Management of Environmental Information in the European Information and Observation Network (EIONET)

Thomas Pick, European Topic Centre on Catalogue of Data Sources (ETC/CDS), Archivstraße 2, D-30169 Hannover, Germany, thom.pick@numis.niedersachsen.de
Hannu Saarenmaa, European Environment Agency (EEA), Kongens Nytorv 6, DK-1050 Copenhagen, Denmark, hannu.saarenmaa@eea.eu.int
Kirsti Lounamaa, TietoEnator, Zaventem, Belgium, kirsti.lounamaa@tietoenator.com

Abstract

The reporting system of the European Environment Agency (EEA) depends on an integrated data flow, driven by the need for indicator data. EEA aims at streamlining this effort by establishing the ReportNet structure. It is the vision of a shared European Environmental Information Infrastructure, used by many networks, furthering standardised data interchange formats and communication protocols.
The paper discusses the EEA data flow and reporting system and presents ReportNet as a strategy to integrate both. ReportNet, the EEA's electronic reporting network, is based on an integrated metadata management system, using DCES as the backbone for content description. It consist of 7 building blocks. The paper focusses on the data dictionary and the registry as the two central metadata management units, with the content being listed in a registry and the data and management elements registered in a dictionary. By automatically generating individual namespaces for each element on the fly and allowing for free combination of these to composites, the system guarantees maximum flexibility, demanded by the heterogeneity of the environmental domain.
It is the diversity of this domain that inhibits the generation of a generic environmental namespace as a general tool for management of environmental information resources and data.
The technical architecture is based on SOAP and XML with individual services being implement in JAVA.

Keywords: *Environmental Information, EIONET, Environmental Reporting*

1 Introduction:

The European Environment Agency (EEA) is a European Community institution with the aim of serving the Community and the Member States with information to support policy making for environmental protection put in the perspective of sustainable development. This is done by collecting and assessing data on the current state of the environment in Europe, with the support of the European Information and Observation Network (EIONET).
Through the provision of timely, targeted, relevant and reliable information to policy making agents and the public, EEA aims at achieving significant and measurable improvement in Europe's environment.
As information is being generated at an exponentially growing rate, while discovery of that information becomes increasingly difficult, EEA faces the challenge of organising the information flow within EIONET.
The necessity for combining the management of information resources with organisational aspects of data flow, lead to the idea of an integrated system for managing information for reporting, the ReportNet [1].

2 Data Flow in EEA's reporting system

The EEA reporting system is driven by the need for

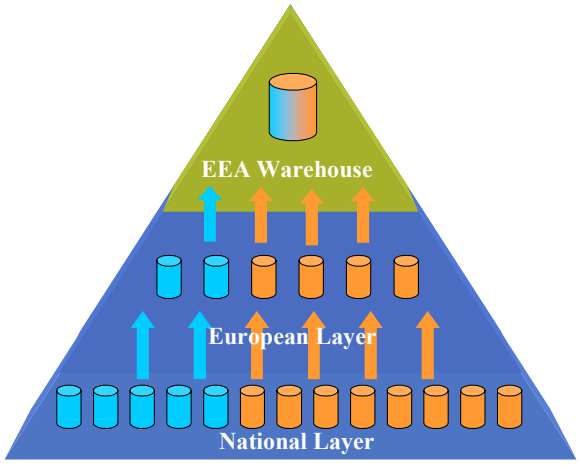

Figure 1: EEA/EIONET data flow

https://doi.org/10.23106/dcmi.952106480

environmental indicator data. The data is being reported from the member states to the agency. The flow of data is strictly hierarchical, bottom up from the national level to the agency (Fig. 1). Currently there is no standardised way of how these data are being communicated. The spectrum ranges from standard Office formats like MS-Excel to individually developed Data Exchange Modules (DEM) that support XML format [2].

The data are being aggregated on their way up and are used as the basis for assessing the state of the environment in Europe. The final product is usually an environmental report produced by the EEA or one of its topic centres. These products are made available through the European Environmental Reference Centre (E2RC), EEA's Web service (http://eea.eu.int). The E2RC is a public information service, recognised throughout Europe as the obvious gateway to easily understandable and efficiently structured environmental information. It provides seamless access to a wide variety of distributed environmental information, in particular information developed through EIONET.

An integrated part of the E2RC is the Directory of EEA/EIONET Information Resources (DIR), a metadata catalogue containing the description of information resources supplied to, maintained in and emanating from EEA/EIONET. The Directory is based on the GELOS (Global Environment Information Locator Service) metadata element set [3] and maintained by the ETC/CDS. All entries are indexed with keywords taken from the general multilingual environmental thesaurus (GEMET), in order to facilitate access across language barriers.

However, this monolithic portrait of EIONET data flow gives a somewhat idealistic impression of the current situation. The EIONetwork is connected to other organisational bodies such as EUROSTAT (European Statistics Office), OECD (Organisation for Economic Co-Operation and Development) and UNEP (United Nations Environmental Program) in a network of networks. This often leads to the duplication of reporting (Fig. 2).

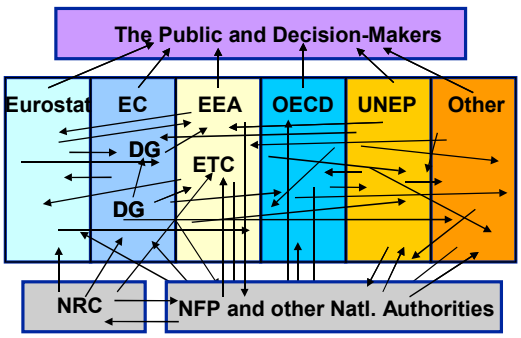

Figure 2: Current Situation With Overlapping Data Flow

EEA aims at streamlining this data flow by establishing the ReportNet structure (Fig. 3).

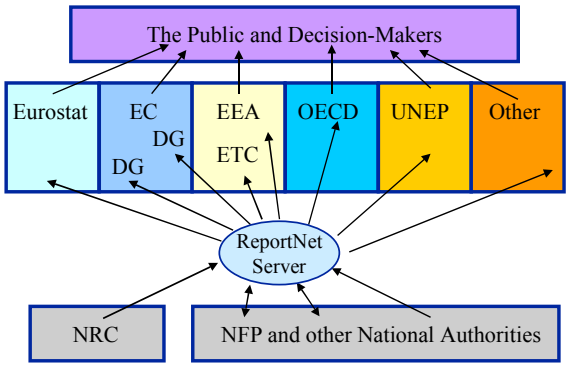

Figure 3: Target Situation With Streamlined Data Flow

3 ReportNet

The ReportNet structure combines the different EIONET activities concerned with reporting into a single integrated system[1].

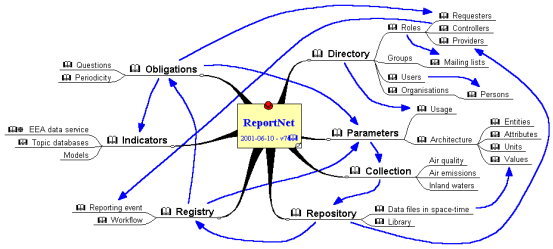

Figure 4: ReportNet Building Blocks

The system incorporates the following modules most of which have been realised in the past as isolated applications (Fig. 4):

- **Parameters** - A data dictionary describing Environmental namespace in XML Schema language
- **Directory** - The official EIONET Directory of contact information
- **Obligations** - The sources of reporting obligations like Legislation, conventions, etc.
- **Registry** - An inventory of the widely distributed repository files.
- **Repository** – The (distributed) pool of information resources.
- **Indicators** - To be calculated for fulfilment of the obligation on the basis of the content in the reported files and models.
- **Collection** - The (distributed) collection of measurement data, partly aggregated.

https://doi.org/10.23106/dcmi.952106480

4 The ReportNet Metadata Concept

Obviously, every ReportNet building block touches on meta data issues. However, in the context of this paper, we would like to focus on the discussion of the Directory, Repository, Registry and Parameter Services.

4.1 The Repository

Hundreds of files are annually created in each country in the reporting process. They contain the answers to the 25 000 or so questions defined in obligations, in the form of values for the parameters needed for the assessment of the state of the environment in various file formats.

All EIONET documents live in CIRCLE (Centre of Information Resources for Collaboration on Environment), a simple document sharing system for Internet. As documents in CIRCLE are organised into Interest Groups (IG) for access by groups of users, CIRCLE can also be called group collaboration software.

Document management on CIRCLE is enhanced by additional services, such as directory of users, meetings, discussions, email, and search.

CIRCLE is based on the European Commission's CIRCA (Communication and Information Resource Centre Administrator) software, but is tailored for the use of EIONET [4].

4.2 The Directory

The Directory is the place where the official EIONET contact information is stored, containing details about actors (or agents in DCMI lingo), their function within the network and their address. It is being built on CIRCA's LDAP system. The decision on what meta data elements to use for best describing contact information is still open, but the system is flexible enough to accommodate changes. Dublin Core does not offer a generic element set here, but refers to other system like e.g. vCard, bridging nicely into the MS-Office world.

Aside from the mere contact information, the actors function, relating to the network's data flow is of greatest importance. ReportNet will connect to the CIRCA system using SOAP [5].

4.3 The Data Dictionary (Parameters)

The data dictionary is designed as a meta data registry for the definition of elementary data objects. These are in general environmental parameters like e.g. ozone, mercury, population density of a certain species. They are used as building blocks for constructing individual XML schema. The XML schema language shall be used to validate the allowed values and value domains of data elements in XML instance files i.e. in data sets to be registered into Repositories. The standard specifies the basic attributes of data elements, in other words the metadata set for describing data elements.

The conceptual model of the Data Dictionary is based on ebXML [15].

The dictionary discriminates two object types: a) the data element as the basic object, independent of any specific context (e.g. longitude, latitude, population) and b) aggregates as a combination of attributes, who's compound characteristics define a specific concept (measuring station, environmental indicator).

Each element is defined as its individual namespace in XML Schema [6]. The dictionary allows for combining these 'on-the-fly' to composites that fit the respective environmental domain (e.g. air, water, waste) or informational target situation (e.g. report, data collection). While guaranteeing maximum flexibility this requires high maintenance efforts on the dictionary side.

```
<xsd:element name="{name of element_with_values_measured_in_units}" type="{unit}">
    <xsd:annotation>
        <xsd:documentation>{Identifying attributes 1.3-1.6}</xsd:documentation>
        <xsd:documentation>{Definitional attribute 2.1}</xsd:documentation>
        <xsd:documentation>{Relational attributes 3.1-3.4}</xsd:documentation>
        <xsd:documentation>
            <xhtml:tr>
                <xhtml:th>Form of representation</xhtml:th>
                <xhtml:th>Quantitative value</xhtml:th>
            </xhtml:tr>
            <xhtml:tr>
                <xhtml:th>Unit</xhtml:th>
                <xhtml:th>{unit}</xhtml:th>
            </xhtml:tr>
        </xsd:documentation>
    </xsd:annotation>
```

Figure 5: XML Representation for a Data element with values measured in units

The example in figure 5 demonstrates how some of the basic element attributes specified in ISO/IEC 11179 can be presented using XML schema language. The example is a fragment of XML schema that could be used to validate an element "longitude" whose value must be a decimal number and allowable value range between –180 - +180 degrees.

https://doi.org/10.23106/dcmi.952106480

4.4 The Registry

This is the central database of the ReportNet content. It is the inventory of the widely distributed repository files and tracks what has been reported to whom and when, on what legal basis, etc..

Each dataset in the Repository is associated with two metadata files: content description and content schema. The content description contains the metadata about the whole dataset, while content schema describes each individual data item in the dataset and the hierarchical structure of the data.

The content descriptions will follow Dublin Core standard (+ EIONET extensions) and will be presented in some XML syntax: RDF or XML Schema.

The content schema will follow ISO/IEC 11179 standard and will be presented in XML Schema.

The Content Registry generates an index and a catalogue of descriptions of the repository content through harvesting the repository and fetching content description (in RDF). The Content registry categorises the content of the RDF files using an existing taxonomy. This information is stored in the Content Registry catalogue.

In addition the registry provides content management services: it helps users to fill in content description metadata, while automating this process as much as possible. Furthermore it provides search and browse functionality.

4.5 Architecture

The system architecture is the same for both the Data dictionary and the Content registry (differences only occur in the database and the user interface).

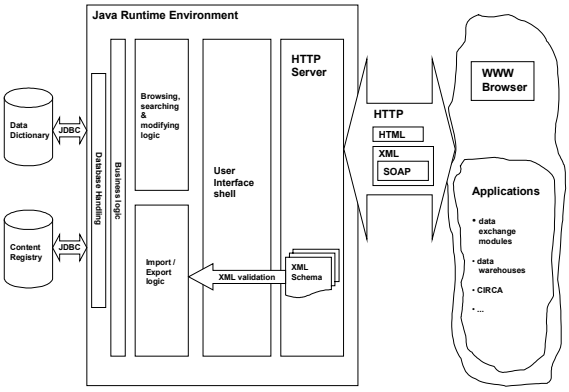

Figure 6: System Architecture

It consists of three main tiers (Fig. 6):

- Presentation layer for providing the user interface
- Application layer (business logic) for implementing the META functionality
- Data layer for the META data

Logically, the application software can be divided into 5 parts:

- *Database handling logic* - responsible for performing the basic database operations, like SELECT, UPDATE, INSERT and DELETE operations.
- *Business logic* - responsible for suiting and feeding the imported/updated data for the Database handling logic
- *Browsing, searching and modifying logic* - responsible for mediating the user – or application input data between the user client or application and the Business logic.
- *Import/Export logic* - responsible for parsing and extracting imported data from XML format and mediating it to the business logic. For the case of export, this logic generates XML format output for the user/application requesting the data.
- *User interface shell* - responsible for performing, keeping and maintaining the communication between the browsing-, searching-, modifying-, importing -and exporting logic and clients interfaced with HTTP-based formats.

On the lowest level, *HTTP Server* handles communication means between the system and it's clients whether they are humans with a WWW browser or applications providing HTTP based communication formats, like HTML, XML and SOAP, based on the latter. As the HTTP Server will also have to provide Java servlets for the clients, it must have compatibility with a Java Servlet Running Engine supporting JSDK 2.2 (Java Servlet Development Kit) and it is going to run in Java Runtime Environment.

XML Schemas, representing the data definitions in the dictionary, will also be provided by the HTTP Server via URL interface. XML Schemas are also used for validating an XML document when such is to be imported or registered in the Content registry. The block arrow with a title of XML Validation serves for that.

The clients are generally divided into 2 groups:

- Humans as users – interfaced with the system by WWW browsers
- Applications as users - interfaced with the system by HTTP based formats, like HTML, XML and inter-application communication

https://doi.org/10.23106/dcmi.952106480

format SOAP (Simple Object Access Protocol).

5 Dublin Core and Environmental Information Resources

As with other domains, the environmental information resources become increasingly available through the internet. In fact, most EIONET resources are circulated through e-EIONET, the telematic network of the organisation, and/or are being made available to the public through the E2RC and EEA's website. The information resources are of several types with type document and type data accounting for more than 90%. Currently the document type resources are described with the GELOS metadata element set. This element set is compliant with DCES [7].

DCES arrived at a point that it is robust enough to use it for EIONET metadata needs and EIONET is adopting DCES for description and discovery of its information resources.

DCES is geared towards discovery of resources on the internet, which makes it less appropriate for the description of data. However, it includes enough flexibility to extend it for these resources and those that are (not yet) online. Applications can build around Dublin Core Statements [8] which follow the subject, predicate, object grammar of sentences. This has a 'binary flavour' and can be readily implemented by using RDF schema language [9]. DCMI has generated two namespaces for general use, DCES [10] and the Dublin Core Qualifiers [11].

For the description of data then (this is a central part of the ReportNet initiative) each sub-domain (like Air Quality) may define their respective namespace and combine these with the generic DC namespaces through an RDF schema.

The Swedish EnviroNet System was the first system to use DCES for the description and discovery of environmental information resources on the internet [12]. They have walked the well travelled path of extending DCES and DCQ with the 'missing' elements and qualifiers, while fitting them with their domain specific themes. With ReportNet the EEA chose a different approach. By defining an individual namespace for each data element, maximum flexibility is possible to cover the unnumbered sub-domains that together make up the environmental domain.

In this light, the efforts of the XML-EML.ORG forum [13] for constructing an Environmental Mark-up Language should be taken with a grain of salt. In our view, the diversity of the environmental domain inhibits the generation of a generic environmental namespace or even an environmental mark-up language.

Conclusions:

With the exponential growth of information on the state of the environment in Europe and world-wide, the European Environment Agency faces the challenge to integrate its reporting system. EEA meets this challenge with the ReportNet structure, a mechanism for bridging the gap between best needed and best available information.

The ReportNet building blocks rely on an integrated metadata management system for information discovery and data flow control. EEA relies on DCES for discovery. In addition it develops the data dictionary for data flow management. It is a registry of namespaces for parameter elements and management events. By assigning each element its own namespace it guarantees maximum flexibility, demanded by the heterogeneity of the environmental domain.

It is the diversity of this domain that inhibits the generation of a generic environmental namespace as a general tool for management of environmental information resources and data.

References:

[1] H. Saarenmaa – ReportNet, 2000. http://eea.eionet.eu.int:8980/irc/eionet-circle/eionet-telematics/info/data/en/reportnet05/.

[2] B.A. Bannink - Air Quality Annual topic update 1999, http://reports.eea.eu.int/Topic_report_72000/en/toprep07_2000.pdf, 2000.

[3] G8 - Environment and Natural Resources Management project. The Global Environment Information Locator Service Metadata Element Set,1999. http://www.mu.niedersachsen.de/cds/etc-cds_neu/download-general.html

[4] H. Saarenmaa - The CIRCLE document sharing system, 2001, http://eea.eionet.eu.int/EIONET_Tech/CIRCA

[5] SOAP/1.1 submission, and XML protocol requirements working draft. XML Protocol Activity, W3C Architecture Domain, http//www.w3.org/2000/xp.,

https://doi.org/10.23106/dcmi.952106480

[6] W3C – XML Schema Specification, http://www.w3.org/XML/Schema, 2000

[7] T. Pick – Mapping the Gelos Element Set to the Dublin Core Element Set, http://www.mu.niedersachsen.de/cds/etc-cds_neu/gelos_dc.html, 2000.

[8] T. Baker – A Grammar to Dublin Core, http://www.gmd.de/People/Thomas.Baker/DC-Grammar.html, 2000.

[9] W3C – Resource Description Framework, http://www.w3.org/RDF/, 2000.

[10] Dublin Core Metadata Initiative – Dublin Core Metadata Element Set, Version 1.1: Reference Description, http://dublincore.org/documents/dces, 1999.

[11] Dublin Core Metadata Initiative – Dublin Core Qualifiers, http://dublincore.org/documents/dcmes-qualifiers/, 2000.

[12] S. Hammarsten, A. Moshiri, M. Sjö – The Swedish EnviroNet, http://smn.environ.se/miljonat/english/om-smn/om-smn.htm, 2000.

[13] H-K. Arndt, O. Günther (eds) – Environmental Markup Language (EML). Proceedings of Workshop 1, Berlin, 1999.

[14] International Organisation for Standardisation, http://www.iso.ch/

[15] ebXML – Initiative for Electronic Business Framework, http://www.ebxml.org/

https://doi.org/10.23106/dcmi.952106480