

CLEAR Info: Publishing in INSPIRE data formats

Version 0.2
21st October 2013

This study has been made possible with Life+ funding from the European Commission.



Changelog

Version	Date	Editor	Notes
0.1	21/10/2013	Stuart Williams	Initial draft.
0.1	10/03/2014	Stuart Williams	Revised in response to comments

CLEAR Info: Populating INSPIRE spatial objects from WIMS Data

Summary

This report explores the extent to which water quality and discharge consent information held in the Environment Agency Water Information Management System (WIMS), and imported into CLEAR info, can be published as INSPIRE compliant data using the thematically defined INSPIRE data specifications.

It is generally the case that WIMS holds more data than can be expressed using the INSPIRE data specifications as they are. In addition, some fields defined for INSPIRE feature and data types require more data than is available in WIMS. Fortunately in most case of there being insufficient data in WIMS to populate an field in the relevant INSPIRE structure, its presense is usually optional.

In order to fully express the data held within WIMS it will be necessary to create application schema for WIMS or CLEAR Info that extends the INSPIRE defined application schema by importing those that are relevant and subclassing from data and feature types defined within them.

This is particularly true for permit and organisational information where there are gaps between the data that WIMS/CLEAR Info hold and what can be expressed with the native INSPIRE data specifications.

Water quality information can be readily expressed using the INSPIRE data specifications drawing on the Enviromental Monitoring Facilities data specification used in conjunction with ISO 19156 Observations and Measurements (O&M) as described in INSPIRE technical guidlines for the use of of O&M.

Discharge sites and discharge outlets can both be presented as instances of *EnvironmentalManagementFacility* which may be arrange in a hierarchical fashion that a Discharge Site may be a parent facility to a Discharge Outlet.

The INSPIRE *ActivityComplex::Permission* can be used to indicated the existence of a permits and some of its details. However, it is not adequate for the expression of the full detail of a permit as held in WIMS. For that further specialisation is required.

Except for *ResponsibleAgency*, INSPIRE's treatment of organizational entities and individuals is to formulate them as data types. As a data type, they have no object identifier and the only basis for the identity of two such organisations is a comparison of their properties - they are basically structured literals. CLEAR Info will to need to develop its own application schema elements in this area. More generally, both EU and W3C (and no-doubt others) provide vocabulary for the idenification and description of people and organisations. The W3C Organization Ontology provides for a richer expression of time varying role holding and organisational sub-structure.

Finally the report discusses approaches and tooling for transforming a data into INSPIRE formats.

The table below summarises the major alignments found between WIMS/CLEAR Info content and INSPIRE application schema define feature and datatypes.

WIMS/CLEAR Info content	INSPIRE feature or data type
Sampling Points	<i>EF::EnvironmentalMonitoringFacility</i>
Samples	<i>OM::SF_Specimen</i>
Measurements	<i>OM::Measurement</i>
Discharge Sites	<i>US::EnvironmentalManagementFacility</i>
Discharge Outlet	<i>US::EnvironmentalManagementFacility</i>

Permits	<i>ActivityComplex::Permission</i> needs specialisation
Regulatory Organizations (eg. EA itself)	<i>ResponsibleAgency</i>
Operating Companies/Organisations	Not well addressed by INSPIRE

Contents

Changelog.....	ii
Contents.....	iii
Abbreviations	v
Tasks Summary from Task sheet (To be deleted).....	Error! Bookmark not defined.
1 Introduction.....	6
1.1 INSPIRE	6
1.2 CLEAR Info	Error! Bookmark not defined.
2 INSPIRE, WIMS and CLEAR Info.....	7
2.1 Water-Quality Monitoring	9
2.1.1 Environmental Facilities - Spatial Object Types.....	10
2.1.2 Environmental Facilities - Observations and Measurements.....	13
2.2 Discharge Agreement/Consents	15
2.3 Utility and Governmental Services	19
3 Transforming Data to INSPIRE formats.....	20
3.1 Specialising INSPIRE application schema.....	24
3.2 Publishing INSPIRE compliant data.....	24
3.3 Method and process.....	25
4 Annex I - WIMS Data Model and Data Sample.....	27
4.1 Out of Scope Data.....	27
4.2 In Scope Data.....	28
4.2.1 CodeLists and Controlled Vocabularies	31
4.2.2 Identifiers.....	31
4.2.3 Managing Change	33
5 Annex II CLEAR Info Datasources.....	35

Abbreviations

API	Application Programming Interface
CAS	Chemical Abstracts Service
CEH	Centre for Ecology and Hydrology
CPS	Catchment Planning System
D2RQ	A platform for accessing databases as RDF graphs
DRN	Detailed River Network
EA	Environment Agency
ETL	Extract, Transform and Load
GCM	INSPIRE Generic Conceptual Model
ICM	Integrated Catchment Management
INSPIRE	EU Directive on "Infrastructure for Spatial Information in the European Community"
ISO	International Organization for Standardization
NERC	Natural Environment Research Council
NGR	National Grid Reference
OGC	Open Geospatial Consortium
ONS	Office of National Statistics
OS	Ordnance Survey
R2RML	RDB to RDF Mapping Language
RDB	Relational Database
RDF	Resource Description Framework
SPARQL	SPARQL Protocol and RDF Query Language
URI	Uniform Resource Identifier
WFD	EU Water Framework Directive
WFS	OGC Web Feature Service
WMS	OGC Web Map Service
WMTS	OGC Web Map Tile Service
WIMS	Water Information Management System

1 Introduction

The EU Life funded CLEAR Info project seeks to integrate environmental and organisational information held by a number of Environment Agency systems in order to demonstrate the benefits of a clearer, more integrated presentation of the information held by the an Environmental regulator. The project aims to test to what degree the Data Collation Cube can be used to create INSPIRE compliant data sets, to enable more effective implementation of EU legislation.

This report focuses primarily on data contributed into the Clear Info data collation tool from the Environment Agency's Water Information Management System (WIMS) which contains water quality monitoring information and discharge agreements (also known as consents).

It shows that much of the data imported from WIMS can indeed be published in INSPIRE compliant form. In particular it presents an alignment of the data contributed by WIMS with the thematic application scheme defined by the INSPIRE community. However, there is more information available from WIMS that can be expressed using the INSPIRE schema as-is. and the INSPIRE Generic Conceptual Model provides for this by allowing an application, Clear Info in this case, to create its own application schema by importing and extending one or more of the INSPIRE schema.

The report also discusses approaches to the transformation of data into INSPIRE compliant formats and its publication principally in the form of an OGC Web Feature Service which is the core technology used to provide the INSPIRE download service.

1.1 INSPIRE

Data available from WIMS covers two key areas related to the Public Register.

- Water quality monitoring
- Discharge Agreement/Consents

Some of this data falls within the scope of EU INSPIRE directive themes. INSPIRE is motivated by a need to provide pan-European access to environmental information in order to improve decision making on environmental matters.

*"This Directive should apply to spatial data held by or on behalf of public authorities and to the use of spatial data by public authorities in the performance of their public tasks. Subject to certain conditions, however, it should also apply to spatial data held by natural or legal persons other than public authorities, provided that those natural or legal persons request this."*¹

This implies that any spatial datasets held by or on behalf of a public authority fall within the scope of the INSPIRE directive. The directive defines spatial data and spatial data sets as:

'spatial data' means any data with a direct or indirect reference to a specific location or geographical area;

'spatial data set' means an identifiable collection of spatial data;

WIMS data contains point position information for sampling points, and discharge outlets. It also contains representative point position or bounding box (envelope) information for discharge sites and controlled waters. Both the water quality monitoring and the discharge agreement/consent data held within WIMS make reference to these spatial entities.

¹ <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:EN:PDF>

1.2 CLEAR Info

The CLEAR Info² EU project integrates information from approximately 20 data source (see Section 5). including WIMS. It uses Microsoft SQL Server Integration Services (SSIS) to perform a number of data clean up, normalisation, enrichment with more authoritative data and de-duplication. This initial 'import' normalises the data into 4 broad categories: **Parties**, **Sites**, **Interactions** (such as permits), and **Transactional Data** (such as samples and measurements).

Some of the properties used to describe each of these entities align along the dimensions of a multi-dimensional hypercube (more than 20 dimensions), or 'cube' for short. Some of the dimensions have a hierarchical organisation. Analytics are run over the resulting structure to create aggregations over the hierarchical dimensions of the 'cube' - which effectively pre-computes answers to questions like "give me a count of the compliance breaches in the South West region, by operating companies, on a quarterly basis for the last three years".

2 INSPIRE, WIMS and CLEAR Info

INSPIRE themes and data specifications for data covered by WIMS and CLEAR Info are:

- Environmental Monitoring Facilities³
- Production and Industrial Facilities⁴
- Government and Utility Services⁵
- Area Management/Restriction/Regulation Zone and Reporting Units⁶
- Hydrography⁷

In addition the following framework specifications are relevant:

- Base Models - Activity Complex⁸
- Base Models - Coverage Types⁹

Along with the following

- Draft Guidelines for the use of Observations & Measurements and Sensor Web Enablement - related standards in INSPIRE Annex II and III data specification development¹⁰
- Guidelines for the encoding of spatial data¹¹

From a technical point of view INSPIRE builds on a foundation provided by the ISO 19xxx series of specifications produced jointly by ISO TC 211 (<http://www.isotc211.org/>) and the Open Geospatial Consortium, the OGC (<http://www.opengeospatial.org/>). Much of the encoding and representation of spatial information is not apparent in the foreground INSPIRE applications schema or data specifications, they are covered largely by this foundation which includes:

- ISO:19107 Spatial Schema
- ISO:19109 Application Schema
- ISO:19115 Metadata
- ISO:19136 GML (Geospatial Markup Language)
- ISO: 19156 Observations and Measurements

² http://www.environment-agency.gov.uk/CLEAR_Info

³ http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_EF_v3.0rc3.pdf

⁴ http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_PF_v3.0rc3.pdf

⁵ http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_US_v3.0rc3.pdf

⁶ http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_AM_v3.0rc3.pdf

⁷ http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_HY_v3.0.1.pdf

⁸ http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/D2.10.3_Activity_Complex_v1.0rc3.pdf

⁹ http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/D2.10.2_CoverageTypes_v1.0rc3.pdf

¹⁰ http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/D2.9_O&M_Guidelines_v2.0rc3.pdf

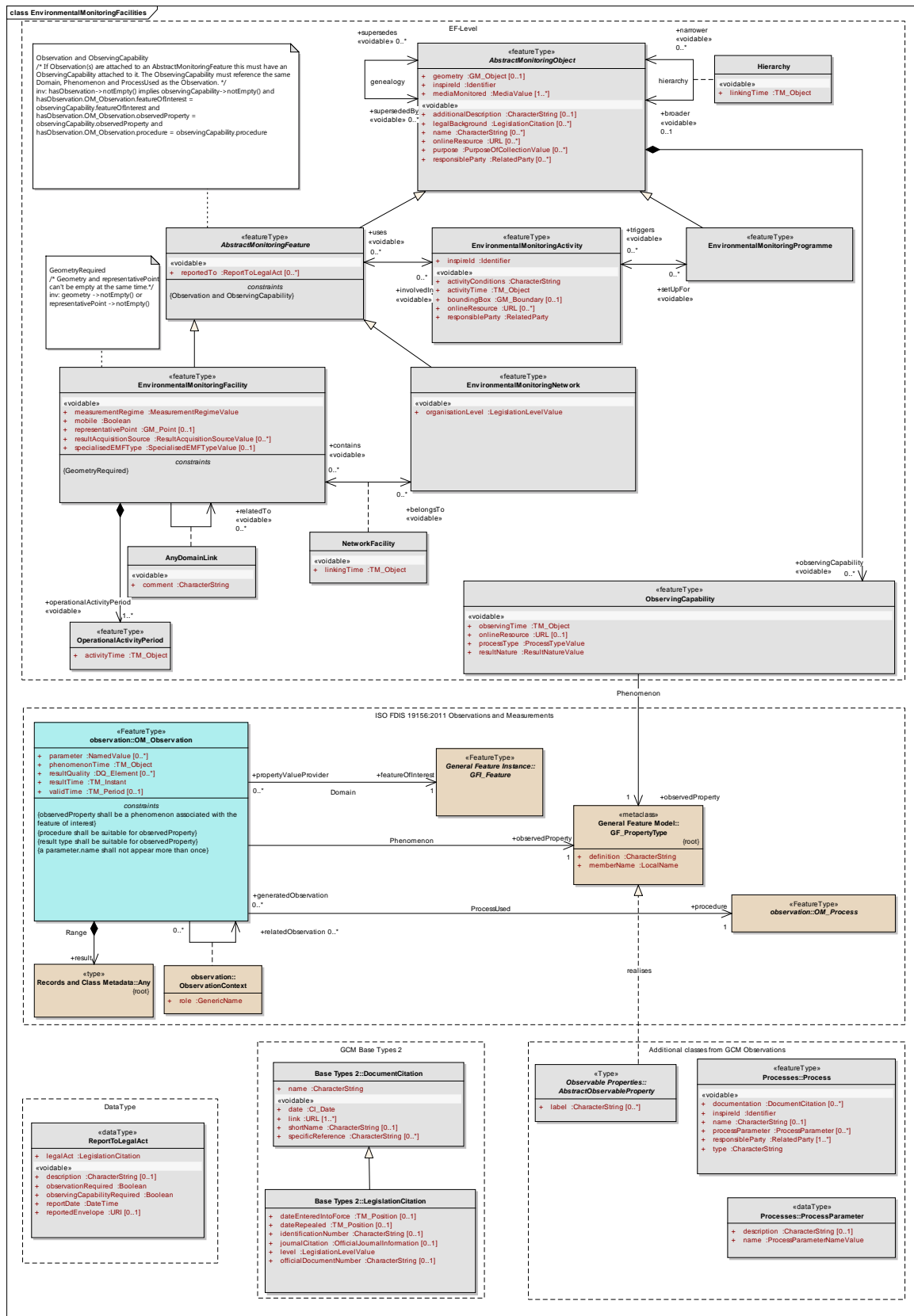
¹¹ http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/D2.7_v3.3rc3.pdf

The diagrams in the following sections are taken from the consolidated UML model for the INSPIRE data specifications, the latest draft of which can be found at <http://inspire.ec.europa.eu/data-models/>. The CLEAR Info project intends to make the data that it publishes available as an INSPIRE data set, harmonised to align with the relevant INSPIRE data specifications. INSPIRE specifies a download service¹² which can be implemented using ATOM or as an OGC Web Feature Service (WFS).

¹² http://inspire.jrc.ec.europa.eu/documents/Network_Services/Technical_Guidance_Download_Services_v3.1.pdf

2.1 Water-Quality Monitoring

The diagram below summaries the INSPIRE Environmental Facilities data specification:



2.1.1 Environmental Facilities - Spatial Object Types

This section makes an assessment of how well the WIMS held sampling-point information is aligned with the information necessary to populate INSPIRE spatial-objects defined by the INSPIRE Environmental Monitoring Facilities data specification. It considers the data required to populate instances of a given featureType. It also highlights fields from the WIMS data that would require specialisation (through subclassing in derived application schema) in order to be expressed as part of a published spatial-object.

The application schema (aka data specification) for the INSPIRE Environmental Monitoring Facilities theme defines four spatial object types (*featureTypes*):

INSPIRE EF featureType	WIMS Entities	Comment
EnvironmentalMonitoring Facility	sampling point	
EnvironmentalMonitoringActivity	Not represented in WIMS	WIMS sampling point carry TYPE (agriculture, freshwater, saline water, waste...) and CLASS (water/effluent volumes); WIMS samples carry PURPOSE and PROJECT_CODE attributes. From these it may be possible to infer existence and membership of sampling networks and sampling programs. <i>EnvironmentalMonitoringActivities</i> are short-term in comparison with <i>EnvironmentalMonitoringPrograms</i> . They may occur as part of a program.
EnvironmentalMonitoringNetwork		
EnvironmentalMonitoringProgram		

At the time of writing, it is not clear that WIMS holds sufficient information to create instances of the *EnvironmentalMonitoringActivity*, *EnvironmentalMonitoringNetwork* and *EnvironmentalMonitoringProgram* featureTypes. WIMS held TYPE, CLASS, PURPOSE and PROJECT attributes of a sampling point **may** carry sufficient information to infer the existence and membership of sampling networks and sampling programs. However, no explicit indication was found in the WIMS schema.

All four of the featureType classes have some common attributes, embodied in the abstract class *AbstractMonitoringObject*, for:

- a) inspire spatial object identification (*inspireId*)
- b) associating an optional geometry (*geometry*)
- c) building hierarchical structures (*broader*, *narrower*)
- d) representing lineage (*supercedes*, *supercededBy*)

and for associating

- e) zero or more legislative citation (*legalBackground*)
- f) zero or more purposes (*purpose* with values from a currently open ended codelist - see codelist diagram below)
- g) one or more monitored media (*monitoredMedia* with values from a currently closed INSPIRE governed codelist - see codelist diagram below eg. air, soil, water, waste...).
- h) zero or more responsible parties (*responsibleParty*)
- i) zero or more related resources (*onlineResource*)
- j) zero or more a names (*name*)
- k) zero or more additional (text) descriptions (*additionalDescription*)

The INSPIRE Environmental Monitoring Facilities data specifications contains the following definitions:

EnvironmentalMonitoringProgram:

Framework based on policy relevant documents defining the target of a collection of observations and/or the deployment of *AbstractMonitoringFeatures* on the field.

Usually an Environmental Monitoring Programme has a long term perspective over at least a few years.

EnvironmentalMonitoringActivity:

Specific set of *AbstractMonitoringFeatures* used for a given domain in a coherent and concise

timeframe, area and purpose. Usually the information collected is treated as one time step in a long term monitoring programme. It is a concrete realisation of a given EnvironmentalMonitoringProgramme.

EnvironmentalMonitoringNetwork:

Administrative or organisational grouping of EnvironmentalMonitoringFacilities managed the same way for a specific purpose, targeting a specific area. Each network respects common rules aiming at ensuring coherence of the observations, especially for purposes of EnvironmentalMonitoringFacilities, mandatory parameters selection, measurement methods and measurement regime.

EnvironmentalMonitoringFacility:

A georeferenced object directly collecting or processing data about objects whose properties (e.g. physical, chemical, biological or other aspects of environmental conditions) are repeatedly observed or measured. An environmental monitoring facility can also host other environmental monitoring facilities.

Of these, EnvironmentalMonitoringFacility most closely fits the nature of a sampling point in its role of making repeated measurements and observations.

Focussing on WIMS sampling points cast as *EnvironmentalMonitoringFacility*, its attributes and associations can be populated as follows:

EnvironmentalFacility Attribute or association	WIMS Sampling Point Table: SAMPLING_POINTS
inspireId	URI or other namespaced identifier derived from SMPT_USER_REFERENCE. If versioned, <i>inspireId.versionId</i> may be derived from the date/time of the last modification of the corresponding record probably, likely to be the value carried in SMPT_DATE_STAMP. If http URI are used as INSPIRE namespaces, and http URI can be formed by the concatenation of <i>{inspireId.namespacedId}{inspieid.localId}[:{inspireId.versionId}]</i>
Geometry	A GM_Point derived from SMPT_GRID_REF or SMPT_NORTHING and SMPT_EASTING.
mediaMonitored	<i>water</i> , specifically < http://inspire.ec.europa.eu/codeList/MediaValue/water >
additionalDescription	One or more narrative comments including SMTP_SHORT_NAME (or SMTP_LONG_NAME) SMPT_COMMENTS, SP_SHORT_DESC, SP_SAMPLERS_DIRECTIONS
Name	SMTP_LONG_NAME (or SMTP_SHORT_NAME)
legalBackground	Not directly represented in WIMS.
Purpose	May be derivable from some combination of SMPT_TYPE, SMPT_CLASS and/or an aggregation of SAMP_PURPOSE_CODE for samples taken at the sampling point. Purpose codepoints are established outside of INSPIRE (extensibility=any).
onLineResource	IF http: URI are used as inspireId , at least a self-referencing URI can be given eg: something like: < http://environment.data.gov.uk/water-quality/{orig-installation}/id/sampling-point/{smtp_user_reference} > Where <i>{orig-installation}</i> signifies the origination WIMS installation for sampling-point.
responsibleParty	One or more of: a) SMPT_RESP_ID resolved to a PARTY, ROLE, ORGANISATION and interval the role is held b) SP_RESP_AUTH_OFFICER resolved to a PARTY c) SP_OUTSIDE_OFFICER resolved to a PARTY d) SP_CONTACT_NAME and SP_ADDRESS_1-5 e) SMPT_REGION, SMPT_AREA_CODE, SMPT_SUB_AREA_CODE as proxy for an EA area/regional organisational unit. Formulated as a <i>RelatedParty</i> . (see below)
reportedTo	This association is intended to link a reporting obligation/requirement and its supporting legislation. Not represented in WIMS data.
measurementRegime	Coded value of either <i>periodicDataCollection</i> or <i>demandDrivenDataCollection</i> specifically: < http://inspire.ec.europa.eu/codeList/MeasurementRegimeValue/periodicDataCollection > or < http://inspire.ec.europa.eu/codeList/MeasurementRegimeValue/demandDrivenDataCollection > Likely to vary with SMPT_TYPE
Mobile	false
representativePoint	As <i>geometry</i> , a GM_Point derived from SMPT_GRID_REF or SMPT_NORTHING and SMPT_EASTING.

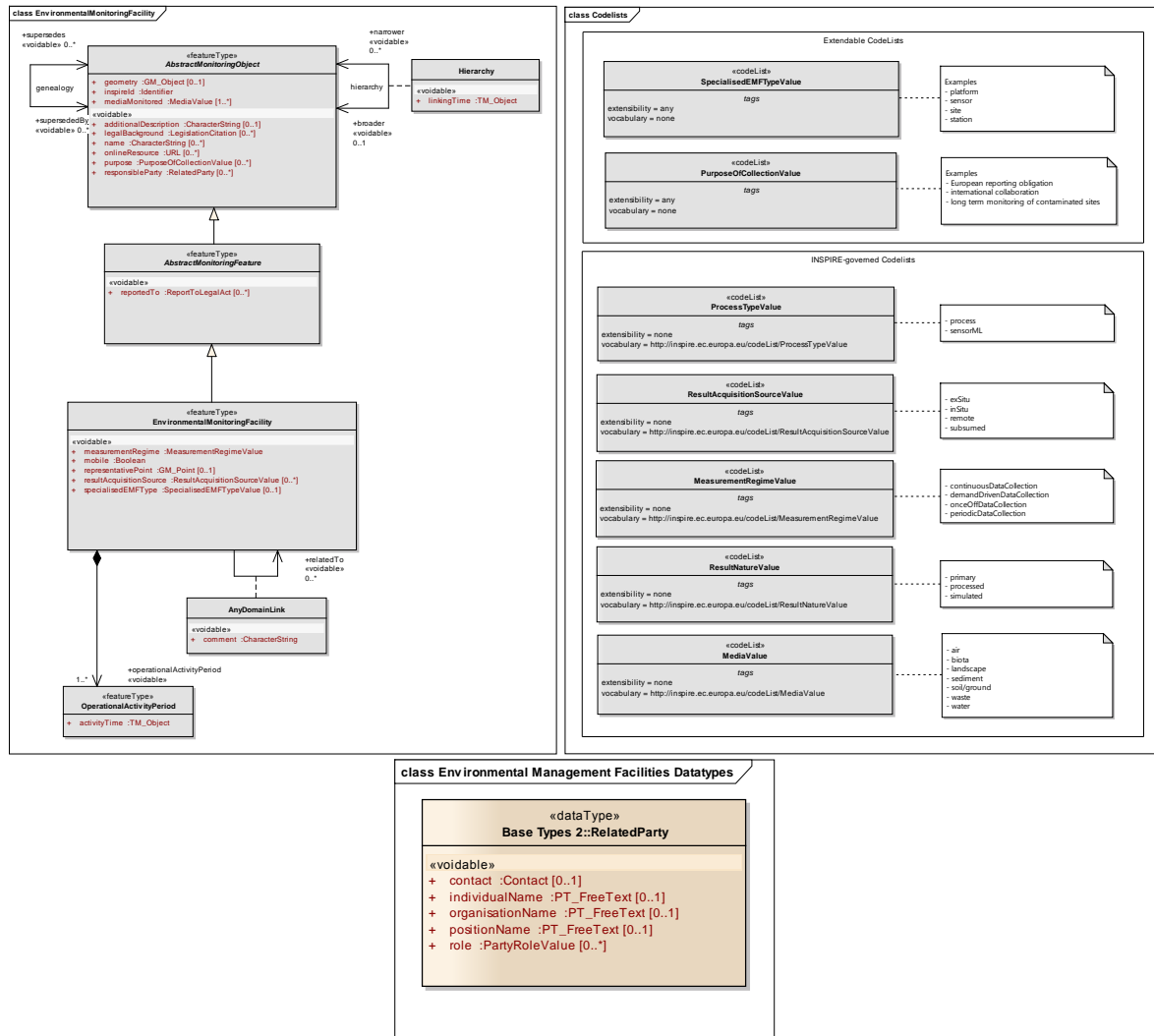
EnvironmentalFacility Attribute or association	WIMS Sampling Point Table: SAMPLING_POINTS
resultAcquisitionSource	Values are from a data provider managed codelist. Recommend values to include are 'ex-situ', 'in-situ', 'remote' and 'subsumed'. 'ex-situ' applies when a specimen of a SampledFeature (eg. water within a controlled water) is taken. 'in-situ' applies when the sampling feature and sampled feature are co-located. 'remote' applies when the sampling and sampled features are the same feature. 'subsumed' applies when the value reported is derived from child features. 'ex-situ' seems the most appropriate value for WIMS samples because samples are taken from the sampling-point for analysis elsewhere. Alternatively, the attribute can be omitted.
specialisedEMFType	A data provider managed code for the facility type. Seems appropriate to create a codelist based on SMPT_TYPE.
operationalActivityPeriod	Conveys the interval (ultimately a TM_Interval), possibly open, over which the EnvironmentalMonitoringFacility (aka. sampling point) has been up and running. Not directly represented in the SAMPLING_POINT table. SMPT_STATUS may indicate operational/no-operational. SAMPLING_POINT journal tables may contain dates from which operational intervals may be derived.

Several facets of the WIMS record of a sampling point are not expressed in this projection as an INSPIRE *EnvironmentalMonitoringFacility* e.g:

The *EnvironmentalMonitoringFacility* featureType can be specialised to incorporate WIMS information from fields expressed by the base featureType i.e.:

SMPT_STATUS, SMPT_HYDROLOGICAL_REF, SMPT_UPSTREAM, SMPT_CONTROLLED_WATER_ID, SMPT_USER_STAMP, SMPT_DATE_STAMP, SMPT_CONS_USER_REF, SMPT_FILING_REF, SMPT_LAST_SAMPLED, SMPT_HYDRO_REF, SMPT_HYDRO_DIST, SMPT_SIDE, SMPT_URN_ORIG, SMPT_AREA_CODE, SMPT_SUB_AREA_CODE, SMPT_CATCH_AREA, SMPT_COUNTRY, SMPT_DC_REF, SMPT_REFERENCE_DATUM, SMPT_SPPC_CODE, SMPT_SPS_CODE, SMPT_GEO_CODE, SP_REGION_ID, SP_ARCHIVE, SP_SHORT_DESC, SP_MAT_CODE, SP_SPT_STATUS, SP_RESP_AUTH_OFFICER, SP_LOCAL_GOV_T_AREA, SP_SAMPLE_COMP_REPORT, SP_OTHER_OFFICER, SP_OUTSIDE_USER, SP_CONTACT_NAME, SP_ADDRESS_1-5, SP_SAMPLER_DIRECTIONS, SP_COMPUTE_STATS, SP_LAB_REF_TYPE

However, many of these fields are unpopulated in the small amount of example data examined.



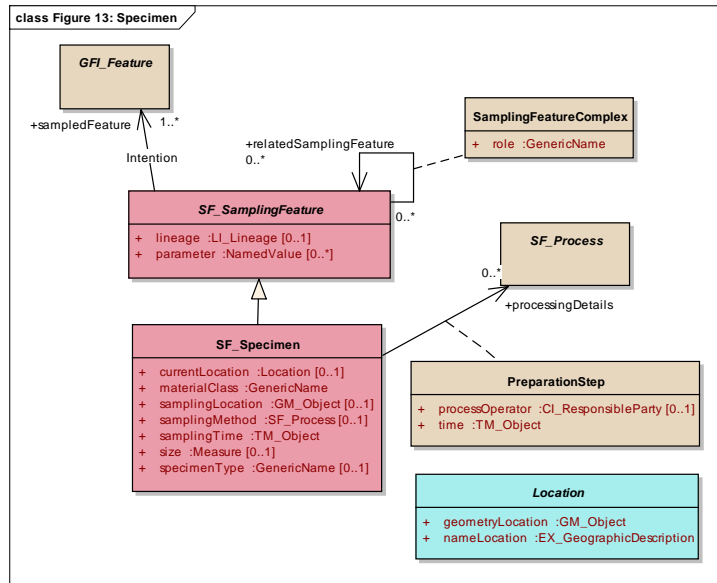
2.1.2 Environmental Facilities - Observations and Measurements

INSPIRE has adopted use of ISO 19156 Observations and Measurements for the reporting of observations and measurements, which includes the process of taking samples (*SF_Specimen* in ISO19156) and measurements (*OM_Measurement* in ISO 19156) taken directly on some feature of interest, or indirectly on a specimen taken at a feature of interest.

INSPIRE provides separate guidelines¹³ for the use of ISO19156 Observations and Measurements alongside other data specifications such as Environmental Monitoring Facilities.

In INSPIRE, WIMS samples can be cast as specimens, *SF_Specimen*, whose *samplingLocation* corresponds with the *representativeLocation* of the sampling point (as an *EnvironmentalMonitoringFacility*). Measurements become *OM_Measurements* which assess some *observedProperty* (a determinand) of *featureOfInterest*, in this case the sample (a *SF_Specimen*), which is presented as a *result*. The specimen/sample can also have a *sampledFeature* which in this case should be the controlled water being sampled. It may be possible to regard the sampling point (*EnvironmentalMonitoringFacility*) as a *relatedSamplingFeature* of the sample (*SF_Specimen*) in order to make a more definite link between the sample and the sampling point from which it was taken. Failing that a reference to the related *EnvironmentalMonitoringFacility* can be included in a specialised (subclassed) *SF_Sample*.

¹³ http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/D2.9_O&M_Guidelines_v2.0rc3.pdf

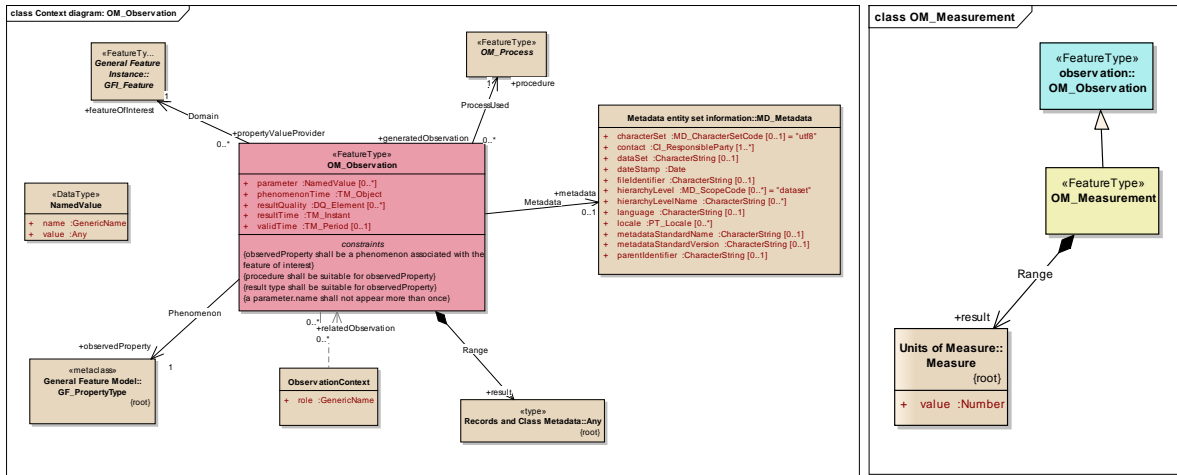


Casting a WIMS sample as an *SF_Specimen*, its fields and associations can be populated as follows:

<i>SF_Specimen</i> field or association	WIMS Sample Table: SAMPLES
inspireId	URI or other namespaced identifier derived from SAMP_ID
currentLocation	Omit: candidate data if available is SAMPLES:SAMP_SAMPLE_ADDRESS_1-4, however this is not a good fit with Location datatype which is a union of <i>GM_Object</i> (a geometry object) or <i>EX_GeographicDescription</i> which resolves to be an <i>MD_Identifier</i> which doesn't 'fit' the data in WIMS.
materialClass	A name from a controlled vocabulary, <i>GenericNames</i> are hierarchically namespaced names. This would require a set of namespaced names based of the materials listed in the SAMPLE_MATERIAL table referenced from SAMPLES::SAMP_MATERIAL
samplingLocation	Copy <i>geometry</i> or <i>representativePoint</i> from corresponding sampling point (as <i>EnvironmentalMonitoringFacility</i>) referenced by SAMP_SMPT_USER_REFERENCE
samplingMethod	SAMP_MECHANISM references an entry in SAMPLING_MECHANISMS which enumerates a number of sampling mechanisms. Typically, <i>SF_Process</i> which has no defined attributes or associations, may be subclassed to add details about the corresponding process. (including for example responsible parties).
samplingTime	Use SAMP_SAMPLE_DATE and SAMP_SAMPLE_TIME expressed as a TM_INSTANT with an iso8601 date and time fields.
size	Omit: Not present in WIMS data.
specimenType	A name from a controlled vocabulary. O&M spec. example offers "Polished section; core; pulp; solution." Either omit or reference EA, DEFRA or 3rd party maintained controlled vocabulary.

SF_Specimen can be specialised (sub-classed) to incorporate fields that include WIMS sample information not expressed in the base sampling feature i.e.:

SAMP_RESP_ID, SAMP_SOURCE, SAMP_LAB_REF_NO, SAMP_PURPOSE_CODE, SAMP_STATUS, SAMP_NOTES, SAMP_COMP, SAMP_RECEIPT, SAMP_AREA_CODE, SAMP_SUB_AREA_CODE, SAMP_CONFIDENTIAL, SAMP_ACTION_TAKEN, SAMP_ANALYSTS_COMMENTS, SAMP_CONFIRMATION_COMMENTS, SAMP_SCHEDULE_SYSTEM_ID, SAMP_SCHEDULE_RUN_NUMBER, SAMP_SCHEDULE_SAMPLE_ID, SAMP_SAMPLE_ADDRESS_1-4, SAMP_LAB_RECEIVED_DATE_TIME and SAMP_PROJECT_CODE.



Casting a WIMS Measurement as an *OM_Observation* or as an instance of its immediate sub-class, *OM_Measurement*, its fields and associations can be populated as follows:

<i>OM_Observation</i> or <i>OM_Measurement</i> attributes and associations	WIMS Measurement Table: MEASUREMENTS
<i>featureOfInterest</i>	The <i>SF_Specimen</i> that represents the water sample on which the measurement is conducted - referenced by MEAS_SAMPLE_ID
<i>observedProperty</i>	A determinand cast as a property the featureOfInterest referenced by MEAS_DETERMINAND_CODE. [I'm clarifying how this works 'formally' because we're not making the determinand a attribute or association of a specimen/sample (skw)]
<i>result</i>	An <i>OM_Measure</i> whose <i>value</i> corresponds to MEAS_RESULT and whose <i>uom</i> (unit-of-measure) corresponds with the unit-of-measure associated with the observedProperty/determinand.
<i>processUsed</i>	An <i>OM_Process</i> derived from the analysis method code carried in MEASUREMENTS::MEAS_ANAL_METH_CODE. The definition of <i>OM_Process</i> is an abstract class intended to be specialised to provide a description and any other relevant process details.
<i>phenomenonTime</i>	The time (as a <i>TM_Instant</i>) at which the observed property applies to the feature of interest. Candidate values are the date/time at which the corresponding sample was taken SAMPLES::SAMP_SAMPLE_DATE, SAMPLES::SAMP_SAMPLE_TIME or SAMPLES::SAMP_DATE_STAMP; or MEAS_DATE_STAMP or MEAS_ANALYSIS_DATETIME
<i>resultTime</i>	The time (as a <i>TM_Instant</i>) at which the measurement result became available: Candidate values are MEAS_DATE_STAMP or MEAS_ANALYSIS_DATETIME

OM_Measurement can be specialised to include information from other WIMS MEASUREMENTS table fields eg:

MEAS_SIGN, MEAS_LIMITS, MEAS_TEXT_RESULT, MEAS_RESULT_STATUS, MEAS_LAB_COMMENT, MEAS_TO_BE_USED_IN_STATISTICS, MEAS_USED_IN_STATISTICS, MEAS_EXCEEDED_COMPLIANCE, MEAS_LAB_ID, MEAS_PRIORITY.

2.2 Discharge Agreement/Consents

This section indicates how it may be possible to represent discharge agreements using the featureTypes and dataTypes provided by INSPIRE. However, the expression of the WIMS held detail of a permit is complex - and will bear some experiment.

Within WIMS the principle data that relates to discharge agreements or consents cover

- Discharge Sites
- Discharge Outlets
- Consents/Agreements
 - Permitted Effluent flows (at outlets)
 - Determinand Levels at sampling points
 - Other coded conditions

Discharge sites and outlets align can be aligned with either:

- *ProductionSite*, *ProductionFacility*, *ProductionInstallation*, *ProductionInstallationPart* hierarchy from "Production and Industrial Facilities"(PF), or;
- *EnvironmentalManagementFacility* (e.g. a sewage treatment works) from "Utility and Governmental Services" (US); or
- *Holding*, *Site*, *WaterManagementInstallation* and *InstallationPart* from "Aquacultural and Agricultural Facilities" (AF).

ProductionFacility, *EnvironmentalManagementFacility* and *Holding* all derive from a base class of *ActivityComplex*. Unfortunately their respective specialisations diverge beneath *ActivityComplex*. However, both *EnvironmentalManagementFacility* (US) and *Site* (AF) have a *permission* field for associating a *Permission* (a datatype associated with the *ActivityComplex* featureType).

The information in the WIMS DSI_TYPE field of the DISCHARGE_SITES table may provide a basis on which to distinguish different derivatives of *ActivityComplex*. For the purposes of this report, we will use *EnvironmentalManagementFacility* to serve as a base featureType for both Discharge Sites and Discharge Outlets.

Unfortunately INSPIRE's *ActivityComplex::Permission* data type does not have sufficient expressive power to fully express the permit restrictions represented by the WIMS data. In general, INSPIRE users/publishers, e.g. national regulators may extend INSPIRE application schema (data-specifications) to meet their local needs. Generally, this will involve importing the relevant INSPIRE application schema and subclassing feature and data types to address the local requirements. Extending INSPIRE application schema is discussed in Appendix F of the INSPIRE Generic Conceptual Model¹⁴.

A WIMS discharge consent:

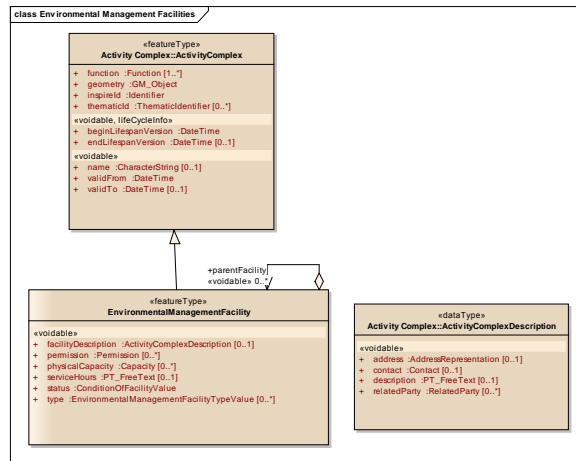
- is associated with a discharge site and a responsible consent holder
- restricts multiple effluent flow at one or more outlets associated with the consent.
- specifies permitted levels of determinands associated with each effluent. These can be specified to vary by month-of-year.
- specifies zero or more standard conditions associated with each effluent.

An INSPIRE *ActivityComplex::Permission* cannot carry this level of detail without specialisation. However, it can be used to convey the existence of the permit, its related parties and the interval over which the permit applies. INSPIRE seeks to attribute discharges to categories of economic activity¹⁵ - which may be determinable from data held about the permit, its outlets and effluents. The *permittedCapacity* fields of a *Permission* can be used to express permitted effluent levels at outlets, although the connection with a physical outlet cannot be represented without further specialisation of *ActivityComplex::Capacity*. However, it is challenging to express the permitted determinand levels associated with each effluent using *ActivityComplex::Permission*.

¹⁴ http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/D2.5_v3.4rc3.pdf

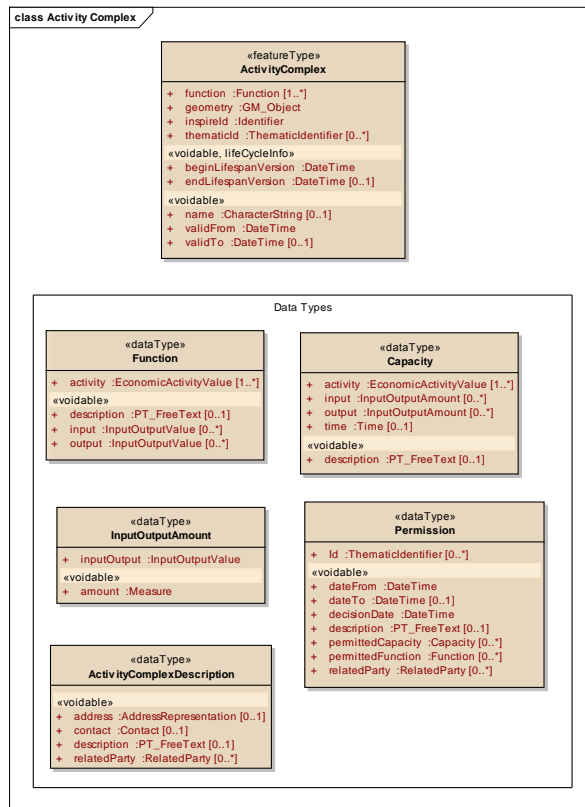
¹⁵ EU INSPIRE has preference for the use of NACE codes:

http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-07-015/EN/KS-RA-07-015-EN.PDF



Given the above, framing a discharge site as an *EnvironmentalManagementFacility*, its fields can be populated as follows:

Discharge Sites as EnvironmentalManagementFacility: attributes and associations	WIMS Discharge Sites Table: DISCHARGE_SITES
function.activity	The economic activity associated with the facility may be indicated by the entries DSI_TYPE. This may also indicate a more specialised subclass of either <i>EnvironmentalManagementFacility</i> or <i>ActivityComplex</i> . The EU preference is for the use of NACE activity codes see http://ec.europa.eu/competition/mergers/cases/index/nace_all.html http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-07-015/EN/KS-RA-07-015-EN.PDF
Geometry	Either a more detailed GM_Object (say a polygon) indicating a site outline (if available outside of WIMS) or a GM_Point object derived from the sites national grid reference, DSI_NGR to serve as representative point.
inspireId	Within WIMS there is a no managed identifier for a discharge site. The 'nearest' thing to an identifier for the site is its' national grid reference DIS_NGR. DIS_NGR is also the field used to form a link between a discharge site and any discharge agreements/consents applicable to that site. Formulate inspireId as URI from a base namespace URI assigned to CLEAR Info for spatial-object identifiers combined http://clearinfo.environment-agency.org.uk/discharge-site/ST6790099200 where http://clearinfo.environment-agency.org.uk/discharge-site/ serves as <i>inspireId.namespaceId</i> , and ST6790099200 serves as <i>inspireId.localId</i> .
thematicId	Either omit, or reuse the <i>localId</i> part of the inspireId.
Name	Derived from DSI_LNAME
facilityDescription.address	Formulated from DSI_ADD1 to DSI_ADD4, DSI_POST_CODE and DSI_COUNTRY.
Permission	Zero or more Permission values covering each of the discharge outlets, their effluent flow and permitted determinand levels - more below.
Status	??
Type	A coded value from http://inspire.ec.europa.eu/codelists/WaterManagementFacilityTypeValue which may be derivable from DSI_TYPE.



Within INSPIRE permissions (consents, permits etc.) are modelled by the datatype, *ActivityComplex::Permission*. As discussed above, it is not possible to represent the full detail of a WIMS discharge consent using this structure as is.

ActivityComplex::Permission attributes	Source WIMS Data
Id	An agreement identifier derived from AGREEMENTS::AGR_APL_NUMBER and AGREEMENTS::AGR_VERSION
dateFrom	The effective start date of the consent derived from AGREEMENTS:AGR_EFFECTIVE.
dateTo	The date that the consent was revoked from AGREEMENTS::AGR_REVOCATION
decisionDate	The date that the consent was issued derived from AGREEMENTS::AGR_ISSUED
description	Omit
permittedCapacity	Should contain a Capacity for each outlet associated with the permitted site. The capacity should be formulate as indicated in the following table.
permittedFunction	Omit

Each outlet associated with a site gives rise to a *Capacity* value for *Permission.permittedCapacity* in the table above. The Capacity for a given outlet may be populated as follows:

Activity::Capacity attributes	WIMS Data
Activity	The Economic activity associated with the Capacity. In the absence of more fine grained detail we suggest repeating the activity associated with the site. <i>EnvironmentalManagementFacility.function.activity</i> see above.

Activity::Capacity attributes	WIMS Data
Inputs	Omit: WIMS does not carry data on inputs. The treatment code, EFF_TMEN_CODE may be indicative of some input to the effluent, but amounts are not indicated.
outputs	<p>For each Effluent flow associated with an outlet provide an <i>InputOutputAmount</i>.</p> <p><i>InputOutputAmount.inputOutput</i> requires as coded value of type <i>InputOutputValue</i>¹⁶⁾ to indicate the nature of the discharge.</p> <p><i>InputOutputAmount.amount</i> requires an expression of the permitted amount of discharge. EFF_MAX_DAILY is probably the most appropriate value, however max rates, mean rates are also recorded in WIMS.</p>

A conceivably different approach would be to provide a *permittedCapacity* for each outlet at a site and for the *outputs* of that *Capacity* to be enumerated over the determinand specifics for all of the effluents associated with the outlet. This assumes that the code-list of determinands can be formed such that its members are all instances of an *InputOutputValue*. This may require a parallel hierarchy to the coding of determinand as properties for the expression of WIMS Measurements as *OM_Measure* or *OM_Observation* (see previous section 2.1.2)

Note that there is no scope to associate a *permittedCapacity* with a particular outlet. If this level of detail is required, discharge outlets could be instantiated as a subordinate *EnvironmentalManagementFacility* of the discharge site, and *Permission* used to express the detail of the restrictions at that outlet. However, this would challenge the notion of identity associated with a consent. The *permission.id* could, perhaps, be seen as a reference to the overarching consent rather than the more fine grained permission. Alternatively, *permission.id* could be derived from the WIMS consent number and the outlet number (which is relative to the consent in any case).

2.3 Utility and Governmental Services

The 'Utility' part of this data specification is primarily focussed on the representation of utility networks and is not considered further in this report since WIMS does not contain information about Utility networks.

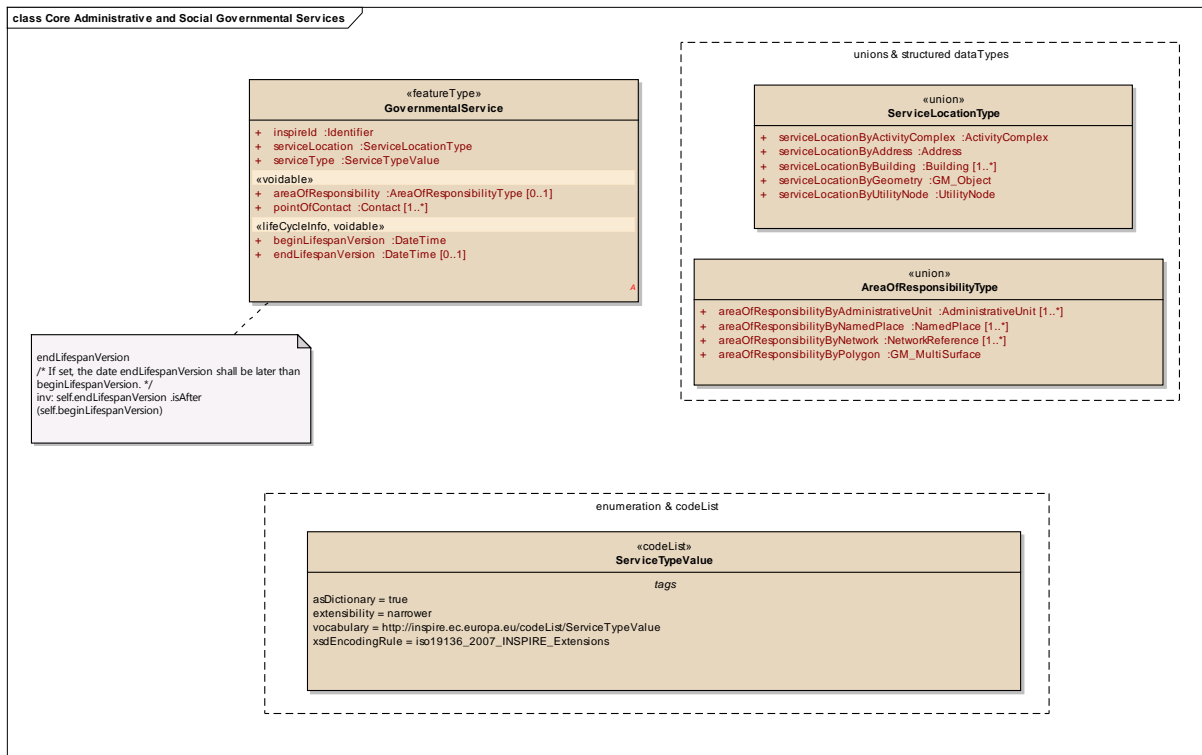
Utility and Governmental Services also includes Environmental Management Facilities which were discussed in the previous section, 2.2 which shows how WIMS Discharge site may be cast as instances of Utility and Governmental Services *EnvironmentalManagementFacility* featureType

Finally, Utility and Governmental Services provides a means to express the provision of public services, the geographic area in which service is provided, the location and contact details of the service providing agency. It seems plausible to use the *GovernmentService* featureType to capture the notion of Environment Agency Regions, Areas and Sub-Areas particularly if geometries can be associated with each of these entities.

WIMS carries very little detailed information about EA regions, areas and sub-areas beyond a code and a narrative description. Ideally, a common coding of all these entities should exist across the agency and indeed be available to anyone making reference to the corresponding entities. Several systems within the Agency make reference to the agencies organisational structure and to the bounded or thematic areas that those organisational entities are responsible for. Maintaining common reference data such as this, as a shared assess across multiple datasets will: improve the consistency of the data from multiple sources; reduce maintenance burden by maintaining the information is on place and by avoding the need to maintain 'cross-walks' between different information systems that use different means to refer to the same thing.

¹⁶ The Activity Complex data specification includes this note on InputOutputValue

"NOTE Depending on the thematic scope it can contain different values including terms as Registered Pollutants, Waste, Processed Products, leakage, etc."

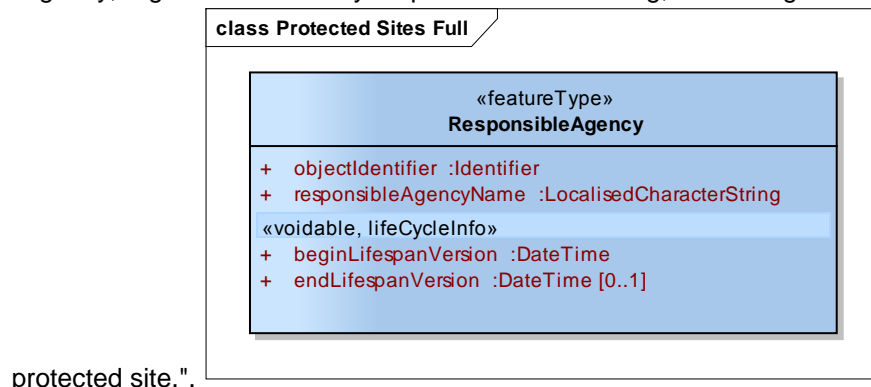


2.4 People and Organisations

The CLEAR Info project currently links WIMS sites to the names of the ParentCompanies responsible for the site. Given the potential ambiguities of names it is desirable to assign unique identifiers to identify corporate entities or responsible parties.

The INSPIRE data specifications do not appear to provide a *featureType* that can be used for creating objects that represent organisations. The closest *featureType* found in the consolidated data specifications UML Model¹⁷ is *ResponsibleAgency* within the Protected Sites data specification¹⁸. However, *ResponsibleAgency*, as defined carries little more than the name of

"The agency, organisation or body responsible for selecting, describing and designating the



It may be necessary for CLEAR Info, in extending any INSPIRE data specifications used to publish INSPIRE compliant data, to also include a *featureType* for organisational entities and organisational structure so that common references (associations) can be made to the same organisational entity.

¹⁷ <http://inspire.jrc.ec.europa.eu/data-models/>

¹⁸ http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_PS_v3.1.pdf

The INSPIRE Generic Conceptual Model¹⁹ defines a number of base data types, including a data type for describing *RelatedParty*. Formulated as a datatype this does not make a related party and object in its own right. It is more like a 'structured' literal, repeated on each occasion that it is used, rather than an object that can be the target of a shared reference.

Several INSPIRE application schema make use of the *RelatedParty*, base type, in particular:

- *AbstractMonitoringFacility.responsibleParty*
(EF - a superclass of *EnvironmentalMonitoringFacility*)
- *EnvironmentalManagementFacility.facilityDescription.relatedParty*
- *Permission.relatedParty*

All use the *RelatedParty* base datatype to provide information about parties that are responsible for, or related to, a particular entity.

Elsewhere, amongst the ISO 19xxx foundation schema, the schema for ISO 19115 includes the datatype *CI_ResponsibleParty*. However, while it is similar in nature to *RelatedParty*, is typically used in citation metadata about a dataset or data item to express attribution and role with respect to the publication (as opposed to, say an action or event described by the publication such as the taking of a sample or the making of a measurement).

The INSPIRE base data types and thematically defined feature and data types are weak for expressing the organisational relationships that CLEAR Info seeks to express. While the base datatypes *RelatedParty* and *Contact* provide means to express information about a related party, they do not give means to express a shared identifier which can serve as the basis of common reference. Conversely, the definition of the featureType *ResponsibleAgency* seems to intended for identifying the Agency responsible for the designation of a site as having some protected status, as opposed to the operator of the site. In particular it would be difficult to use *RelatedParty* or *ResponsibleAgency* as defined by INSPIRE as the basis a 'catalog' of organisational entities and their inter-relationships. This may be deliberate on the part of INSPIRE so as to avoid taking on a world modelling problem

Elsewhere, outside of INSPIRE the "W3C Organization Ontology"²⁰ is well suited to expressing organisational structure and role holdings that vary with time.

Other activity elsewhere in the EU has led to the creation of the "iSA Core Person Vocabulary"²¹ which can be used to express information about individual people and legal entities such as organisations. However, it is less expressive than the W3C Organization Ontology in that it does not provide for the expression of role holding relationships between people and organizations or for expressing organizational sub-structure.

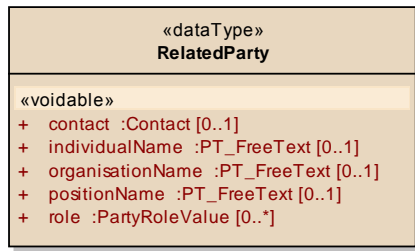
Either could be taken as a basis for defining an application schema for CLEAR Info to cover party (people and organization) related aspects.

¹⁹ http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/D2.5_v3.4rc3.pdf

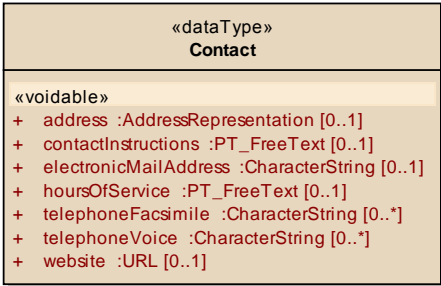
²⁰ <http://www.w3.org/TR/vocab-org/>

²¹ https://joinup.ec.europa.eu/asset/core_person/home

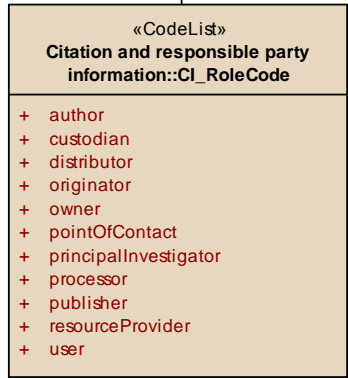
class RelatedParty

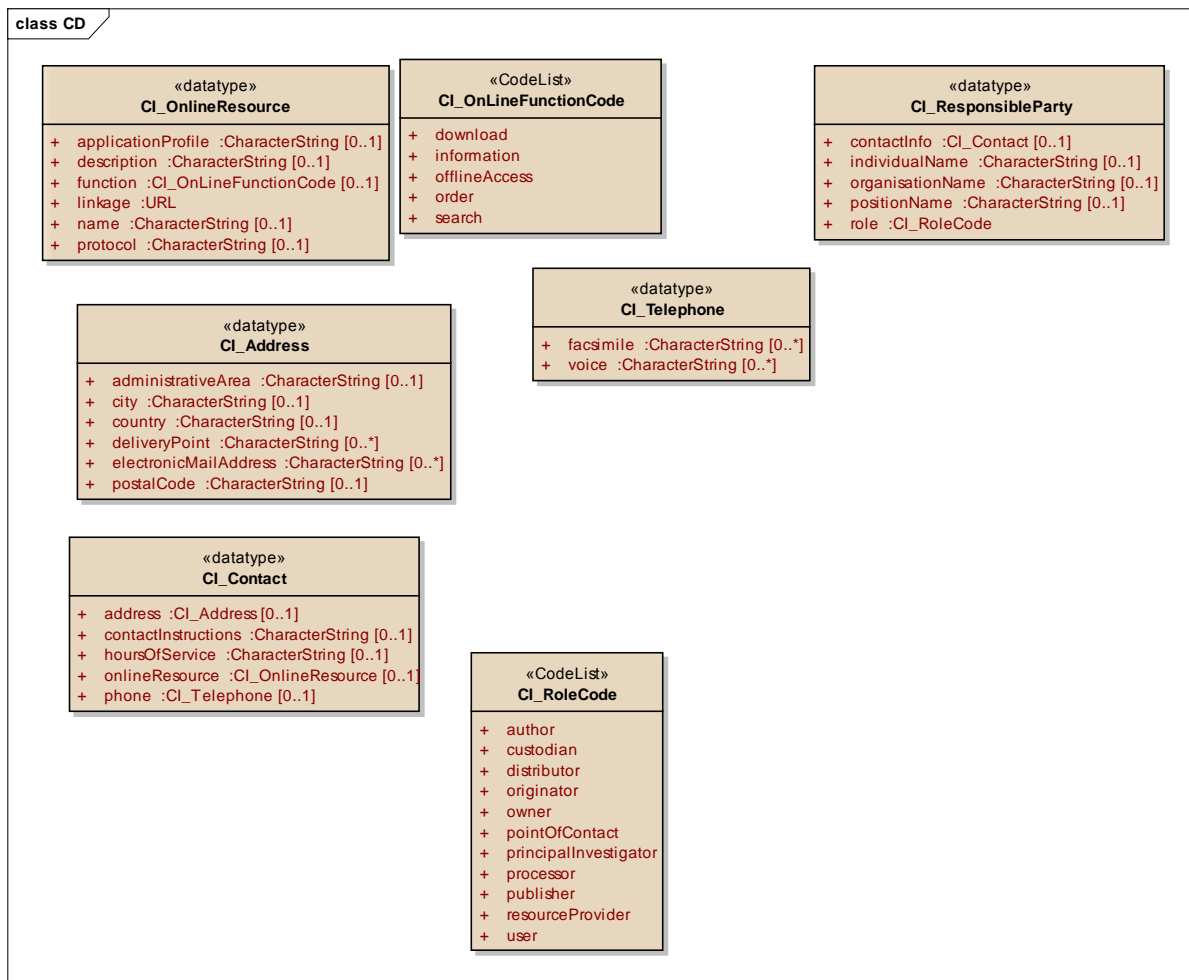


individual, organisation or position name shall be provided
 /*At least the individual, organisation or position name shall be provided.*/
 inv: individualName->notEmpty() or organisationName->notEmpty() or positionName->notEmpty()



The allowed values for the PartyRoleValue code list comprise the values of the following code lists or other code lists specified by data providers:
 - Role Code (CI_RoleCode): Functions performed by a responsible party (EN ISO 19115:2005/AC:2008).
 - Role of a Related Party (RelatedPartyRoleValue): Classification of related party roles.





3 Transforming Data to INSPIRE formats

INSPIRE is centred on the publication and access to spatial datasets, principally by public-sector bodies. At a technical level it provides a series of thematically organised data specifications and specifies a number of network services for discovering, accessing, viewing and downloading data.

The INSPIRE data specification is built on a foundation of OGC GML 3.2.1²² and ISO 19xxx series specifications that it references together with the INSPIRE Generic Conceptual Model²³.

INSPIRE network services are based on:

INSPIRE Service	OGC Specifications	Output Format
Discovery	OGC CSW 2.0.2 (Catalog Service Web) ²⁴ OGC CSW 2.0.2 ISO Metadata Application Profile	ISO 19139 XML encoding of profile of the ISO 19115/19118 metadata standards.
View	OGC WMS 1.3 (Web Mapping Service) ²⁵ Also ISO 19128 Or alternatively OGC WMTS 1.0.0 (Web map Tile Service) ²⁶	MapTile images in various formats

²² http://portal.opengeospatial.org/files/?artifact_id=20509

²³ http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/D2.5_v3.4rc3.pdf

²⁴ <http://www.opengeospatial.org/standards/cat>

²⁵ <http://www.opengeospatial.org/standards/wms>

²⁶ <http://www.opengeospatial.org/standards/wmts>

INSPIRE Service	OGC Specifications	Output Format
Download	OGC WFS 2.0 (Web Feature Service) ²⁷ Also ISO 19142 May be extended in future to cover OGC WCS (Web Coverage Service) ²⁸	GML 3.2.1

The INSPIRE data specifications are captured as UML based conceptual schema. The INSPIRE drafting teams use "Enterprise Architect" by Sparx Systems²⁹ as the tool for capturing the models. Current versions of both the Enterprise Architect source files and the derived XML Schema are available from: <http://inspire.jrc.ec.europa.eu/data-models/>.

3.1 Specialising INSPIRE application schema

Specialising INSPIRE applications schema involves deriving subclasses of INSPIRE *featureTypes*, *dataTypes* and *propertyTypes*. This process adds attributes or associations to existing base classes (*featureTypes*, *propertyTypes*, *dataTypes*).

Appendix F of the INSPIRE Generic Conceptual Model provides an example of how an INSPIRE application-schema (data-specification) can be extended³⁰

Specialisation can occur at two levels. Either at the conceptual schema level - in which case "Enterprise Architect" is likely to be the tool of choice; or by the creation of GML application schema expressed as an XML schema document which typically are derived by the application of rules in the GML specification to the UML conceptual model (application schema). Deployment tooling typically uses (or generates) the XML schema form of GML application-schema. At the time of writing, it is not known where Sparx EA is used to generate the reference INSPIRE XML schema documents or whether some other tooling is used to generate them a representation maintained using EA. This has a bearing on whether/how EA can be used to generate GML/XML application schema from UML application schema that are specialisations of the INSPIRE consolidated model.

3.2 Publishing INSPIRE compliant data

When building a new INSPIRE information system from 'scratch' tooling such as Snowflake GO Loader³¹ can be used to create database table schema corresponding to the feature and property types within the supported application-schema (generally GML application schema of which the INSPIRE application schema are a particular concrete set). Tooling like GO Loader can be used to then populate a database from imported GML conforming to the supported application schema. This database is then used to present INSPIRE view (WMS/WMTS) and download (WFS/WCS) network services.

However, in the situation where data is already held in a database and the requirement is to present that data in an INSPIRE compliant form, the task is to map from the database schema to the appropriate INSPIRE application schema (or derivatives thereof if specialisation was needed i.e. if subclassing of feature, property and data types, is needed to adequately express the available data). This is similar to the manual mapping exercise described in the preceding sections. At scale some tooling and automation is required.

Snowflake GO Publisher³² can be used to perform this task. It provides an environment in which feature and property type attributes and associations can be populated from database fields. Computations can also be done to derived output elements from multiple fields. FME, available in the UK from 1Spatial³³ can also be used to perform similar transformations.

²⁷ <http://www.opengeospatial.org/standards/wfs>

²⁸ <http://www.opengeospatial.org/standards/wcs>

²⁹ <http://www.sparxsystems.com.au/>

³⁰ http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/D2.5_v3.4rc3.pdf

³¹ <http://www.snowflakesoftware.com/products/goloader/>

³² <http://www.snowflakesoftware.com/markets/inspire/create-inspire-data/>

³³ <http://www.1spatial.co.uk/products-services/fme/>

One other question that needs to be addressed is whether the INSPIRE publication should be by live access to the applications database (CLEAR Info in this case) or whether one or more INSPIRE compliant datasets are exported from CLEAR Info and then installed within a server, such as GeoServer³⁴, that implements the OGC WMS, WFS and WCS services.

In the latter case a normal ETL like process is performed which produces an extracted copy of a dataset in INSPIRE compliant form. That data set can be made available as is, as a large web download or it can be loaded into a server, such as GeoServer, an open-source platform that implements the core WMS, WMTS, WFS and WCS services that underpin the INSPIRE network services.

For direct access to a WFS without having to extract, transform and load, Snowflake offer a product called GO Publisher WFS³⁵ which introduces a mapping and WFS service layer between the 'network' and an existing database. It is not clear whether this offering extends to supporting a WCS which would be useful in presenting both the water quality samples and measurements archive and the results of the aggregation processes.

One other thing that should be borne in mind is that INSPIRE is largely about spatial datasets. Most of the spatial information available in WIMS is point positions. To be really useful INSPIRE spatial-objects usually include some non-trivial geometry component. Clearly this could exist for entities such as EA regions, areas and sub areas; hydrological catchments; controlled waters; water courses; and discharge sites. If geometries are available and can be merged into CLEAR Info they would provide much more useful spatial-objects.

[Note: these comments are based solely on information available on the relevant web sites. No product evaluations have been conducted, neither does the author have experience of using any of the products mentioned - other than some experimental use of GeoServer for a previous project.]

3.3 Method and process

Modelling:

1. Identify best fit candidate INSPIRE featureTypes for the entities represented by the data being published.
2. If necessary create a 'local' application schema that imports the appropriate INSPIRE application schema and extends INSPIRE feature, property or data types to extend them with the attributes and associations necessary to express the available data. This is a balance between using INSPIRE as is, and limiting the expression to what the INSPIRE schema can accommodate and using INSPIRE as a base and extending it to give fuller expression to the data being published.
3. Generate GML/XML application-schema for the 'local' application schema.

Transformation/Publishing ETL

This will requires the use of a tool such as FME or Snowflake GO Publisher (there may be others) to generate a mapping from the available data base tables and fields to instances of feature, property and data types with their attributes and associations as specified by the application schema in use.

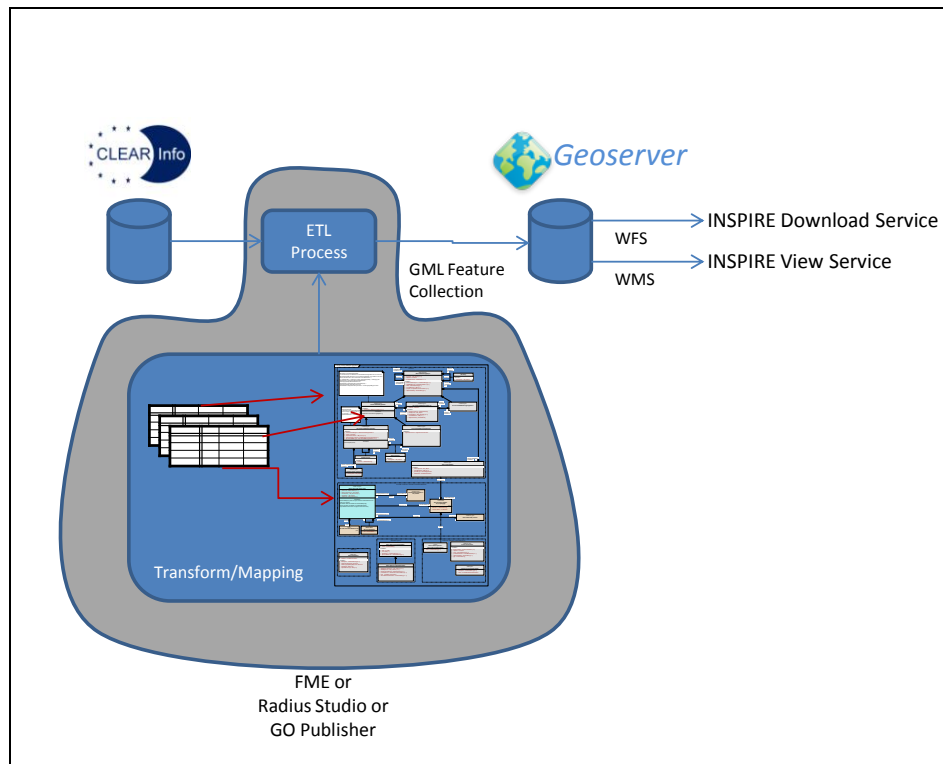
Once constructed this mapping is used to export and transform data from the database into, typically, a FeatureCollection (typically either a GML FeatureCollection or a WFS FeatureCollection) which contains the transformed features and can itself be thought of as a spatial-dataset.

The resulting GML can be made available for simple web based download, however given its likely size and the potential for users to have a more focussed interest in the data, it can also

³⁴ <http://geoserver.org>

³⁵ <http://www.snowflakesoftware.com/products/gopublisher/wfs/>

be loaded into a server that supports the INSPIRE view(WMS) and download (WFS/WCS) services. GeoServer is a free, open-source platform that could serve in this role.



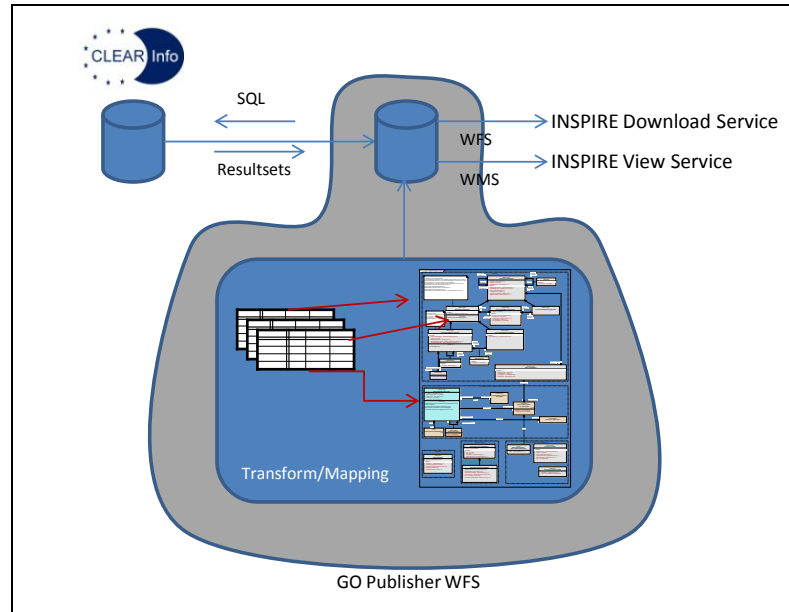
Transformation/Publishing - Direct Access

Again, typically, a tool is required to map from database tables and fields into the instances of feature, property and datatypes. The difference here is that rather than extract, transform and load the data from one system into another, a service interface (as WFS interface) is build which turns WFS request into database queries (SQL) and then formats the database resultsets using the transformation to deliver the resulting FeatureCollections.

The advantage of this approach is that it completely avoids the extract and load phases of the ETL process. It effective performs an on-demand ETL in response to service requests. A possible disadvantage is that it creates a query load on the 'main' database cluster.






At the time of writing the only product found that supports this functionality is "GO Publisher WFS". It may be possible to manually create database mappings for GeoServer but that is likely to be a cumbersome process. Once a WFS is available it can be used as a basis for providing a WMS that renders features from the WFS onto a base map - and then renders map tiles for use in, say, an OpenLayers driven web application front-end.

It may be possible to use Geoserver to directly map the relevant CLEAR Info database tables in a similar fashion to GO Publisher WFS, although the database to application schema mapping task is likely to require more effort because good visual tooling is probably unavailable.



4 Annex I - WIMS Data Model and Data Sample

Our investigation of the WIMS data model is based on the attached diagram (PDF files) and subsequent conversations with Mark Selwood and Dave Sampson at EA.

		
WIMS_ERD.PDF	wims.pdf	WIMS MIDAS Table Relationships1.pdf
WIMS Schema	MIDAS ERD	WIMS-MIDAS Table Relationships
		
WIMS_Example_Data.xls	WIMS_Tables_13-09-02.xls	
WIMS Sample Data	Additional WIMS Sample Data	

4.1 Out of Scope Data

The first page of the WIMS Schema diagram details a number of tables that are used to produce formal letters, typically in response to applications or non-compliance situations revealed by measurements made on samples. These tables are not present in the MIDAS ERD and are deemed out of scope for this report.

The other main difference between the MIDAS ERD and the WIMS Schema is that the latter includes a number of journaling tables (*_JNL) used for tracking changes. The corresponding 'main' tables represent the current state of the corresponding entities.

4.2 In Scope Data

The main tables in the WIMS schema are listed below.

- Journal a journal table or tables used to maintain some form of change history
- Spatial entities have a spatial/geographic context
- Admin Org entities reflect aspects of organisational structure (specifically EA organisational structure)
- Admin Proc entities reflecting procedural aspects associated with the processing of an application or actions arising from a non-compliant sample.
- Mon entities directly involved in the monitoring of water quality.
- Reference Data entities that are thought to be slowly changing and potentially relevant to multiple data sets and or shared between multiple WIMS installations.
- INSPIRE Data Spec indicates any relevant INSPIRE Data specification for the corresponding entities.

Entity Type	Journal	Spatial	Admin Org	Admin Proc	Mon	Ref Data?	INSPIRE Data Spec(s)	Comment
Samples	Y				Y	N	<p>"Environmental Monitoring Facilities" via D2.9_O&M_Guidelines</p> <p><i>OM_Observation</i></p> <p>The act of taking a water sample can be thought of as making an observation. The sample is representative of the water at some location at some instant (or some interval)</p> <p>c.f. Link Data: <i>qb:Observation</i> <i>qudt:Unit, qudt:QuantityValue etc.</i></p>	Records about samples of materials taken at sampling points on given date/time,, for some purposed relate to a responsibility associated with an agreement (ie. permission)
Measurements	Y				Y	N	<p>"Environmental Monitoring Facilities" via D2.9_O&M_Guidelines</p> <p><i>OM_Observation</i> <i>OM_Measurement</i></p> <p>The application of some method to assess some determinand wrt to a particular sample can be thought of as making an observation on a sample (which stands proxy for water at some location at some time or interval as above).</p> <p>c.f. Link Data <i>qb:Observation</i> <i>qudt:Unit, qudt:QuantityValue etc.</i></p> <p>also</p> <p>ICM data cubes <i>classificationItem and classificationValue.</i></p>	A measurement of some determinand wrt to some sample.

Entity Type	Journal	Spatial	Admin Org	Admin Proc	Mon	Ref Data?	INSPIRE Data Spec(s)	Comment
Sampling Points	Y	Y			Y	Y	"Environmental Monitoring Facilities" <i>EnvironmentalMonitoringFacility</i> <i>EnvironmentalMonitoringNetwork</i> e.g. sampling points (stations) for similar purpose/intent eg. bathing water, waste water treatment etc. could be considered members of 'networks'	Sampling points of various kinds where samples are taken.
Determinands					Y	Y	D2.9_O&M_Guidelines <i>observedProperty</i> <i>GFL_PropertyType</i> c.f. ICM data cubes <i>classificationItem</i> and <i>classificationValue</i> .	Entails a substance being monitored, units of measurement in use plus indications about whether measurement values are computed (e.g. aggregated from other measurements) and maximum time required to make a determination.
Applications				Y			?? doesn't appear to be considered by INSPIRE	Record structure of the permissible levels of discharge of a given determinand allowed at a given outlet under a permit. Includes provision for levels of vary on a monthly basis (recurring annually)
Agreements/ Permissions	Y	N		Y	?		INSPIRE base models <i>ActivityComplex::Permission</i>	A variant of effluent specifics (monthly varying, allowable discharge levels of a given determinand at a given outlet) at water treatment plants (typically) working under the Urban Waste Water Treatment directive.
Discharge Sites	Y	Y		Y	?	N	"Agricultural and Aquacultural Facilities": eg. <i>Holding</i> <i>Site</i>	Discharge sites associated with an application or an agreement(permission). Broad area that has outlets where particular effluent discharges are permitted
Outlets	Y	Y			?	?	"Production and Industrial Facilities" <i>ProductionSite</i> <i>ProductionFacility</i> <i>ProductionInstallation</i> "Utility and Government Services" <i>EnvironmentalManagementFacility</i> (cf. sewage works) Generally <i>ActivityComplex::ActivityComplex</i> from INSPIRE base models	The points within a discharge site where effluent enter the river (or ground water?) system.
Responsibilities	Y		Y	Y	?		FoundationSchema: ISO:19115:2003 <i>Cl_ResponsibleParty</i>	A role held by a party on behalf of a organisation from some interval of time.
Parties (inc. Staff)			Y			?	c.f. <i>Linked Data:</i> <i>foaf:Agent</i> or <i>org:Agent</i> and <i>org:Role</i>	
Organisations			Y	Y		?	FoundationSchema: ISO:19115:2003 <i>Cl_ResponsibleParty</i> c.f. <i>Linked Data:</i> <i>foaf:Agent</i> or <i>org:Agent</i> and <i>org:Role</i>	
Holders	Y		Y	Y				N-M relation between Agreements and responsibilities (held by parties on behalf of organisations)
Regions		Y	Y			Y	"Utility and Government Services"	Organisational partitions with responsibility for given geographic region.
Areas		Y	Y			Y	<i>GovernmentalService</i>	Spatial and organisational sub-division of a region.

Entity Type	Journal	Spatial	Admin Org	Admin Proc	Mon	Ref Data?	INSPIRE Data Spec(s)	Comment
Sub-areas		Y	Y			Y	<i>AreaOfResponsibility(byPolygon)</i> Potentially used in the role of a <i>ReportingUnit</i> in some contexts.	Spatial and organisational sub-division of an area
Catchments (Hydrological)	Y	Y	Y			Y	"Hydrography - Physical" <i>DrainageBasin ?</i>	The spatial extent over which water enter into or is captured by a controlled water.
Controlled Waters	Y	Y				Y	"Hydrography, TransportNetworks, Area Management" various classes eg: <i>SurfaceWater</i> <i>WaterCourse</i> <i>River</i> <i>WFDLake</i> <i>WFDRiverOrLake</i> <i>WFDSurfaceWater</i>	Some fragment of the river network with which sampling points, outlets and discharge sites can be associated.

4.2.1 CodeLists and Controlled Vocabularies

Embedded within the columns of many tables are references that suggest the presence of controlled vocabularies or codelists (visible in UIs as menu-pick lists). However, the detail of these lists is not evident from the schema alone.

Examples include:

Column	Table(s)	Comment
RESP_ROLE_CODE	Responsibilities	Indicates the role held by a party for some interval
AGR_FULL_STATUS	Agreement	??
AGR_CHARGE_STATUS	Agreement	??
AGR_REVIEW_CODE	Agreement	??
AGR_HISTORIC_STATUS	Agreement	??
AGR_PERMIT_TYPE	Agreement	??
AGR_TYPE	Agreement	??
DETE_CODE	Determinands	
DETE_UNIT	Determinands	
EFF_SPT_CODE	Effluents	Sampling Point Type Code
EFF_TMEN_CODE	Effluents	??
EFF_FLOW_CODE	Effluents	??
EFCO_STCO_CODE	Effluent_Conditions	??
MEAS_ANAL_METHOD_CODE	Measurements	
MEAS_RESULT_STATUS	Measurements	
MEAS_UNIT_CODE	Measurements	
SAMP_PURPOSE_CODE	Samples	
SAMP_MECHANISM	Samples	
SAMP_STATUS	Samples	
SAMP_MATERIAL	Samples	
SAMP_PROJECT_CODE	Samples	
SMPT_TYPE	Sampling_Points	Sampling Point Type Code
SMPT_STATUS	Sampling_Points	
SMPT_CLASS	Sampling_Points	
SMPT_SPPC_CODE	Sampling_Points	
SMPT_SPS_CODE	Sampling_Points	
SMPT_GEO_CODE	Sampling_Points	
SP_MAT_CODE	Sampling Points	Sampled Material Code
SP_SPT_STATUS	Sampling_Points	
DSI_TYPE	Discharge_Sites	Discharge Site Type
OL_CODE	Outlets	
OL_REN_CODE	Outlets	

In additions, although not marked as foreign keys, intuition suggests that many table columns make reference to entities in other tables eg. EFF_SPT_CODE is probably a reference to a Sampling Point used to monitor compliance of effluent levels associated with the corresponding agreement.

Many of these controlled vocabularies or code lists seem, whilst largely static, may be occasionally extended, or the status of their use may change. These are potential candidates for being managed in a Registry along with other reference data elements.

Access to sample data from most WIMS tables will enable a clearer assessment.

4.2.2 Identifiers

Strictly there are approximately 8 WIMS installations each serving different regional areas of EA. Whilst many/most coded lists are coordinated across installations eg. determinand codes, EA region/area/sub-area coding, status and method codes etc. there are significant groups of entities (sampling points, samples and measures) whose key identifiers are managed on a per-installation basis and need to be disambiguated (by region) in order to avoid a clash. This can present problems if responsibility for a sampling point changes from one region to another (due to region/area boundary change) and information about it needs to be maintained in by the WIMS installation associated with the region that takes on responsibility for the sampling point. Informal conversations indicate that mostly this is not a problem, but there seem to be a couple of region/area combinations where sample point identifier clashes are possible (SAMPLING_POINTS.SMPT_USER_REFERENCE). Similar considerations apply to WIMS generated identifiers for most entity types - SAMPLES, MEASURES, EFFLUENTS, EFFLUENT_SPECIFICS, EFFLUENT_CONDITIONS, AGREEMENTS, OUTLETS, DISCHARGE_SITES, RESPONSIBILITIES etc.

This problem could be obviated if some indication of the WIMS installation assigned the identifier were incorporated as part of the identifier - records could then be moved between installations and reference between installations with fewer problems.

Entity/Table	Entity ID Field	Example Identifier(s)	Comment	ID Scope
SAMPLING_POINT_TYPES	SPT_CODE	AB	2 character coding of a 2-3 level type hierarchy. 2nd and 3rd levels flattened into second character	global
CONTROLLED_WATERS	CONT_USER_REF	44/01A/001/01	Arbitrary depth, hierarchical coding based around ordering of tributaries and river stretches (probably?)	global
CONTROLLED_WATER_TYPES	CWT_SHORT_CODES	AF	Two character code. Unstructured	global
SAMPLING_POINT_CLASSES	SPC_CODE	0	Single letter or digit character code.	global
REGIONS	SHORT_CODE	SW	Two (letter) character mnemonic code.	global
AREA	ARE_CODE	1	Single character letter or digit code. Flat across all regions. Area Descriptions include region name, but no reference for area to region or vice-versa!	global
REGION_SUB_AREA	SAR_ARE_CODE, SAR_SUB_AREA_CODE	SAR_ARE_CODE=A SAR_SUB_AREA_CODE=3	Composite id: Two part coding. Subarea code scoped by area code.	global
CATCHMENTS	CATC_HYDRO_AREA_ID	44/02-	Hierarchical letter/digit code. Appears to be the stem of controlled water identifiers.	global
DISTRICT_COUNCILS	DC_REF	0J	Two character letter/digit coding of local authorities. For linked data should be resolved to OpenlyLocal or OpenDataCommunities (DCLG funded) Local authority URI.	global
RESULT_CODES	DETE_CODE, RESULT	DETE_CODE=8347 RESULT=3	Composite Id: Single character result code scoped by relate determinand.	global
DETERMINANDS	DETE_CODE	8347	Numeric code (simple counting sequence?) Approx 8.5k entries.	global
SAMPLE_MATERIAL	SMC_CODE	1AZZ	4 character code, mixed letters and digits.	global
PARTIES	P_ID	TC01	4 character code, mixed letters and digits. Note some are individuals, some organisations, some organisational roles (The Manager, White Hart Inn...)	local?
RESPONSIBILITIES	RESP_ID	22697	Numeric code.	local?
ROLES	ROLE_CODE	04	2 digit numeric code (approx 42 listed)	global?
PURPOSES	PURP_CODE	CA	2 character (letter) codes (approx 30 listed)	global
SAMPLING_MECHANISMS	SM_CODE	M	Single character letter or digit code (approx 35 listed).	global
UNITS	UNIT_CODE	005	3 digit numeric code Approx 46 listed. For linked data should be aligned with QUDT Units - where possible - hmmm.. megaLitre per tide!	global

Entity/Table	Entity ID Field	Example Identifier(s)	Comment	ID Scope
SAMPLING_POINTS	SMPT_USER_REFERENCE	70111015	Multi character code, mostly numeric. Have seen leading alpha possibly as a means to disambiguate otherwise clashing numerics following migration from a different installation	local?
SAMPLES	SAMP_ID	778370	Numeric	local?
MEASURES	MEAS_SAMPLE_ID MEAS_DETERMINAND_CODE	MEAS_SAMPLE_ID=778425 MEAS_DETERMINAND_CODE=0111	Composite id:	local?
AGREEMENTS	AGR_APL_NUMBER AGR_DSI_NGR	AGR_APL_NUMBER=001229/PE/01 AGR_DSI_NGR= SX9408089470	Three part mixed character application numbers. Linkage to discharge sites seems to be by NGR.	local?
EFFLUENTS	EFF_AGCO_ID EFF_AGCO_VERSION EFF_OL_REF EFF_NUM	EFF_AGCO_REF=001229/PE/01 EFF_AGCO_VERSION=1 EFF_OL_REF=1 EFF_NUM=1	Composite id: Effluents keyed by agreement, outlet(ordinal) and effluent (ordinal).	local?
EFFLUENT_CONDITIONS	EFCO_AGCO_ID EFCO_AGCO_VERSION EFCO_OL_REF EFCO_EFF_NUM	EFCO_AGCO_ID=001632/PW/01 EFCO_AGCO_VERSION=1 EFCO_OL_REF=1 EFCO_EFF_NUM=1	Composite Id: Effluents keyed by agreement, outlet(ordinal) and effluent (ordinal).	local?
EFFLUENT_SPECIFICS	ESP_AGCO_ID ESP_AGCO_VERSION ESP_OL_REF ESP_EFF_NUM ESP_NUM	ESP_AGCO_ID=001632/PW/01 ESP_AGCO_VERSION=1 ESP_OL_REF=1 ESP_EFF_NUM=1 ESP_NUM=2	Composite Id: Effluents keyed by agreement, outlet(ordinal), effluent (ordinal) and specific (ordinal)	local?
OUTLETS	OL_AGCO_ID OL_AGCO_VERSION OL_REF	OL_AGCO_ID=001229/PE/01 OL_AGCO_VERSION=1 OL_REF=1	Composite id: Agreement and outlet(ordinal)	local?
DISCHARGE_SITES	DSI_NGR	ST6790099200	The NGR for a representative point at discharge site seems to be it's principal identifier.	global
HOLDERS	HLD_RESP_ID HLD_AGR_NUMBER HLD_AGR_VERSION	HLD_RESP_ID=4162 HLD_AGR_NUMBER=010118 HLD_AGR_VERSION=1	A relational table between and responsibility and an agreement (seemingly truncated agreement number (to first part?))	Object Property ?

4.2.3 Managing Change

As indicated above, a change history is maintained for many of the entity and relationship tables listed above. It is worth being clear as to how much of the journalled history of an entity is of long-term interest to potential users of the data.

It is likely that changes in table entries for a given sample or measurement occur during the course of the analysis of the sample. While it is known from the Bathing Water Quality monitoring that sample assessments can be revised or withdrawn (typically within a week of initial publication) if changes never occur more than say 3 weeks after initial publication, then most of the dynamic journalled behaviour will relate to tracking the accumulation of measurements related to a sample in order to finalise and assessment result. There seems little public value in the added complexity of publishing the change history in such cases.

Likewise, the Sampling_Points table contains a SMPT_LAST_SAMPLED field. This table is journalled. If the mere act of taking a sample at that sampling point leads to a new journalled entry for the sampling point, which duplicates all its other field values then the size of the corresponding journal table will increase with routine sampling. Other changes in the content of the sampling_point table are expected to be much less frequent - and are worthy of being journalled in order that an accurate history can be reconstructed in the face of change (eg, change of operator of a facility, or

organisational responsibility for the sampling point, or its position if that is possible without identity change).

Broadly, some requirements based assessment needs to be carried out as to whether a single-present value publication of each journalled table is adequate or whether some (user) important capability would be lost in so doing. This is worth considering for entities whose table entries only change very early in their lifecycle and which remain invariant from that point on.

5 Annex II CLEAR Info Datasources

Received from Katie Hughes EA CLEAR Info Project.

Source System	Description	Permitting / MDS	Monitoring / transactional	Relevance to Regulation	What
CCS	Compliance Classification Scheme			1 Scheme for classifying non-compliance with permit conditions for activities regulated by the EA.	CCS is a means of assessing and classifying the seriousness (potential environmental damage) of any non-compliances with permit conditions at Environment Agency permitted sites. It contains all instances of permit breaches and the classification score, and records actions. Information on water quality, water resources, Waste permits and Installation (industry) permits and emissions
NALD	National Abstraction Licensing Database		1	1 Details of water abstraction licenses and actual abstraction quantities	Details of water abstraction licences, water abstraction points and actual abstraction quantities. Some information is confidential
OPRA	(Neil's database for calculating OPRA scores)			1 Produces a site's compliance banding (A= good - F = Bad), operator performance (attitude) attribute and calculates charging	This system produces risk score for permitted sites. It looks and a sites permit compliance, operator performance, sensitivity of the location etc. also used to calculate permit charging.
PAS	Permit Administration System		1	This data is updated used continuously to track National Permitting's Progress in determining PPC (EPR) Applications and variations. It is used by NPS TL's and the 4 NPS Managers to work plan to the KPIs and other regulatory targets. Contains permit information of Installations (larger permitted sites).	Permit statuses, Operator addresses, Sectors, details of applications for permits and variations that have been made and either not yet processed or completed, etc. IPPC information is downloaded on a weekly basis for internal and external customers. Details of Permits and Variation histories, Copies of Permits and applications etc used to be stored. (data will be migrated into IR system within lifespan of the project)
RATS	REGIS Appended Tonnage System			1 The purpose of RATS is to collect, store and report information about the types and quantities of waste handled by licensed waste management facilities. The type of waste can be recorded using either the UK National Waste Classification or the European Waste Classification. The origin of the waste can be recorded to district authority level.	Details of waste processed by permitted sites (e.g. construction and demolition waste) – general waste category & tonnage, plus (in some cases) origin of inputs and destination of outputs. + any sites accepting waste - cement kilns, CORUS, paper & pulp
REGIS	REGulation Information System for Waste Management		1	Waste management information system	Details of permitted waste management sites – address, operator, licence number, waste types, etc.; also data on registered exemptions. (data will be migrated into IR system within lifespan of the project)
CRC	Carbon Reduction Commitment		1	1 Energy usage for large organisations in the private and public sectors	Applies to parent company level for organisations that have a Half Hourly electricity Meter and use more than 6,000 Mega Watt Hours through all HHM annually. Companies report on energy consumption across their estate and Carbon emissions
WIMS	Water Information Management System		1	1 Holds environmental monitoring and water quality permitting information	Water discharge consents and water quality information
HAZWASTE database 1 - Registrations			1		Hazardous waste registrations for premises that produce or hold more than 500kg of hazardous waste in a year.
HAZWASTE database 2 - Consignee returns				1	The system support our regulatory requirements in the Hazardous Waste Regulations 2005. Data from Consignees (i.e. waste management sites) that received hazardous waste. Each waste load received is summarised and supplied to the FSC
ETSWAP	Emissions Trading Scheme (Workflow Automation Project)		1	1 Details of permits and emissions of GHGs	Annual CO2 emissions and credit trading data for EU ETS scheme. Only large companies in specific sectors need to adhere to the scheme. E.g. Energy, cement and lime, glass, paper and pulp, ceramics and aviation.
COMAH	Control of Major Accident Hazards			1 Details of COMAH sites and potentially hazardous substances	
high public interest site.				1	
NIRS	National Incident Recording System			1 Recording of reported incidents. Details of all environmental incidents, including pollution incidents at regulated and unregulated sites.	Could be of particular interest to investor companies about environmental risk of a sector without sharing any specific details. Could also be used working with a specific parent company to look at their incidents and complaints. The Incident Communication Service/RCC Wales log incidents on NIRS received from other responders (e.g. the fire service), operators (e.g. water industry), the general public, EA staff etc. The investigating officer completed the NIRS record. Data recorded for pollution incidents includes the source premises type, source address (if known), if the site is regulated, pollutant type, and cause.
PIEDC	Electronic Data Collection for Pollution Inventory			1 Allows Operators to submit Integrated Pollution Control (IPC), Pollution Prevention and Control (PPC), Radioactive Substance Regulation (RSR), Waste Management Licensing (WML), Sewerage Undertaker	Annual release returns via the web. These include emissions to air, land, water or sewer.
NED	National Enforcement Database			1 Maintenance and reporting of national enforcement actions	Details of enforcement activity (all, not just waste) – enforcement notices, fines, cautions, prosecutions, etc. Pulls data from NIRS



121206 KH List of datasets V2.0.xls