Informing you on ambient air quality
in the Belgian Regions

Air quality data and e-reporting:

using the HUMBOLDT Alignment Editor (HALE) for data transformations in a legacy context

Olav Peeters
Belgian Interregional Environment Agency (IRCEL – CELINE)
peeters@irceline.be

12/03/2014

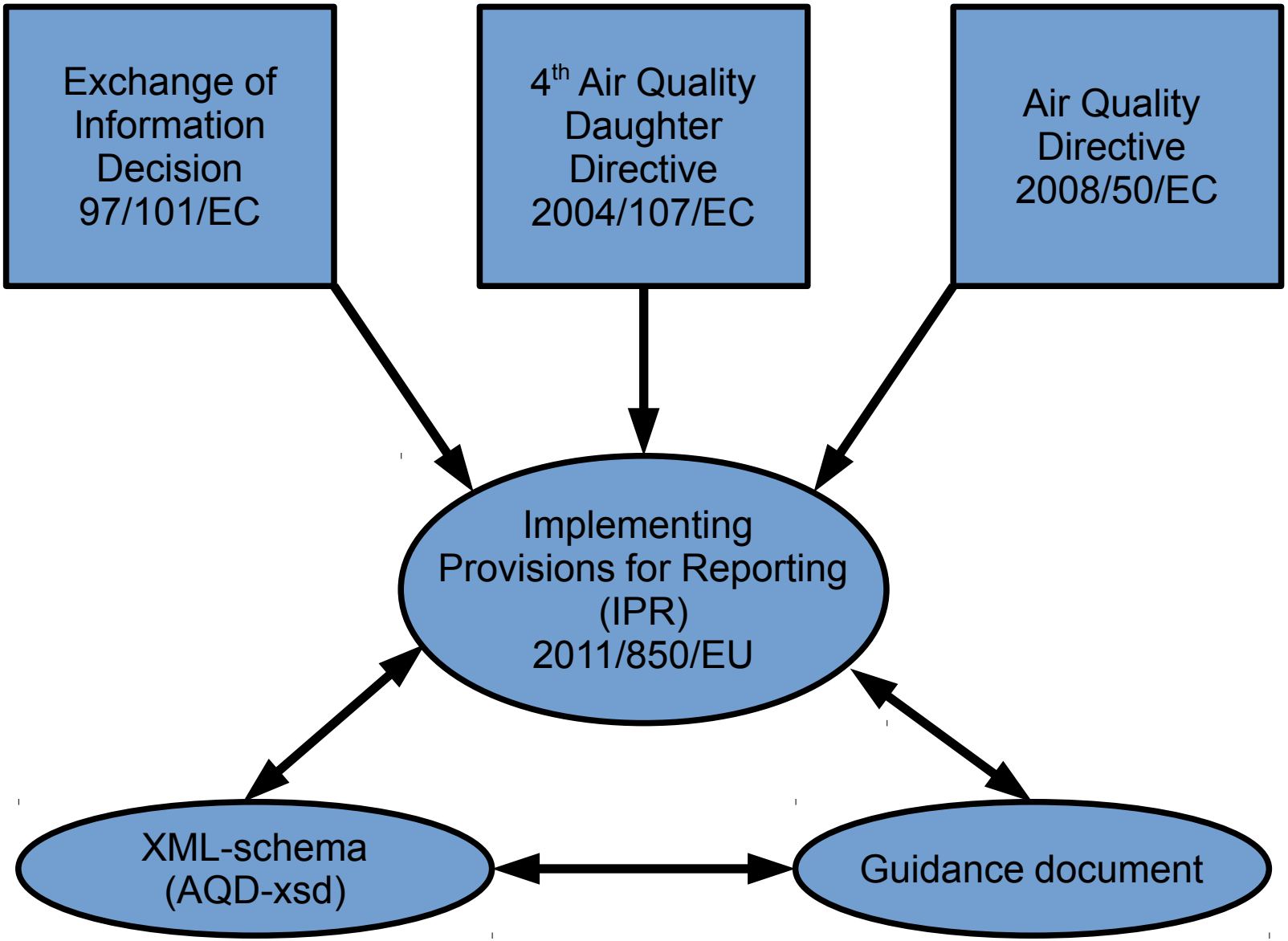
eENVplus

VMM
VLAAMSE MILIEUMAATSCHAPPIJ

Outline

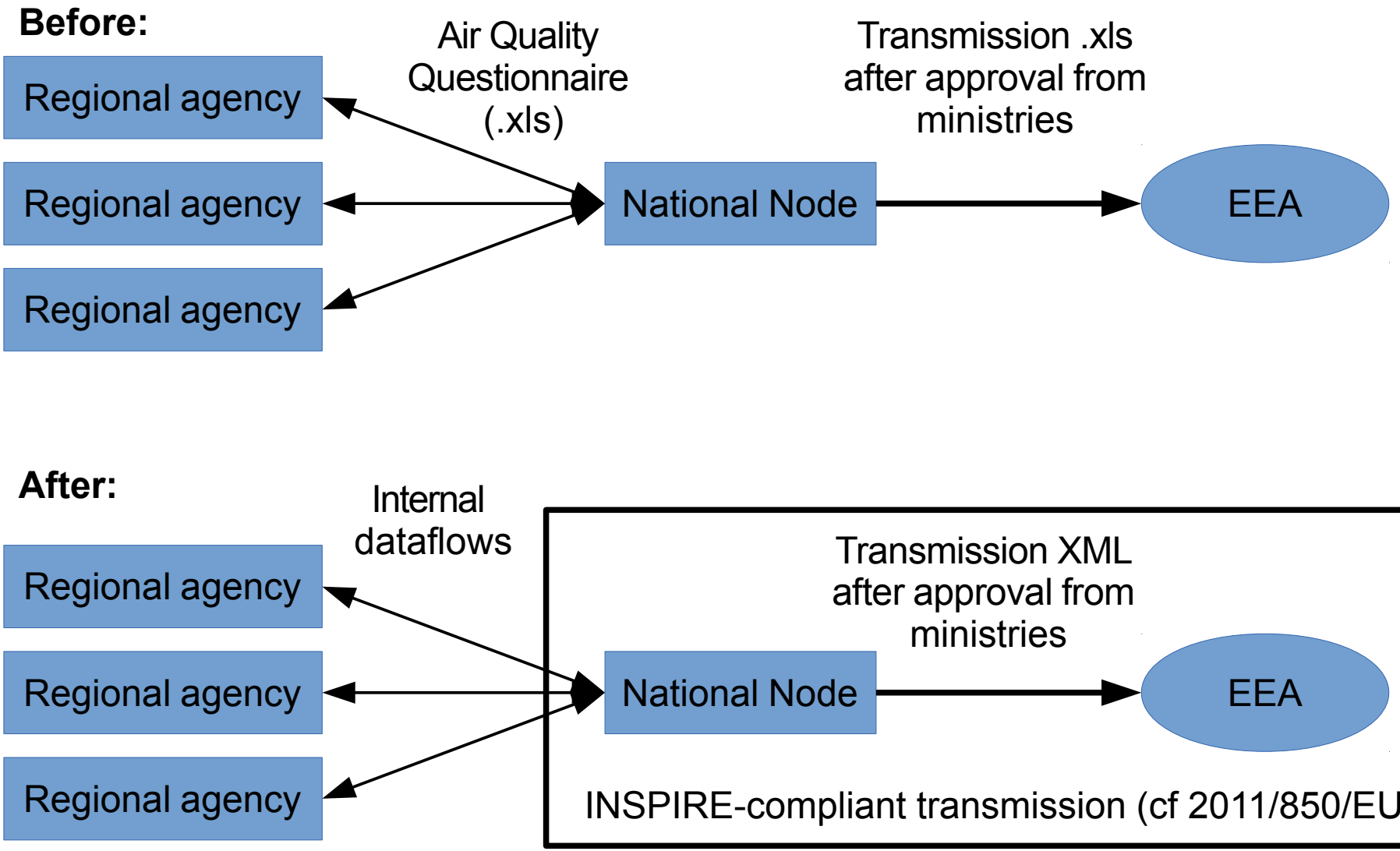
- the context of e-reporting under the Air Quality Directive
- possible implementation via services
- the HUMBOLDT Alignment Editor
- The service stack (cf the eENVplus project)

The relevant directives



more info:
http://aqportal.eionet.europa.eu/

IPR under INSPIRE



The dataflows involved

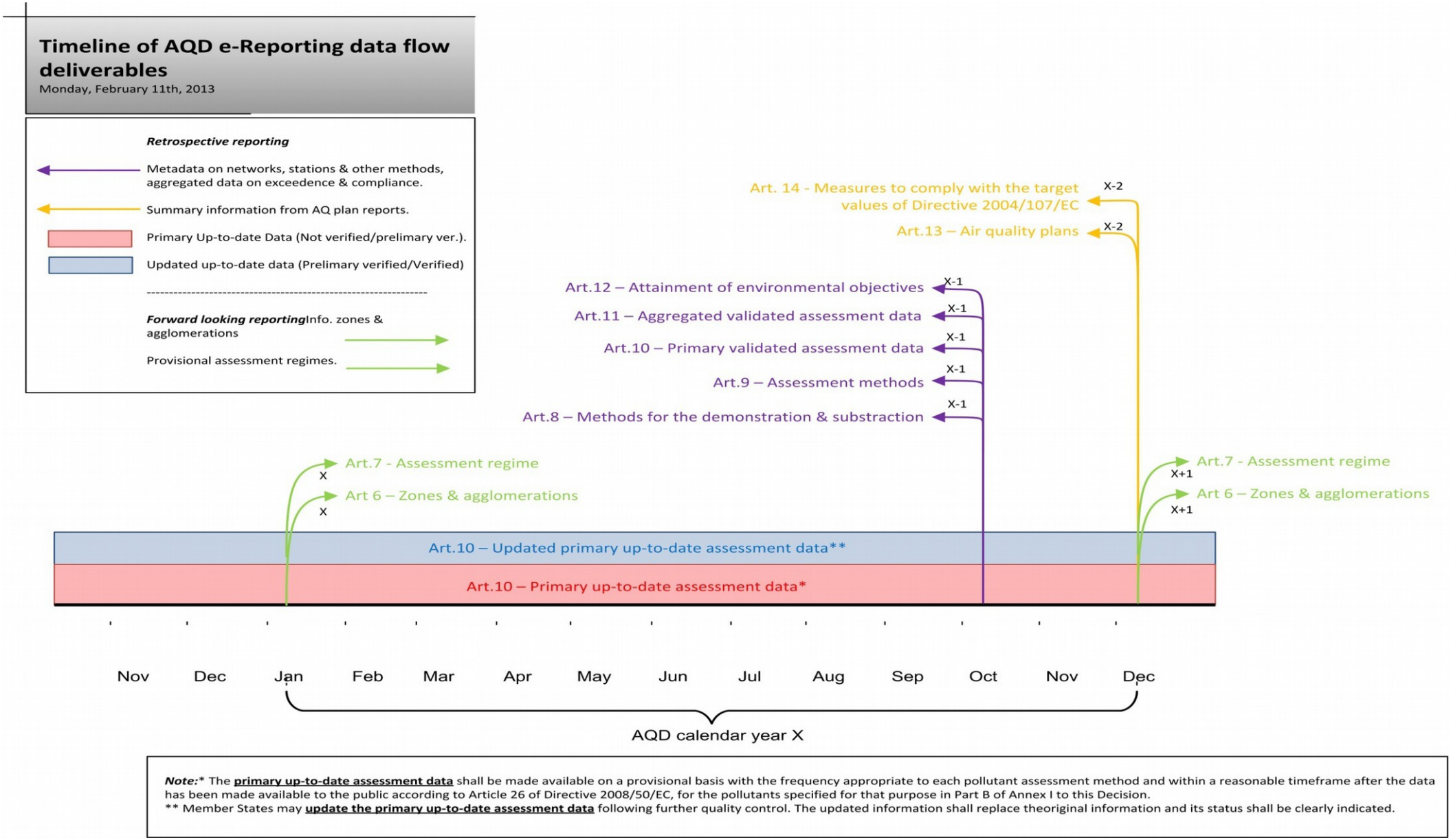
(cf Implementing Provisions for Reporting (IPR) 2011/850/EU)

INPIRE Data Theme	Content
III.11.AM	Dataset B – "zones and agglomerations"
III.11.AM	Dataset C – "assessment regime"
III.11.AM	Dataset D – "assessment methods"
III.7. EF	Dataset E1a – “primary validated assessment data – measurements”
III.13 AC	Dataset E1b – “primary validated assessment data – modelled”
III.7. EF	Dataset E2a – “primary up-to-date assessment data – measurements”
III.13 AC	Dataset E2b – “primary up-to-date assessment data – modelled”
III.7. EF	Dataset F1a – “aggregated data - primary validated measurements”
III.13 AC	Dataset F1b – “aggregated data - primary validated modelled”
III.7. EF	Dataset F2 – “aggregated data - primary up-to-date measurements”
III.11.AM	Dataset G – “attainment of environmental objectives”
III.11.AM	Dataset H – “air quality plans”
III.11.AM	Dataset I – “source apportionment”
III.11.AM	Dataset J – “scenario for the attainment year”
III.11.AM	Dataset K – "measures"

+ Dataset A: a header transmitted with every separate submission

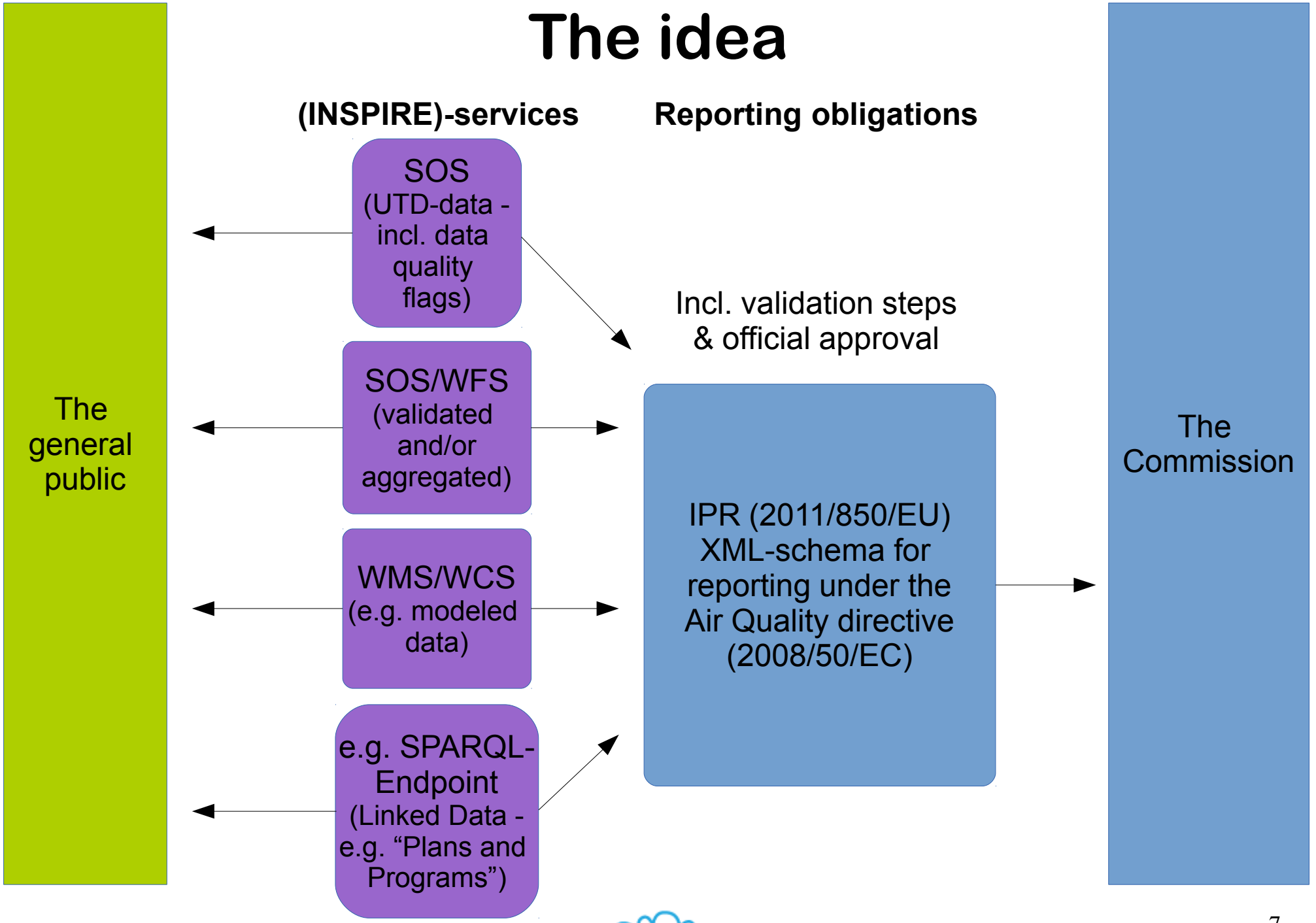
AM = Area Management Restriction Regulation Zones and Reporting units
EF = Environmental Monitoring Facilities
AC = Atmospheric Conditions

Timing



Note: The **primary up-to-date assessment data** shall be made available on a provisional basis with the frequency appropriate to each pollutant assessment method and within a reasonable timeframe after the data has been made available to the public according to Article 26 of Directive 2008/50/EC, for the pollutants specified for that purpose in Part B of Annex I to this Decision.
** Member States may **update the primary up-to-date assessment data** following further quality control. The updated information shall replace theoriginal information and its status shall be clearly indicated.

The idea


(INSPIRE)-services
Reporting obligations
The general public
SOS (UTD-data - incl. data quality flags)
SOS/WFS (validated and/or aggregated)
WMS/WCS (e.g. modeled data)
e.g. SPARQL-Endpoint (Linked Data - e.g. “Plans and Programs”)
Incl. validation steps & official approval
IPR (2011/850/EU) XML-schema for reporting under the Air Quality directive (2008/50/EC)
The Commission

Why?

- after the initial effort to set up the system, it might save resources
- ensure coherence between reported and public data
- INSPIRE compliant e-reporting is the best way to ensure that up to date quality data will be updated regularly
- Open Data initiatives (cf Aarhus Convention & PSI Directive)
- MS Excel spreadsheets are not an open standard data format
- automated quality control

Automated quality control e-reporting

Feedback: AutomaticQA result for file AQD_dataflowB_test_20131118.gml: Check air quality zones

Back to envelope

Subject:	AutomaticQA result for file AQD_dataflowB_test_20131118.gml: Check air quality zones
Posted automatically on:	18 Nov 2013 15:28
Task:	Automatic quality assessment
Referred file:	AQD_dataflowB_test_20131118.gml

Check air quality zones - Dataflow B

B1	Total number of AQ zones	22
B3	The number of zones designated with coordinates via the ./am:geometry element	22
B4	The number of zones designated with coordinates via the ./aqd:LAU element	0
B8	All gml:id attributes, ef:inspireId and aqd:inspireId elements shall have unique content	All Ids are unique
B9	./am:inspireId/base:Identifier/base:localId shall be an unique code for network starting with ISO2-country code	All Ids are unique
B14	./am:name/gn:GeographicalName/gn:nativeness attribute xsi:nil="true" nilReason="unknown"	No unknown values found
B15	./am:name/gn:GeographicalName/gn:nameStatus attribute xsi:nil="true" nilReason="unknown"	No unknown values found
B16	./am:name/gn:GeographicalName/gn:sourceOfName attribute xsi:nil="true" nilReason="unknown"	No unknown values found
B17	./am:name/gn:GeographicalName/gn:pronunciation attribute xsi:nil="true" nilReason="unknown"	No unknown values found
B21	./am:geometry/gml:Polygon/gml:exterior/gml:LinearRing/gml:posList the srsDimension attribute shall resolve to "2" to allow the x & y-coordinate of the feature of interest	All srsDimension attributes resolve to "2"
B30	./am:legalBasis/base2:LegislationCitation/base2:name value shall be "2011/850/EC"	All values are valid
B31	./am:legalBasis/base2:LegislationCitation/base2:date value shall be "2011-12-12"	All values are valid
B32	./am:legalBasis/base2:LegislationCitation/base2:link value shall be "http://rod.eionet.europa.eu/instruments/650"	All values are valid
B35	./aqd:residentPopulation shall be an integer value GREATER THAN 0 (zero)	All values are valid
B37	./aqd:area the value will be a decimal number GREATER THAN 0 (zero)	All values are valid
B42	Where ./aqd:LAU has been used then the reference must point to a concept in the list of LAU2 or NUTS	All values are valid

http://cdr.eionet.europa.eu/be/eu/aqd/b/

Automated report generated after submission to the CDR

It's allot easier to check the quality of datasets if they are transmitted via XML instead of spreadsheets and shapefiles

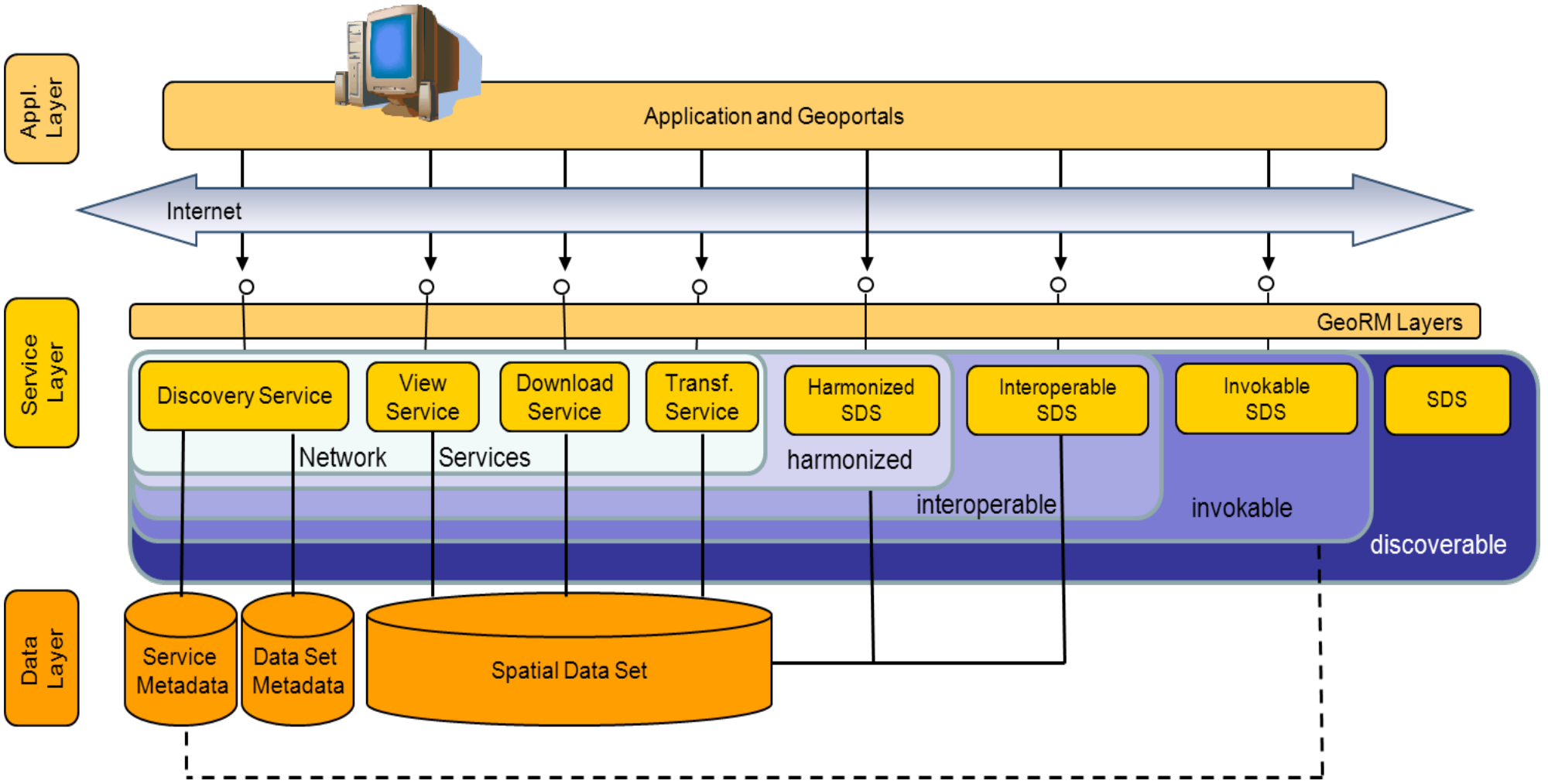
The Air Quality Zones

The Air Quality Zones
as we have always reported them
by means of a shapefile

When we zoom in on some of the borders between administrative units ...

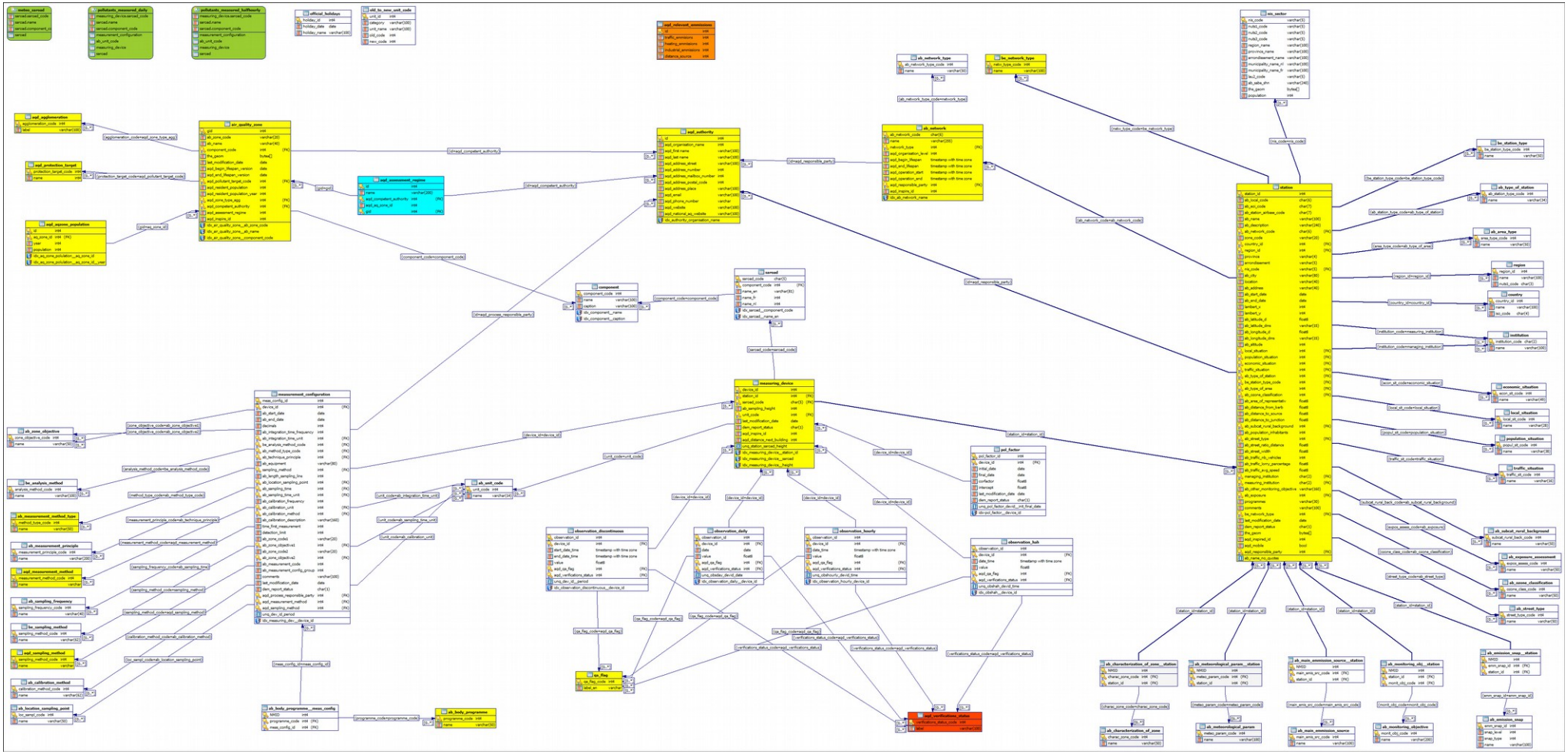
How?

Service Oriented Architecture (SOA)

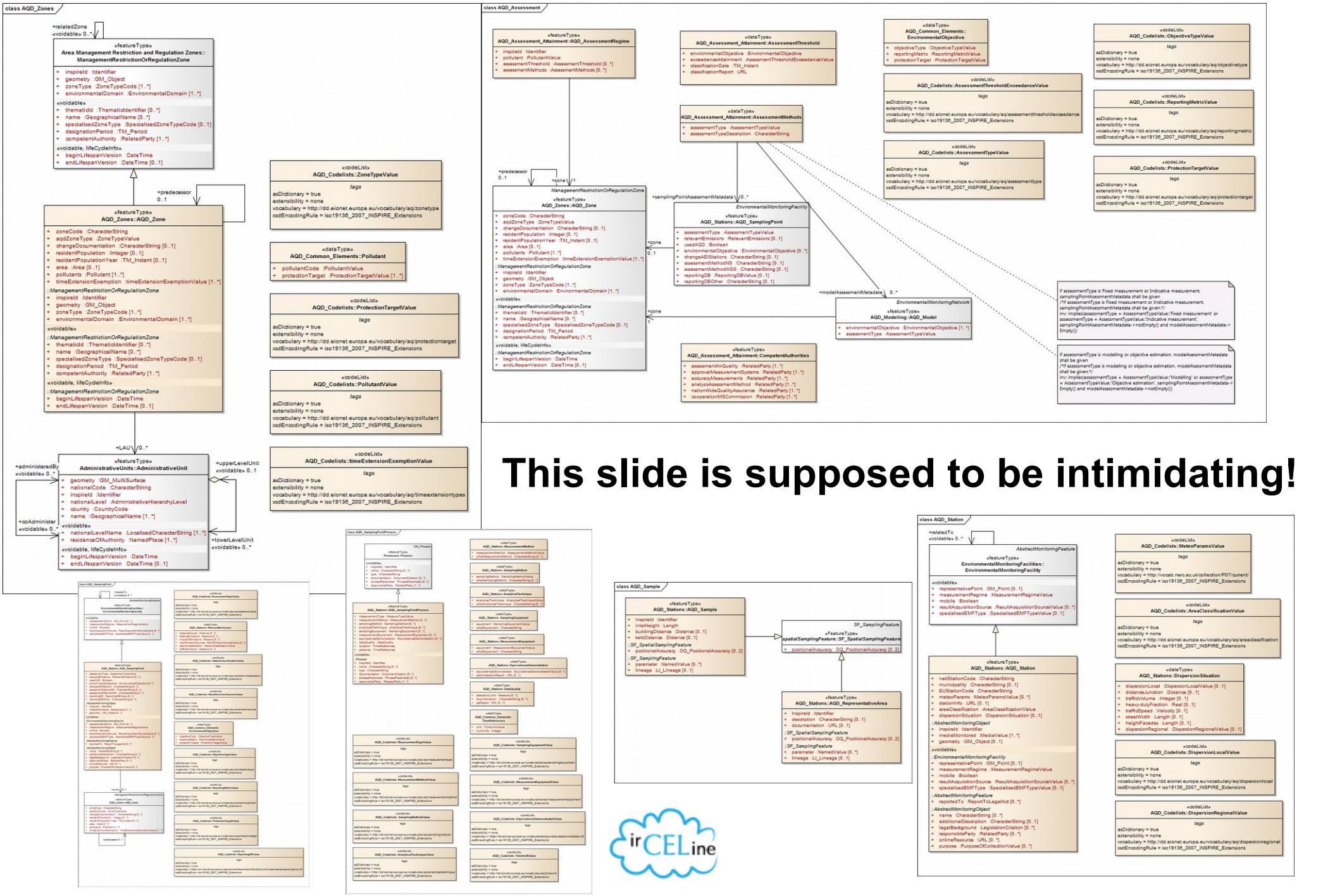


Source: JRC

A legacy database



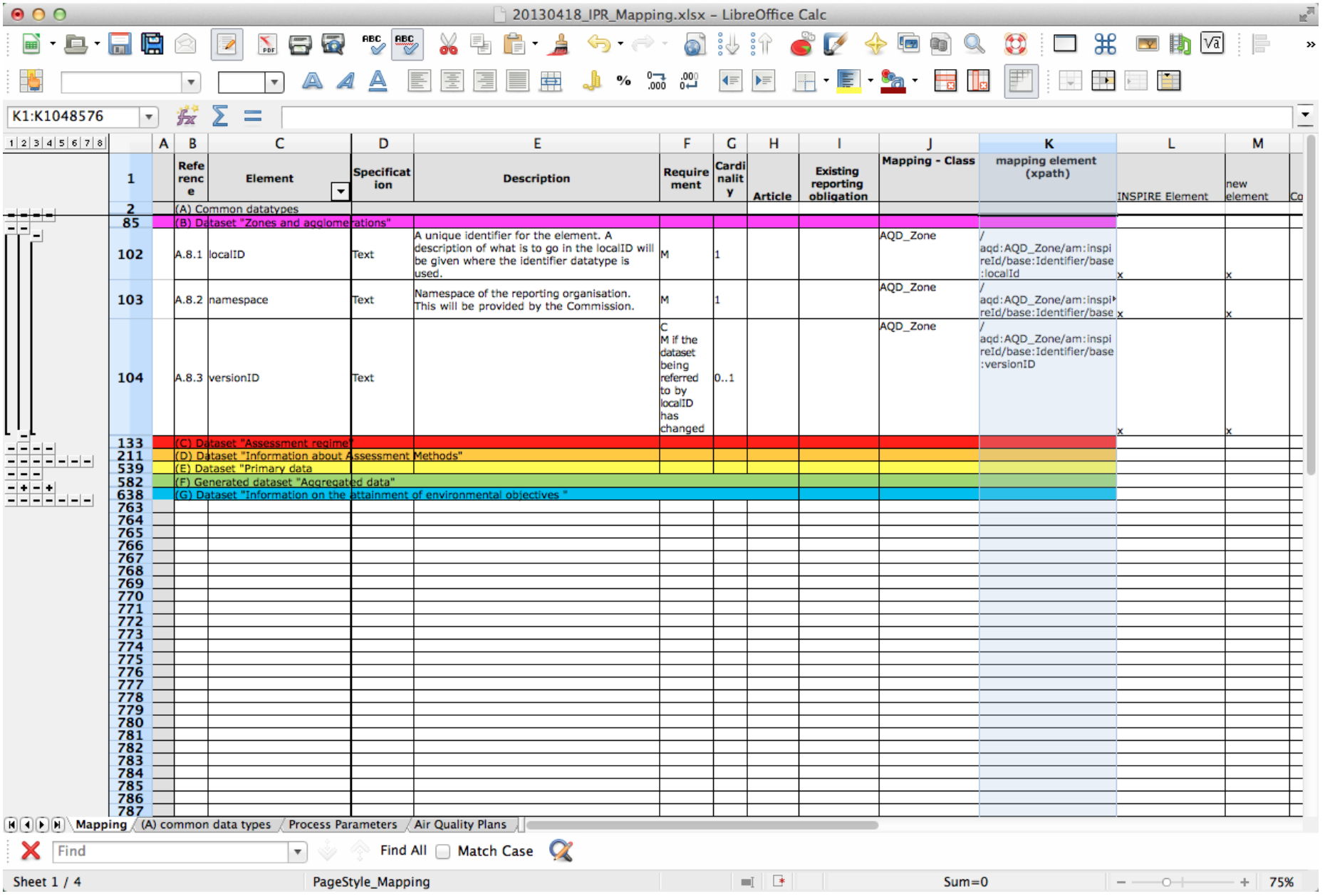
Some data model



This slide is supposed to be intimidating!

irCELine

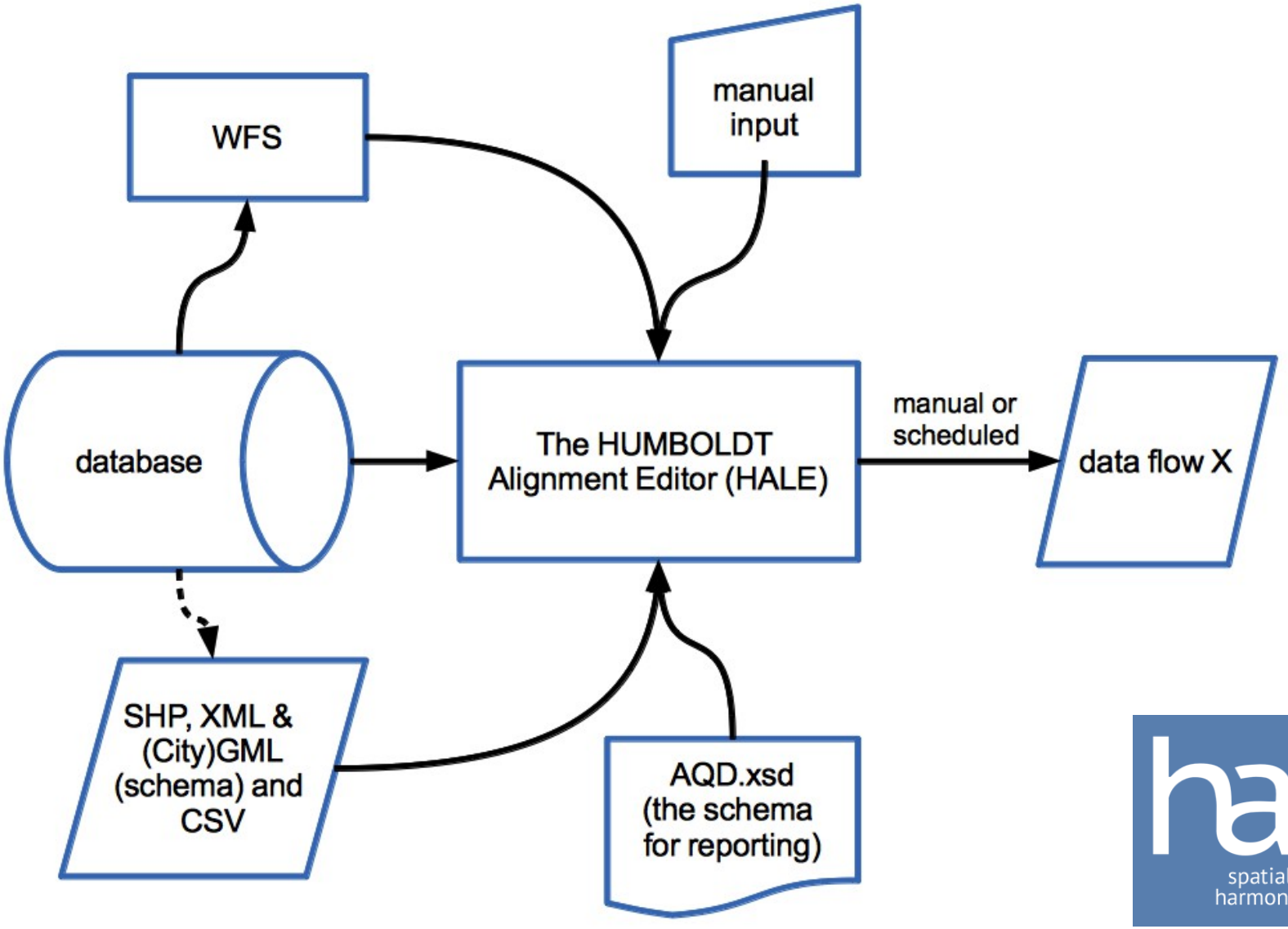
The mapping table (thanks Giacomo!)



	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Reference	Element	Specification	Description	Requirement	Cardinality	Article	Existing reporting obligation	Mapping - Class	mapping element (xpath)	INSPIRE Element	new element
2		(A) Common datatypes											
85		(B) Dataset "Zones and agglomerations"											
102		A.8.1	localID	Text	A unique identifier for the element. A description of what is to go in the localID will be given where the identifier datatype is used.	M	1			AQD_Zone	/ aqd:AQD_Zone/am:inspireId/base:Identifier/base:localId	x	x
103		A.8.2	namespace	Text	Namespace of the reporting organisation. This will be provided by the Commission.	M	1			AQD_Zone	/ aqd:AQD_Zone/am:inspireId/base:Identifier/base	x	x
104		A.8.3	versionID	Text		C M if the dataset being referred to by localID has changed	0..1			AQD_Zone	/ aqd:AQD_Zone/am:inspireId/base:Identifier/base:versionID	x	x
133		(C) Dataset "Assessment regime"											
211		(D) Dataset "Information about Assessment Methods"											
539		(E) Dataset "Primary data											
582		(F) Generated dataset "Aggregated data"											
638		(G) Dataset "Information on the attainment of environmental objectives "											
763													
764													
765													
766													
767													
768													
769													
770													
771													
772													
773													
774													
775													
776													
777													
778													
779													
780													
781													
782													
783													
784													
785													
786													
787													

Possible set-up with an ETL

extract, transform, and load



hale spatial data harmonisation

irCELine

HUMBOLDT Alignment Editor (HALE)

- For Linux (32 and 64bit), Mac (requirement of Java 7 from version 2.7+) and Windows (32 and 64bit)
- The product of a FP6 project
- Good video tutorials and guidance (see website)
- Conceptual Schema Transformer (CST) based, but can also be used as XSLT-editor (cf XSLT- extension of e.g. GeoServer or some other WFS-server)
- Internal schema validation
- Possible to document transformation internally, etc.
- Very helpful developers + forum

Why use HALE?

- Intuitive GUI for mapping data from you database to the schema (with debugging, logging, etc.)
- You can execute HALE from the command line (meaning you can do a crontab on Linux), e.g.:

  [hale@hale ~]$ HALE -nosplash -application de.fhg.igd.hale.fme.app.exec -project <URI-to-project> -source <URI-to-source-data> -out <Path-to-target-file>

  Optional parameters:

  - reportsOut <Path-to-report-file> (Write report of transformation to a file)
  - validate (Enable XML validation)
  - format <format> (Its either 'GML' or 'XML', with 'GML' set as default)
  - root <root-element-name> (The name of the root element to use when using 'XML' as form
  - root-ns <root-element-namespace> (The namespace of the root element to use if using 'XM as format)

- NO closed source dependencies (like MS Access, cf AQU

The website

community.esdi-humboldt.eu/projects/hale

data harmonisation panel — HUMBOLDT Harmonisation Toolkit » HUMBOLDT Alignment Editor

Zoeken:

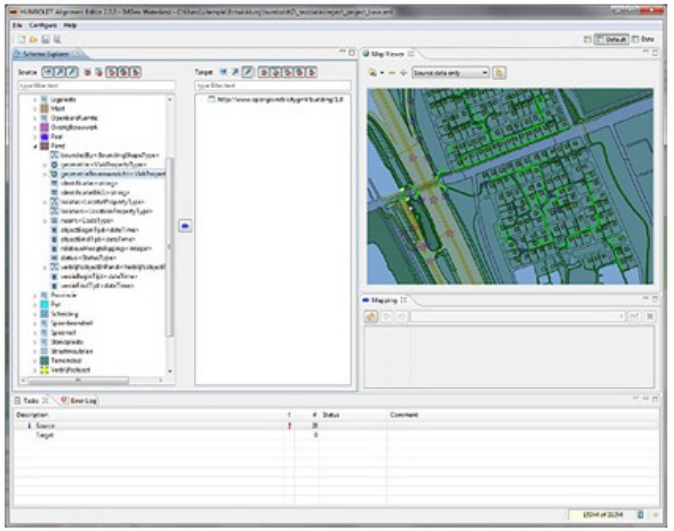
Home | Projecten | Help | Inloggen | Registreer

Overzicht | Activiteit | Roadmap | Issues | Nieuws | Wiki | Forums | Bestanden | Repository

Overzicht

HALE is a tool for *defining and evaluating conceptual schema mappings*. The goal of HALE is to allow domain experts to ensure *logically and semantically consistent mappings* and consequently transformed geodata. Furthermore, a major focus is put on *documentation of the schema transformation process* and its impacts, e.g. in the form of lineage information attached to the resultant transformed data.

Download | User documentation | Sources



Issue-tracking

- **Bug**: 11 open / 158
- **Feature**: 59 open / 322
- **Support**: 1 open / 8

Leden

Manager: **Simon Templer**, **Thorsten Reitz**
Developer: **Andrea Antonello**, **Andreas Burchert**, **Anna Pitaev**, **Christian Malewski**, **Christopher Greening**, **Jachym Cepicky**, **Jan Jezek**, **Jan Kolar**, **Josef Bezdek**, **José Ignacio Gisbert Sanus**, **Kai Schwierczek**, **Kevin Mais**, **Marian de Vries**, **Mark Doyle**, **Patrick Lieb**, **Robert Gregor**, **Sebastian Reinhardt**, **Willem Weiss**
Reporter: **Stefan Thorn**

Laatste nieuws

HALE 2.8.0 connects to your PostGIS database
We just released HALE 2.8.0 to bring you several new features, including PostGIS source data support, sample sets for large source data sets, a nicer classification UI and more.
Toegevoegd door **Thorsten Reitz 12 dagen** geleden

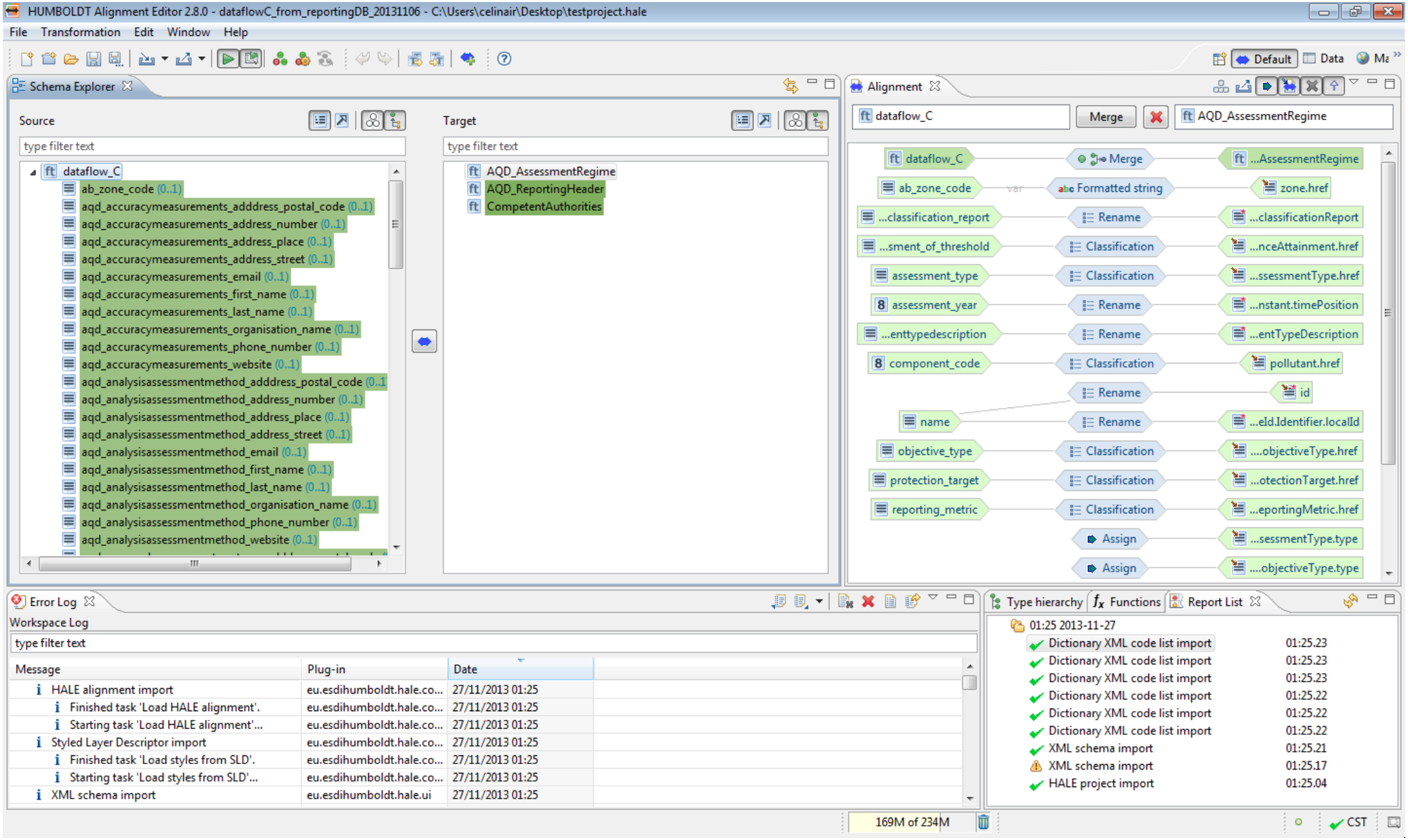
HALE 2.7.0 writes (Geo)JSON and brings web template projects
In this release, we have including a few long-time pending features such as multi-level joins, but also added new formats and cool collaboration features.
Toegevoegd door **Thorsten Reitz ongeveer 1 maand** geleden

HALE 2.6.0 brings integrated data harmonisation workflows
The new version directly integrates with XSLT and FME data transformation processes to bring you improved workflows.
Toegevoegd door **Thorsten Reitz 5 maanden** geleden

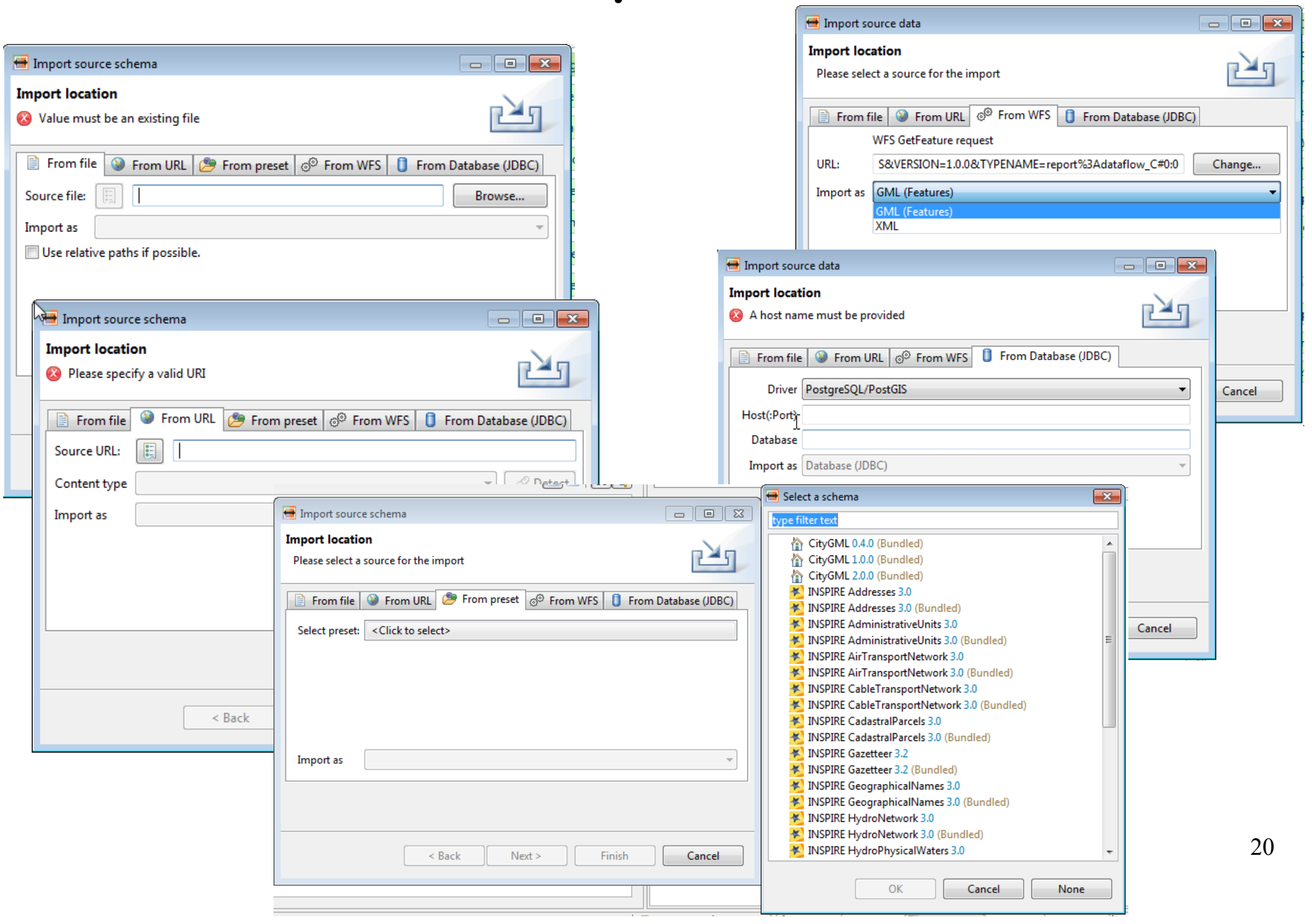
HALE 2.5 released - A Milestone
A year has passed since the last stable HALE release - version 2.1.2 in December 2011 - and much has happened since then.
Toegevoegd door **Thorsten Reitz 12 maanden** geleden

Load a complex schema in a few seconds!

The GUI


HUMBOLDT Alignment Editor 2.8.0 - dataflowC_from_reportingDB_20131106 - C:\Users\celinair\Desktop\testproject.hale
File Transformation Edit Window Help
Schema Explorer
Source
type filter text
dataflow_C
ab_zone_code (0..1)
aqd_accuracymeasurements_adddress_postal_code (0..1)
aqd_accuracymeasurements_address_number (0..1)
aqd_accuracymeasurements_address_place (0..1)
aqd_accuracymeasurements_address_street (0..1)
aqd_accuracymeasurements_email (0..1)
aqd_accuracymeasurements_first_name (0..1)
aqd_accuracymeasurements_last_name (0..1)
aqd_accuracymeasurements_organisation_name (0..1)
aqd_accuracymeasurements_phone_number (0..1)
aqd_accuracymeasurements_website (0..1)
aqd_analysisassessmentmethod_adddress_postal_code (0..1
aqd_analysisassessmentmethod_address_number (0..1)
aqd_analysisassessmentmethod_address_place (0..1)
aqd_analysisassessmentmethod_address_street (0..1)
aqd_analysisassessmentmethod_email (0..1)
aqd_analysisassessmentmethod_first_name (0..1)
aqd_analysisassessmentmethod_last_name (0..1)
aqd_analysisassessmentmethod_organisation_name (0..1)
aqd_analysisassessmentmethod_phone_number (0..1)
aqd_analysisassessmentmethod_website (0..1)
Target
type filter text
AQD_AssessmentRegime
AQD_ReportingHeader
CompetentAuthorities
Alignment
dataflow_C
Merge
AQD_AssessmentRegime
dataflow_C
Merge
...AssessmentRegime
ab_zone_code
var
Formatted string
zone.href
...classification_report
Rename
...classificationReport
...sment_of_threshold
Classification
...nceAttainment.href
assessment_type
Classification
...ssessmentType.href
assessment_year
Rename
...nstant.timePosition
...enttypedescription
Rename
...entTypeDescription
component_code
Classification
pollutant.href
Rename
id
name
Rename
...eId.Identifier.localId
objective_type
Classification
....objectiveType.href
protection_target
Classification
...otectionTarget.href
reporting_metric
Classification
...eportingMetric.href
Assign
...sessmentType.type
Assign
....objectiveType.type
Error Log
Workspace Log
type filter text
Message
Plug-in
Date
HALE alignment import
eu.esdihumboldt.hale.co...
27/11/2013 01:25
Finished task 'Load HALE alignment'.
eu.esdihumboldt.hale.co...
27/11/2013 01:25
Starting task 'Load HALE alignment'...
eu.esdihumboldt.hale.co...
27/11/2013 01:25
Styled Layer Descriptor import
eu.esdihumboldt.hale.co...
27/11/2013 01:25
Finished task 'Load styles from SLD'.
eu.esdihumboldt.hale.co...
27/11/2013 01:25
Starting task 'Load styles from SLD'...
eu.esdihumboldt.hale.co...
27/11/2013 01:25
XML schema import
eu.esdihumboldt.hale.ui
27/11/2013 01:25
Type hierarchy
Functions
Report List
01:25 2013-11-27
Dictionary XML code list import 01:25.23
Dictionary XML code list import 01:25.23
Dictionary XML code list import 01:25.23
Dictionary XML code list import 01:25.22
Dictionary XML code list import 01:25.22
Dictionary XML code list import 01:25.22
XML schema import 01:25.21
XML schema import 01:25.17
HALE project import 01:25.04
169M of 234M
CST

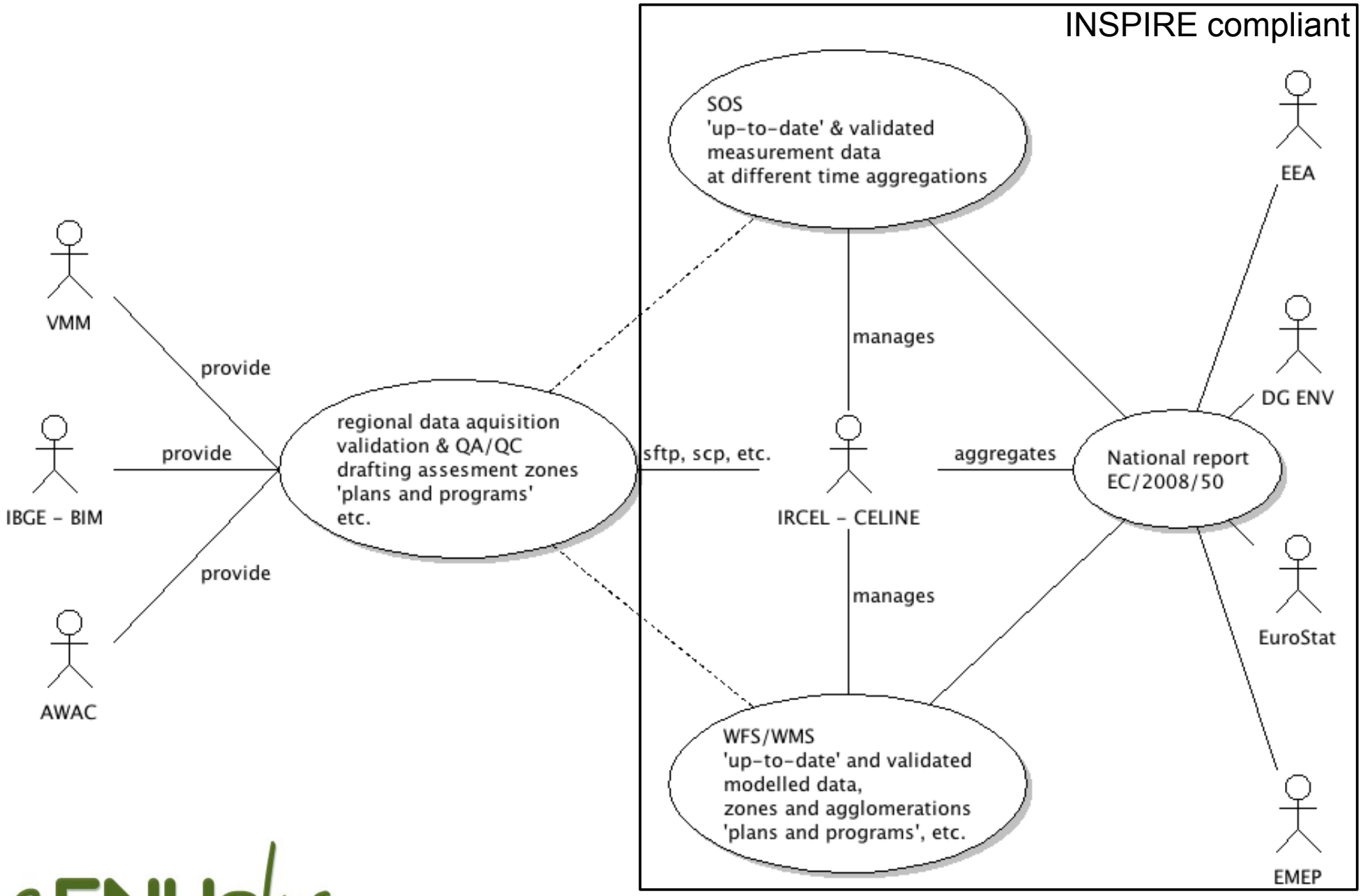
From file, URL, preset, WFS or DB



Some links to more information

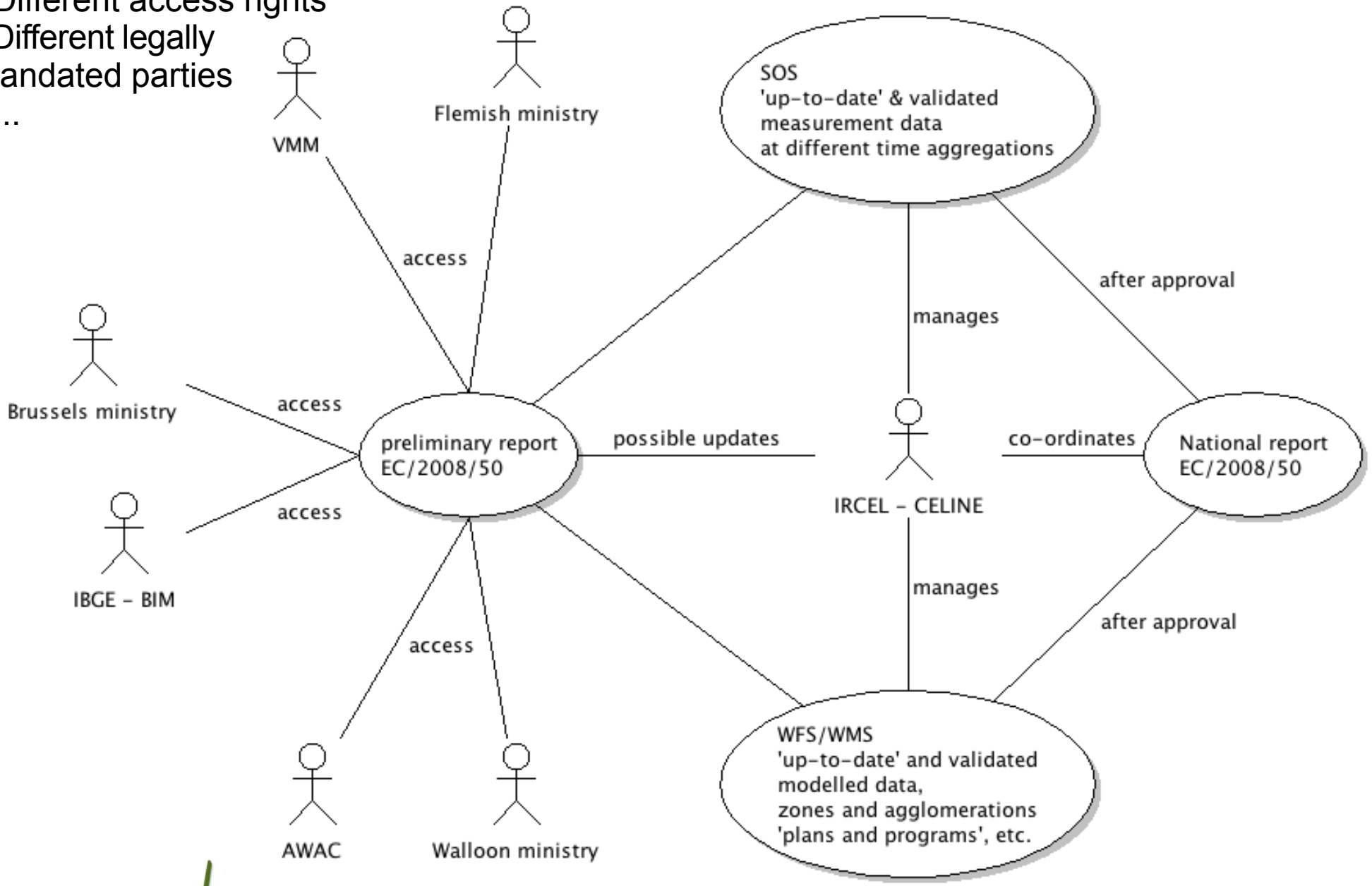
- http://www.dhpanel.eu/humboldt-framework/hale/ (data harmonisation panel)
- http://www.esdi-community.eu/projects/hale (project site - wiki, forum, tutorials, issues, downloads software, etc.)
- https://github.com/igd-geo/hale (source code)
- http://blog.dhpanel.eu/2013/11/09/inspire-ken-all-presentations-and-videos-now-publicly-accessible/ (recent presentations, incl. demo)

Reporting service

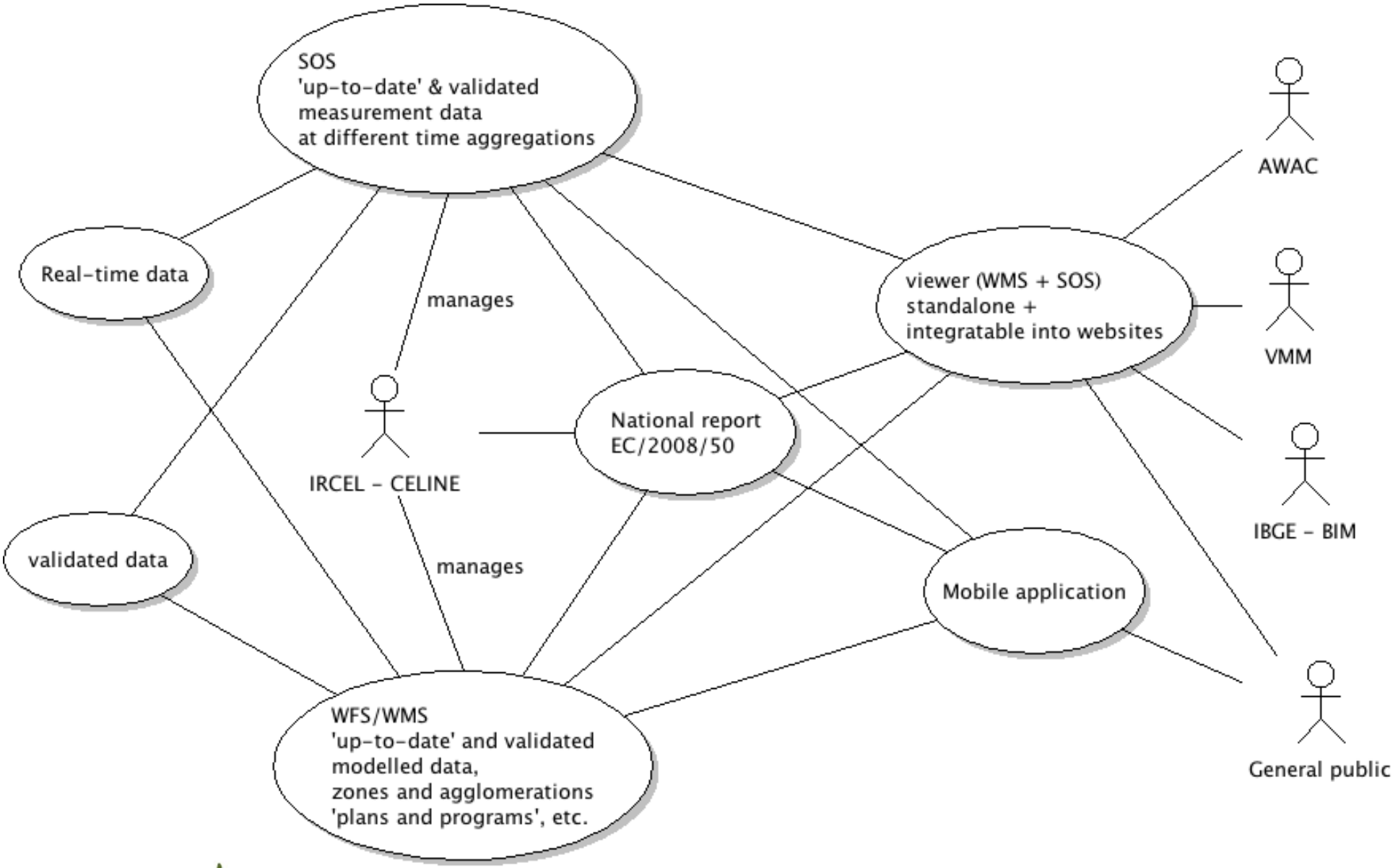


Validation tool

- Different access rights
- Different legally mandated parties
- ...



Public information



Conclusion

- e-reporting certainly adds considerably to the technical complexity of reporting
- no single tool will be able to solve everything
- the HUMBOLDT Alignment Editor (HALE) is a very handy LEGO block in the stack (excellent standards compliance very useful transformation functions, incl. Groovy scripts)
- HALE can help simplify several steps in setting up INSPIRE and/or IPR compliant services (cf XSLT for a WFS server)
- INSPIRE compliant e-reporting is crucial to ensure that interoperable quality data will be updated in the future

Informing you on ambient air quality
in the Belgian Regions

Thank you!

Olav Peeters
Belgian Interregional Environment Agency (IRCEL – CELI
peeters@irceline.be