an island.

Numbers 000100 and 100100 are respectively the first ID of true catchments for continents and islands. If the series starts with coastal catchments, their ID will be 000001 or 100001, for example.

5.3.6 *Special featuring*

Topology basically comprises simple information: 'A flows to B' or 'segment X goes from upstream node to downstream node'. Lack of upstream nodes indicates if it is a spring segment, absence of downstream nodes that it is an outlet segment.

Both forms of information are sufficient for most calculations; the issue here is time and computer limitations. To improve the functionalities of ECRINS, much 'pre-cooked' supplementary information was prepared.

From the outlet FEC, it is possible to journey upstream using recursive algorithms that become quickly very demanding in terms of computation time. Such procedures ([12]) are extremely effective, but can be resource consuming when the depth ([13]) is great and the number of calls large, and may outpace machine capacity or programme settings.

([11]) 0 to Z, making 36 different seas within a system.
([12]) A recursive procedure refers to itself, by contrast with an iterative procedure. Iterative procedures are generally preferred because they do not need a stack to store the results when the procedure is going to the deepest level, them going up backwards. Many problems (e.g. computing a factorial) can be solved either by iterative or recursive methods. By contrast, others (e.g. the Tower of Hanoi (Dynamic Drive, 2012)), and in this case, the river tree, are much simpler to solve with recursive algorithms, whose programming is significantly simpler, despite their running limitations.
([13]) Depth designates the number of branching levels.

Figure 5.2 Illustrating CodeArbo structure

Query1

BASINNAME	Exutoire	Outlet	ZHYD	NextDown_ID	Code_Arbo
	-1	ASO0000233	A020000004		A020000004
	0	ASO0000233	A020000005	A020000004	A0200000041
	0	ASO0000233	A020000007	A020000004	A0200000042
	0	ASO0000233	A020000009	A020000007	A02000000421
	0	ASO0000233	A020000010	A020000009	A020000004211
	0	ASO0000233	A020000011	A020000009	A020000004212
	0	ASO0000233	A020000020	A020000011	A0200000042121
	0	ASO0000233	A020000013	A020000020	A02000000421211
	0	ASO0000233	A020000017	A020000013	A020000004212111
	0	ASO0000233	A020000035	A020000013	A020000004212112
	0	ASO0000233	A020000022	A020000020	A02000000421212
	0	ASO0000233	A020000034	A020000011	A0200000042122
	0	ASO0000233	A020000037	A020000007	A02000000422
	0	ASO0000233	A020000076	A020000037	A020000004221
	0	ASO0000233	A020000100	A020000037	A020000004222

To make allowances for ECRINS users not familiar with the recursive processes, a special code that permits users to carry out the same calculations with a flat algorithm has been installed. The CodeArbo structure (developed by one of us, GLG, as a component of NOPOLU Système 2) is very simple. The root is the FECs ID of the outlet FEC. This root is padded to the left at each level with '1' or '2', representing the first or second detected branch accordingly. The rank has no meaning other than being first or second to be detected during the computation process.

Figure 5.2 illustrates (IDs are from ECRINS β) the composition of CodeArbo. Root is FEC A020000004 ([04], ID is indicated by the two rightmost characters in this example), whose CodeArbo is limited to this ID, as root. FEC [04] receives FECs [05] and [07], whose CodeArbo becomes respectively A0200000041 and A0200000042. FEC [07] receives FEC [09] whose CodeArbo is then A02000000421 (CodeArbo of FEC [07] plus a '1'. The most upstream FECs are FEC [17] and FEC FEC [35], whose CodeArbo has a depth of 5. To compute cumulated characteristics of any FEC, it is then sufficient to pick all the FECs whose CodeArbo start by its own CodeArbo.
For example, the total area upstream FEC [11] is the sum of FEC elementary areas, provided their CodeArbo starts with A020000004212 (all upstream FECs, itself included) or having any character more than this (itself plus just upstream). In SQL, this is coded …LIKE 'A020000004212*' in the first case and … LIKE 'A020000004212?' in the second. Such programming is accessible to any beginner.

Due to the nested level of the ECRINS drainage system, the field CODE_arbo was adapted to a memo type instead of text type, since memo fields can contain an unlimited number of characters.

In order for ECRINS to provide the appropriate attributes, the following calculation of tree code and ancillary data has been divided into three stages that can be carried out independently of the building of a FEC data set.

It takes more than 24 hours to compute the CodeArbo over ECRINS. The time needed per basin ranges from a few seconds to several hours, depending on the depth of branching, which in the very first approximation depends on the number of FECs in the basin (it depends, in fact, on the number of confluences). The possibility of computing up to a certain depth of branching directly depends on the programme stack size that stores the successive calls to the kernel recursive algorithm. Unless calculations have been carried to check, it is practically impossible to foresee the exact depth that will be reached during computations. A new CodeArbo must be recomputed if any change is made in the FEC topology. If only an FEC area has been modified, only the third step is required to produce accurate figures. This may happen when the shoreline is changeable; at the moment, coastal FECs may have their area updated but the topology is retained.

The computations are carried out with WERC, being an MS Access® application; it is therefore subject to MS Access® standard limitations that are set in the msaccess.exe programme. MS Access® has a reserve stack size of 1 MO and allows committing it by pages of 4 kO. A very interesting document ([14]) (in French) offers tips on how to increase the stack, not an obvious MS Access® setting (by contrast with locks per file, etc.). Changing the stack size calls for editing the compiled executable.

This is achieved thanks to utilities found on the Web; the author of [...] proposes a utility tested both by Pöyry and the EEA, that was used to determine the best suited settings.

There is no optimum way to carry this out. Adding some call counters to the recursive algorithm will take up space and lower the performance. Based on experience and documentation, a simple rule of thumb has been applied.

Calculations were carried out by growing number of FECs to process. With basic settings, the Rhine (~ 3 200) FECs were computed without error. The following catchments (the Neva, ~ 4 200) failed and presented an 'Out of stack space' error. Analysis of the MS Access® stack settings indicates that 3 715 calls can be performed with the baseline stack setting, reinforcing the hypothesis that the number of calls equals the number of FECs.

Since the largest catchment (the Volga) has ~ 23 000 FECs, the required settings would be 1 Mo*(24 000/3 715) = 6.46 Mo, rounded off to 8 Mo, and a stack commitment of 32 Ko instead of 4 Ko.

To achieve the best settings, the computation of the CodeArbo has been separated from other computations and a restart option inserted. The computation itself requires identification of all the outlets (WXOID), setting the corresponding FEC as outlets (field Exutoire=True) and finding all the upstream FECs of each of the FECs that are outlets, with the path from downstream to upstream. This is carried out as a first step and results in completing the LogArboDone table that drives computations. Table LogArboDone contains the outlet ID and outlet FEC ID, and the number of FECs upstream of each outlet, to sort the table from the smallest basins (3 FECs) to the largest basin (Volga, 23 000 FECs).

This table is scrutinised from smallest to largest ([15]), date and time of computation stored and flag Is_done set to True when the computation has been successfully carried out. At the end of the process, the depth of each basin is computed (as number of branching, see next section). If a basin cannot be computed, for any reason, its depth is set to joker -9 and not further computed; the process skips to the next basin not yet computed.

If the process is stopped, for any reason, it is possible to restart automatically from the last point of computing. The process can be stopped as many times as required. This restart procedure was needed because of the duration of the process.

Once all basins are computed, the application updates depths for all basins.

The third stage can be carried out on (preferably) finished or unfinished calculations. Only those basins whose CodeArbo is not set to joker ('*') have their FEC cumulated (all upstream each) and final (upstream and itself) areas computed.

5.4 Basin and aggregation watershed rationales

The requirements for aggregation watersheds are many. First of all, the WFD institutes the 'RBD' that is a management unit more or less fitting to hydrographical boundaries, and opens up the possibility of designating 'subunits' that apportion the RBDs. For many assessment purposes, sub-basins of different sizes are required. For example, 'water accounts' demand reasonably large catchments that could possibly be aggregated at the RBD level.

For many statistical purposes, the aggregates are not strongly bound to hydrology. Stratified statistics can, for example, be carried out at country or region level.

With exception of the 'natural basins', no aggregation watersheds constructions requires rivers to be set from segments.

([14]) http://www.developpez.net/forums/d731589/logiciels/microsoft-office/access/contribuez/vba-augmenter-taille-pile-dexecution/, information provided by G. le Gall (Pöyry), checked and implemented.

([15]) Precisely to allow estimation of the current capacity of the stack file and to avoid the need to recompute all, once the stack reserve and committing have been changed.

5.4.1 *Source data and overlaying precautions*

With exception of the aggregates which are potentially defined by internal data, most aggregates are the fruit of processing between external delineations (country, region, RBD, etc.) and the FEC layer. Since the source and FEC layers share no common geometry, the means of matching consists either in finding if the FEC centroid (barycentre) falls into the source element to match to, or if one of the objects captures more or less of the other object, by intersecting them.

At this point, special precaution must be taken for two reasons.

- The FEC polygon (and this applies to lakes as well) whose centroid is sought, may have such a shape that the barycentre could fall outside of it.
- The ArcGIS® 'feature to point' facility proposes to force the point inside, which may produce unpredictable results. In some cases, even with a regular polygon, the centroid forced inside is set on the edge of the polygon, resulting in potential mismatches.

The solution to this twin source issue is to build the centroid layer in the following way:

- create all centroids **without** the inside forcing; all centroids are hence true barycentres;
- check which of these centroids falls outside their source polygons;
- replace the misplaced centroids by centroids forced inside.

With this procedure, the number of points on the edge is minimised and no point may create a mismatch, which is essential in lakes processing.

5.4.2 *Types of aggregates*

The aggregates are divided into three categories that are processed in quite different ways. The aggregation entities are sets of FECs that either constitute or do not constitute hydrologically connected sets. The aggregation entities comprise several types of objects.

First, the aggregates based on ECRINS fundamental features are those that need no external information to be set.

Second, the aggregates that are just transferred from overlay to ECRINS are quite simply integrated: country and region fall under this category. In future, additional aggregates could be considered, which are not necessarily part of the FECs table, despite being systematically populated with FEC IDs.

The identification is simply W_CTxxxxxx and W_Rxxxxxxx, where xxxxxx and xxxxxxx are the two characters of the ISO code of the country and the NUTS code of the region. The region is defined from the average size of French/Spanish NUTS 2, taking the appropriate level in other countries, even non-EU ones. When an administrative reference layer is completed, more consistent relationships will be established.

The third category of aggregates requires considering overlays between ECRINS and the source feature data set, plus processing to ensure hydrological continuity. The most relevant case is the building of **FRBDs** that use a true RBD-to-FEC matching, then a complex procedure of reallocating FECs so that FRBDs are built best matching the original RBDs, after restructuring hydrological continuity.

5.4.3 *Processing rationales*

Making aggregates requires swapping between WERC and ArcGIS®, and eventually post-processing the outcomes to attribute them with appropriate versioning.

The process comprises distinct steps.

- For each aggregate, process it and populate C_Zhyd table with the aggregate ID, for each aggregated FEC.
- Dissolve the FECs previously earmarked with an aggregate and get the feature class of the aggregate.
- Post-process the aggregated feature classes and group them in the final data set.

The organisation of the process hence requires placing the features classes in a buffer PGDB, named Wk_Agg that contains both the aggregates and their source, if any (e.g. the RBDs used to allocate the FECs to a RBD before processing).

The post-processing copies the buffer PGDB Wk_Agg into a working PGDB Wk_AggWK since the structure of the tables will be changed; this organisation (similar to the one used for lakes) ensures the source aggregates remain untouched. Once complete, the working PGDB is copied as final data set and erased.

Versioning is inserted at the post-processing phase, since keeping attributes along the dissolving process is unnecessary and complicates the development. Versioning has a maximum of two IDs: the ECRINS version and the aggregate source version. The same ECRINS version may result in different FRBD aggregates if the delineation of RBDs is changed.

5.4.4 *Implementation of procedures in WERC application*

The implementation in WERC has been achieved considering the more or less difficulty of the procedure: checking the existence of a post-processed C_Zhyd feature class table whose calculation number matches the latest calculations, and having all the target fields created.

The second point is that the computation of the FRBDs must be concluded before any processing needed by aggregates that are apportionments of the FRBDs.

Last but not least, since all the aggregates will have to be post-processed (some post-processing being purely cosmetic), the possibility of processing all as populations of C_Zhyd was considered, allowing a systematic breakdown of feature classes to create the feature classes aggregates.

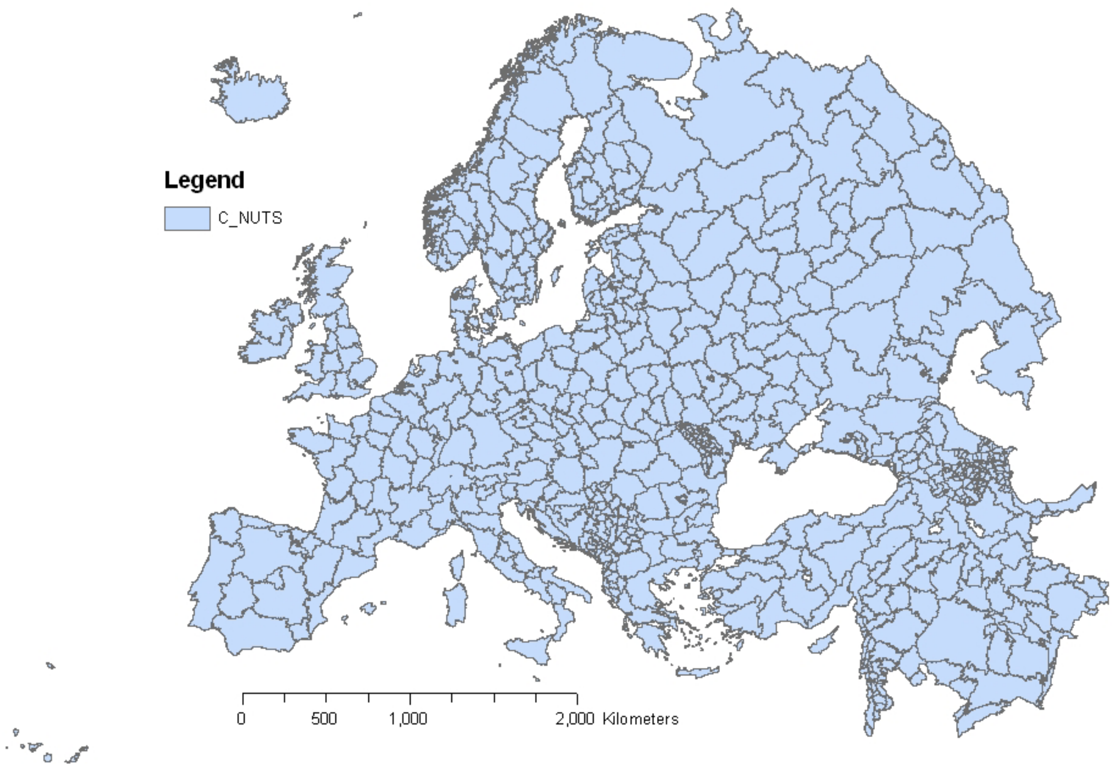
5.5 Building procedure for large basins based on shores

5.5.1 *Procedure rationales*

The 'shore basins' (field letter 'B') are just the sets of FECs belonging to basins that empty to the sea of the same shore. In this case, the shore concept is slightly extended to allow a hierarchical identification of the basin. A shore basin (making a C_B feature class) is composed of all the basins pouring out into the same sea from the same terrestrial mass. For example, the basins from France and England pouring out into the British Channel are two distinct basins, albeit having the same prefix for the shore.

The identification of shore basins follows ECRINS nomenclature: WBcsxxxxxx where cs are respectively the ocean system and the sea code. The chain xxxxxx is numeric and contains the island id (rightmost part of the island ID in ECRINS) or '_____C' for continents. The current number of islands in ECRINS is 37 990, making it possible to append 999 999 — 37 990 = 962 009 new islands before saturating the code. In this case, another marking could be added by counting islands per sea, hence making the possibilities practically infinite.

Map 5.3 Map of NUTS selected for building ECRINS C_NUTs layer



5.5.2 *Computing large basins 'B'*

The building is fully automated and comprises three steps:

- creating the codes with the appropriate WERC function, and inserting the WB IDs in C_Zhyd;
- dissolving the basins under ArcGIS®;
- post-processing and placing the C_B feature class in the target database EcrAgg.

5.6 Building procedure for countries and regions

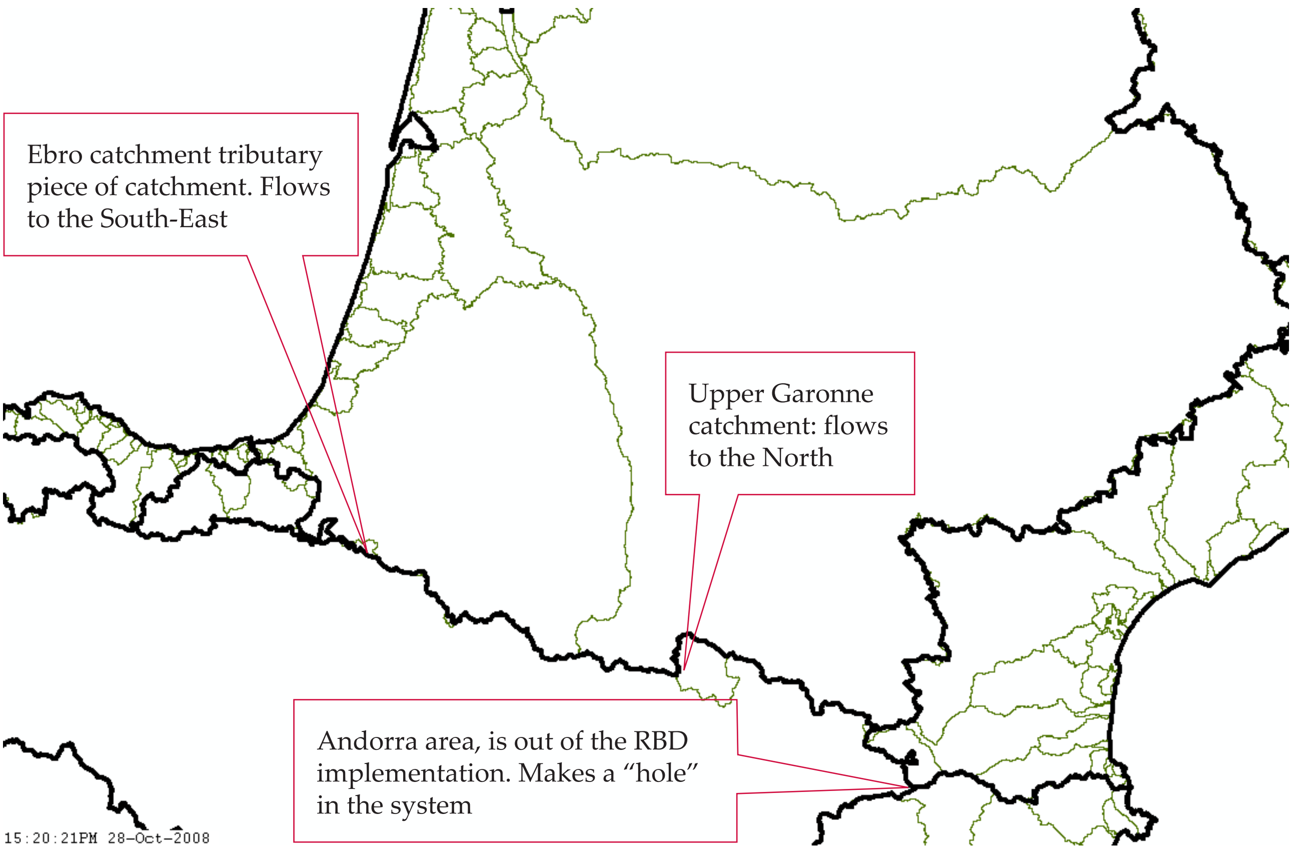
5.6.1 *Procedure rationales*

Country and NUTS limits are overlaid to the FECs layer, and the FEC having a majority of its area is allocated to this type. NUTS are statistical units that do not match precisely the administrative delineations and that exist only in the EU.

To preserve consistency, the NUTS are computed first, and the countries composed of NUTS. NUTS are not comparable across Europe since their definition depends on the political structure of the country. The objective of placing administrative equivalents of watersheds is, however, driven by statistical considerations. Hence comparable NUTS level were selected to fall into the narrowest range of sizes.

To this end, the NUTS were taken from the appropriate level so that homogeneous areas would be used. Despite precautions, in some countries (Moldova, Serbia and Albania, for example) no appropriate level could be found. This should be mitigated after making the administrative mirror of FECs, from an administrative perspective, as planned in 2012.

Map 5.4 Discordance between RBD and basin: the upper Garonne (France)

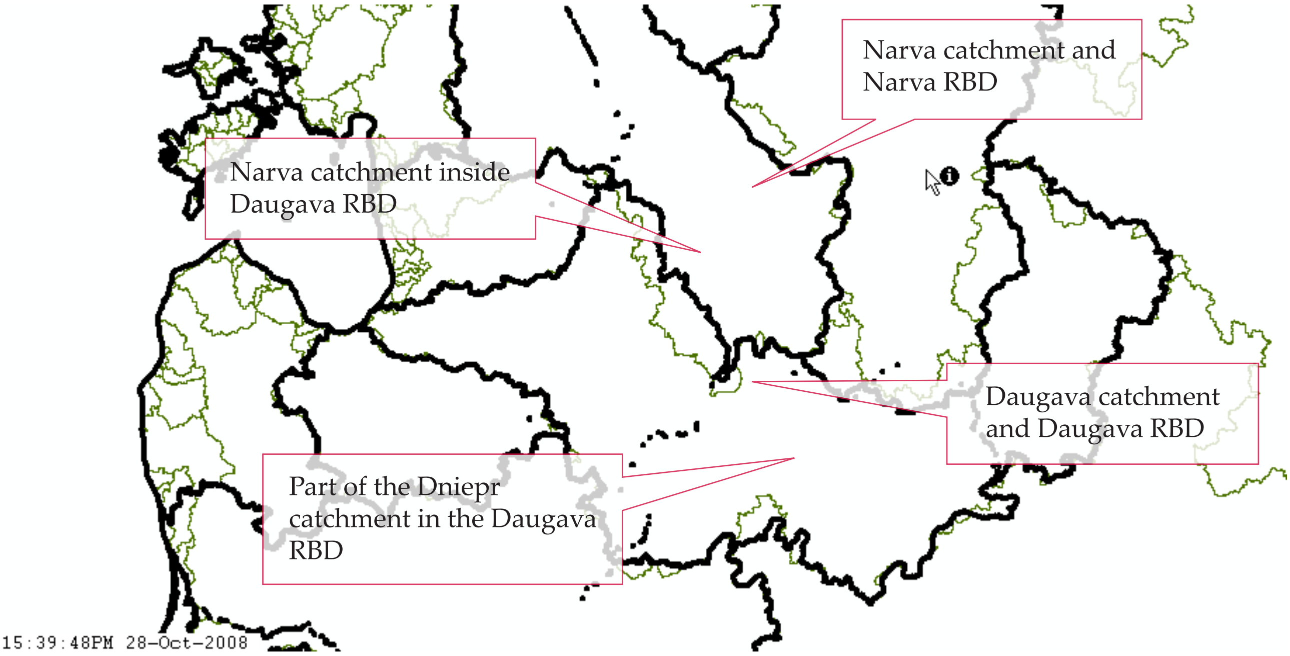


Note: Limits of the clipped international RBDs are in bold black, and limits of FEC basins (FEC envelopes) are in thin green. The callouts pinpoint the problems.

In no place is there an exact match between hydrography and management units: however, some discrepancies are excessive.

Source: EEA processing, map drawn from ECRINS β.

Map 5.5 Complex 'RBD to non-RBD' basin relationships: examples



Source: EEA processing, map made from ECRINS β.

No hydrological constraint has been imposed; it would only further complicate the issue.

5.6.2 *Computing NUTS' and countries' FEC equivalence*

The building is fully automated and comprises the following steps:

- build the appropriate pseudo-NUTS layer and overlay with the FECs, taking the largest share of FECs in a pseudo-NUTS;
- populate C_Zhyd with the NUTS equivalent in a field and combine the NUTS as countries to populate C_Zhyd;
- dissolve both sets of FECs as feature classes C_Ctry and C_Reg;
- post-process and place the feature classes in the database EcrAgg.

5.7 Building procedure for functional river basin districts (FRBDs)

5.7.1 *Procedure methodology*

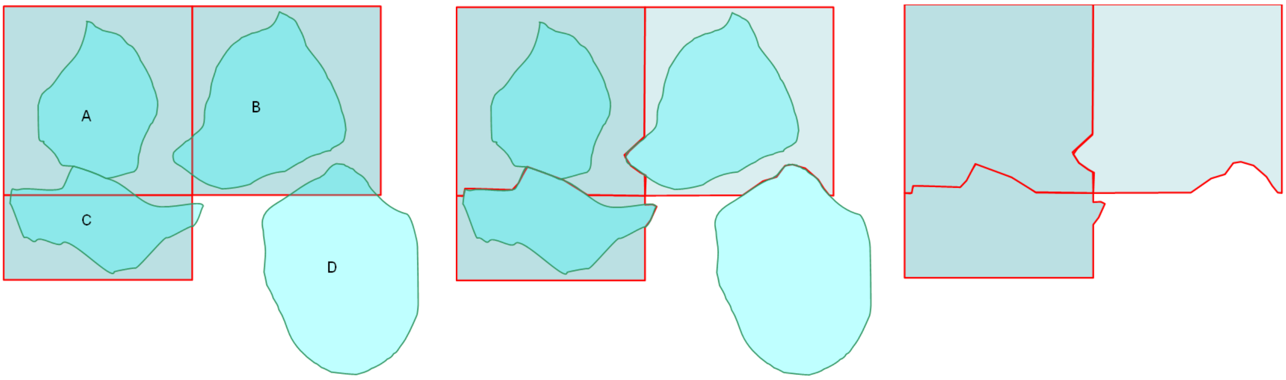
The RBD limits do not come from the same geometry as the FECs: hence the only way to match both and identify which FEC falls into a certain RBD is to place FEC centroids inside the international clipped RBDs. Once this is done, the C_Zhyd table is populated by the joining Ids, hence populating the FRBD_BD field (_BD stands for 'baseline data'). In this case, hydrological consistency or its absence is irrelevant; this aggregate is purely descriptive, or may help in case the most exact match between ECRINS and the WFD management areas is required.

In the second stage, the FRBDs are created by checking the consistency with basins that cannot be obtained from the placement of centroids , and correcting hydrological anomalies that result from the fact that national RBDs are management units and not watersheds.

In the last stage, consistency statistics are carried out. The main problems encountered are the differences between RBD and basin, and RBDs covering different basins with no hydrological relationships.

Map 5.4 and Map 5.5 fully exemplify the different types of problems encountered. In the first examples, the result of the FRBD-building procedure is to reallocate elements of basins to the 'natural' RBD. The issue is simple since all the territory belongs to some RBD.

Figure 5.3 Principles applied to allocate/dis-allocate a FRBD to a FRBD



Note: In box 1, RBDs are represented as squares with red-lined borders. Basins may expand fully inside (A), over several RBDs (B, C) and expand outside the RBD-covered domain (C, D). In cases B, C and D, the RBD outlines must be adjusted. This is done considering as a rule a larger share of basin inside the RBD that results (boxes 2 and 3) in either distorting RBD outlines (even expanding on areas not covered by the RBD (basin C)), or in excluding the basin from the domain covered by the RBD (basin D).

France and Spain have not created an international district for the Garonne basin, hence the upper reaches of the Garonne basin (some 550 km^2) are attached to the Ebro district. This is not acceptable from hydrological point of view; the water from this area flows towards Toulouse and Bordeaux, not in the direction of Barcelona.

A small part of the Ebro catchment related to Spain, in the upper Pyrenees, is administratively in France despite flowing towards Spain. This may be corrected automatically.

A territory not submitted to the WFD (Andorra) is kept outside the RBD delineation. This would make a 'hole' in the system, as related catchments are attached to the functional district in which it lays.

The next example is taken from the area between the Baltic States and Belarus or Russia. The hydrological system is very complex and poorly captured by the WFD that applies to Baltic States, and not to Belarus or Russia.

The relationship between hydrography (from CCM, from which ECRINS derives) and the RBD is indeed weaker in the eastern part of Europe. The main issue is that non-EU catchments such as the Dnieper are part of the Daugava RBD, for example.

The procedure to properly allocate the FECs to the FRBDs is therefore as follows, using the most likely RBD for hosting a basin: FECs from the outlier basin are transferred to the neighbour RBD or discarded from the RBD assignment. It is illustrated by in Figure 5.3.

To implement these common sense rules, the following algorithm is implemented in WERC. It requires computing two ratios for each basin:

- ratio A = area of basin inside each baseline RBD to the total area of basin;
- ratio B = area of basin over all baseline RBDs concerned to the total area of basin.

The rules are:

- if ratio A = 1, then basin is fully inside the RBD (this is the common case, case A);
- if ratio A< threshold and ratio B > threshold, it means that the basin is allocated to the RBND which has the largest extent (cases B and C);
- if ratio A < threshold and ratio B <threshold, all basins are excluded from the RBD (case D).

The example of the Garonne is typically case B (despite four different RBDs being involved). The Daugava RBD mixes several cases (that are addressed basin by basin, however); the Dnieper basin is hence excluded.

5.7.2 *Mitigating heterogeneity in provided RBDs*

The RBDs have been produced by Member States using different methodologies. The delivered shapefiles providing the delineation may extend over the seashores or process different international bodies. The differences in the second case are corrected stepwise with subsequent deliveries. Should there be FECs not allocated to a RBD when they should be, a correcting and integrating process has been implemented in WERC.

In the event that FECs are part of catchments that expand over a lake, these FECs cannot be selected if the lake is not included in the delineation of RBDs as provided by the Member State. This adds another source of errors to the functional RBDs delineation; the only way to process this is manually add the missing areas to the international RBD polygons.

5.7.3 *Processing*

Once all data are ready, the process proper consists in the following.

- Setting all former settings related to the FRBD to joker value ('#Joker#'), because processing of nulls/empty raises problems and errors.
- Updating the fields FRBD and FRBD_BC to the same RBD ID, from the joined table. Field FRBD_BC (baseline catchment) is the raw spatial joining result. The former FRBD shall be corrected afterwards. The field retaining the best geographical match is used to keep trace of the closet match, which is not also the best functional match in some cases.
- Correcting the ID inside FRBD to the most likely RBD. This is done by selecting the basins according to the procedure described at the end of the previous section.

Once the C_Zhyd table is populated, it can be subdivided under fields FRBD and FRBD_BC to create respective candidates for C_FRBD and C_FRBD_BC feature classes.

5.7.4 *Post-processing*

As the other feature classes are produced, both candidates have to be post-processed. Post-processing consists in all cases in copying the candidate feature classes in the target final database and populating the table with adequate attributes.

Post-processing is carried out in the same way as is preprocessing: selecting the type of work and checking the check box.

For reasons of simplicity, the same name is kept; changes involve placing the RBD English name, setting the area in km^2 and suppressing useless service fields.

5.7.5 *Results*

The procedure fills a table of differences between the source RBDs and the computed FRBDs. This table, named LogFRBDDeltas, stores the differences and computes them as a percentage.

The analysis of differences is performed using both a raw percentage of difference that ranges between 12 % and 32 %, and a weighted percentage of differences whose absolute values range between 0 and 0.053 %, with a cumulated difference of 0.75 %.

The distribution of differences is limited to very small and specific areas; it also shows potential uncertainties in the FEC data set of Greek islands. Other discrepancies that cannot be analysed directly are placed in those RBDs sharing catchments with Belarus, Russia or Ukraine. A more subtle assessment of common areas should be carried out in these cases in subsequent ECRINS versions.

5.8 Building procedure for sub-basins based on Strahler levels

5.8.1 *Prerequisites and rationales*

The main purpose of ECRINS is to provide appropriate aggregation catchments that can be used for statistical and water accounting purposes.

Figure 5.4 Recording RBD vs FRBD differences: sample

LogRBDDeltas

F_RBDID	RBDID	RBD_areakn	FRBD_areakm2	Deltap100	RBDname	FRBD_areakm2_Corr	Deltap100Co	calcNum
WFD0000001	WGD0000001	156764.56	156468.89	0.19	Loire, Brittany and Vendee coastal waters	156819.66	-0.04	30
WFD0000002	WGD0000002	97712.52	97652.69	0.06	Douro	97713.98	0	30
WFD0000003	WGD0000003	83119.82	82993.68	0.15	Tagus and Western Basins	82882.85	0.29	30
WFD0000004	WGD0000004	66988.56	67208.4	-0.33	Guadiana	67219.41	-0.34	30
WFD0000005	WGD0000005	7919.5	7903.92	0.2	Neagh Bann	7903.92	0.2	30
WFD0000006	WGD0000006	12375.48	12286.84	0.72	North Western	12286.84	0.72	30
WFD0000007	WGD0000007	18316.86	17921.63	2.16	Shannon	17837.03	2.62	30
WFD0000008	WGD0000008	116739.99	117303.81	-0.48	Adour, Garonne, Dordogne, Charente and coastal waters …	116681.97	0.05	30
WFD0000009	WGD0000009	40828.98	40790.05	0.1	Minho and Lima	40703.75	0.31	30
WFD0000010	WGD0000010	11596.1	11514.89	0.7	Vouga, Mondego and Lis	11514.89	0.7	30
WFD0000011	WGD0000011	10071.53	9851.18	2.19	Sado and Mira	9851.18	2.19	30
WFD0000012	WGD0000012	3836.1	3766.81	1.81	Algarve Basins	3766.81	1.81	30
WFD0000013	WGD0000013	68010.15	67653.48	0.52	Andalusia Atlantic Basins	67633.85	0.55	30
WFD0000014	WGD0000014	12219.82	11719.97	4.09	Western	11804.57	3.4	30
WFD0000015	WGD0000015	13080.55	13055.33	0.19	Galician Coast	12997.15	0.64	30
WFD0000016	WGD0000016	3357.96	3361.54	-0.11	Cavado, Ave and Leca	3361.54	-0.11	30

The FRBDs do not meet this criterion since their size range is not suited to this purpose.

Two approaches have been used:

- building a set of sub-basins depending only on the structure of watersheds (described below);
- building a set of sub-basins based on the size of the watersheds of main affluents (addressed in a next section).

Once the functional RBDs are defined and produced, the other aggregation catchments that must be part of the FRBDs can be defined and prepared. Their definition, as indicated in Section 3.3.3, is fully dependent on the functional RBDs. The methodological principle has been inspired by the building of FECs. First of all, 'envelopes' suited to the sub-basins are defined and then populated. The difference is that in this case, within the defined envelopes, the aggregation catchments will result from the status quo (the envelope is kept as such); different envelopes are aggregated or broken down into smaller components. Moreover, the 'envelopes' have to be dynamically defined, thus adding some complexity to the process.

It is expected that very small aggregates can result from the process, when:

- watersheds are contained in a limited area, e.g. islands: the island area is the upper ceiling of the aggregation catchment size;
- the aggregate is a set of watersheds between two larger unbreakable envelopes, for example a small RBD or a set of coastal catchments between two large basins/RBDs.

By contrast, too large aggregation catchments may result from a poorly branched river system where the Strahler levels are low despite a huge catchment area. This is mainly the case in long and narrow basins with a 'fishbone' structure, and in the large oriental plains where the river density is low compared to the size of the basin.

This second case is quite difficult to address because there is no obvious solution. The issue and approaches are discussed separately for the sub-basins and the sub-catchments, because the latter depend on the way the former are built.

The entity is fundamentally a FRBD, within a single sea facade and within an island (if appropriate) (a set of basins meeting the same criteria where no RBD exists) and the sub-entity is a partition of the entity.

The general rule is that sub-basins can derive from Strahler 7 or 6 (with exceptions, because of the problems mentioned above).

By contrast with FECs, there is no natural envelope to sub-basins. Defining the FEC basin envelope as potential envelopes has favourable processing consequences but practically forbids envisaging overlapping classes of source catchments for making the aggregates.

FRBD delineation constraints are analysed as follows:

- if the FRBD area is smaller than the lower range boundary of the target area, it is a sub-basin by itself;
- if the FRBD is within the area threshold boundaries, it may be apportioned or not;
- if the FRBD is larger than the upper range boundary of the target area, it is quasi-certainly apportioned.

The reason for the terms 'may' and 'quasi-certainly' is that the decision depends on the internal structure of the FRBD, and on the presence of relevant objects inside the FRBD, on the one hand, and on rounding algorithms used to compute the number of catchments into which the set of watersheds must be apportioned, on the other hand.

5.8.2 ID settings

The sub-basins based on watershed structure are named 'sub-basins Strahler', nicknamed SS. This refers to the class of sub-basins, sous-secteurs in French nomenclature and in line with NOPOLU Système 2 syntax.

Since SS are potentially to be strictly embedded inside larger aggregates (to define), whose total number is quite low, the following ID structure was eventually implemented:

Prefix 'WSS' and numeric part, comprising 7 numeric characters, padded to the left with zeroes.

5.8.3 Definition and implementation of primary selection algorithms

The aggregation catchment production for sub-basins is carried out in the following way.

Data preprocessing is required to draw up a statistical table organised to provide area distribution of Strahler 4, 5, 6, and 7 levels per entity and sub-entity, named StatC45And67. This table is prepared only if there is a change in the source data sets.

First, the entities are clustered by groups meeting the criteria of 'entity'/all basins; system and sea IDs, island ID, and sorting order. Then each cluster can be processed, with the rules applying only at the cluster level.

The sorted list of basins (FEC envelopes) within a cluster is then processed and the cumulated area of basins along the sorting order is computed in the following way.

1. If the basin has an area below the lower boundary of the threshold range, its area adds to the previous ones, to be aggregated into one or several sub-basins.
2. If the basin has an area that is between the lower and upper boundary of threshold range, it is marked as a singleton and constitutes a sub-basin by itself, and the sum of areas is reset to 0.
3. If the basin is larger than the upper boundary of threshold range, it is marked to be broken down into smaller sub-basins, and the sum of areas is reset to 0.

Case no 1 is processed in a way comparable to the production of coastal basins to create coastal FEC envelopes. A procedure computes the rounded number of sub-aggregates and then aggregates the objects. The aggregation is quite sensitive to the threshold range; a narrow range often results in one more aggregate than computed. In practice, the catchments' cumulated size distribution is very erratic. The programme processes by computing the average size of the sub-aggregate (total divided by rounded number) and gives a tolerance which is the average percentage of tolerance entered by the user.

To decide on a cut, the programme checks if the next area will fall between the tolerance limits. If they are narrow, it tends to reject the latest catchment to include if it is foreseen larger than the previous ones. At the end of the process, the last foreseen sub-aggregates may therefore transcend the upper tolerance value, making it necessary to create a supplementary sub-aggregate.

Once processed, the sub-aggregates are given a sub-basin ID and populated with their FECs to update the C_Zhyd table.

In case no 2, the item is given a sub-basin ID and populated with FECs to update the C_Zhyd table.

Case bullet 3 is the most complex because it requires breaking down the item (the basin) into smaller objects that may be either Strahler level 6 or 7 according to their size, and size distribution vs the target threshold range. At the end, the selected objects are given sub-basin IDs and populated with their FECs to update the C_Zhyd table.

Processing the FRBD basins or non-FRBD basins is identical in principle; the difference lies in processing data proper.

To make it possible to subdivide the entities, statistical and summary tables must be available, as described in the next sections.

During the creation of sub-basins, it transpired that some produced sub-basins are too large and cannot be further subdivided under the general algorithms because they comprise a single Strahler 6 that cannot be further disaggregated. Whether such sub-basins are acceptable or not depends on the threshold values; however, no simple solution has yet been implemented for this class of sub-basin.

An important issue is proposing the most appropriate threshold ranges. The relative flexibility of the approach is the reason for the development of this computation algorithm. As a rule of thumb, SS in the range 10 000 km² to 50 000 km² are good starting points. The procedure set in WERC is inspired by the building of FECs: a seed threshold is given, completed by upper and lower values (as percentages). For example, if a threshold of 20 000 and percentages + 20 % / -30 % are proposed, a calculation range of 16 000 to 26 000 is used in calculations.

If a functional RBD extends over two different sea facades (this is the case, for example, in the Loire Brittany RBD whose shores are on the Atlantic (the Bay of Biscay), Atlantic proper and British Channel), the limit between the seas is an impassable border for basin and for sub-basins Strahler. Similarly, an island cannot aggregate to a continental part of the RBD, for obvious reasons.

5.9 Natural sub-basins based on rivers

5.9.1 Rationales

The 'sub-basins Strahler' presented earlier were used in the first computation of the accounts and for stratified assessments of water quality. Their main advantage was the immediate availability;

by contrast, the rather odd size distribution was a serious disadvantage.

Building 'natural sub-basins' like those that were taught in primary school seemed to be the best option. The question was raised with the ETC/LUSI in 2009 and 2010, but despite efforts to move forward, it could not be appropriately addressed until a sounder definition was elaborated. With the development of ECRINS v1.x, the means of processing them was presented and a first delivery, not automated, was produced in late March 2011.

The main rationale is subdividing the largest basins into smaller units; generally, these smaller units correspond to the drainage areas of the bigger tributaries inside the river basin. The goal is to get a map with a natural subdivision of the basins that is also homogeneous in size across Europe, allowing for analysis and comparison among homogeneous units that retain a natural meaning. The final average size of these units, should be roughly in the area of 20 000 km^2 to 25 000 km^2.

In order to maintain consistency with the WFD management units, FRBDs are considered as the units to be subdivided in the areas where the WFD applies. Where the WFD doesn't apply, the subdivision is focused only on big continental river basins. The coastal and small river basins between big river basins are grouped together, taking into consideration the regional sea in which they flow.

Only FRBDs (where the WFD applies) or river basins (where it doesn't) that drain a surface bigger than 40 000 km^2 are subdivided because they are likely to make two sub-basins. FRBDs with a surface smaller than the threshold are normally considered as independent sub-basins; there are some exceptions, like Scheldt river, for instance, where the morphology and size of its basin and the other smaller basins in the FRBD point towards making the river independent.

Taking this into account, there are 38 FRBDs where subdivided, and 23 river basins where subdivided, where the WFD doesn't apply.

Attempts to use the FRBD 'subunits' were hindered by the fact that many of these are only managerial units and are not hydrologically consistent.

5.9.2 Subdivision method

The elements where the subdivision effort is concentrated are the largest continental river basins. The concept of FRBDs creates some exceptions that are explained later on.

The **method** for subdividing continental river basins is based on the hierarchy of tributaries inside the river basin, and the area they drain. The subdivision is organised hierarchically, in the following way:

- select all river basins draining more than 40 000 km^2;
- identify the tributaries to the main river that drain a surface larger than 10 000 km^2, and create a sub-basin for each of those tributaries (these are the so called 'level 1 sub-basins');
- if level 1 sub-basins drain more than 40 000 km^2, they are subdivided again, using the same criterion, thereby creating level 2 sub-basins;
- this is repeated until all sub-basins are within the area range of 10 000 km^2 to 40 000 km^2; level 3 was only needed in big river basins, especially the Volga and its big tributary Kama;
- each sub-basin is given the name of the tributary that drains it.

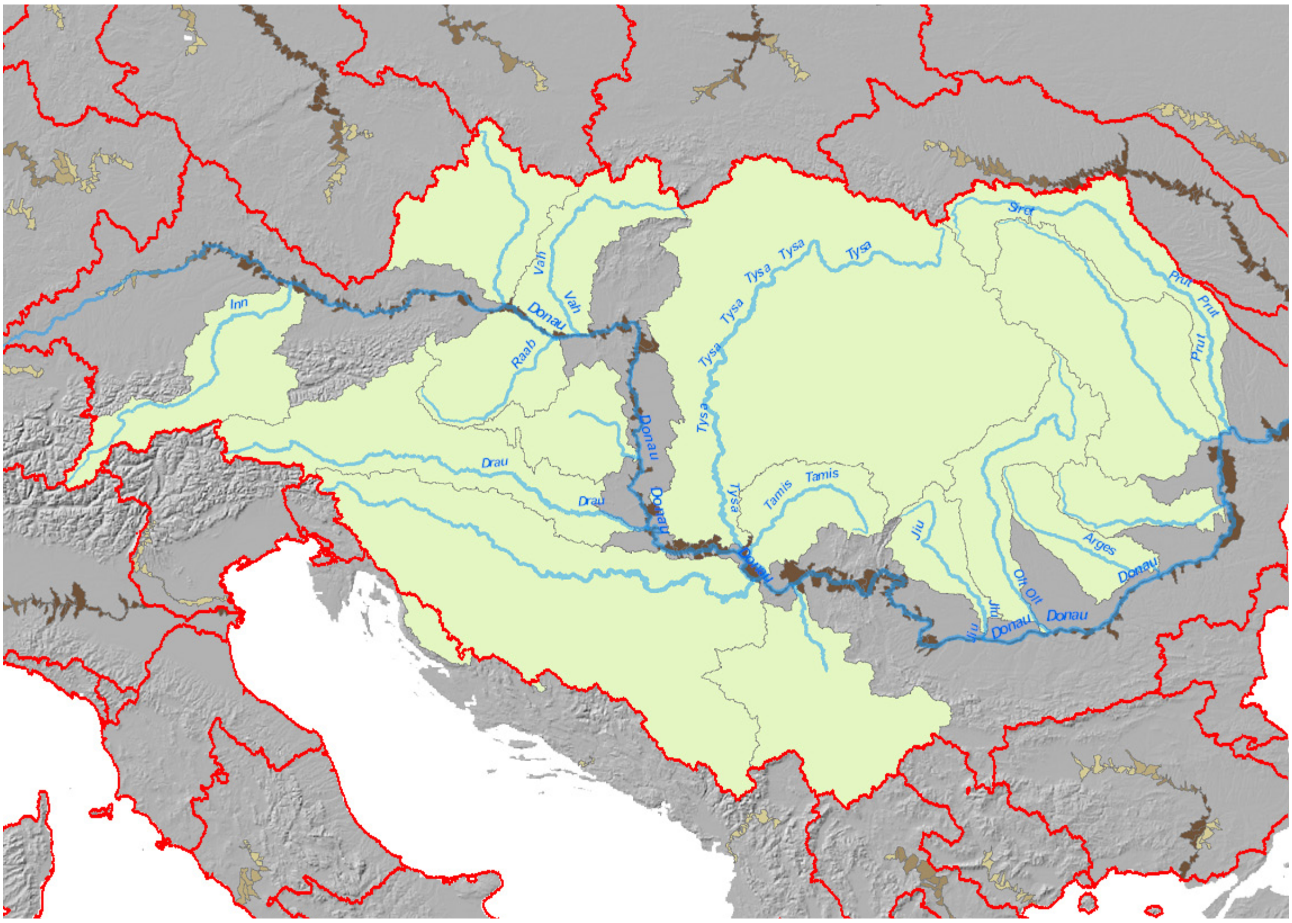
In addition, it's also important to consider that the threshold of 10 000 km^2 leaves the central part of the river (main river or tributary if the latter drains more than 40 000 km^2), plus the areas drained by tributaries draining less than 10 000 km^2, undivided. This central part may be bigger than 40 000 km^2; if this is the case, it is subdivided manually, to meet the following criteria.

- Target size of 25 000 km^2.
- Aim for the subdivision to be in three parts at least (upper, medium, lower), and those parts to be subdivided into more parts if needed (the three parts would be level 1 sub-basins, and the more detailed ones the level 2 sub-basins).
- If possible, use confluences, ideally those where level 1 sub-basins meet the main river.
- Name them using the following convention: '<Name of the river> main x: <list of the small named tributaries> in it', with x being the order number of the subdivision of the main part starting from the head of the river, e.g. Danube main upper 1 — Altmuhl, Lech, Iller.

In FRBDs, some coastal catchments and small basins could be included. In such cases, the following applies.

- Basins draining a surface smaller than 10 000 km^2 and coastal catchments are grouped together using the Strahler sub-basins.
- If the FRBD comprises several independent river basins draining a surface bigger than 10 000 km^2

Map 5.6 Manual subdivision of central part of Danube basins

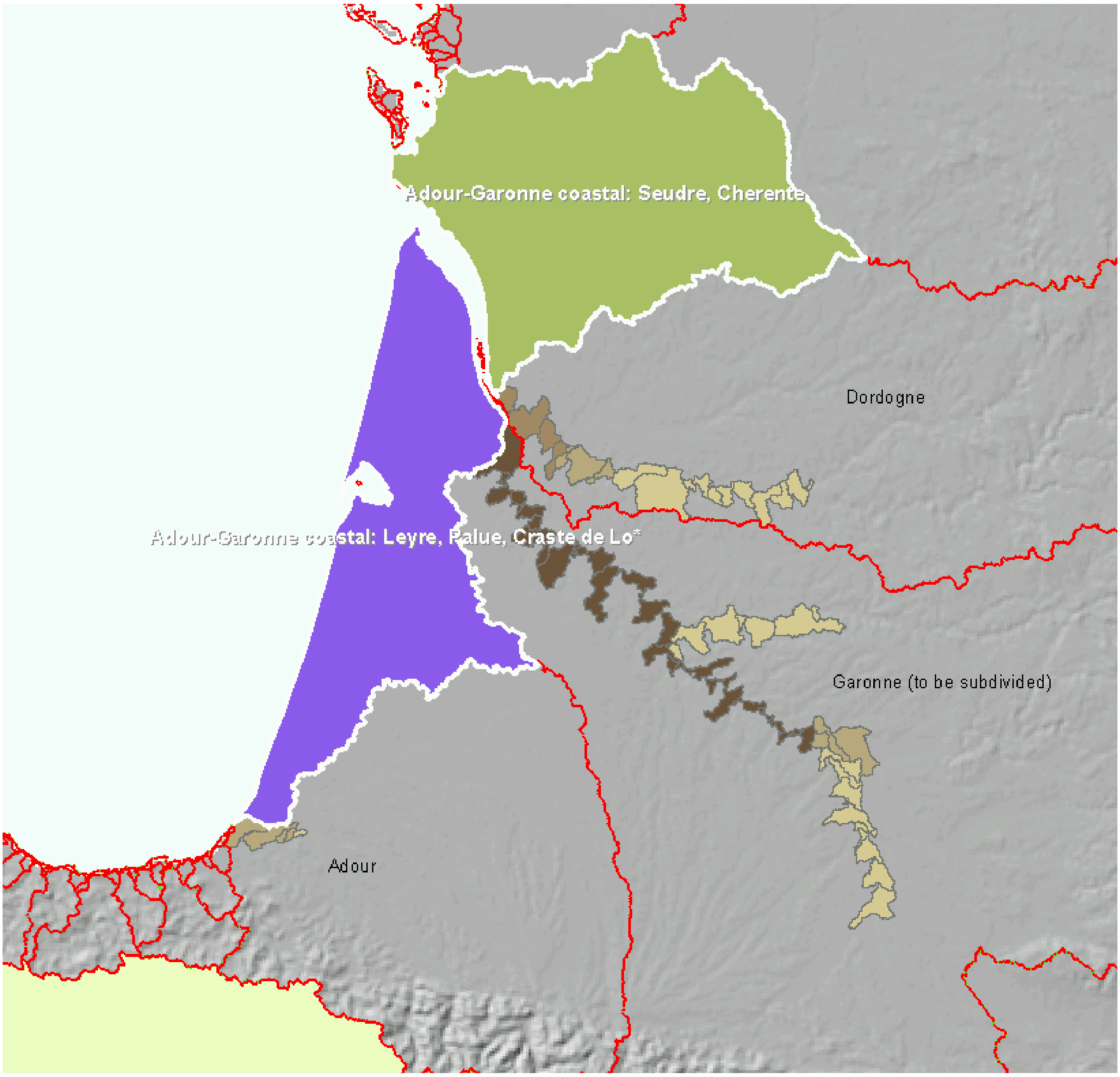


each, each of these river basins is considered as a level 1 sub-basin, to be further subdivided when draining more than 40 000 km². This is the case in the FRBD Garonne, where Garonne, Adour and Dordogne are identified as level 1, and then Garonne is subdivided into continental level 2 sub-basins.

- Subdividing is only possible when the tributary flowing inwards is identified in ECRINS as a named river. There are cases when a tributary draining more than 10 000 km² exists, but it's not named. In this case, its sub-basin can be determined, and it's named as 'name of the river it flows into: Tributary x', x being the order of the tributary starting from the head of the named river it flows into. If this sub-basin drains more than 40 000 km², it cannot be further subdivided, as the course of the main river is unknown. For this reason, there are complete basins or big sub-basins that could not be subdivided, especially towards the east (e.g. the Tigris and Euphrates are not named, for the time being).

Although the process was first conceived as a manual one, some steps have been automated, by taking advantage of some elements of the ECRINS data model. In particular, the detection of tributaries (to the main river, or to other tributaries) draining a surface bigger than the threshold (e.g. 10 000 km²) is carried out automatically. In this process, two fields from the FECs data model are crucial:

Map 5.7 Coastal level 1 sub-basins of Garonne FRBD, subdivided using Strahler sub-basins and named manually



6 Drains, routes and rivers

6.1 Preliminary settings

6.1.1 Rationales for 'rivers', 'routes' and 'drains'

In contrast with catchments, rivers are twin-faced objects. First and foremost, they are geometrical features that can be derived from existing maps. Secondly, they are cultural objects: it is the name given to a river that makes it a river and makes all its stretches be earmarked with this name (16). This cultural dimension determines to a large extent what is a 'main drain' or a 'route', understood as being the candidate match of a 'main river' as reported and processed under the WFD.

Before ECRINS had names to call the rivers and computed routes, contained only 'drains'. A drain is the channel that carries water produced in a catchment. From the terminal outlets, and stepping backwards to the first segment of all twigs from that trunk, it is possible to define, with simple logic (the largest drained area, then the longest set of segments, etc.) drains that are primary, secondary, and so on. However, drains are not yet rivers, since they have no name.

Because of the cultural dimension of the river concept and the many gaps in naming, alongside the urgent necessity to deliver relationships between WFD 'main rivers' and ECRINS, the concept was developed of 'main drains', that are those segments connecting FECs. The main drains create a simplified backbone of European rivers, which is adequate for water accounting, for example. The drains internal to FECs are hence called 'secondary drains'. The main drains indeed sketch the main river systems, and when it was suggested that ECRINS would be the best candidate for hosting the 'main rivers and main lakes', as reported under the compliance provision for the WFD, naming the most rivers possible became a necessity, since Member States only define true rivers as 'main rivers' .

A 'route' is a set of segments forming a single line from a spring to an end; the end may be the sea or another route with which it makes a confluence. Routes are pseudo-rivers, since they are constructed using purely topological considerations. Routes must be systematically attributed to all segments. Attempts to make routing match to existing naming failed, and was abandoned; attendant advantages did not outweigh the immense complexity of the issue. Automatic routing takes the longest drain from a spring to the sea.

The 'river' concept adds a supplementary dimension to ECRINS. Rivers constitute a specific type of route defined by the given name. In most cases, true rivers and routes are identical in most senses. For example, the river Loire from its spring to its mouth has a one-to-one relationship between its named segments and routed segments. By contrast, the river Weser, which is made from the confluence of rivers Werra and Fulda, does not present a match between route segments and river segments.

ECRINS β had already generalised the notion of drains and rivers, creating two distinct identifiers that are attached to river segments: CEGENELIN identifies all routes (named or not named) and River_ID identifies segments sharing the same river name.

In ECRINS, the 'name' attached to a river must be understood as unique identifier: it is composed of an ID, managed by the application and used for modelling on the one hand, and a name(when possible) that facilitates communication with the public, that may also be the source of ID creation. Naming is an extremely complex operation with the ultimate target of providing a unique identifier, completed by a name of common understanding; its aliases and hopefully the different river IDs provided by different sources make it possible to update attribute information in ECRINS from

([16]) 'Name' should be understood in its semantic sense: 'Der Rhein' and 'Le Rhin' are the same name despite spelling differences.

joins between IDs, and to avoid further ArcGIS® processing as far as possible.

Instances where a river changes names many times along its course (regardless of translation issues) form an issue with no conceptual solution for the time being, and they make it difficult to envisage a simple relationship between route and name. In such cases, however, the existence of two different IDs is fully justified.

6.1.2 Information sources

The possibility of featuring the rivers' calculable layer from existing maps instead of using the intermediate 'drains' has been tested, but the improvement in precision of their geometry compared to CCM drains was countered by several factors: the considerable gaps in their topological attribution, difficulties in connecting them to catchments, the unexpected differences in density across Europe between existing and available sources and, last but not least, the insurmountable issues related to licensing.

Hence it seemed preferable to first develop a drain layer that matches the watershed layers, and from this, by referring to the drains with their given name, to derive a layer of rivers. These considerable difficulties were the reason for abandoning the project aiming to use the ERM as a candidate reference layer, and led to the ECRINS project kick-off in 2008.

6.2 Drains and rivers data model

6.2.1 Complementary definitions

As already mentioned, ECRINS is being built in a complex administrative environment in which many hydrological terms are used with a specific accepted sense. The key definitions related to catchments and watersheds were provided in the first part of this document; Table 6.1 explains the complementary definitions related to the current section.

6.2.2 Conceptual definition of 'main drains'

The conceptual model dealt primarily with catchments, and defined the FECs as the smallest areal grains of the system. These FECs are all related by the downstream FEC ID or by the outlet ID. The FECs belong to a composite layer comprising three different types of watersheds, when dealing with river issues.

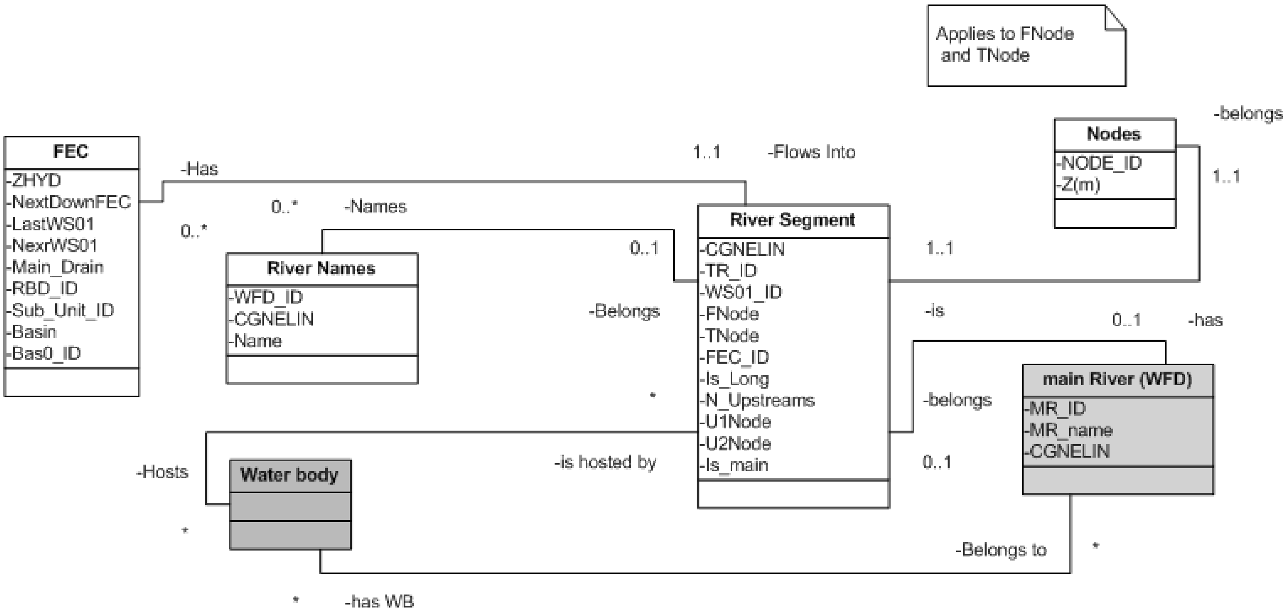
1. A coastal FEC may have several candidate main drains since it may have several outlets.
2. Upstream FECs are defined as having no watershed for which they are downstream FECs. The main drain is by decreasing order of priority during the process:
   (a) the river defined by the longest drainage line towards the next FEC downstream;
   (b) ii) if existing, the set of vertices belonging to the named river that pursues its flow downstream;
   (c) ii) in remaining cases, the set of segments that drain the largest area. In further updates, the lowest rank route in these FECs will be the candidate main route start.
3. Intermediate FECs are those in which the main drain is immediately defined by the entry and exit river vertices. In intermediate FECs, the main drain may or may not be named: this is not a selection factor since connectivity is the primary selection factor. The practical setting of these main drains is the largest Strahler order equal to or larger than the FEC Strahler level.

Limiting the river system to the main drains has the big advantage of freeing the system of candidate main rivers and setting aside all those rivers that develop inside a FEC. This simplification speeds up the calculation of all transfers between watersheds, as well as river fragmentation by dams, for instance.

Table 6.1 Complementary definitions of terms used in ECRINS river building

Term	Definition
River identifier	(River ID) in CCM is the single identifier that links a named river to its name. Hence rivers not named have no river ID. The route ID, named CGENELIN (for a common generic line) matches the river ID when a name exists, since the name has the highest priority in defining the route. The syntax of these IDs is built to make the category unambiguous
node	Nodes are defined in Table I.1, page 13
vertice	A river segment is a polyline, that may be reduced to a single geometrical segment. A vertice is the smallest straight segment building the polyline. Vertices are used to name river segments and have only been in use since ECRINS v1.x.

Figure 6.1 Conceptual model for river layers in ECRINS v1.x



Note: The river aliases are not represented in this picture.

Source: EEA.

The counterpart of this definition is that main drains are sensitive to the design of FECs: when a FEC changes, the main drains must be recomputed as well, whereas routes do not need to be. This is why, in the versions of ECRINS, starting with v1, drainage segments, routes, rivers and main drains are decoupled. The consequence is that any object (dam, monitoring station, water body, etc.) is attached to a river segment secondarily earmarked as belonging to main drain or not. The segment keeps its ID, hence making it unnecessary to snap the object again.

In the production flow, assigning main drains and rivers is consequently decoupled as well.

Main drains represent the downstream route of water across the FECs: it is therefore the only possible support for assessing stratified statistics from monitoring at any stage of naming. Any monitoring station not situated on a main drain cannot be kept for analysis at basin level; otherwise its upstream drivers could not be computed.

Once main drains are defined, secondary drains can be defined as those that are non-main drains.

6.2.3 Conceptual model design and definitions

Following the generic rules relating river stretches, rivers, FECs and the WFD objects, a provisional data model has been set. This data model shows that the main drains result from a selection of river segments, sorted as 'main drains' or secondary drains. Implicitly, the 'main rivers' should all be a subset of the main drains if they mimic actual rivers, since they normally drain over 500 km^2, and the average FEC is 62 km^2. This assumption has not been verified to date since some countries have declared as 'main rivers' sets of drains that are not even part of ECRINS drains, being too small. The main improvement is the placement of all segments in a single feature class, flagging the 'main drain' segments only. In the previous version, there were 18 features for drains and 1 for main drains, complicating the processing.

The conceptual model expresses that:

- segments may or may not have names, or may be considered as long rivers (the longest set of stretches from the mouth) having or not having a name;
- river segments exist in strict association with FECs;
- the fact a set of stretches is a main drain results from the selection procedure, and not from the data model.

The important difference between ECRINS β and ECRINS v1.x lies in the change to the structure of drains and rivers. In ECRINS β, rivers were not significantly inserted. By contrast, rivers instead of drains are the key featuring elements of ECRINS v1.x. The change is not that obvious for the layperson user of ECRINS, but the structural changes are extremely important in the data model and in the processing.

The accurate naming of rivers has an important ancillary point of interest. Since a river is a set of segments having the same name, all named segments should be set in continuity (this is not stating the obvious when one considers that neither ECRINS nor the source maps for acquiring the names are free of errors). If the continuity of names is breached in some places, there is a high likelihood of topological errors resulting from inaccurate geometry or information. Assuming that geographical maps are of higher geometrical quality than is ECRINS, in which topological errors are likely to be present, naming is a potential checking method for such errors. However, if many different rivers have the same name in a single basin, this synonymy is flagged as a potential topological error.

All former CCM river segments have been recoded with an ID insensitive to any change (window, connecting to another segment than that initially considered, etc.). When topological errors are corrected, the ID of segments deleted shall be lost and segments inserted will receive a brand-new ID.

This change in the production method allows, as an ancillary service, the population of river segments independently of the ECRINS version. If drains are added, their ID should of course be possible. To maintain the links with the CCM source, the new ID for river segments is extremely simple: the prefix is 'Z', as before, followed by the source CCM ID of the Strahler level 1 catchment of origin, padded to the left with zeroes. The largest number of Strahler level 1 items is 1 371 375, demanding 7 characters. The space provided for the segments is 9 characters wide, allowing 999 999 999 different IDs, thus leaving more than 990 million of different identifiers free, many more than needed, even when enlarging the ECRINS domain to include EEA neighbours and increasing the resolution.

6.3 Routes, drains and rivers processing steps

Populating ECRINS tables in relation to feature classes in the EcrRiv.mdb data set is carried out with a series of functions implemented in WERC that are detailed in the WERC manual.

The processing of the ECRINS drains system is made in distinct classes of processes.

- The preparatory steps consisted in identifying all segments with ECRINS syntax after reallocating windows. This process is carried out only once until a full revision is carried out. It includes the table of nodes as well.
- Creating the segments layer and setting all required relationships with the FECs layer. In the source CCM, there was a 0/1-to-1 cardinality between segments and catchments (a catchment may have 0 or 1 river segments). The aggregation of CCM elementary catchments as FECs disrupted this relationship to make FECs more usable. The ECRINS cardinality is 0/1 to many.
- Populating the segments with the necessary information on distance to the sea, for example, something not populated in CCM. Distance to the sea constitutes vital information for river naming, for example, and for all further calculations.
- Creating provisional main drains until routes and names are populated.
- Creating routes.
- Linking available names to drains and matching named sets of segments to rivers.
- Updating the final databases with the appropriate tables and feature classes.

Appropriate functions have been implemented in WERC to carry out the different procedures. A new function has been implemented to clean the segments data set from the 'triple points' that severely jeopardised the assignment of names and resulted in unexplained abnormalities in the computation of routes and water accounts in ECRINS β. During the naming process, a great deal (7 090 groups) of nodes having more than two upstream were detected. One node received up to five upstream segments.

Cleaning the data set occurs as follows.

- Having identified the triple points (any segment that is downstream of more than 2 segments), and the segments converging to the point, a procedure scrutinises the segments upstream of each segment of the triplet and counts them.
- The candidates for deletion are the smaller set of segments in the triplet (usually a single segment vs many segments), or in the event the decision is not straightforward, the shortest set of segments.
- First, deleted segments are stored in a log table, and then their upstream nodes are deleted and also stored in the log table.

This can be an extremely long, and to some extent, risky operation. Candidates for deletion are selected using a recursive procedure applied to each of the elements of a triplet, and such a procedure cannot flag more than two upstream segments. In the unlikely case of chains of triplets occurring, inappropriate deletions might occur. This is the reason for storing the deleted segments in log table. Final checking (repeating the procedure) revealed that no discrepancies remained, except for some sets of drain segments which became orphans, being severed from their downstream drainage system. This discrepancy shall be corrected in version 1.1.

6.4 Building provisional main drains

River routing and naming constitute a long process ([17]) that is further complicated by the imperative need to use data reported by the Member States under the WFD first and foremost to populate river IDs and river names. This is discussed at a later point.

To ensure modelling capabilities in the meantime, the assigning of main drains has been carried out independently; this procedure is in line with the decoupling described in the introductory sections.

Provisional main drains concern only two categories of FECs: coastal FECs and upstream FECs. Intermediate FECs have their main drains defined once only within a FEC calculation. Coastal FEC main drains can be changed if the main river inside the coastal FEC is different. Similarly, for upstream FECs, if the first rank river is not the longest drain inside this FEC, the main drain will be adjusted accordingly.

6.4.1 *Data prerequisites and processing*

Main drain assignment is part of post-processing of the C_Tr feature class. This is carried out by a special process in WERC, comprising different steps for completing the segment and nodes.

Node and segment data sets are populated with a code indicating if the node is mouth ('M'), intermediate ('N') or source ('Y'). In principle, this code is provided in the CCM source, along with the distance to mouth from any node and for the segments. In fact, only 50 % of such data were actually populated in the source CCM: a special procedure has been developed to compute all values for all nodes and segments. During this process, a strange error in node identification was detected that indicates the CCM dataset had been likely been manually modified ([18]).

The relationship between FECs and segments is required as well. This relationship is set with a special function, implemented with the FEC post-processing. Coastal FECs are a special case because of the one-to-many cardinality between FEC and basin in this case.

The processing is new to ECRINS v1.x. It is based on totally decoupling the main drains computation from the river naming procedure.

Any difference between coastal and source FEC processing is found only in the organisation of calculations. Coastal FECs are processed independently, whereas source FECs are processed by basin to facilitate the follow-up of calculations. These calculations take approximately three hours using a powerful machine having 4 GB RAM (Windows XP Pro).

6.5 Route and name making

6.5.1 *Route and name identification*

Routes and names both designate either dummy rivers (purely topological) or real rivers (something between topology and culture). Both are composed of sets of segments from a starting point to an end point, and make up a single line. In both cases, updates may change the code and it is important, considering the long calculation time needed, that changes somewhere in a basin do not impact identification in other basins.

The solution is to attribute to each basin a range of numbers, allocated in such a way that enough are available. The number of routes is ceiled by the number of springs. After removing spurious segments, 510 801 springs were found. The number

([17]) Full routing demands about four full days of computation.
([18]) Node ID Y000488472 was indicated as the downstream node of a segment of an affluent of a river in the Caspian Sea. Errors on 640 segments resulted from this value. This node belongs to a segment in the Dnieper catchment. After checking, the ID was corrected with Y000488742. The inversion of figures 'and 7 makes it likely that manual correction had occurred with miskeying. After correction, and resetting of segment type that depends on the U/D relationships, all segments were successfully computed. This example demonstrates that redundant procedures have to be implemented when dealing with topological systems.

of rivers is more difficult to estimate; it was assumed that the number of Strahler level changes in a basin could be a good proxy. Taking the largest value, this was rounded to the next multiple of 10n with a simple formula.

- Start value is below 10, below 100, etc.
- If 2 times start value > ceil, ceil is 2 times ceil; if 4 times start value > ceil, ceil is 3 times ceil.

For example, 20 potential IDs result in an allocation of 100, and 50 potential IDs result in an allocation of 300. Allocated ranges are set by decreasing the number of potential IDs so that the big ranges come first and the small ones after, to increase readability. This allocation method proved to be operational and was hence applied.

6.5.2 Route making

Routes are made in a systematic way, while trying to minimise calculation time that is nevertheless considerable. Basins are scrutinised and considered as two categories.

1. Continental basins. These basins are expected to have a single outlet. All river segments in a basin are earmarked, making their identification easy. The spring most distant to the sea makes the basin route level, by identifying all segments downstream of it. Some continental basins, however, have a single outlet made of two converging segments, thereby making two level 1 routes that are processed with the second procedure mentioned for coastal basins.
2. Coastal basins may have one or many outlets:
   (a) basins with single outlet as processed as continental basins;
   (b) basins with many outlets are processed differently; the set of segments having each outlet as starter is identified with a recursive algorithm, and then the most distant spring is taken to define the route as level 1.

All segments for level 1 are stored and marked as already used; the upper levels are then processed.

Level 2 comprises all the routes that have a downstream segment of level 1. Hence, all segments in level 1 (per river in case of multiple outlets) are scrutinised and the starting segments of level 2 candidates are used to seek for their upper segments, then the most distant is taken as the starting point, and so on.

Subsequently, all routes of level 2 are taken, and used to identify the level 3 candidates, and so on, with upper levels until all segments in the basin have been marked.

When the basin is completed, the log table V_BasRivID is updated so that the basin is finished. This table is also updated each time a route is made so that the route ID is checked against the allocated range of routes.

The process allows either restarting and then computing only for the remaining non-routed segments, or recomputed basins (just setting that it has not been populated in the log table).

6.6 Methodology for naming drains and creating rivers

6.6.1 Basic remarks

Basically, drain naming seems to be a simple process. Given a source map, in which rivers are represented by lines and populated with names, the method consist in matching the source map river lines with the target river segments and transferring the name attribute from source to target. Naming must be understood as capturing every river name (as represented on **maps**) and transferring it to a set of segments (as managed in the ECRINS GIS) so that the representation of the river in ECRINS mimics the understanding of the river by the user of the map. A river in ECRINS is a unique object, the 'name' being the apparent part of the identification of the river, which in parallel is ensured by a river ID.

The simple concept is very complex in practice, for the following reasons.

- Geometry of source and target are radically different both in structure and in drawing; they may differ for many other reasons including generalisation, errors, selection of rivers by the producer of the source, undefined river naming syntax and spelling (accents, quotes, etc.).
- Attributes population in the source data set may be incomplete, odd, varying in density and quality. In some cases, sources are contradictory: one source names B river A and the other reciprocally (with more complex mismatches).
- River naming is the result of a cultural process dating back centuries; it reflects linguistic issues (translation in different languages) of local habits (rivers having no spring, the upper part having a different name that is not a translation, etc.), plus linguistic issues such as the word that means 'river' being incorporated in the river name.

- The same names in the same basin used to identify different rivers constitute a tricky problem, not fully solved in certain cases, especially if different rivers with the same names are close enough to match the same ECRINS target.
- Segmentation of rivers may result in segments populated, unpopulated, or populated with ambiguous attributes (next to confluences, for example).
- The presence of lakes and reservoirs may break the continuity of naming the ghost river inside the reservoir having no name (as river) in the exploited source (see Figure 6.4, page 74).

The naming procedure must be as automated as possible and should provide a reasonable estimate of the likelihood of attributes' transfer success. This likelihood is primarily expected to produce candidate topological errors (e.g. source of river A actually flows to river B in ECRINS).

6.6.2 *Specific processes for building rivers in ECRINS*

The final methodology eventually applied to produce ECRINS v1.0 is the fruit of many trials and errors that resulted in extensive programme development. The first methodological developments were carried out by the ETC/LUSI, complemented by the EEA, revised and partly developed by Pöyry consultants; the final implementation was developed by the EEA, with the support of the ETC/ LUSI to match sources and ECRINS targets.

During the development of the methodology, the issue of incorporating the WFD waterbodies was raised, making it necessary to elaborate the method. In fact, river naming was understood under a wider sense of 'transferring attributes from source linear feature A to target linear feature B': waterbodies are parts of 'rivers' with a very small number of segments that are placed along a stretch of connected segments.

The final methodology helped distinguish segment naming and river making as two distinct, albeit interdependent, processes.

1. First, segment naming, using ECRINS topology, is a process commonly used with the transfer of any attribute (e.g. water body IDs) to segments.
2. Second, river making from the named segments is a separate process that is obviously based on the previous one, but presents specific problems.

6.6.3 *Simplified summary of segment naming process*

The developments related to this first process were carried out by Pöyry, under framework contract. Understanding the historical development of the final methodology highlights the respective contribution of different actors and the progress resulting from 'trials and errors'. Although effective, the final procedure is not optimum, and is under revision. The successive steps tested are set out below.

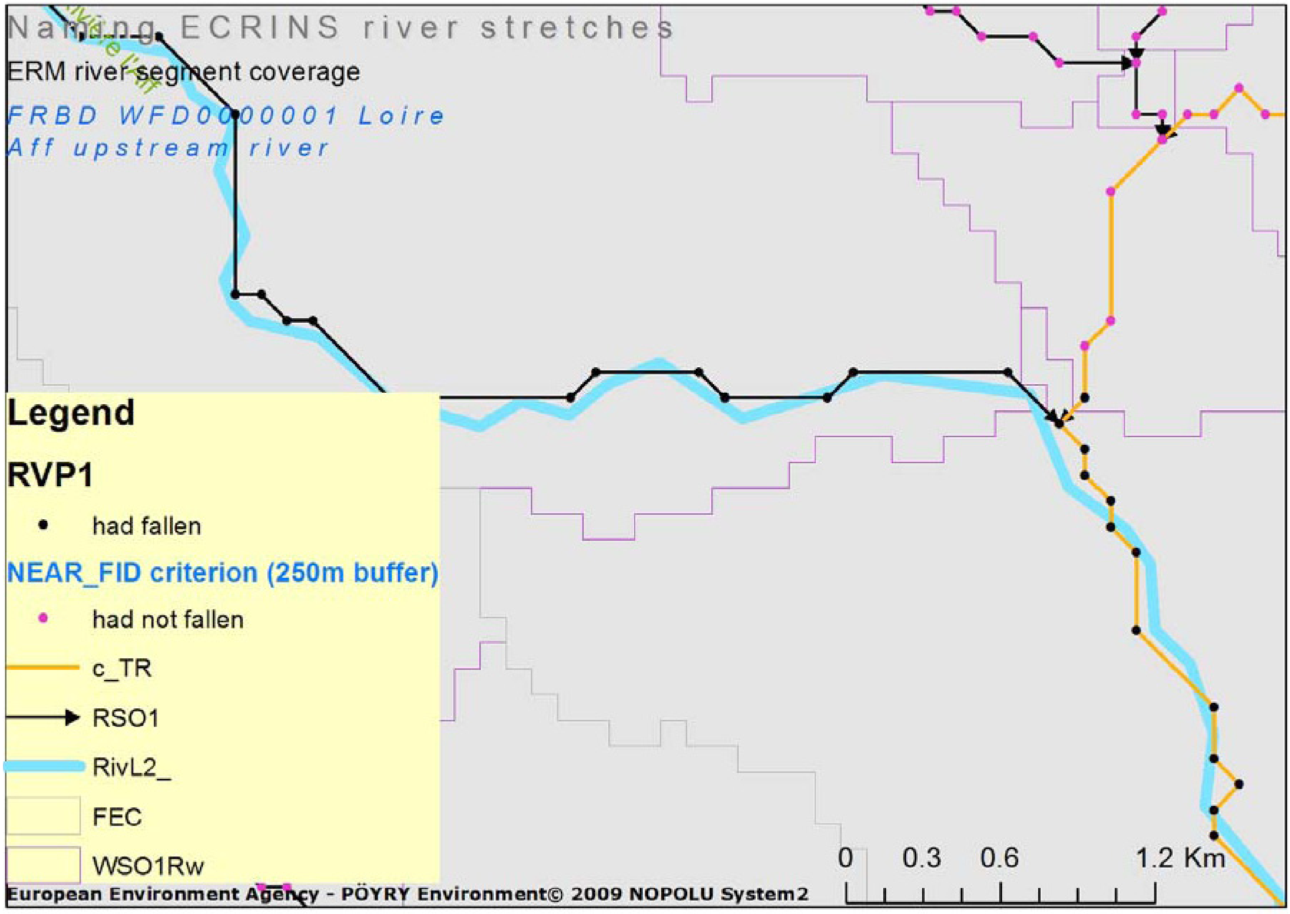
1. Matching source segments to target segments, with a buffer. The yield was more or less acceptable because even close lines may mismatch if they cross frequently, for example, but the accuracy was very poor and hence the results could not be used as source data for river making.
2. Matching source segments to target nodes, the latter being buffered. This refinement solved some problems experienced with the former attempt, especially at confluences, which could be sorted out and analysed as such in case of many hits (expected at confluences), but the overall yield was poor.
3. The final methodology is a blend of the 'line to node' approach, replaced by the 'line to vertices point' that increases the number of hits and allows statistical assessment of likelihood. This methodology is described in a special report (Pöyry, 2010).

The third step could achieve successfully its goals because it is largely backed by sophisticated post-processing of the GIS production. The consultant developed a scenario-based approach that allows screening the GIS matches and improving the likelihood of segment naming using ECRINS topology. The procedure was later applied to collect waterbodies identifications that in turn contributed to segment naming.

The segment naming procedure is carried out in two stages. First, a complex ArcGIS® procedure overlays source ERM data sets and ECRINS drains. The ECRINS drains are transformed from segments between nodes into vertice points, making it necessary to break down the drains layers by sub-catchments — otherwise ArcGIS® collapses. In a second stage, driven by a MS Access® procedure, a table of named segments is built.

When matching rivers to segments, the 'near' function of ArcGIS® is used; this function provides the nearest object from a target feature class and the

Figure 6.2 Vertices and nodes



Note: ECRINS segments are broken lines (black = secondary, yellow = main). Vertices are dots on the lines, and nodes (between segments) are circled in red. The light blue line is a river from a sample map. Vertices in black have found a match, whereas the red ones have not, since the ECRINS segments have no counterpart in the source map: only one river in this source is present in the displayed sample.

The source map is the French Art. 13 delivery from BD Carthage, simplified.

Source: NOPOLU naming application; Pöyry, 2010.

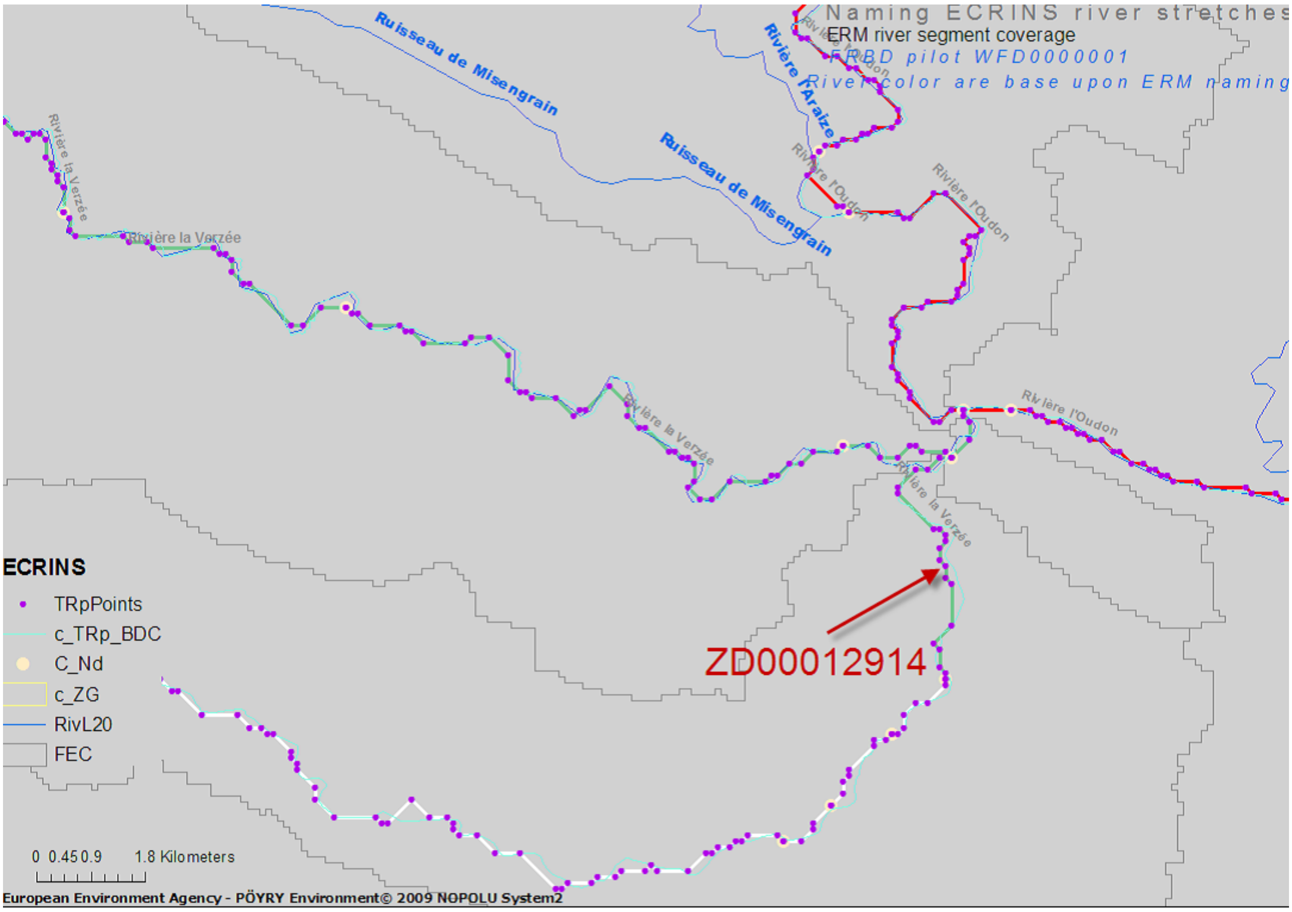
distance of matching, within a capture radius. In the example, this was set to 250 m (one millimetre on a 1:250K map).

This method provides a much larger number of hits than the node-to-river would do: only one node would match the blue river on the sample Figure 6.2. The number of hits is very important when eventually deciding what belongs to what. This is exemplified in Figure 6.3. The segment marked ZD00012914 is erroneously proposed to be part of 'Rivière la Verzée'. The reason is that the first vertices on this segment are in the capture radius of the actual 'Rivière la Verzée'.

However, the true river has many more vertices captured, as shown by the turquoise segments. The relative proportion of vertices capturing the name on both segments allows choosing the appropriate candidate. The method is not perfect, since vertices identify changes in polyline direction: if the segment is straight, no vertices exist, something that can result in erroneous assignment.

It may occur, however, that different segments, obviously belonging to different rivers, have the same name, depending on the data source. This cannot be processed at the segment naming level and is addressed at the river building step.

Figure 6.3 Segment naming: example



The full process is indeed long and complex, owing to the large volume of data to be processed in some systems.

6.6.4 River building principles

River construction consists in using the named segments and ECRINS topology to create sets of segments having a starting point (not necessarily a 'spring' since not all rivers have a spring, because of naming habits) and ending point.

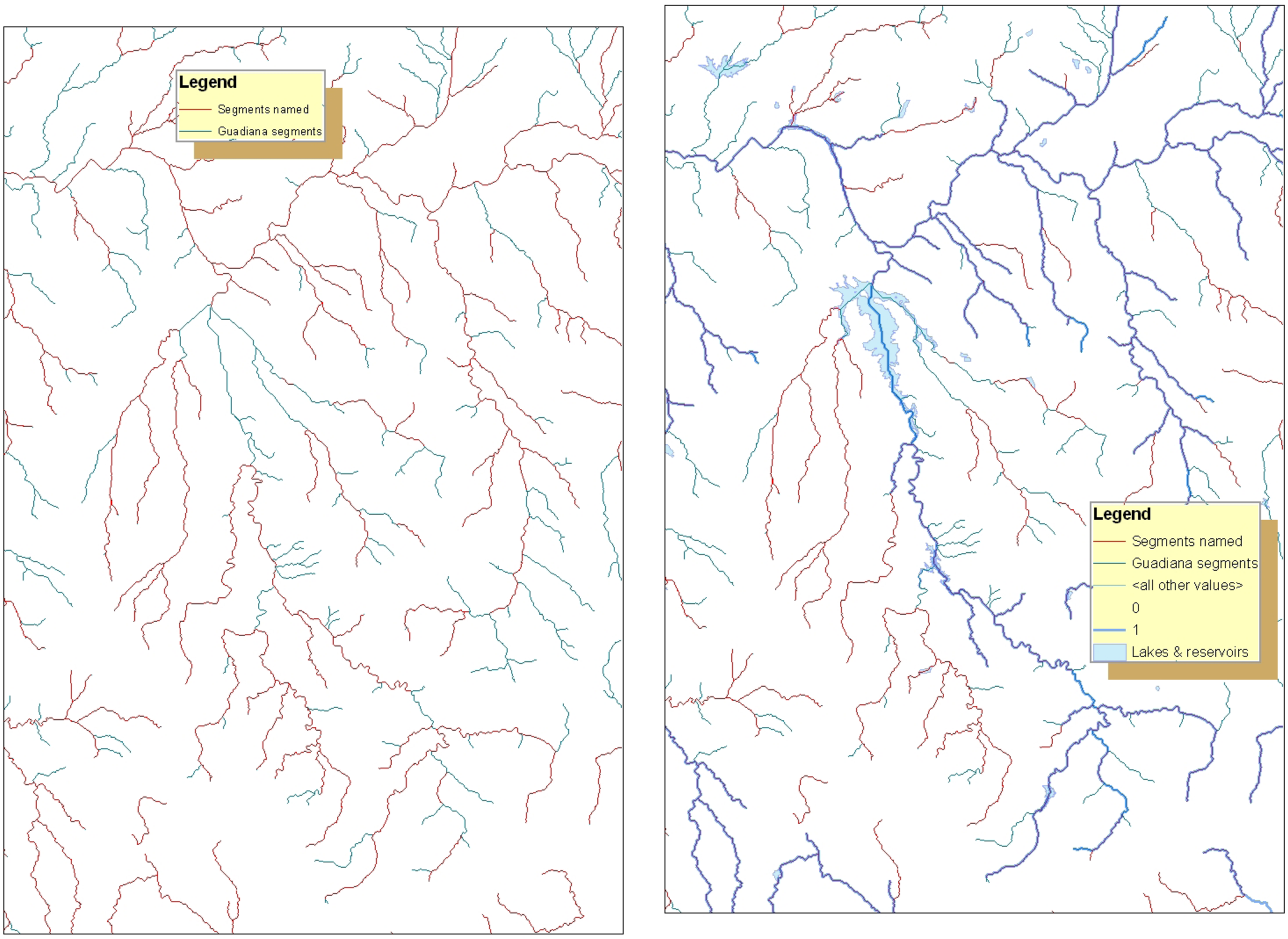
Processing is driven by the available data source: Figure 6.4 (left) indicates the source data. Red segments are those having information (name or ID), regardless of whether this data is appropriate or not, and the green ones are the non-documented segments. A large number of gaps can be observed: documented segments have no hook at lower river level.

The right side of the picture shows the result of the processing. Named rivers are in dark blue.

Many 'documented' rivers could not be exploited; by contrast the huge gap in the reservoir in the upper centre of the figure has been bridged because confluence with the Guadiana river that flows roughly horizontally was found, and the upper segments had the same name as the lower ones.

The general processing is based on the same logic as is the routing of segments; the difference is that naming of segments is the factor for setting a true river instead of distance to the sea. The ending point is necessarily either a terminal recipient or an already named river. To this end, the process scrutinises all basins having named segments and created rivers from the principal one, whose level is 1, and appends level 2 rivers that have a common node with the level 1 river, then takes all the level 2 rivers in this basin and create the level 3 rivers that have a common node with the level 2 rivers, and so on. Indeed only segments finding a match with level L can be the starts for rivers level L+1. A missing L+1 level inhibits the creation of all the levels L+n that depend on it.

Figure 6.4 Stages in building rivers from named segments



Source: Sample result of EEA processing.

The processing of rivers in each level is crucial since named segments can be used only if they have a candidate target for their attachment. The rationale is as follows.

- The most downstream segments having no downstream segment in each basin are identified. If the basin is coastal, many parallel rivers may exist; otherwise, if the catchment is continental, only a single outlet is allowed.
- Each most downstream segment is checked for the names and IDs attached, and all segments sharing the same names and IDs are extracted.
- Then all IDs and names for all the segments having the same IDs as the first selected segments are selected, firstly select more candidate segments, and secondly to enrich the list of synonyms ([19]).
- Collected segments are made unique (duplicates are no longer useful) and are sorted by decreasing distance to the sea, based on the assumption that the most distant are candidate starting points. Segments are populated with the source of naming, to make it possible to choose concurrent segments thanks to scoring based on the naming source.
- Contiguity checking is performed from spring to mouth in several steps. Different cases may crop up.

([19]) For example, the river Loire last segment is proper to ECRINS and outreaches French Loire downstream extend. Hence this segment has not captured the French naming for the Loire, whereas it has a CCM naming. Taking all segments with 'Loire' + CCM ID for the Loire also selects upstream segments with same ID, the French 'Fleuve la Loire' and the French ID (plus in this case another ERM ID + another spelling of the name can be captured). Hence the upper reaches incorrectly set by CCM are corrected.

(a) A single segment has been collected; the process is closed for this river that is made of a single segment.
   - Two segments are collected. Different cases occur.
   - Two segments are in line; the river process is closed.
   - Two segments are 'V' shaped (they have a common junction to the downstream river); one is selected by scoring and the other set as error.
   - Two segments are not in line; the process for several segments applies.

(c) A pile of two or more segments is collected; again, several cases are explored.
   - Segments are fully connected from top to down (verification is carried out in this way because adjacent segments with more downstream confluence are immediately eliminated); this is the most frequent case encountered.
   - A gap is detected during the up-to-down checking. A maximum of x (forced less than 10, default set to 5) consecutive segments are tentatively inserted to bridge the gap. If the gap is bridged, checking continues until the next gap or the end of the pile.
   - An unbridged gap raises the possibility of topological error (the upstream segment in fact connects to another set of segments), or results from lack of naming. Lack of naming occurs if the reaches are in a lake, and hence are no longer named as a river. In this case, an attempt to bridge the gap is made from downstream, and all the segments that exist (not selected yet) in the sub-basin upstream of the upmost segment in line from the mouth are selected. The former is flagged as topologically suspect; the latter makes a pile of connected segments.

5. In all cases where the number of segments (supposedly connected) is three or more, there are many possibilities of 'Y' types of piles. The issue is selecting the appropriate 'Y' branch because both have segments sharing the candidate name for the river. Selection is made by scoring (sum of scores for each segment). The scoring method hence takes into account the trust in the source and the number of segments documented. Rejected segments are flagged as rejected (not to be selected in next steps, albeit not flagged as topological errors).

The final pile is named, and all segments updated, along with the table of names and aliases. The complete process requires more than 10 hours and yields more than 22 000 rivers.

The number of river IDs can be counted; this is not the accurate number of rivers set, since some are incorrect. If a pile of three segments has respective names in source 'A', 'B' and 'C', each being named such by different source, the application cannot decide if 'A' is the alias of 'B', or if 'B' is actually a different river. Such an instance has occurred.

6.7 Future developments

The process has been sophisticated to the maximum affordable level and subsequent upgrades can result only from improvement in the quality of source data. Considering the application's flexibility, such improvements would mostly involve removing some incorrect names or populating one segment in the set with appropriate aliases.

The possible improvements address two different perspectives.

- Eliminating naming errors or uncertainties.
- Capturing more rivers. It is important to consider that populating **one** single segment may capture many supplementary rivers owing to the level-by-level approach. Figure 6.4 shows this.

The experience gained from routing rivers is now exploited, jointly with the outcomes of building rivers. The idea is to use the routes as possible masks for making rivers. In this case, the lack of connecting segments could be overcome. The algorithm could be refined when processing the next version.

Some radically different methodologies are being explored: the most promising is derived from routes and analysing all data sources (for example, the name of lake outlets, the name of rivers supporting a dam) as supplementary information to the vertice-to-segment name capturing process.

7 Lakes

7.1 Rationales

7.1.1 *Baseline information*

Lakes constitute a specific class of objects: there are natural, semi-natural or truly artificial lakes that can be related to river systems with outlets and in most cases inlets, depending on both hydrography and resolution of the river data sets.

Lakes (natural or not) are one of the three categories of water sparing (with aquifers and ice/snow) that turn an effective rainfall into a resource with a lag between weeks and centuries. Hence, lakes are key components of the water accounting procedure, and must be both inserted and connected in ECRINS and populated with hydrographical information.

In some circumstances, lakes can be true endorheic systems and hence have no outlet: Van, Issyk Kul, etc. (Pourriot and Meybeck, 1995). Some are very small (e.g. lake Pavin in France, that does, however, have a small outlet and hence is not truly endorheic), others larger, resulting from changes in the catchment system (e.g. lake Trasimeno in Italy), and still others may be huge inland seas. For example, the Caspian Sea is hydrologically speaking a 'lake' since it is not connected to the global ocean; likewise, the Dead Sea is both the deepest water body in the world and the most famous endorheic system.

This chapter reports the rules, production method and results relating lakes to their catchments, to the rivers, and when relevant, to dams.

Lake information comes from several sources that make the integration quite complicated. Moreover, requirements related to lakes make this multiplicity of sources extremely important because:

- the environmentally critical lakes are firstly those created or managed by damming, making it necessary to relate dams with the lake (a lake may have several dams; only one is considered the 'main dam');
- the important information attached to the lake (name, creation or modification date, and its hydrographical characteristics (volume, mean depth, maximum depth, etc.)) are not immediately evident from the geometrical information, whereas proxy river lengths or catchment area are directly produced by GIS processing.

Lakes participate in the production of relevant information as water reservoirs and water storage: water ageing, retention (settlement ponds for sediments, nutrients, etc.), ecological areas, and obstacles to fish migration. In this latter case, the dam is the main obstacle.

Since lakes slow down water movement and offer a larger area than rivers do to sun radiation and atmosphere, lake water tends to be warm. Where large number of small lakes have been created on the river stretch, especially as result of small dams, river habitats and their temperatures are widely affected.

Lakes are not straightforward defined objects. The objects to be taken into account depend on sources that possess their own definition of lakes. The analysis of cases led to a inserting a new category, a 'reference lake', that is defined in a further section.

However, it is not possible to manage in ECRINS the total number of lakes and small dams: there are practical limits to the details that can be incorporated. This report deals only with the lakes actually inserted as individual objects in the system. A later release could deal with important classes of objects not addressable by GIS tools and that require statistical assessment:

- areal hydrology to consider areas where a significant number of clustered objects result in a 'lumped' lake covering X % of the area and comprising N individuals;

- density of damming inside a FEC, for example, concerning either main drains, secondary drains ([20]) or the rest of river stretches not individually managed (making the 'small rivers' inside a FEC).

7.1.2 Conceptualisation issues

In ECRINS, a lake is a polygon with an ID and, if possible a name. The polygon must be a water mirror having sufficient duration to be considered a lake. However, different reasons can lead to having different 'lakes' not matching one to one considering different supplies. By contrast with rivers where the permanent channel is always unique (or identifiable as such), a large and shallow lake may make several smaller ones that can be considered as different 'lakes' or unique ones. Such a pattern can be the result of drying: the most famous example is the Aral Sea, now broken down into several disconnected pieces of water.

The lakes layer identifies all the still water masses described as such in the source data sets. It is not the role and capacity of ECRINS to provide a definition of a lake. Therefore, all water areas identified as such in CCM (lakes data set), the ERM, WFD reporting under Arts 3, 5, 8 or 13, or CLC water categories are candidates for inclusion or identification as a lake in the ECRINS lake feature class.

In contrast to drains and rivers, there is no strict equivalent to drains in the lake's area. From a geometrical point of view, a lake should be a joined polygon. In real life, separated water masses may have the same lake name, and a unique water body may have different names. The most obvious example is the Great Lakes in North America, where Lake Huron and Lake Michigan are from a single polygon, whereas lakes Superior, St Clair, Ontario and Erie are indeed different polygons.

Similarly to rivers, lakes are also cultural objects which have been given one or several names, even in the same country ([21]). Being cultural objects, a set of ponds can be considered as either a single lake or a series of different lakes. Analytical issues are need to be considered: close ponds are often grouped together by CLC whereas they can be subdivided by the ERM.

A dam in ECRINS is a hydraulic work, generally making a lake. The dam is the whole set of works; for example a dyke and its related sluice constitute a single dam if both works contribute to the same function. By contrast, the main dam and the pass dams are separate objects, because they are distinct, distant and contribute to different functions of the lake they make.

The inclusion proper is carried out provided no copyright denies the possibility of doing so. However, if the lake polygon is common to a copyrighted and a free of rights sources, the feature is considered as free of rights. The use of licensed sources has hence been limited for the production process since it may help determine if a free source of water mass constitutes a lake or not.

The ECRINS general conceptual model (Figure 2.1, page 23; Figure 2.3, page 24) considers the relationships between relevant elements. A lake is a specific object that can be conceptually described as a water body having (or not having) downstream relationships with a river and upstream relationships with none to several rivers, and situated in one (at least) FEC or extending over several FECs.

A lake can be from 0 to 100 % artificial, and to this extent, it is dependent on the dam that created or enlarged it. In the case of artificial lakes, the starting date of commissioning and the ending date become important features of the lake itself: they may even result in a sharp change between no lake and a lake. Natural lakes also have starting and ending dates, which can be recent even for large lakes ([22]). Such information is unfortunately not available and hence is not populated in ECRINS.

Since the information on lakes comes from several sources, special precautions were taken to trace the data source in the final layer; these are reflected in the data model.

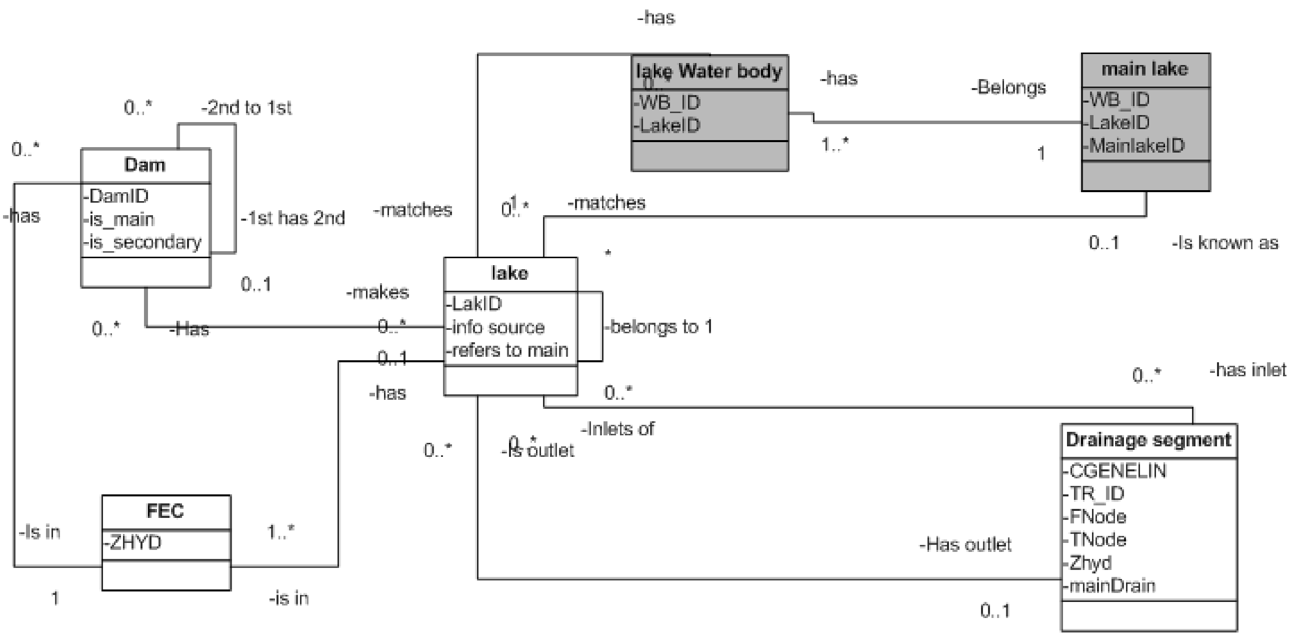
Apart of the normal inclusion of hydrographical objects, lakes in ECRINS must be potential hosts for the 'main lakes' as foreseen by the WFD. The simplest solution is to include in ECRINS all lakes having an area much less than the threshold for 'main lakes'. In practice, this simple rule is not

([20]) As defined in the second report.

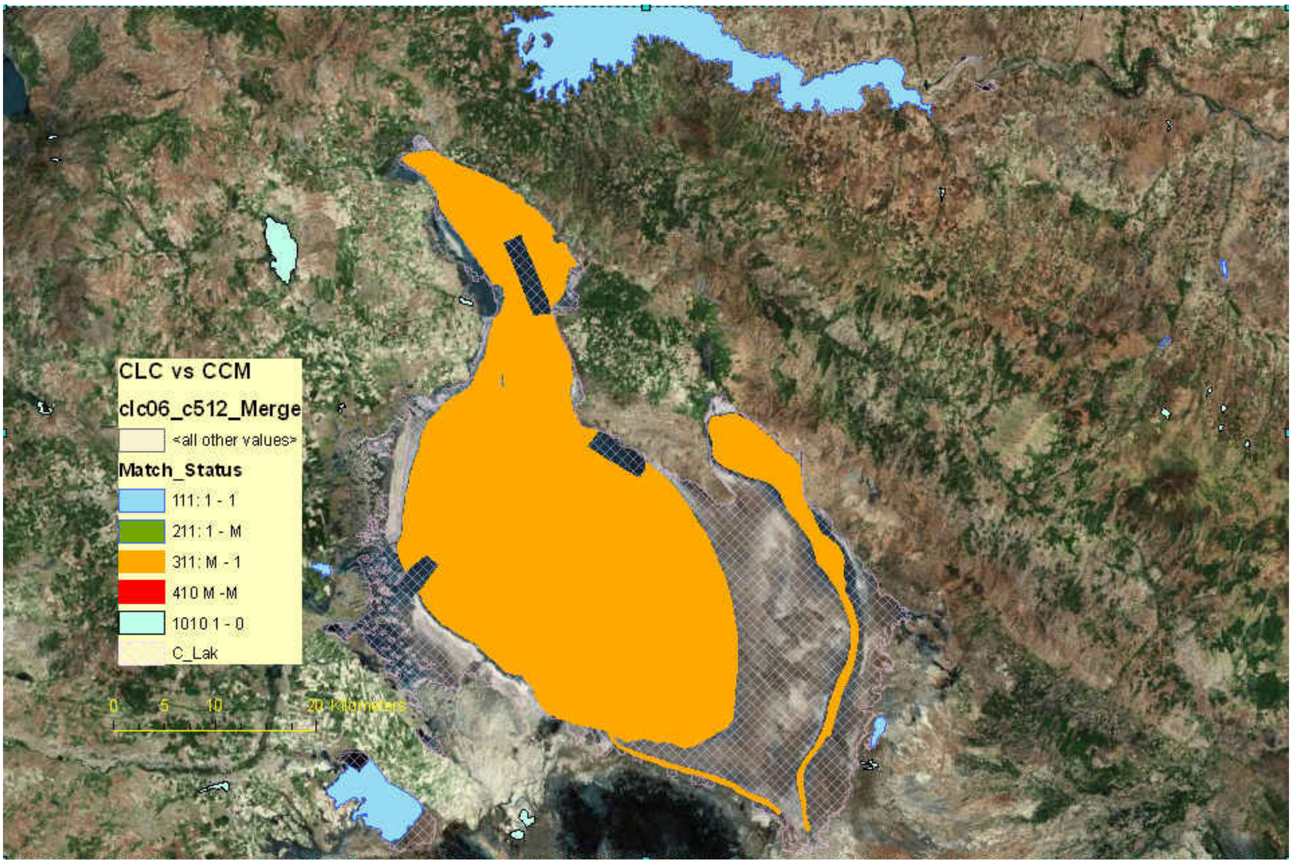
([21]) An popular anecdote among Israelis is that since they have only one freshwater lake, they gave it several names: lake Kinneret, lake of Tiberiad, lake Guinossar and the Galilean sea.

([22]) Lake Sarez in Tajikistan was created by a landslide resulting from an earthquake in 1911. The area of the lake at present is 80 km^2, its length is 60 km, and the volume of water is about 17 billion m^3 (http://www.tajik-gateway.org/index.phtml?lang=en&id=983).

Figure 7.1 Conceptual model of lakes in ECRINS



Map 7.1 'Main lake' concept: Lake Tuz Gölü (Turkey)



Note: Maximum extension of the lake is reported by the C_lak polygon (pink grid). The CLC polygons are orange because they are in a M-1 cardinality with the reference lakes layer.

always that straightforward since the cardinalities between 'lake water bodies'/'main lakes'/lakes are extremely complex and result from the intrinsic definition of lake and the management unit's definition by each Member State.

This conceptual data model was modified stepwise with experience. The key change is the insertion of the 'main lake' concept that is illustrated by several examples, the most obvious being the Tuz Gölü lake in Turkey. This lake is brackish and its extent varies from year to year. It was recorded as single water mass by CMM, and has 3 CLC polygons. Data processing will relate in such cases the smallest polygons to the biggest, keeping it as a reference lake.

The CCM-sourced polygon will be replaced by three lake polygons but its original size preserved as maximum extension and reported as source; the largest CLC polygon is mentioned twice as a reference primary lake.

As indicated in Section 1.2, page 16 and thereafter, not all information related to either lakes or dams is normally contained in the ECRINS data set; it can be found in the ancillary data sets as well.

By convention, the ECRINS lakes layer refers only to the physical design of the lake as it may have existed in the most common situation, disregarding date issues. Date and usage issues are dealt with in the topic database Eldred2, to which the lake features are linked thanks to their ID. In a transition phase, a double ID is maintained since the Eldred2 ID format is not equal to the ECRINS generalised ID format, having been developed three years prior. To facilitate processing, first and last years are indicated in ECRINS to allow the building of accurate maps for certain periods. The same conventions are adopted for dams as well.

In summary, ECRINS presents all lakes having existed or having potentially existed at the moment the layer is released, as practicable from sources. The lakes in the public delivery of ECRINS are limited by their public status.

7.1.3 Methodology for incorporating lakes into ECRINS

By contrast with drains, no complete or accurate lake information was present in CCM. The procedure used was developed in two stages to make it possible to have a lake layer that is as comprehensive as possible, updateable and ready for use.

A the moment the lakes are processed, the prerequisites are that all FECs are available as GIS features, as well as the drains (the field indicating whether these are main or not, being populated) and the aggregation catchments being available as GIS objects integrated in the ECRINS data set.

The work organisation consisted in the following.

1. In the first development of ECRINS, building an internal reference data set based on simple integration of ERM lakes and CCM lakes where the ERM was not present. This stage resulted in identifying the different lake polygons, from both sources under unique IDs, and flagging the data source.
2. For the final use and dissemination of ECRINS, stage 2 was developed at the end of 2011, taking stock of the updates permitted by integration of lake waterbodies, aiming at producing a reliable layer in which only data that is free of charge would be provided. The process in this second stage consisted in:
   (a) upgrading the design of ECRINS lakes table so that multiple sources could be traced and multiple sources be identified;
   (b) designing and operating the procedure in which licensed/incorrect polygons are replaced by appropriate polygons;
   (c) attaching the lake polygon to river stretch and making it easily updatable;
   (d) building the final attachment of the lake to the stretches as inlet stretches (that can be many) and an outlet stretch that is unique, defined or non-existing;
   (e) building the attachment to Eldred2 to link with both the dam on the one hand and with information contained in Eldred2 relating to the lake on the other hand (see specific chapter).
   (f) updating, from relevant sources, the hydrographical information attached to the lake.

To carry out these steps, specific developments have been made under the ArcGIS® and MS Access® application (WERC) process, to log the process phases and output the required data sets.

7.1.4 Decision rules

Since alternate sources can be proposed to those selected, a set of decision rules has been established. The rules are as follows.

- If no exception case is met, a CLC polygon substitutes any existing lake.

- If an Art. 13 lake matches 1–1 with a CLC lake, this Art. 13 lakes substitutes the CLC lake. This exception breaches the first rule because it ensures full geometrical compatibility with waterbodies attached to the lake.
- When a CLC lake matches in 1–M, a primary lake is defined.
- When CLC lakes match in M–1, the largest substitutes the previous lake in ECRINS; others are created from CLC.
- When CLC lakes match in M–M, the rule above applies, repeated until all polygons are included from CLC and eliminated from ECRINS.
- The case of 1–0 is trivial: the CLC lake is added; the reciprocal case results in keeping as private (ERM) or public (CCM) the currently existing lakes.

These rules are modulated according to the CLC category. In particular, there can be a conflict when inserting CLC lakes from 512, and subsequently, 411 categories. It is possible that a source lake is understood by CLC partly as a true lake and its shores described as marshes. The rule in this case is that the source lake is considered as a main lake and the marsh parts as referring to the same main lake. However, this can be applied only if both CLC classes have been processed in the same run.

7.2 Inclusion procedure

7.2.1 Lake reference sources

There is no lake reference source covering all ECRINS. The CCM v2.1 is a lake data set organised as a feature class. It comprises the fields listed and explained in the table below. The total number of lake objects is 89 943. The CCM lake table structure is reported in Annex 4, Table A4.1.

The size and number of lakes, however, is to a large extent spurious, because the CCM lake layer comprises many 'non-lake' lake polygons. These errors are generally the result of incorrect processing of the original data source, which relies first on the 'water layer' of the SRTM data set. In many cases, the number of lakes is locally overestimated because the same water mass has been split into several smaller ones during the process, or underestimated because neighbour lakes have been analysed as a single water mass. By contrast, very different lakes that are very close are identified as a single lake, thereby underestimating the number. All these issues were discussed in the preparatory documents (not published).

The ERM is the primary data source of importance in the building of the lakes data set. The ERM is a cartographic product delivered by EuroGeographics in two separate versions that do not coincide. The first version (nicknamed ERM0) comprises 147 970 polygons and the second (ERM2) counts only 147 430. The number of polygons is not homologous to the number of lakes, since a lake is divided into polygons inside country borders, that do not necessarily match together.

The ERM was understood as having the best geometry that can be expected for the countries covered and was used to identify whether an external source is likely to be a 'lake' or not.

Countries' deliveries under WFD Art. 13 are references since they come from country's geometries. However, their processes are complex, since the waterbodies-to-lakes cardinality covers the whole range from 1–1 to M–M. It constitutes both a reference and a feature source.

7.2.2 Lake features sources

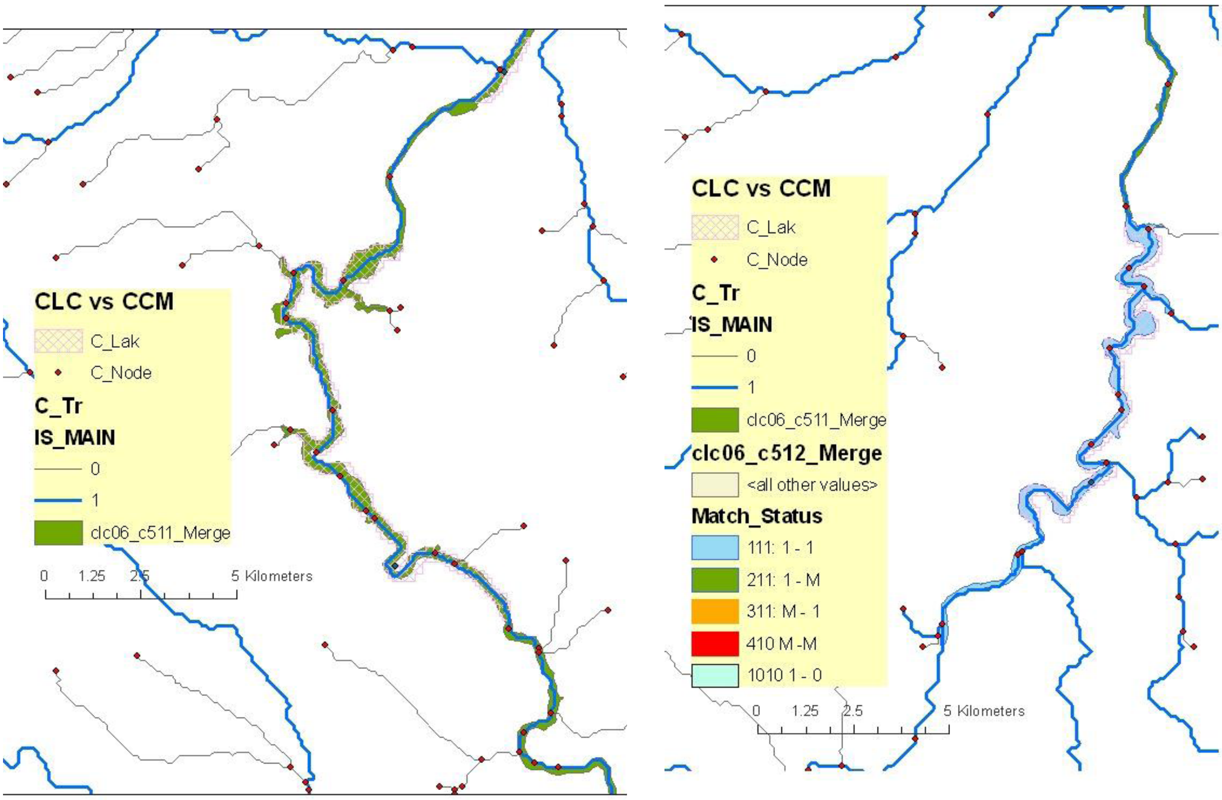
The most complete source of lakes is CLC. CLC 2006 covers the most recent area of Europe, and can be completed for Greece by CLC 2000. Lakes are identified primarily as category 512 (all categories listed in Section 2.2.3, page 27) and 521 (coastal lagoons).

Categories 511 (running waters) and 411 (Inland marshes) are candidates for lakes. The two later categories, comprising some large reservoirs, are found in 511, and potential non-lakes in 411.

Reservoirs are often made by damming a river. If the dammed valley is deep and narrow, the water mirror is not discriminated from the river. This is the case for the Villerest dam and lake in France, on the upper Loire. This CLC practice is somewhat odd since a smaller reservoir, upstream Villerest (Grangent lake) is recorded in category 512. Processing such cases poses difficult issues that can be solved and secondary processed by clipping a segment of the 511 CLC feature when a lake is suspected. This is illustrated in Figure 7.2.

However, CLC building has some exceptions and errors. In principle, all reservoirs in rivers should be identified as at least individualised polygons. This not always the case: series of large dammed reservoirs in Spanish rivers belong to a single river (511) polygon. By contrast, in there countries, a small lake is apportioned between 411 (wetland area), an

Figure 7.2 Lake recorded as river in CLC: Villerest vs Grangent dams



Note: Left side Villerest, right side Grangent. In both cases, the river or lake identified by CLC matches a lake in ECRINS (C_lak); in the Grangent's case, the match is 1–1 and the former CCM lake shall be substituted.

upper reservoir is reported as 511 (river), and the main lake as lake (512).

The number of lakes lumped together as rivers made it necessary to create a temporary class on inclusion, the 'pseudo-inclusion'. When a lake was present on a river as any source polygon and matches a CLC 511 polygon, a dummy CLC polygon is created (Corine ID + arbitrary rank) that is reported as source polygon, despite the former lake polygon not being substituted.

This method indicates to users that the lake is 'under processing' until CLC has been corrected. The number of such cases is around 1 000 over the rounded figure of the final 72 000 lakes in the new layer.

7.3 Final structure of data sets

7.3.1 Main lakes feature class

The main table is the lakes feature class: this is rather simple and contains only main fields. Country, NUTS, river segments and FECs are not reported since their cardinality is 1–many. They are stored in additional tables. The main table structure is reported in Annex 4.

The lake identifier, LakID, follows the syntax indicated in Table 3.2, page 36. As already mentioned, the lakVersion field may have several values in a table to earmark the date when the lake was introduced. When a lake created in stage 1 is substituted, the lakeVersion is the version computed when the inserting is carried out; for example, as CLC processing is done by category and not as a whole, a new version is created per category and set for those lakes substituted.

Lakes are defined by a single polygon proper. To facilitate subsequent uses, the centroid is systematically computed and provided. Since a lake can be represented by a convex polygon (in which the centroid falls) and concave polygons (outside which a true centroid may be placed), a pseudo-centroid is defined as being a point inside the lake polygon and as close as possible to the polygon barycentre or replaced by any point inside the lake polygon, in the case where the barycentre would fall outside. Geographical coordinates (WSG84) are entered into the table as well.

7.3.2 *Inlets and outlets*

The relationship between lakes and river segments allows using all information provided by the topology of catchments and rivers. Hence the lake vs drains relationship has been managed in the most comprehensive way possible. Figure 7.2, page 81 shows the drains and the lakes along with the nodes and facilitates understanding of the notions of inlets and outlets. The process is fuelled by intersecting lakes and river segments, and then processing the results using the segments topology.

It combines different information on rivers and on FECs, this may appear to be duplicated later, and Table 7.1 is kept for ensuring downstream compatibility. It creates the flat tables LakInOut (inlets and outlets) and LInterFEC (apportionment of final lakes per FEC). In the delivered structure, names are V_lakInOut and V_Linterfec for matching NOPOLU Système 2 syntax.

The paired flags (Is_MaDr: is main drain, Is_SeDr: is secondary drain) are mutually exclusive, except if InOut='X', in which case they are both set to 'false'. Fields Ztr and Zfec report respectively the segment (closest node) or FEC elevation, as taken from source data set. Information in ZFec is the average altitude in metre above sea level of the FEC. This last value is purely indicative, since it is computed from all the elementary catchments constituting the FEC, disregarding their relative size and the placement of the lake.

The absence of lake-to-FEC relationship (last line in the displayed sample) is explained by the fact that some lakes are situated in islands that are not part of FEC layers or just don't belong to the elementary catchments layer (endorheic lakes).

7.3.3 *Lakes per country and per NUTS*

Most lakes are inside a single country and a smaller number is inside a single region (nicknamed 'NUTs' and as defined in Table 3.3, page 37).

Table 7.1 Table of 'lakes vs drains' relationships LakInOut: sample

LakInOut

lakID	InOut	SegID	FecID	Is_MaDr	Is_SeDr	ZTr	ZFEC
OC00000002	O	ZE00069784	W3A1013901	☐	☑	44	
OC00000003	X	#Joker#	E020147815	☐	☐	-32768	6
OC00000004	X	#Joker#	W3N8005305	☐	☐	-32768	
OC00000005	X	#Joker#	W3N8005304	☐	☐	-32768	
OC00000008	O	ZJ00076505	W3M4119001	☐	☑	51	
OC00000008	I	ZJ00038209	W3M4119001	☐	☑	51	
OC00000008	I	ZJ00075370	W3M4119001	☐	☑	51	
OC00000009	O	ZL00042967	L020008407	☑	☐	480	110
OC00000010	O	ZL00009677	L020008439	☑	☐	457	81
OC00000010	I	ZL00009915	L020008457	☑	☐	457	91
OC00000010	I	ZL00010234	L020008439	☐	☑	457	81
OC00000010	I	ZL00026230	L020008439	☑	☐	457	81
OC00000010	I	ZL00042967	L020008407	☑	☐	457	110
OC00000011	X	#Joker#	H020089326	☐	☐	-32768	31
OC00000012	X	#Joker#	W3N8005309	☐	☐	-32768	
OC00000013	X	#Joker#	#Joker#	☐	☐	-32768	-3276
OC00000015	X	#Joker#	#Joker#	☐	☐	-32768	-3276
OC00000018	X	#Joker#	H020089456	☐	☐	-32768	27
OC00000019	X	#Joker#	#Joker#	☐	☐	-32768	-3276
OC00000020	X	#Joker#	H020089611	☐	☐	-32768	31
OC00000021	X	#Joker#	#Joker#	☐	☐	-32768	-3276

Source: WERC application, EEA. T_71_ Table_lakInOUt.png

This table stores for each lake, its LakID, information InOut ('O': Outlet, 0/1 record per lake, 'I' for Inlet(s), 0/many records par lake, 'X' no information found for inlets or outlet.

Table 7.2 Structure and sample of table LakperFec

LInterFEC

LakID	Zhyd	LInFECkm2	LakKm2	FECkm2	CF_LIF_times
OC00000002	W3A1013901	0.310008672104375	0.310008687237809	224.32	1.00000004881616
OC00000008	W3M4119001	0.539980370634567	0.53998036592173	110.37	0.999999991272208
OC00000009	L020008407	2.99909571949283E-02	2.99909543459687E-02	109.13	0.999999905006045
OC00000010	L020008439	1.34998522575954	1.67001954115028	57.49	1.00000059724102
OC00000010	L020008407	0.170018809077626	1.67001954115028	109.13	1.00000059724102
OC00000010	L020008457	0.150014508909535	1.67001954115028	41.9	1.00000059724102
OC00000118	H030084718	9.99606039318404E-03	9.99605911216228E-03	299.72	0.999999871847337
OC00000131	W3N8001301	2.85005573902776	2.85005572699895	293.23	0.999999995779446
OC00000134	W3N8001301	4.6599540753795	4.65995397488444	293.23	0.999999978434325
OC00000135	W3N8001301	5.92995697164225	7.34994114266396	293.23	0.999998987191865
OC00000135	W3N8001201	1.41999161510943	7.34994114266396	245.08	0.999998987191865
OC00000136	W3N8001201	2.87998954509685	2.87998952862851	245.08	0.999999994281808
OC00000138	H020090360	2.90997104001781	7.82999310189698	3.74	1.00000007838897
OC00000138	H020090298	2.31004660498421	7.82999310189698	9.77	1.00000007838897
OC00000138	H020090362	1.19000329922158	7.82999310189698	48.38	1.00000007838897
OC00000138	H020090522	1.08998308896175	7.82999310189698	28.32	1.00000007838897
OC00000138	H020090374	0.329988454926541	7.82999310189698	21.06	1.00000007838897
OC00000150	H030084831	5.52002293703078	5.52002291442132	239.85	0.9999999959041
OC00000154	W3N8100101	[illegible]	[illegible]	[illegible]	[illegible]

Source: WERC application, EEA.

Since the source information of the area apportionment results from the process, it has been added to the final table in the event it would be useful for external users of ECRINS.

Country and regions are the ISO 2 character code and the NUTS code with the leading country ID.

7.3.4 Lake apportionment by FEC

This information is primarily needed for the water balances under the water accounting procedure. The System of Environmental-Economic Accounting for Water (SEEAW) requests that the evaporation from free water masses be separated from the evapotranspiration from land. This information can be very useful in dry areas where the multiplication of surface reservoirs may lead to substantial water loss. However, evaporation and precipitation are primarily computed at FEC level, making it necessary to apportion the lakes (and all features that are computed in the same way) by FEC.

This information is computed by intersecting all candidate lakes with the features, given a zero metre tolerance to keep the sum of split areas equal to source lake area, and is FEC version dependent.

Table 7.2 shows the lake area, FEC area and intersection area (field LinFECkm^2). The correction factors, applied by multiplying the sum of LinFECKm^2 by it, yields the total lake area.

7.3.5 Altitude complements

Lake altitude is important information which is seldom populated in the source data sets: the CCM source has ~ 50 % altitude set to joker and the field 'altitude' is missing in the ERM source ([23]). The STRM90_100m, completed beyond 60° N has been used to complement the missing altitude data.

However, computation from STRM often provides erratic results when not enough altitude pixels are captured by the lake polygon: in some circumstances the lake watershed is captured as well, and provides largely incorrect data.

When possible, the altitude at inlet and outlet nodes is taken and the FECs are considered as well, to obtain an outcome with 'believable' altitude. Altitude from the lake is considered when these sources do nay yield results.

([23]) Discussion with the ERM project manager confirmed that this information was not considered important, and hence was discarded, because it is infrequently provided by the EuroGeographics' partner geographical institutes.

The delivery of RDA 30 m DEM instead of the current DEM will significantly improve the computation of elevations; this calculation has not been carried out in the version.

7.3.6 Hydrographical data and naming

From modelling perspective, a lake is first and foremost a water mass having an area and a volume. Ancillary information such as maximum depth and residence time are important, but not essential. Volume, area and mean depth can be completed if the two are provided, but having all three allows for consistency checks.

Hydrographic data is not part of the geographical data sets and is gained from alternate sources. Two sources are currently exploited to populate these variables: country deliveries (all types of deliveries) and open sources, of which Wikipedia is one.

External data sources are evaluated by geography and by lake name as controls.

8 Preview of the future of ECRINS

8.1 Ideal situation

ECRINS is a combined product with modelled features (catchments) and geographical features (rivers and lakes). It creates a fully documented relationship between all these components. The developers of CCM, the base bricks of ECRINS, constructed it entirely using modelling because of the practical impossibility of modelling catchments and using river geometry from scratch, and not having the flow direction. If this had been technically possible, licensing issues would have jeopardised the process, beyond the cost of the product.

However the principle of having modelled catchments as 'ecrins' ([24]) of rivers having accurate geometry and proper topology is still an objective. This objective would eventually combine:

- current and improved ECRINS,
- GIS sources: CLC, WFD, ERM, etc.,
- open attributes sources: Wikipedia, literature, etc.,
- outcomes from the RDA project.

With time, three steps can be considered in the development of ECRINS.

- Current step: ECRINS is fully from model, but matches external information from geographical systems.
- Homologous (see Section 8.2) elements (nodes and connecting segments) are identified between both systems; this can be achieved notwithstanding the licensing issues.
- Based on found homologies, the ECRINS elements that are geographical are replaced by their homologous elements to achieve a fully fledged topological system with good geometry.

8.2 Methodological principles: finding homologous objects

The principle for achieving such a goal is quite simple. Let's assume that in a certain catchment, the basins are correctly delineated. The river system, as represented by the modelled central lines of the rivers, is articulated by the nodes: the nodes are exactly like the body's bone joints. If considering the main rivers and their affluent, as defined through routes and naming, the nodes in systems that connect the same river elements are homologous. For example the node linking the Loire and the Allier has the same function in ECRINS and in ERM.

Applying this concept of comparing anatomy to hydrology is not that frequent, but it is very powerful, because it allows for matching homologous elements even though their geometries don't fully overlay.

A very active cooperation has started between the ERM and the EEA to prepare, on test catchments, this comparison. In parallel, the development of RDA continues and will provide much more accurate DEMs to which ECRINS should snap better.

8.3 Correcting topological errors in ECRINS

ECRINS rivers are 'dug' by model and not by erosion; they do not systematically have the actual placement and relationship with their relatives. A systematic search for topological errors is being carried out and the location of homologous objects will help in this task.

Correcting topological errors is a rather complex task because all elements in ECRINS are related together. Correcting this calls for:

([24]) In French un écrin is the box containing a present, generally a jewel.

- identifying the source point at which to input the correction;
- recomputing all the depending relationships and values.

Changing the A segment destination from segment B to segment C calls for:

- recomputing all B and C descendants with the new Strahler level;
- recomputing all A antecedents and updating their outlet IDs;
- building new relationships at the elementary catchment level and their depending aggregates (Strahler 2, 3, etc.).

Once done, the entire ECRINS needs to be recomputed. This is why such operation cannot be carried out every time an error (generally impacting a very limited upstream area) is detected.

This methodology is suited to limited errors that do not substantially change the delineation of elementary catchments. Otherwise, a different method could be used, that has not been analysed since it is not required for the time being.

As inputs, there is little otherwise practice than taking a segment and attaching it to another node. Programmes under ArcGIS® exist and are being tested. The major issue is not changing the attachment but zooming to the potential error, which requires automation for systematic detection of 'candidate errors', as summarised in the section dealing with river naming.

References

[Author], 2012, 'ArcBruTile — Tile Services in ArcGIS' (http://arcbrutile.codeplex.com/releases/view/55377) accessed 17 March 2012.

Ansorge, C., 2010, 'Report on the Requirements of the INSPIRE Hydrography Data Specification for the EEA European Catchment and RiversNetwork System (ECRINS)'.

BODC, 2012, 'SeaVoX', British Oceanographic Data Centre (https://www.bodc.ac.uk/data/codes_and_formats/seavox/) accessed 16 March 2012.

Bredahl, L. and Sousa, A., 2006, 'European River Catchments, Version 1.01', European Environment Agency.

DynamicDrive, 2012, 'Tower of Hanoi' (http://www.dynamicdrive.com/dynamicindex12/towerhanoi.htm) accessed 16 March 2012.

EC, 2000, Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy (OJ L 327, 22.12.2000, p. 1–73).

EC, 2003a, 'Identification of water bodies', Guidance Document no 2, Common Implementation Strategy for the Water Framework Directive (2000/60/EC), European Communities, Luxemburg (http://circa.europa.eu/Public/irc/env/wfd/library?l=/framework_directive/guidance_documents/guidancesnos2sidentifica/_EN_1.0_&a=d) accessed 16 March 2012.

EC, 2003b, 'Implementing the Geographical Information System Elements (GIS) of the Water Framework Directive (2000/60/EC)', Guidance Document no 9, Common Implementation Strategy for the Water Framework Directive, European Communities, Luxemburg.

EC, 2003c, 'Towards a Guidance on Reporting under the Water Framework Directive', Common Implementation Strategy for the Water Framework Directive (2000/60/EC), European Communities, Luxemburg.

EC, 2008, Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive) (Text with EEA relevance) (OJ L 164 , 25/06/2008 P. 0019 — 0040).

EEA, Collins Bartholomew, 1998, 'The European Rivers and Catchments Database', European Environmental Agency.

EEA, 2012, 'Biodiversity indicators SEBI 2010 among the “best ideas to save nature”' http://www.eea.europa.eu/highlights/biodiversity-indicators-sebi-2010-among-the-2018best-ideas-to-save-nature2019) accessed 17 March 2012.

Eionet, 2012, 'Eionet, Central Data Repository' (http://cdr.eionet.europa.eu/) accessed 17 March 2012.

ETC/LUSI and EEA, 2008, *Building the EEA European Catchment and Rivers Network System (ECRINS) from CCM v2.1. Part 1: Setting and implementing rules an producing the watersheds layer (βv2)*, European Topic Centre on Land Use and Spatial Information, and European Environmental Agency, Copenhagen.

EuroGeographics, 2005, *EuroRegionalMap. Pan-European database at medium scale*, Technical producer, Brussels.

Eurostat, 2012, 'Eurostat, GISCO Geographical Information and maps' (http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco_Geographical_information_maps/geodata/reference) accessed 16 March 2012.

FAO, 2012, 'Greece — A survey on carp and eel culture fisheries in lake Ioannina. A report prepared for the study of fishery methods and requirements in Epirus project. Observations and Accomplishments', Food and Agriculture Organization of the United Nations Corporate Document Repository (http://www.fao.org/docrep/field/003/F9266E/F9266E02.htm) accessed 16 March 2012.

Furnans, J. and Olivera, F., 2000, *Watershed Topology — The Pfafstetter System.*

Füssel, H.-M., 2007, 'Vulnerability: A generally applicable conceptual framework for climate change research', *Global Environmental Change*, (17) 155–167.

IHO, 1953, *Limits of Oceans and Seas*, International Hydrographic Organization, Monte-Carlo.

IHO, 2002, *Names and limits of Oceans and Seas*, International Hydrographic Organization, Monaco.

IHO, 2012, 'International Hydrographic Organization' (http://www.iho.int/srv1/) accessed 16 March 2012.

JRC, 2008, CCM2 *River and Catchment Database for Europe, Version 2.1 Release Notes*, Joint Research Council, Ispra (http://ccm.jrc.ec.europa.eu/documents/JVogt_etal_CCM21.pdf) accessed 16 March 2012.

Leonard, J. and Crouzet, P., 1998, *Lakes and reservoirs in the EEA area*, European Environment Agency, Luxembourg.

Pourriot, R. and Meybeck, M., 1995, *Limnologie générale*, Masson, Paris.

Pöyry, 2010, *Completing ECRINS reference layer to carry out water accounts production on a regular basis in synergy with the WFD implementation. ECRINS naming application at EU scale*, European Environment Agency.

SeaVoX, 2009, 'Polygon data set of the extent of water bodies from the SeaVoX Salt and Fresh Water Body Gazetteer'.

Strahler, A. N., 1957, 'Quantitative analysis of watershed geomorphology', *Transactions of the American Geophysical Union*, (8) 913–920.

UNSD, 2007, *System of Environmental-Economic Accounting for Water (Statistical Commission Background document)*, United Nations Statistics Division, Thirty-eighth session 27 February – 2 March 2007, New York (http://unstats.un.org/unsd/statcom/doc07/SEEAW_SC2007.pdf) accessed 16 March 2012.

UNSD, 2003, *Integrated Environmental and Economic Accounting. Handbook of National Accounting*, United Nations Statistics Division, European Commission, International Monetary Fund, Organisation for Economic Co-operation and Development, World Bank, New York (http://unstats.un.org/unsd/envAccounting/seea2003.pdf) accessed 16 March 2012.

UNSD, 2010, *International recommendations for Water Statistics*, United Nations Statistics Division, New York.

Valero, Antonio, Uche, J., Valero, Alicia, Martínez, A. and Escriu, J., 2007, 'Physical Hydronomics: application of the exergy analysis to the assessment of environmental costs of water bodies. The case of the Inland Basins of Catalonia', in: *Seminari sobre costos i comptes de l'aigua a Catalunya en relacio amb la Directiva marc de l'aigua*, Generalitat de Catalunya, Barcelona, p. 10.

Vogt, J., Soille, P., de Jager, A., Rimaviciute, E., Mehl, W., Foisneau, S., Bodis, K., Dusart, J., Paracchini, M.-L., Haastrup, P. and Bamps, C., 2007, *A Pan-European River and Catchment Database*, European Commission, Office for Official Publications of the European Commission, Luxemburg.

WFD CIS, 2005, *Reporting under the Water Framework Directive. Informal guidance document on compliance reporting using the Water Information System for Europe (WISE)*, Directorate General for the Environment of the European Commission, Brussels.

Annex 1 Data structure of key ECRINS feature classes

Feature class C_Zhyd contains all FEC information. The current definition of the data set has been completed to carry out all applications as developed under NOPOLU. Table A1.1 shows the changes.

Table A1.1 Main table of FECs' detailed data structure in feature class C_Zhyd

Nom	Type	Size	Comments
B	String	10	n/a
Bas0_ID	String	10	ID of the FEC envelope
BASINNAME	String	50	Name of the BAS0_ID
BV	String	10	n/a
CalcNum	Long I	4	Refers to the calculation, i.e. the settings of the ECRINS version (thresholds, used for processing)
Code_Arbo	**Memo**	**n/a**	**n/a**
Ctry	String	10	Country code as ISO2, set to 10 for consistency reasons
DRAINPPAL	String	10	CGNELIN f the main river drain
Exutoire	Boolean	1	TRUE if the FEC is outlet
FRBD	String	10	n/a
FRBD_BC	String	10	n/a
FSU	String	10	n/a
ILD_ID	Long I	4	n/a
Is_Anomaly	Boolean	1	If the FEC has been set as anomaly (discordance WFEfBas3/FEC FC)
Is_BasS3	Boolean	1	n/a
Is_Coast	Boolean	1	n/a
Islandname	String	50	Name of the island, if FEC part of the island
Is_Upstream	Boolean	1	n/a
Libelle	String	50	n/a
MeanElev	Double	8	Mean altitude
Nb_Catch	Integer	2	n/a
nbFeature	Long I	4	Number of elementary objects (basins if coastal, Strahler 1 catchments otherwise)
NextDown_ID	String	10	ID of the FEC immediately downstream
NUTS2	String	10	Region code, set to 10 for consistency reasons
OBJECTID	Long I	4	n/a
Outlet	String	10	n/a
SB	String	10	Aggregation catchment, based on Strahler rank (SS=Sub-catchment Strahler)
Sea_CD	Integer	2	n/a
Shape	Objet OLE	n/a	n/a
Shape_Area	Double	8	n/a
SS	String	10	Aggregation catchment, based on Strahler rank (SS=Sub-catchment Strahler)
Surf	Double	8	FEC area, km^2
Surfc	Double	8	Cumulated area upstream, FEC not included

Nom	Type	Size	Comments
Surffinal	Double	8	Cumulated area, FEC included
System_CD	String	1	Main recipient system
W1LinkD	String	10	Strahler 1 catchment ID linking to current FEC in the downstream one (is not part of this FEC)
W1LinkU	String	10	Strahler 1 catchment ID linking in current FEC in the downstream one (is part of this FEC)
Window_C	Integer	2	Source window
WX02ID	String	10	ID of the Strahler 2 level used to make the CZHYD
WX03ID	String	10	ID of the Strahler 3 level used to make the CZHYD
WXSOID	String	10	ID of the sea outlet
XBaryDD	Double	8	n/a
YBaryDD	Double	8	n/a
ZG	String	10	n/a
ZHYD	**String**	**10**	**FEC ID**
ZU1	String	10	n/a
ZU2	String	10	n/a

Note: Fields are sorted by alphabetic order, not by placement order in the table.

Table A1.2 Main table of segments' detailed data structure in feature class C_Tr

Nom	Type	Size	Comments
Bas0_ID	String	10	ID of the basin in which the segment is located (is a set of FECs)
BURNED	Long I	4	No of pixels burned, i.e. forced at this place (CCM)
CALCNUM	Long I	4	ECRINS version
CATCHMENT_ AREA	Double	8	n/a
CGNELIN	String	10	Route ID
CONFIDENCE	Long I	4	n/a
CONT_PIXELS	Long I	4	No of pixels used for making the segment (CCM)
EXPORTVERS	Long I	4	Versioning ID
FNode	String	10	ID of the upstream node of the segment
Gauging_ Station	String	10	ID of relevant gauging station
Is_2Keep	Integer	2	Management flag
IS_D	Boolean	1	TRUE is segment is outlet — has no downstream
IS_MAIN	Boolean	1	TRUE if segment belongs to a main drain
is_MainCTR	Boolean	1	TRUE if segment hosts a main river from WFD (not updated in v1.0)
IS_Named	Boolean	1	TRUE if River_ID is populated with a river ID
IS_Routed	Boolean	1	TRUE if segment is routed (CGNELIN filled)
IS_U	Boolean	1	TRUE if segment has no upstream (is spring)
L_F2SOURCE	Double	8	Not populated, yet will be distance to source on the same route
L_SEG	Double	8	Segment length, kilometres
L_T2MOUTH	Double	8	Distance to mouth from the upper node, in kilometres
LENGTH	Long I	4	Length of the segment
LONGPATH	String	1	CCM partly documented information; kept for compatibility with ECRINS β
MainChan	Integer	2	NOPOLU reserved field
MAINDRAIN_ CLASS	Integer	2	n/a

Nom	Type	Size	Comments
MAINDRAIN_ ID	Integer	2	n/a
Model	Integer	2	NOPOLU reserved field
Nb_Amo	Integer	2	No of segments upstream
Nb_AmoF	Integer	2	No of segments upstream, after routing
NXDownID	String	10	ID of FEC downstream of FEC where segment is located
NXDownL1	String	10	ID of the Strahler 1 catchment to which segment connects
OBJECTID	Long I	4	n/a
Pente	Double	8	Slope
Quality_ Station	String	10	ID of relevant quality station
River_ID	String	10	If named, ID of river which segment is part of
RIVRANK	Integer	2	River rank, populated only if Is_named is TRUE (otherwise -1)
RouRank	Integer	2	River rank, populated only if Is_routed is TRUE (otherwise -1)
SHAPE	OLE Object	n/a	n/a
SHAPE_Length	Double	8	n/a
STRAHLER	Integer	2	Strahler rank of segment
SurfC	Double	8	Cumulated area upstream
TNode	String	10	ID of downstream node of segment
TR	String	10	ECRINS TR ID
TR1_Amo	String	10	ID of first met upstream segment (random choice)
TR1_AmoF	String	10	ID of upstream segment on the same route
TR2_Amo	String	10	ID of second met upstream segment (random choice)
TR2_AmoF	String	10	ID of upstream segment on the other route
WB_ID	String	10	unused
WINDOW	Integer	2	CCM source window of provisioning, kept for compatibility
WX01ID	String	10	ID of Strahler 1 catchment in which segment is located
WX02ID	String	10	ID of Strahler 2 catchment in which segment is loated
WX03ID	String	10	ID of Strahler 3 catchment in which segment is located
WXSOID	String	10	n/a
ZHYD	String	10	ID of FEC in which segment is located

Note: Fields are sorted by alphabetic order, not by placement order in the table.

Indexes of table C_TR

Table has TR as primary key (no duplicated segment can be accepted) and is indexed on fields Bas0_ID (basin to which this belongs), FNode (upstream node) and TNode (downstream node). Index has been set to WB_ID, but this was found to be of no practical use.

Table A1.3 Main table of nodes' detailed data structure in feature class C_Node

Name	Type	Size	Comments
ELEV	Long I	4	Z altitude of the node
ID	Long	4	CCM integer node ID
Is_2Keep	Integer	2	Working field
LEN_TOM	Long I	4	Working field
LENK_FRS	Double	8	Working field
LENK_TOM	Double	8	Working field
NodID	String	10	Primary node ID
NUM_SEG	Long I	4	n/a
OBJECTID	Long	4	n/a
Shape	OLE Object	n/a	n/a
SOURCE	String	1	n/a
WINDOW	Integer	2	n/a
WSO_ID	Long I	4	n/a
WXSOID	String	10	n/a

Indexes of table C_Node

Table has NodID as primary key (no duplicated node can be accepted)

NodID Ascending

Annex 2 Clustering of elementary catchments according to their Strahler levels

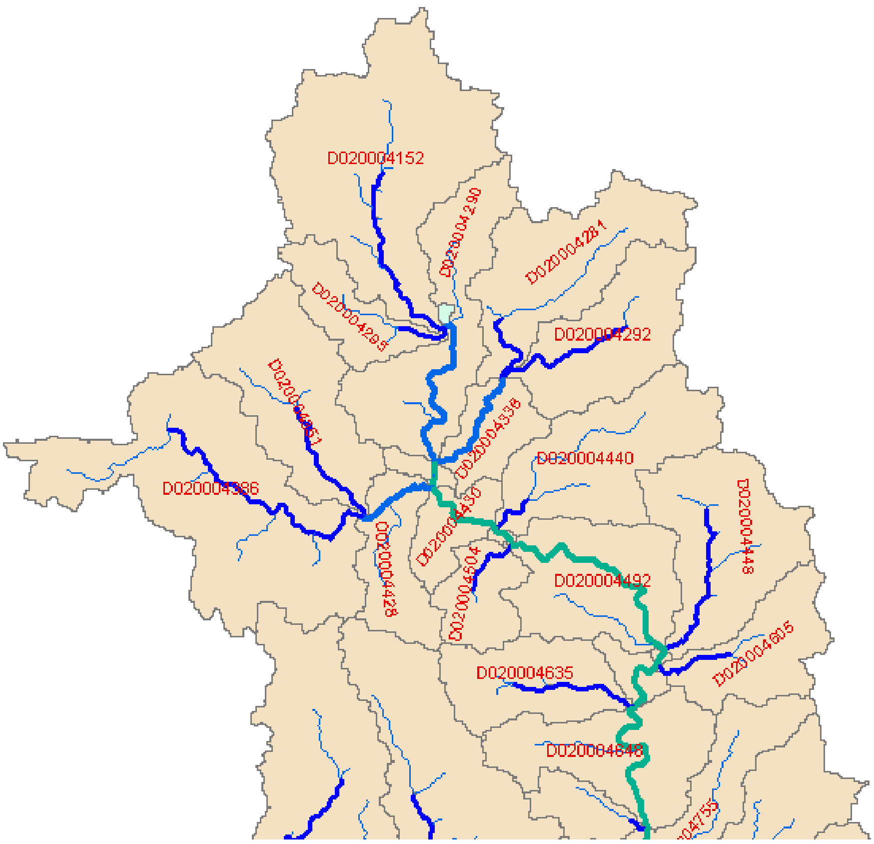
ECRINS is based on the construction of FECs by aggregating catchments of different Strahler levels. Strahler levels do not normally apply to catchments and only refer to rivers. By extension of the concept, the authors of the CCM have applied the Strahler concept to the different clusters of catchments, based on the Strahler levels of the rivers they contain.

The Strahler level of rivers is very simple: two rivers of same Strahler level S that make a confluence create a river with level S+1. If two rivers with a different Strahler level make a confluence, the resulting river has the biggest of the two Strahler levels.

When applied to catchments, two rivers of Strahler level 1 (the most elementary) are contained in a catchment with the Strahler level of 1. The confluence of these two rivers make a level 2 river which is contained in a level 1 catchment, since this is the original delineation of catchments.

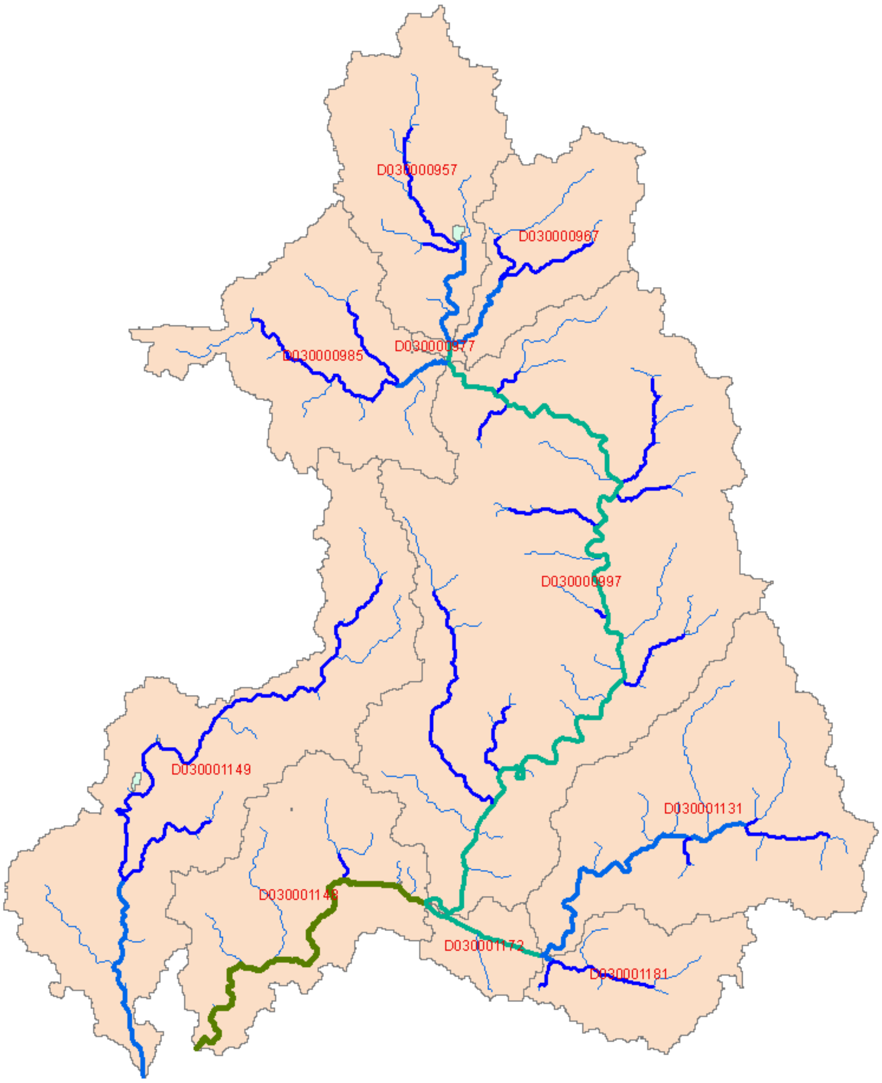
The rule applied by the CCM authors is that a catchment is Strahler order X, if it is the end of a catchment containing a river of Strahler order X **and** having confluence with Strahler X or upper, or if it is a catchment containing a river of Strahler order higher than X, up to the confluence with Strahler river >=X.

Map A2.1 Catchments Strahler level 2 (upper 2 watershed)



Source: EEA, CCM processing for ECRINS β.

Map A2.2 Catchments Strahler level 3 (whole Blavet and Scorff watersheds)



Source: EEA, CCM processing for ECRINS β.

This quite condensed definition is illustrated by Maps A2.1 and A2.2 made from the Blavet river watershed, displaying the different Strahler levels.

Map A2.1 above displays level 2 catchments and their ID (D02… indicate catchments from data provision window D, Strahler level 2). It is clearly displayed that where two level 2 rivers (thin dark blue) join, they make a level 3 river (thick light blue) that is contained in a level 2 catchment. When two level 2 rivers join, they make a level 3 river (thick green) that is also contained in catchment Strahler level 2, until it reaches rivers of either level 2 or 3.

This is confirmed by Map A2.2 below that displays the same rivers with catchments at level 3. The upper catchments level 3 ends when any level 3 river joins another level 3, thus making level 4 rivers contained in level 3 catchments too. At the bottom of the map, in dark green, is a short level 5 river, made by the confluence of two level 4 ones. In all cases, catchments IDs are D03, etc., using the same logic as described with level 2 above.

The watershed of the Scorff River reaches only level 3, and forms a single watershed level 3 from spring to mouth.

Annex 3 Seas and endorheic systems: revised codification

Based on the different 'oceans' mentioned in SeaVoX, the following 'systems' result.

The original CCM codification is reported in Table A3.1.

For the reasons explained earlier (see Section 4.3, page 49) the final coding has been created and applied.

Table A3.1 'Seas' (final recipients), as CCM

NAME	System (single letter)	Sea (inside System, as long integer)
Northern Atlantic Ocean	A	1
Norwegian Sea	A	2
Celtic Sea and Channels	A	4
North Sea	A	5
Baltic Sea	A	6
Caspian Sea	C	1
Orumiyeh Salt Lake	H	9
Dead Sea	J	9
Western Mediterranean Sea	M	2
Eastern Mediterranean Sea	M	4
Black Sea	M	5
Barents Sea East	N	7
Barents Sea West	N	8
White Sea	N	9
Prespa Lake	Q	9
Southern Atlantic Ocean	S	1
Trasimeno Lake	T	9
Van Lake	V	9
Tuz Salt Lake	Z	9

Table A3.2 Final proposal for final ECRINS v1.x recipient coding (oceans and seas)

OCEAN	System letter (ECRINS v1.0)
Arctic Ocean	I
Atlantic Ocean	A
Baltic Sea	B
Indian Ocean	E
Mediterranean region	C
Pacific Ocean	#NA: could be G or H
South China and eastern Archipelagic seas	#NA: could be F
Southern ocean	#NA: could be J

Source: SeaVoX data set processing by EEA, letters suggested by the EEA, as selected in the previous document.

Table A3.3 Continent code letter for endorheic systems

Continent name	Letter code
Africa	T
Antarctica	U
Asia	V
Australia	W
Europe	X
North America	Y
South America	Z

SeaVoX considers 8 systems; it is estimated that IHO made proposals for 10 systems. In all circumstances, some letters are left unassigned. Table A3.2 above proposes use of the 10 first letters to secure further recodification.

An endorheic system ('flows in itself') is by definition not a sea but a lake [12], even though it is named a 'sea'. Few systems are truly endorheic, often depending on the input run-off. The distinction for ECRINS v1.0 is quite academic, the important point being not to assess as a potential resource for outlets what is likely endorheic. For ECRINS v1.0, evaporative systems (e.g. the Dead Sea, the Aral Sea (outside the areas concerned, but could be appended in the event of extending ECRINS to east and south neighbours)), quasi-endorheic systems (e.g. the Caspian Sea, whose terminal recipient is Kara Bogaz, where full evaporation of inputs occurs) or groundwater-related endorheic systems (Neusiedler See, Austria) are considered endorheic since they have no surface outlet.

The CCM has reported some endorheic systems, without systematically pointing them out. The CCM endorheic systems are noted as individual System_cd, which is not relevant for comprehensive consideration because not enough letters could be found.

The proposal is that a System_cd code letter is given to each continent (sorted by alphabetic order). The rationale is that continents are the equivalent of the large oceans and hence should be coded at the same hierarchical level. This letter (i.e. T to Z) is the System_CD for following uses. It has no intrinsic meaning other than that continents are sorted by alphabetical order, and assigned a letter starting from the last in the English alphabet to avoid any confusion with true oceanic systems.

Following this approach, the updated 'ECRINS seas' is identified using the following.

1. A starting letter as in Table A3.2 and Table A3.3. In the case of ECRINS v1.0 , only letters A, B, C, E and I are currently used for seas, and T, V or X for endorheic systems.
2. The trailing code:
   (a) from SeaVoX , with follow-up of this coding in case of shore subdividing for practical purposes if the terminal recipient is a sea;
   (b) an arbitrary number in the case of endorheic systems.

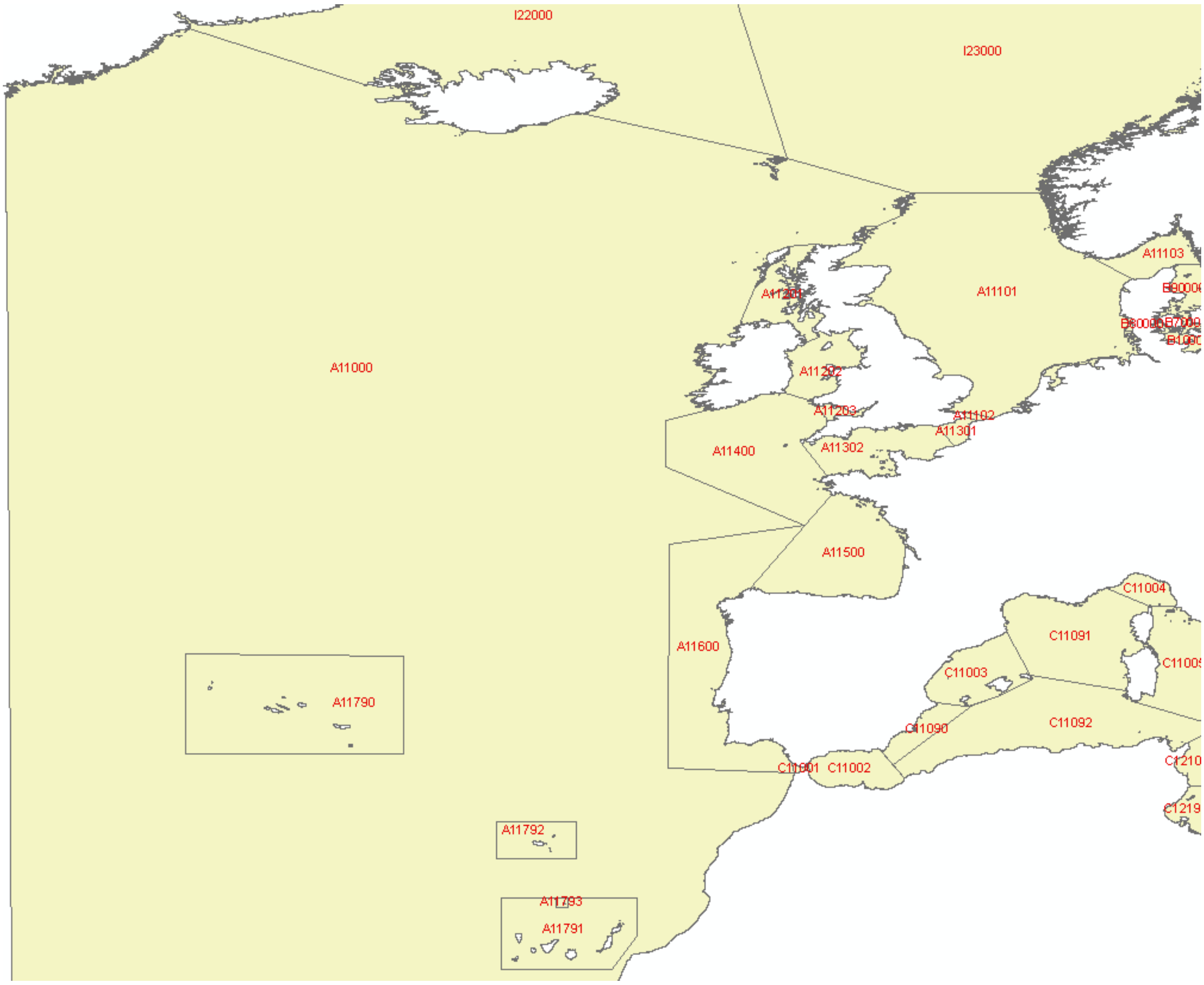
The operational approach is to code the seas (at the level of shores where the rivers pour out) in the following way.

1. Prefix letter as defined above.
2. Trailing code. The numeric form is inspired by the 'levels' that have to be transformed to a five-digit number:
   (a) level 1 is coded 10000 to 90000, or 0 if no level 1 exists;
   (b) level 2 is coded 1000 to 9000, or 0 if no level 2 exists;
   (c) level 3 is coded 100 to 900, or 0 if no level 3 exists
   (d) the last sub-ocean is coded 01 to 99, or 00 if no sub-sea exists.
   (e) the final number is the sum of the different items from level 3 to the sub-ocean item.

The sea identification is to some extent a form of coding since there is an implicit hierarchy in the numbering. Such hierarchy reflects the commonly agreed 'hierarchy' in seas, driven by coding constraints. However, the hierarchy for seas is rather arbitrary. Where controversy might result or if no specific information was provided, the general rule of 'nesting' was applied.

In the endorheic case, a similar coding applies, but there is no implicit hierarchy in the system of coding; the number is purely incremental from 00001 to 99999 per continent.

Map A3.1 Atlantic and North Sea as recoded from SeaVoX, with supplementary delineations



Source: SeaVoX layer, EEA-adjusted

The Atlantic covers a much larger area than that addressed by the EEA. Analysing from the ECRINS and MSFD perspectives, supplementary shores have been defined for the Spanish shore south of the Bay of Biscay, Azores, Madeira and Canarias (hence meeting MSFD requirements) as parts of the main Atlantic. Moreover, to accommodate MSFD organisation, slight adjustments have also been made to the SeaVoX structure. Only areas of EEA/EU relevance are reported in Tables xx, but the coding system can apply to any sea as defined by SeaVoX or the IHO.

The Atlantic, in the SeaVoX context, is a single ocean, secondary apportioned into its north and south, east and west parts. By contrast with the levels proposed in SeaVoX, the sub-oceans between Ireland and Great Britain have been grouped into a new cluster; otherwise, the available numbers for coding would have been insufficient.

Macaronesia, as suggested by the MSFD, has been added; two sub-oceans are created to allow appropriate shore assessment for the relevant archipelagos. Similarly, a shore has been added to include the Iberic peninsula shores.

For example, the Skagerrak can be considered as part of the North Sea *sensu lato* or a higher level sub-ocean of the Atlantic. For practical reasons, the first option was selected. The use of levels 1 to 3 was oriented in the recodification to allow semi-automatic creation of features (level 1 is made by breaking down all those at level 2 having the same level 1 ID, and so on).

By contrast with the SeaVoX polygons, used in Maps A3.2, A3.3, A3.4 and A3.5 below, the codification considers as coded elements all the aggregates at different levels of sea imbrications. For example, the North Sea (A11100) is the aggregate of Skagerrak (A11103), Dover straits (North Sea side) (A11102) and North Sea central (A11101), as displayed in Table A**3.4**.

Shaded features have to be produced *de novo* (e.g Canarias islands shores) or built by dissolving their components (e.g. English Channel and its dependencies)

The Baltic Sea subdivisions have to be slightly adjusted to considerer the eastern, southern and western shores, as depicted in the next figure and

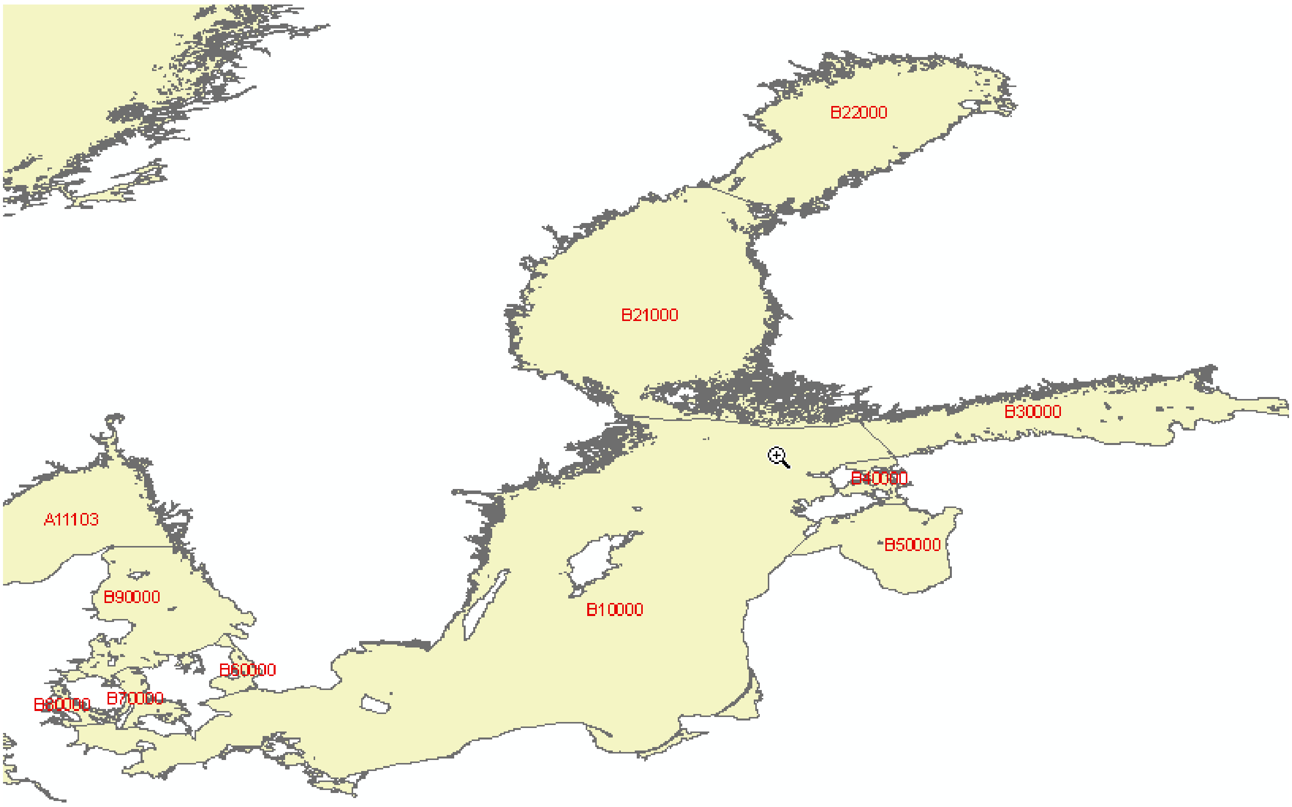
Table A3.4 Recodification of Atlantic Ocean to take into account MSFD delineations from SeaVoX source data sets and ECRINS building requirements

ECRINS code	Name	Level 1	Level 2	Level 3
A11000	North-east Atlantic Ocean	North of the equator	East 40W	n/a
A11001	North Atlantic ocean central	North of the equator	East 40W	n/a
A11100	North Sea	North-east Atlantic Ocean	East 40W	North Sea
A11101	North Sea central	North-east Atlantic Ocean	East 40W	North Sea
A11103	Skagerrak	North-east Atlantic Ocean	East 40W	North Sea
A11102	Dover straits, North Sea side	North-east Atlantic Ocean	East 40W	North Sea
A11200	Inner seas between Ireland and Great Britain	North-east Atlantic Ocean	East 40W	n/a
A11201	Inner seas off the west coast of Scotland	North-east Atlantic Ocean	East 40W	Inner seas between Ireland and GB
A11202	Irish sea	North-east Atlantic Ocean	East 40W	Inner seas between Ireland and Great Britain
A11203	Bristol Channel	North-east Atlantic Ocean	East 40W	Inner seas between Ireland and Great Britain
A11300	English Channel and its dependencies	North-east Atlantic Ocean	East 40W	n/a
A11301	Dover straits, Channel side	North-east Atlantic Ocean	East 40W	English Channel and its dependencies
A11302	English Channel proper	North-east Atlantic Ocean	East 40W	English Channel and its dependencies
A11400	Celtic Sea	North-east Atlantic Ocean	East 40W	n/a
A11500	Bay of Biscay	North-east Atlantic Ocean	East 40W	n/a
A11600	Iberic peninsula shores	North-east Atlantic Ocean	East 40W	n/a
A11700	Macaronesia	North-east Atlantic Ocean	East 40W	n/a
A11790	Azores islands shores	North-east Atlantic Ocean	East 40W	Macaronesia
A11791	Canarias islands shores	North-east Atlantic Ocean	East 40W	Macaronesia
A11900	*Gulf of Guinea (for memory)*	*North-east Atlantic Ocean*	*East 40W*	*n/a*

tables. Moreover, the logic of nesting in the SeaVoX layers is unclear, and hence has been changed slightly to allow nested a ID, without affecting the geometric limits.

In the current version, the central Baltic was not further subdivided since this was not necessary for the production of ECRINS v1.0.

Map A3.2 Baltic Sea recoded from SeaVoX and completed



Source: SeaVoX layer, EEA-adjusted

Table A3.5 Recodification of Baltic Sea to take into account MSFD delineations from SeaVoX source data sets and ECRINS building requirements

ECRINS code	Name	Level 1	Level 2	Level 3
B10000	Baltic central proper	Central Baltic	n/a	n/a
B20000	Gulf of Bothnia	Gulf of Bothnia	n/a	n/a
B21000	Bothnian Sea	Gulf of Bothnia	Bothnian Sea	n/a
B22000	Bothnian Gulf	Gulf of Bothnia	Bothnian Gulf	n/a
B30000	Gulf of Finland	Gulf of Finland	n/a	n/a
B40000	Sound Sea	Sound Sea	n/a	n/a
B50000	Gulf of Riga	Gulf of Riga	n/a	n/a
B60000	The Sound	The Sound	n/a	n/a
B70000	Storebaelt	Storebaelt	n/a	n/a
B80000	Lillebaelt	Lillebaelt	n/a	n/a
B90000	Kattegat	Kattegat	n/a	n/a

The Mediterranean region constitutes a rather complex sea, with many subdivisions that are hydrologically consistent. The applied principles of identification are as follows:

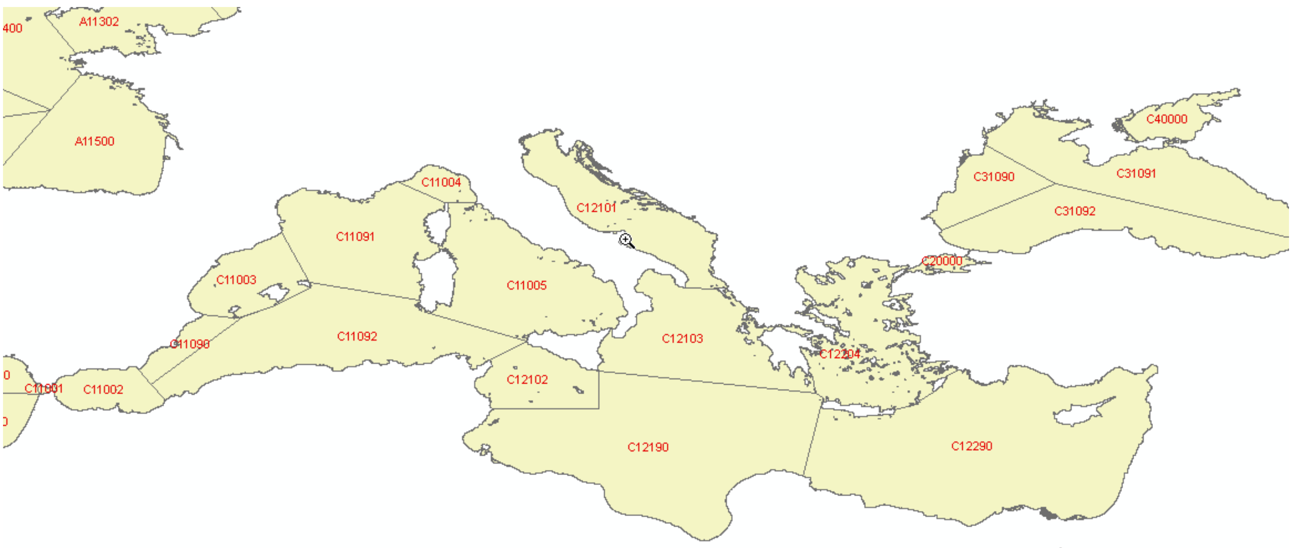
- level 1 embedded respectively the Mediterranean proper (1), the Sea of Marmara (2), the Black Sea (3) and the Sea of Azov (4), following SeaVoX approach;
- level 2 separated the western Mediterranean (11) and the eastern Mediterranean (12);
- level 3 separated the Ionian-Central basin (121) and the Levantine basin (122).

The differences between the SeaVoX seas and the MSFD ones does not lie in the approximate subdivision of seas that can aggregate. The key difference is the proposed delineation of the Levantine basin from the prolongation of the west Peloponnese–western Crete instead of from the junction point between the Ionian–Aegean seas in the SeaVoX proposal. In essence, the result is the same, since SeaVoX delineates the Ionian Sea using the Aegean Sea delineation!

To fulfil computation requirements, other subdivisions have to be inserted: it is not possible to compute coastal catchments in the event that a 'sea' would be in the middle of a shore (as is the Balearic Sea, for example). To create the appropriate shores, the western Mediterranean has been subdivided, as has the Black Sea.

Hence the following recodification is used.

Map A3.3 Mediterranean, recoded from SeaVoX and completed map name



Source: SeaVoX layer, EEA-adjusted.

Table A3.6 Recodification of Mediterranean to take into account MSFD delineations from SeaVoX source data sets and ECRINS building requirements

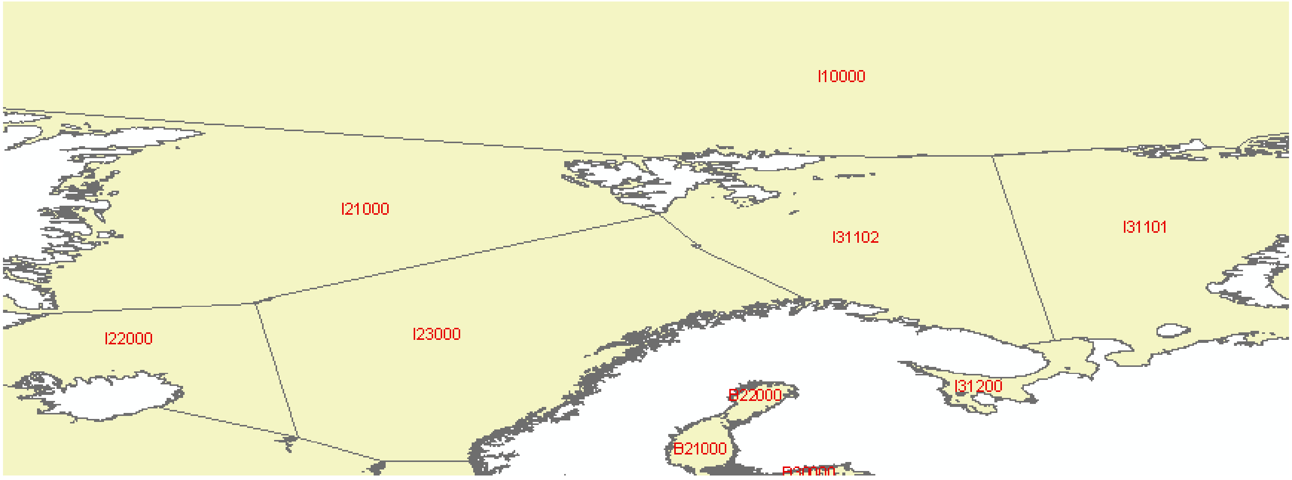
ECRINS code	Name	Level 1	Level 2	Level 3
C11000	Mediterranean western basin	Mediterranean Sea	Western basin	n/a
C11001	Strait of Gibraltar	Mediterranean Sea	Western basin	n/a
C11002	Alboran sea	Mediterranean Sea	Western basin	n/a
C11003	Balearic sea	Mediterranean Sea	Western basin	n/a
C11004	Ligurian sea	Mediterranean Sea	Western basin	n/a
C11005	Tyrrhenian sea	Mediterranean Sea	Western basin	n/a
C11090	Murcia Gulf shore	Mediterranean Sea	Western basin	n/a
C11091	Golfe du Lion shore	Mediterranean Sea	Western basin	n/a
C11092	Maghrebian shore	Mediterranean Sea	Western basin	n/a
C12000	Mediterranean, eastern basin	Mediterranean Sea	Eastern basin	n/a
C12100	Mediterranean, Central-Ionian basin	Mediterranean Sea	Eastern basin	Central-Ionian basin
C12200	Mediterranean, Levantine basin	Mediterranean Sea	Eastern basin	Levantine basin
C12101	Adriatic Sea	Mediterranean Sea	Eastern basin	Central-Ionian basin
C12102	Strait of Sicily	Mediterranean Sea	Eastern basin	Central-Ionian basin
C12103	Ionian Sea	Mediterranean Sea	Eastern basin	Central-Ionian basin
C12190	Tunisian — Libyan shores	Mediterranean Sea	Eastern basin	Central-Ionian basin
C12204	Aegean sea	Mediterranean Sea	Eastern basin	Levantine basin
C12290	Eastern Mediterranean shores	Mediterranean Sea	Eastern basin	Levantine basin
C20000	Marmara Sea	Marmara Sea	n/a	n/a
C30000	Black Sea	Black Sea	n/a	n/a
C40000	Sea of Azov	Sea of Azov	n/a	n/a
C31090	Western shores of the Black Sea	Black Sea	Subdivision of Black Sea proper	n/a
C31091	Russian shores of the Black Sea	Black Sea	Subdivision of Black Sea proper	n/a
C31092	Turkish shores of the Black Sea	Black Sea	Subdivision of Black Sea proper	n/a

Shaded elements in Table A3.6 are created and strictly aggregated into the higher level, hence making any difference with either the SeaVoX or the MSFD proposals.

The Arctic sea has limits at the Atlantic and has portions of outlets from northern-European Russia.

The sub-parts are mainly the Norwegian Sea and the Barents Sea. The Barents Sea, from the ECRINS point of view, poses a problem since its shores are separated into eastern and western parts by the White Sea. Hence it is split into two smaller seas, west and east shores, to circumvent problems in defining the coastal catchments (a catchment might happen to be split into two parts on each side of the White Sea). This resulted in the creation of 'Barents Sea central' to fall in with the hierarchy.

Map A3.4 Arctic Ocean delineations



Source: SeaVoX layer, EEA-adjusted.

Table A3.7 Recodification of the Arctic Ocean to take into account MSFD delineations from SeaVoX source data sets and ECRINS building requirements

ECRINS code	Name	Level 1	Level 2	Level 3
I10000	Arctic Ocean	n/a	n/a	n/a
I20000	Arctic–Atlantic border seas	n/a	n/a	n/a
I21001	Norwegian sea	Arctic–Atlantic border seas	Norwegian sea	n/a
I22001	Icelandic sea	Arctic–Atlantic border seas	Icelandic sea	n/a
I23001	Greenland Sea	Arctic–Atlantic border seas	Greenland Sea	n/a
I31000	North Siberian shore seas	n/a	n/a	n/a
I31000	Barents Sea	North Siberian shore seas	Barents Sea	n/a
I31100	Barents Sea central	North Siberian shore seas	Barents Sea	Barents Sea central
I31101	Barents sea eastern shores	North Siberian shore seas	Barents Sea	Barents Sea central
I31102	Barents Sea western shores	North Siberian shore seas	Barents Sea	Barents Sea central
I31200	White Sea	North Siberian shore seas	Barents Sea	White Sea

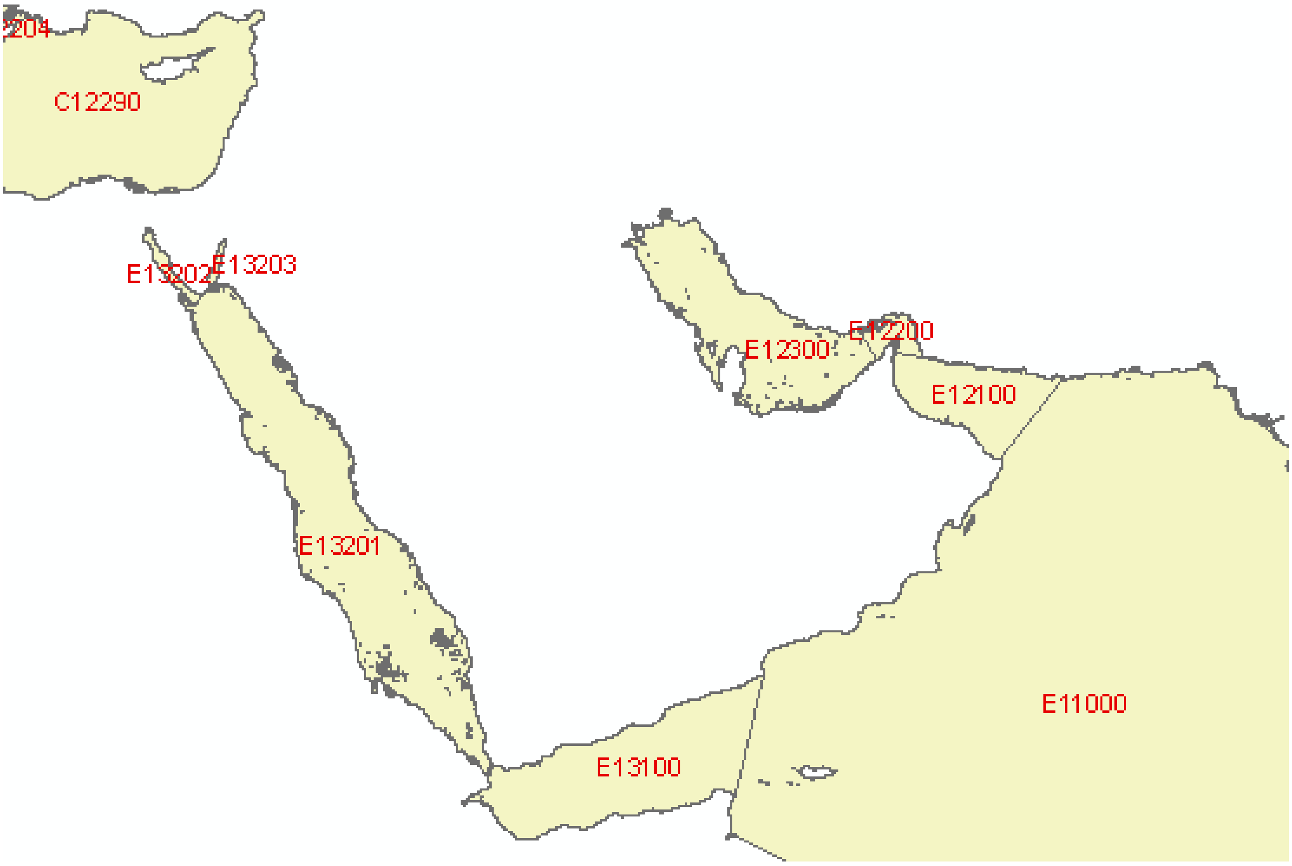
Note: Splitting the Barents Sea into two distinct shores is necessary for coastal FEC computations.

The Indian Ocean is concerned at the margin.

The CCM rivers from Turkey (the Tigris and Euphrates) flow out into the Persian Gulf. For more consistency, the coding based on SeaVoX includes the Red Sea to cover the possibility of addressing Egypt, as a Mediterranean EEA neighbour, in the future.

To follow the codification system, an 'Arabian system' has been set at level 1; the Red Sea and the Persian Gulf have been set at level 2. The recodification is somewhat artificial, but has the advantage of allowing aggregates and permitting further codification fully in line with the delineations proposed by SeaVoX. The different codes are shown in Table A3.8 below.

Map A3.5 Indian Ocean delineations



Source: SeaVoX layer, EEA-adjusted.

Table A3.8 Partial recodification of the Indian Ocean to take into account MSFD delineations from SeaVoX source data sets and ECRINS building requirements

ECRINS code	Name	Level 1	Level 2	Level 3
E10000	Arabic seas system	Clustering the Arabic sea proper plus its adjacent seas	n/a	n/a
E11000	Arabic sea	Clustering the Arabic sea proper plus its adjacent seas	Arabic sea proper	n/a
E12000	Persian Gulf and its annexes	Arabic seas system	Persian Gulf and its annexes	n/a
E12100	Gulf of Oman	Arabic seas system	Persian Gulf and its annexes	Gulf of Oman
E12200	Strait of Hormuz	Arabic seas system	Persian Gulf and its annexes	Gulf of Oman
E12300	Persian Gulf	Arabic seas system	Persian Gulf and its annexes	Gulf of Oman
E13000	Red Sea and its annexes	Arabic seas system	Red Sea and its annexes	n/a
E13100	Gulf of Aden	Arabic seas system	Red Sea and its annexes	Gulf of Aden
E13200	Red sea proper	Arabic seas system	Red Sea and its annexes	Red Sea
E13201	Red sea central	Arabic seas system	Red Sea and its annexes	Red Sea
E13202	Gulf of Suez	Arabic seas system	Red Sea and its annexes	Red Sea proper
E13203	Gulf of Aqaba	Arabic seas system	Red Sea and its annexes	Red Sea proper

The codification proposed involves creating four features from the existing ones:

1. E10000 by breaking down all features E1*;
2. E12000 by breaking down all features E12*;
3. E13000 by breaking down all features E13*;
4. E13200 by breaking down all features E132*.

Coding and delineation of endorheic systems has been carried out using the same rationale.

Table A3.9 Continent letter code for endorheic systems

Continent name	Letter code	Number	Lake name	CCM code	Remarks
Africa	T	n/a	n/a	n/a	n/a
Asia	V	00001	Caspian Sea	C +	371 000 km², altitude 27.6 m. Only the west bank is in CCM. Is Volga river recipient
		00002	Orumiyeh lake (Iran)	H +	5 200 km²; west of Caspian Sea
		00003	Van lake (Turkey)	V +	3 755 km², altitude 1640 m
		00004	Tuz Lake (Turkey)	Z +	1 600 km², altitude 905 m, very shallow
		00005	Dead Sea (Israel/ Jordan)	J +	1 050 km², (shrinking permanently)
		00006	Jordan superficial systems	J +	n/a
Europe	X	00001	Prespa Lake (Balkans: former Yugoslav Republic of Macedonia, Greece, Albania)	Q	Prespa is only partly endorheic: since the Great Prespa Lake sits about 150 m above Lake Ohrid, which lies only about 10 km (6 miles) to the west, its waters run through underground channels in the karst and emerge from springs which feed streams running into Lake Ohrid
	X	00002	Trasimeno (Italy)	T +	128 km², altitude 256 m
		00003	Neusiedler Sea (Austria, Hungary)	NR	Is partly endorheic. Inclusion as terminal system to be considered later; has man-made outlet
		00004	Lake Velence (Hungary)	NR	Is partly endorheic. Inclusion as terminal system to be considered later; has man-made outlet
		00005	Rahasane Turlough (Ireland)	NR	May be too small to be considered
		00006	Larnaca Salt Lake (Cyprus)	NR	Is truly endorheic and a Nature 2000 site, buts potentially has no catchment area
		00007	Ioannina Lake	NR	Quite complex system, owing to drainage and works. It seems that it has an outlet and that this outlet is not the CCM outlet. See FAO (2012) for more information

The proposed sea delineation and coding matches the SeaVoX current delineation of seas, used to qualify metadata related to the identification of samples, as well as what is known from the former IHO approach (excluding adjustments to the Mediterranean and updates of the English Channel), and proposes a more effective method for building ECRINS v1.0 catchments and sea outlets.

Compared to the ICES proposal, the differences are:

- Channel, Irish, Scottish and Celtic seas identified as such;
- Macaronesia, as partition of the Atlantic, identified as two distinct systems (otherwise catchment-making would be jeopardised);
- Mediterranean Sea eastern basin divided into Ionian–Central and Aegean–Levantine, but the

apparently 'quick and dirty' limit proposed by MSFD WG had to be modified to match the available Aegean Sea limit;

- Sea of Marmara stands alone and not part of the Aegean sea;
- Black Sea subdivided to make it possible to place river outlets;
- White Sea separated from Barents Sea; Barents sea split into two shores for ECRINS purposes;
- Faeroes sea not identified as such; could be added if required;
- Iceland coast split between Atlantic and Greenland seas;
- Bay of Biscay and Iberian coast separated (the latter added);

The ICES-proposed coding was not considered because it is not practical for automation; it is replaced by operational code allowing different levels of aggregation in line with SeaVoX 'levels', which were in certain cases modified.

These differences are merely details, especially considering the possible identity between ICES proposal since the reporting on the map is not accurate enough.

The WERC application, as designed with the original System + sea on a compound character + integer (0-9) code had to be maintained, especially since some entities' ID must have terminal recipient information. To this end, the complete sea code has been turned into a terminal ID, assuming that the second position could be changed to a character with minimum adjustment of the WERC programme. Letters plus numbers provide 36 possibilities per system — wide enough for seas, possibly not so for endorheic systems. The equivalence is reported in Table A3.10 below.

Table A3.10 Final equivalence between shores and FEC coding

OCEAN	SUB_OCEAN	EcrinsCD	System_CD	Sea_CD
Atlantic Ocean	Azores Islands shores	A11790	A	C
Atlantic Ocean	Bay of Biscay	A11500	A	A
Atlantic Ocean	Bristol Channel	A11203	A	6
Atlantic Ocean	Canarias Islands shores	A11791	A	D
Atlantic Ocean	Cape Verde Islands shores	A11794	A	G
Atlantic Ocean	Celtic Sea	A11400	A	9
Atlantic Ocean	Dover Strait	A11301	A	7
Atlantic Ocean	Dover Strait	A11102	A	2
Atlantic Ocean	English Channel	A11302	A	8
Atlantic Ocean	Iberic Peninsula shores	A11600	A	B
Atlantic Ocean	Ilhas Selvagens shores	A11793	A	F
Atlantic Ocean	Inner seas off the west coast of Scotland	A11201	A	4
Atlantic Ocean	Irish Sea	A11202	A	5
Atlantic Ocean	Madeira Islands shores	A11792	A	E
Atlantic Ocean	North Sea	A11101	A	1
Atlantic Ocean	North-east Atlantic Ocean (40W)	A11000	A	0
Atlantic Ocean	Skagerrak	A11103	A	3
Baltic Sea	Bay of Bothnia	B22000	B	2
Baltic Sea	Bothnian Sea	B21000	B	1
Baltic Sea	Central Baltic Sea	B10000	B	0
Baltic Sea	Gulf of Finland	B30000	B	3
Baltic Sea	Gulf of Riga	B50000	B	5
Baltic Sea	Kattegat	B90000	B	9
Baltic Sea	Lillebaelt	B80000	B	8
Baltic Sea	Sound Sea	B40000	B	4
Baltic Sea	Storebaelt	B70000	B	7
Baltic Sea	The Sound	B60000	B	6
Mediterranean Region	Adriatic Sea	C12101	C	8

Table A3.10 Final equivalence between shores and FEC coding (cont.)

OCEAN	SUB_OCEAN	EcrinsCD	System_CD	Sea_CD
Mediterranean Region	Aegean Sea	C12204	C	C
Mediterranean Region	Alboran Sea	C11002	C	1
Mediterranean Region	Balearic Sea	C11003	C	2
Mediterranean Region	Eastern Mediterranean shores	C12290	C	D
Mediterranean Region	Golf Du Lion shore	C11091	C	6
Mediterranean Region	Ionian Sea	C12103	C	A
Mediterranean Region	Ligurian Sea	C11004	C	3
Mediterranean Region	Maghrebian shore	C11092	C	7
Mediterranean Region	Murcia Gulf shore	C11090	C	5
Mediterranean Region	Russian shores of the Black Sea	C31091	C	G
Mediterranean Region	Sea of Azov	C40000	C	I
Mediterranean Region	Sea of Marmara	C20000	C	E
Mediterranean Region	Strait of Gibraltar	C11001	C	0
Mediterranean Region	Strait of Sicily	C12102	C	9
Mediterranean Region	Tunisian–Libyan shores	C12190	C	B
Mediterranean Region	Turkish shores of the Black Sea	C31092	C	H
Mediterranean Region	Tyrrhenian Sea	C11005	C	4
Mediterranean Region	Western shores of the Black Sea	C31090	C	F
Indian Ocean	Arabian Sea	E11000	E	1
Indian Ocean	Gulf of Aden	E13100	E	5
Indian Ocean	Gulf of Aqaba	E13203	E	8
Indian Ocean	Gulf of Oman	E12100	E	2
Indian Ocean	Gulf of Suez	E13202	E	7
Indian Ocean	Lakshadweep Sea	E0	E	0
Indian Ocean	Persian Gulf	E12300	E	4
Indian Ocean	Red Sea	E13201	E	6
Indian Ocean	Strait of Hormuz	E12200	E	3
Arctic Ocean	Arctic Ocean	I10000	I	0
Arctic Ocean	Barents Sea eastern shores	I31101	I	4
Arctic Ocean	Barents Sea western shores	I31102	I	5
Arctic Ocean	Greenland Sea	I21000	I	1
Arctic Ocean	Iceland Sea	I22000	I	2
Arctic Ocean	Kara Sea	I33000	I	7
Arctic Ocean	Norwegian Sea	I23000	I	3
Arctic Ocean	White Sea	I31200	I	6
Caspian Sea	Caspian Sea	V00001	V	0
Dead Sea	Dead Sea	V00006	V	4
Orumiyeh Lake	Orumiyeh Lake	V00002	V	1
Sea of Galilee	Sea of Galilee	V00007	V	5
Tuz Lake	Tuz Lake	V00004	V	3
Van Lake	Van Lake	V00003	V	2
Prespa Lake	Prespa Lake	X00001	X	0
Trasimeno Lake	Trasimeno Lake	X00002	X	1

Annex 4 Lakes table structure

Source data sets have each their own structure, as depicted in Table A4.1 and Table A4.2.

Table A4.1 CCM lakes layer structure (as in publication)

Name	Type	Size	Comment
ID	Double	8	Lake-unique ID in the source CCM ID format (numeric)
NAME	Text	100	Lake's name, in original language or blank
LGE_ID	Text	2	Language used for naming the lake (may be blank or not)
WSO_ID	Double	8	Outlet code, as in the source CCM ID format (numeric)
WINDOW	Long Integer	4	Source window of placement (may be inaccurate considering reallocation to processing windows)
AREA	Double	8	Lake area in squared metres
PERIMETER	Long Integer	4	Lake's perimeter in metres
UPSTREAM_AREA	Double	8	Lake's catchment area, as computed from level 1 catchments, not fully documented. Unit is unknown (not indicated in v2.0, not indicated in the v2.1 release)
SYSTEM_CD	Text	1	Sea domain
SEA_CD	Long Integer	4	Sea inside sea domain (see Brochure, Part 1)
COMM_CD	Long Integer	4	Code of starting river for Pfafstetter coding
PFAFSTETTER	Double	8	Pfafstetter coding
UPDATED_BY	Text	15	Name of the updater (name)
UPDATED_WHEN	Date/Time	8	Date of update
LKE_TYPE	Text	1	(*) Code of the lake type. The coding is unclear: In the CCM brochure, the lakes types (field LKE-TYPE) have the following codes: N(atural), P(it), D(ammed lake), R(eservoir); O(xbow) (???), L(agoon), U(nknown) By contrast, in the V2.1 lakes data set, the following codes are found: B, N, P, R, Y
ALTITUDE	Long Integer	4	Centroid of the polygon altitude, from the DEM
SHAPE	OLE Object	n/a	Binary of the object shape
OBJECTID	Long Integer	4	Automatic ID
SHAPE_Length	Double	8	Automatic perimeter
SHAPE_Area	Double	8	Automatic area

(*) Following a discussion with A. de Jager from the JRC, transpired that this is dummy information.

The final data sets have a specific ECRINS structure reported in the next tables.

Table A4.2 Lakes feature class structure C_lak

Name	Type	Size	Comments
Alt_Score	Integer	2	Quality of altitude value (0: , 1: ; 2:)
Altitude	Long Integer	4	Altitude of lake at standard volume in metres above sea level, or blank
Area	Double	8	Standard lake area in km^2, as from shape_area
Area0km2	Double	8	Minimum lake area (or filled with Area)
AreaXkm^2	Double	8	Maximum area lake, or blank
SourceXarea	Integer		Is one of the values of the Is_, or a sum if multiple (no double counting)
comm_cd	Long Integer	4	CCM source data
EcrinsVersion	Long Integer	4	ECRINS version (tells if lake relationships are
ERMID	Text	10	ERM
is2Create	Integer	2	Management flag
is2Keep	Integer	2	Management flag
Is2Substitute	Integer	2	Management flag
IsCCM	Boolean	2	Indicates if lake is from CCM data source (management) (0/1)
Is_art13	Boolean	n/a	Indicates is lake is/was from Art.13 (0/2)
Is_ERM	Boolean	n/a	Indicates if lake is/was ERM present (0/4)
Is_511	Boolean	n/a	Indicates if lake is/was clc511 (0/8)
Is_512	Boolean	n/a	Indicates if lake is/was clc512 (0/16)
Is_521	Boolean	n/a	Indicates if lake is/was clc521 (0/32)
Is_411	Boolean	n/a	Indicates if lake is/was clc411 (0/64)
Is_Wiki	Boolean	n/a	Indicates if lake is recorded in Wikipedia (and was hence updated from) (0/128)
Is_Other	Boolean	n/a	Indicates if lake is recorded in other sources (and was hence updated from) (0/256)
LakSScore	Integer	n/a	Scores the source adding the sources (1–511, since at least one category!)
lakID	Text	10	Main identifier of the lake. Follows ECRINS syntax (see below)
lakPrimID	Text	10	Identifier of lake that is primary to this lake (or itself)
LakVersion	Long Integer	4	Lake data set update number
langerm	Text	7	Code of language for name
lke_type	Text	1	Lake type code (not filled yet)
NAMA2	Text	50	Lake's secondary name, as from either CCM of ERM or other source
name	Text	100	Lake's primary name
Nb_Ctry	Integer	2	Number of countries involved in the lake (1: national lake)
OBJECTID	Long Integer	4	n/a
perimeter	Long Integer	4	Standard lake perimeter, as from shape_length
PerimeterX	Long integer	n/a	Perimeter corresponding to AreaXkm2
sea_cd	Long Integer	4	CCM source data
Shape	OLE Object	n/a	n/a
Shape_Area	Double	8	ArcGIS® information

Name	Type	Size	Comments
Shape_Length	Double	8	ArcGIS® information
system_cd	Text	1	CCM source data
Upstream_ area	Double	8	Upstream area (= area of inlet river) in km^2, or blank
volavghm3	Double	8	Average volume in hm3 (= million m3), or blank
volmihm3	Double	8	Minimum volume in hm3 (= million m3), or blank
volmxhm3	Double	8	Maximum volume in hm3 (= million m3), or blank
window	Long Integer	4	CCM window (management)
Wnb	Integer	2	n/a
Zavgm	Double	8	Average depth, in metres
Zmim	Double	8	Minimum depth, in metres
ZMxm	Double	8	Maximum depth, in metres

Note: This structure is common to other tables used in the process before final copying into the final results database.

Shaded cells represent fields added to take into account the CLC lakes source.

It should be noted that there is a change in meaning of the field ERMID (ECRINS reference Management ID) source ID of the polygon. When CLC is the origin syntax is clcID(_rank)_Yyyy, where yyy are three last digits of the reference CLC year _ _rank is optional and relates only to pseudo-substitution.

For example, EU-123456_001_Y006 indicates that the polygon is a pseudo-substitute, with rank 1 coming from CLC 2006 item EU-123456.

Annex 5 Important working data sets in building ECRINS

The FEC construction (Section 5.1, page 54) is dependent on statistics of catchments of Strahler levels X within Strahler levels X+1.

Several tables are produced at different steps of the processing.

The table structure of the statistics for Strahler 2 and 3 is shown in Figure A5.1 below.

The computation of statistics is time-consuming (~ 1 hour) and needs updating only if a new export is performed. The application checks consistency between each of the 18 windows to be scrutinised and the current export version. If the first window has a different export version, it can be changed at this point. Any other differences are flagged as errors and cancel the process.

Figure A5.1 Structure of aggregated catchment statistical table to prepare FEC building

StatFEC2_3vs4

Field Name	Data Type	Description
ID4	Text	Level 4 watershed ID
level	Number	2 or 3, waterhsedID belonging to ID4 for the level
NbItems	Number	Nb of watersheds inside the ID4
Ar_mean	Number	average area of the items
ArQ10	Number	10% of the items have their area less or equal to
arQ50	Number	50% of the items have their area less or equal to
ArQ75	Number	75 of the items have their area less or equal to
ArQ90	Number	90% of the items have their area less or equal to
ArMX	Number	maximum are of the items (is 100% quantile)
Is_ForBas0	Yes/No	the item is to be inserted. Exclusive between levels 2 and 3 for same ID4
ID4Down	Text	ID of level 4 upstream
ID4D	Text	ID of level 4 downstream
ID1Down	Text	ID of level 1 downstream: linker
ID1UofDown	Text	ID of level 1 upstream: linker, this is the last level 1 inside the group that connects to ID1Down
ID2U	Text	ID of level 2 most downstream inside the level 4 and connecting downstream
ID2D	Text	ID of level 2 to which the curent level 4 connects to
ID3U	Text	ID of level 3 most downstream inside the level 4 and connecting downstream
ID3D	Text	ID of level 3 to which the curent level 4 connects toto which the curent level 4 connects to
exportversion	Number	

European Environment Agency

EEA Catchments and Rivers Network System ECRINS v1.1
Rationales, building and improving for widening uses to water Accounts and WISE applications

2012 — 111 pp. — 21 x 29.7 cm

ISBN 978-92-9213-320-7
doi:10.2800/51667

TH-AK-12-007-EN-N
doi:10.2800/51667

European Environment Agency
Kongens Nytorv 6
1050 Copenhagen K
Denmark

Tel.: +45 33 36 71 00
Fax: +45 33 36 71 99

Web: eea.europa.eu
Enquiries: eea.europa.eu/enquiries

ISBN 978-92-9213-320-7

9 789292 133207

Publications Office

European Environment Agency