

*Prestudy*

# **Reportnet Live**

— general principles

**Prepared by**  
Søren Roug

**September 2008**  
**Version 0.1**

European Environment Agency 

# Version management

No	Date	Changes	Author
0.1		Initial version	Søren Roug

# Contents

<b>1. SCOPE/INTRODUCTION.....</b>	<b>4</b>
<b>2. CORE PREMISES.....</b>	<b>4</b>
<b>3. OPERATIONAL STEPS.....</b>	<b>4</b>
3.1. SERVICE AGREEMENT.....	4
3.2. RECEIVING THE DATA AT EEA.....	4
3.3. ACCUMULATION DATABASE.....	5
3.4. PERIODIC DATASET GENERATION.....	5
<b>4. POTENTIAL TIE-INS WITH REPORTNET.....</b>	<b>6</b>

# 1. Scope/introduction

During the discussions on implementing the WISE dataflows, we slowly distilled the principles behind our one object reporting dataflow; Ozone. This was surprisingly difficult to do, because we tended to slip into describing how the Ozone dataflow worked and thereby giving the impression that all other live dataflows must behave the same way.

We realised that we need a term to distinguish between the general principles and the Ozone dataflow. “Reportnet Live” was the term we landed on.

## 2. Core premises

Reportnet Live has a number of built-in assumptions of what is sent through it.

- Data reporting has a periodical cycle of minutes, hours or a few days.
- Dataflows are continually updated and have no real deadlines. This contrasts a hypothetical “*AlarmNet*,” where messages go through very seldom.
- Data object reporting is only set up to handle small partial deliveries. Those partial data are accumulated in a database on the receiver side who decides how the partial data is filtered and aggregated.
- The legal or voluntary instruments define a service agreement between provider and receiver.
- Both ends of the infrastructure are constantly monitored and maintained by system administrators who are in contact with each other.
- Deliveries can come in different protocols and formats.
- Deliveries can be more than required and at a higher frequency than required.
- Deliveries will always be based on a web-service approach because it is based on a system to system.

## 3. Operational steps

### 3.1. Service agreement

When establishing a new live dataflow, it is decided in advance what needs to be collected and how often. Then this framework is presented to each member state for discussions on how to achieve it.

In the end a contract is set up where agreements are made on how the transfer is done (SOA such as SOAP or REST, FTP, automated email,...), data format (OGC GML, homemade, country standard), responsibilities on frequency, error follow-up, maximum down time and for what the data can be used for at EU level.

### 3.2. Receiving the data at EEA

No matter what is sent, a set of data items are required for all objects. Some of them can be determined implicitly – e.g. through the service agreement, others must somehow be specified explicitly in the delivery.

- Flow identifier – to decide on the receiving end how to process it

- File format – There can be several file formats for the same dataflow, but one country usually uses just one format.
- Timestamp or coverage period – deliveries can be delayed or they can be updates to previously sent deliveries. The timestamp is used to figure out which is which. It can also be used to check the system. If the service agreement says there will be a delivery once an hour, it could be used to detect a missing delivery.
- Sender – e.g. station identifier

Then there is the payload. The sending entity can send more than actually needed, for instance if the station collects more than the mandatory measurements. The system operates on the principles of “receiver makes right”.

In practice all deliveries are converted into XML if they aren’t already, then an XSL stylesheet is used to harmonise all the XML formats into a single XML format containing only the data needed. This is called orchestration.

All the original deliveries are kept – in the order they arrived, so that if something is wrong with the orchestration, the process can be rerun.

The next step is to do a quality check on the data. This is also done with XSL and is called *filtering*. If the delivery fails the QA, then a rejection notification is sent to the administrators of both the sending and receiving end.

Notifications are also sent if a sender hasn’t sent anything in a pre-specified period.

### **3.3. Accumulation database**

The deliveries are then inserted into a database where they are accumulated. The database has query mechanisms that allow online viewers to retrieve the latest updates. It will normally be implemented as an update into an application specific database that is organised more optimally towards the application’s needs.

Note, that this can only be a snapshot of what is known at the moment. The countries can typically resend data – even data born months ago.

### **3.4. Periodic dataset generation**

Some live dataflows have a parallel Reportnet dataflow. The prime example is the Ozone dataflow, where the countries are also required to produce monthly and summer statistics on events. These statistics can in most cases (depending on what is reported) be generated from the live dataflow. In ozone the system would generate the monthly statistic one month after the last day of the month in question, to allow the deliveries to stabilise. Then there are two options:

1. To upload the generated dataset to CDR, and then let it go through the normal workflow. It would preload the envelope with the data, and then ask the country to release the envelope. Doing so would imply accepting the generated dataset matches what the country wants to deliver

2. Make a repository that works the same way as CDR (following the principles for Reportnet/SEIS nodes), but with a workflow tailored to generated data. I.e. with a simpler acceptance procedure for the country.

In both options, the datasets are then picked up by Reportnet and progressed through the Reportnet pipeline, which amongst other things means that it is possible to see the deliveries on ROD.

## **4. Potential tie-ins with Reportnet**

Reportnet has some services that could be used by Reportnet Live. First of all, Reportnet has a notification system that keeps track of subscriptions. This could be used in the orchestration step and the periodic dataset generation to simplify the sending of notifications.

Content Registry could be used to track more closely what is available in the accumulation database. It is somewhat ironic that we ask the member countries to make their databases available for SEIS, but we don't do it for our own databases.

There could be better mechanisms in CDR to receive the periodically generated datasets.