

Existing and proposed CAFE reporting and data exchange systems

LOT3 – United Kingdom

Report to JRC

Restricted Commercial
ED43757
Issue 1
July 2009




Title	Existing and proposed CAFÉ reporting and data exchange systems LOT 3 – United Kingdom
Customer	DG-JRC, Joint Research Centre, Institute for Environment & Sustainability
Customer reference	Contract Number - 384306
Confidentiality, copyright and reproduction	This report is the Copyright of the European Community DG Joint Research Centre (EC) and has been prepared by AEA under contract to the EC dated 26th November 2008. The contents of this report may not be reproduced in whole or in part, nor passed to any organisation or person without the specific prior written permission of the EC. AEA accepts no liability whatsoever to any third party for any loss or damage arising from any interpretation or use of the information contained in this report, or reliance on any views expressed therein.
File reference	ED43757
Reference number	AEAT/ENV/2832 – Issue 1

AEA group
 329 Harwell
 Didcot
 Oxfordshire
 OX11 0QJ

t: 0870 190 6423
 f: 0870 190 6318

AEA is a business name of AEA Technology plc

AEA is certificated to ISO9001 and ISO14001

Author	Name	Tony Bush Xingyu Xiao Geoff Broughton Trevor Davies
Approved by	Name	Alan Collings
	Signature	
	Date	2 July 2009

Executive summary

This report presents deliverables D1.1 and D2.1 for LOT3 under the JRC's INSPIRE-CAFE contract¹. LOT3 relates to a generic example of a centralised CAFE compliance network of 50+ monitoring stations and refers directly to the UK. Deliverables D1.1 and D2.1 relate to:

Deliverable D1.1 – A report on the implementation of CAFÉ reporting and data exchange

Deliverable D2.1 – A report on the architecture for a prototype INSPIRE-CAFÉ reporting system

The key objectives for these deliverables are to document the existing CAFE reporting systems in operation within the UK, providing;

- a high level description of these in terms of data-flows, decision making and the overall data inputs and outputs
- a description of the state of compliance of existing systems relative to INSPIRE concepts and draft CAFE implementation provisions
- a high-level description of how existing systems are to be adapted to incorporate INSPIRE concepts and services
- a description of the proposed prototype system architecture

This work has identified that the main areas of non-compliance for the UK's existing reporting and data exchange systems relative to INSPIRE implementation rules may be summarised as follows;

1. Data specifications and formats to promote interoperability
2. Services to support download
3. Services to support discovery of data

In order to close this compliance gap it has been proposed to implement the INSPIRE compliant harmonized data specification for measurement datasets (previously defined by the JRC's Environmental Monitoring Facilities project) within the prototype INSPIRE-CAFE reporting system for the UK. It has been recognised however, that this data specification may require extending to accommodate CAFE specific information instances for measurement data in some cases.

An INSPIRE compliant XML based data specification and schema for reporting of CAFE assessment information, delimitation of CAFE zone's and supporting documentation (supplementary assessment data in the form of modelled outputs) will also be development and implemented. This specification shall include mandatory INSPIRE metadata elements and other ISO based metadata elements where these add value to the data specification.

Existing UK air quality download services are compliant, at a high-level, with INSPIRE concepts for such services. However, an important gap relative to INSPIRE for existing download services relates to file formats used. As a result, transformation tools and processes to migrate existing datasets to the data specifications outlined above will be developed and implemented to achieve interoperability.

Metadata to support discovery services is, in general, missing from the UK's existing systems. A CAFE specific metadata profile relevant to such services has already been the subject of a collaborative report within this contract. This profile will be used within an automated system to generate up to date metadata to support discovery services for all UK CAFE data themes. Metadata will be registered with INSPIRE registries.

¹ Development and demonstration of technical IT solutions for data exchange and reporting under the CAFE Directive using INSPIRE services.
Contract Number - 384306

Table of contents

1	Introduction	1
1.1	Structure of this report	1
1.2	Scope of INSPIRE services relevant to this pilot study	1
1.3	Scope of CAFE data themes for reporting and data exchange	2
2	Existing CAFE reporting and data exchange in the UK	3
2.1	Air quality monitoring data	3
2.1.1	Reporting official validated measurement data via the DEM data submission	4
2.1.2	Reporting near real-time data	5
2.2	CAFE assessment information and zone datasets	11
3	Implementation of INSPIRE concepts	15
3.1	Data specifications	15
3.1.1	Measurement datasets reported via the DEM	15
3.1.2	CAFE assessment information and zone datasets	16
3.2	Discovery metadata	17
3.3	Download services	18
4	Architecture options for the prototype	21
4.1	Data transformations	23
4.1.1	Data schema transformation	23
4.2	Additional archive processes to support INSPIRE services	25
4.2.1	Processes to support INSPIRE compliant download services	26
4.2.2	Processes to support discovery metadata services	27
4.2.3	UK INSPIRE-CAFE data repository file structure	28
5	Summary and conclusions	31
5.1	Key gaps in the existing UK systems relative to INSPIRE IR	31
5.2	Summary of steps to close the gaps	32

Appendices

Appendix 1	Extended DEM data specifications incorporating draft CAFÉ IP and with preliminary mappings to the harmonised EMF data specification
Appendix 2	List of acronyms

1 Introduction

This report presents deliverables D1.1 and D2.1 for LOT3 under the JRC's INSPIRE-CAFE contract². LOT3 relates to a generic example of a centralised CAFE compliance network of 50+ monitoring stations and refers directly to the UK. Deliverables D1.1 and D2.1 relate to:

Deliverable D1.1 – A report on the implementation of CAFÉ reporting and data exchange

Deliverable D2.1 – A report on the architecture for a prototype INSPIRE-CAFÉ reporting system

In relation to deliverable D1.1, this report aims to document the existing CAFE reporting systems in operation within the UK. It provides a high level description of the implemented work and data-flows, decision-making and the overall data inputs and outputs. The report presents how existing procedures are to be adapted to incorporate INSPIRE concepts and services and highlights the data, software and process gaps, which are to be closed in order to achieve this. In addition, potential problems and implementation issues are identified.

In order to achieve resource efficiencies under the INSPIRE-CAFE work programme the architecture of the proposed prototype system is closely coupled to the architecture of the existing UK systems which are well developed and administered. As a result, and in order to draw upon the benefits of a single document to report on existing and future prototype reporting practices, we have chosen to report on Deliverables D1.1 and D2.1 within the same document.

For Deliverable D2.1 we have presented, at a high level, the prototype system architecture, its building blocks and their functionality. Data formats and methodologies (services and processes) to be implemented are also described, including the embedded quality control procedures relevant to the draft CAFE Directive Implementing Provisions³ (CAFE IP) for quality control verification, versioning and resubmission of CAFE datasets.

1.1 Structure of this report

Section 1 of this report provides overall context on the structure and scope of this report and pilot study, its aims and objectives. Information relevant to Deliverable D1.1 on existing UK systems for CAFE reporting and data exchange, data, software and process gaps and closure mechanisms are presented in Section 2 and 3. Information relevant to Deliverable D2.1 on the propose prototype architecture is presented in Section 4. A summary and conclusions of the report is provided in section 5.

1.2 Scope of INSPIRE services relevant to this pilot study

Although the full range of generic INSPIRE services relate to those for data discovery, viewing, mapping, downloading and interoperability, the terms of reference for this pilot study, and clarifications on these raised at the INSPIRE-CAFE kick-off meeting, have identified the core INSPIRE services for implementation in this pilot study. The core INSPIRE service concepts to be implemented for the reporting of CAFE data themes within this pilot study are as follows:

1. Creation of metadata for discovery services making it possible to locate spatial data sets and services, on the basis of the content of the corresponding metadata and to display the content of the metadata
2. Transformation services, enabling spatial data sets to be transformed with a view to achieving interoperability

² Development and demonstration of technical IT solutions for data exchange and reporting under the CAFE Directive using INSPIRE services. Contract Number – 384306.

³ Draft Implementing Provisions for reporting under Directive on ambient air quality and cleaner air for Europe (Data exchange Group - Preparatory material), DEG internal discussion document, January 2009. File reference - DEG Working Document_IP Draft Last09.doc.

3. Download services, enabling copies of spatial data sets, or parts of such sets, to be downloaded and, where practicable, accessed directly

1.3 Scope of CAFE data themes for reporting and data exchange

Section 3.2.1 of the tender specification's Technical Annex⁴ identifies the following generic CAFE data themes on air quality information as within the scope of this pilot study:

1. Air quality monitoring data I; provisional near real-time air quality monitoring data relating to exceedences of information and alert thresholds according to the draft CAFE IP
2. Air quality monitoring data II; validated (fully quality controlled) monitoring data.
3. Air quality monitoring metadata; information on monitoring sites, measurement configuration etc. currently requested under the Council Decision 97/101/EC on Exchange of Information and reported via the Data Exchange Module (DEM)
4. Assessment (compliance/exceedence) information; the main annual reporting requirement which combines all different types of ambient air quality assessment performed (monitoring, modelling, objective estimation) in a prescribed format and declares (non) compliance with limit and target values, as well as relation to other assessment thresholds⁵
5. Information on delimitation of zones and related metadata; where 'zone' means part of the territory delimited by that Member State for the purpose of air quality assessment and management. This is also currently requested as part of the Council Decision 97/101/EC on Exchange of Information

All of these generic data themes are the subject of existing reporting obligations or systems under CAFE and its forerunners. However, with the creation of CAFE and related revisions to legislation and instruments regulating air pollution, (compliance checking, reporting and dissemination to the public), the existing reporting mechanisms have been put forward for change via new implementing provisions. In particular the new CAFE Directive extends the scope of monitoring and reporting of levels of PM_{2.5}. These revised CAFÉ IP will be accommodated in this pilot study.

Section 3.2.1 of Technical Annex also identifies that the following data themes are outside the scope of this pilot study:

1. Data collection and reporting of closely related air pollution information (emission inventories, industrial emissions-EPER, transport related emission information etc.) as well as air pollution related health effects.
2. Provisional up to date air quality information to the public and summer ozone reporting.
3. Exchange of information on air quality management (air quality plans and programmes).

We have noted that reporting obligation for pollutants regulated by the 4th Daughter Directive on PAH and heavy metals⁶ are outside the scope of the CAFE IP but within the scope of reporting under the existing assessment information theme(s) outlined above (in bullet 4 above in the existing Questionnaire reporting system). For the purpose of this study we have taken this to mean that pollutants regulated by the 4th Daughter Directive are beyond the scope of the pilot study.

⁴ TA_SEIS_CAFE_080312.pdf

⁵ The CAFÉ 'Questionnaire'

⁶ DIRECTIVE 2004/107/EC http://eur-lex.europa.eu/LexUriServ/site/en/oj/2005/l_023/l_02320050126en00030016.pdf

2 Existing CAFÉ reporting and data exchange in the UK

Details of the work and data-flows, decision making process and input / output data for existing CAFÉ reporting and data exchange systems in the UK are presented in this section. Each of the five generic CAFÉ data themes is addressed.

2.1 Air quality monitoring data

For historical reasons two parallel systems for reporting and data exchange of validated and near real-time data exist within the UK. These are commonly referred to as:

1. The HIS QA/QC System
2. The UK Air Quality Information Archive⁷

The high level functionality of the two systems is distinct, even if the interdependencies are not, and summarised below.

The HIS system has been developed over more than 25 years by AEA as an off-line tool to assist in the ratification (validation) of the UK's centrally managed measurement datasets. The system manages the transition of the data from provisional (raw) data to ratified (validated) data, the latter being of sufficient quality for dissemination to the Commission and the European Environment Agency (EEA) as official measurements for information exchange and compliance reporting (checking) purposes via the current Data Exchange Module (DEM) and CAFÉ Questionnaire mechanisms. The HIS system currently acts as the main interface between ratified, official UK datasets and the DEM for reporting of datasets to AIRBASE under the ReportNet mechanisms. The system also provides an interface for existing reporting obligations to summer ozone concentrations under the Ozone Directive⁸.

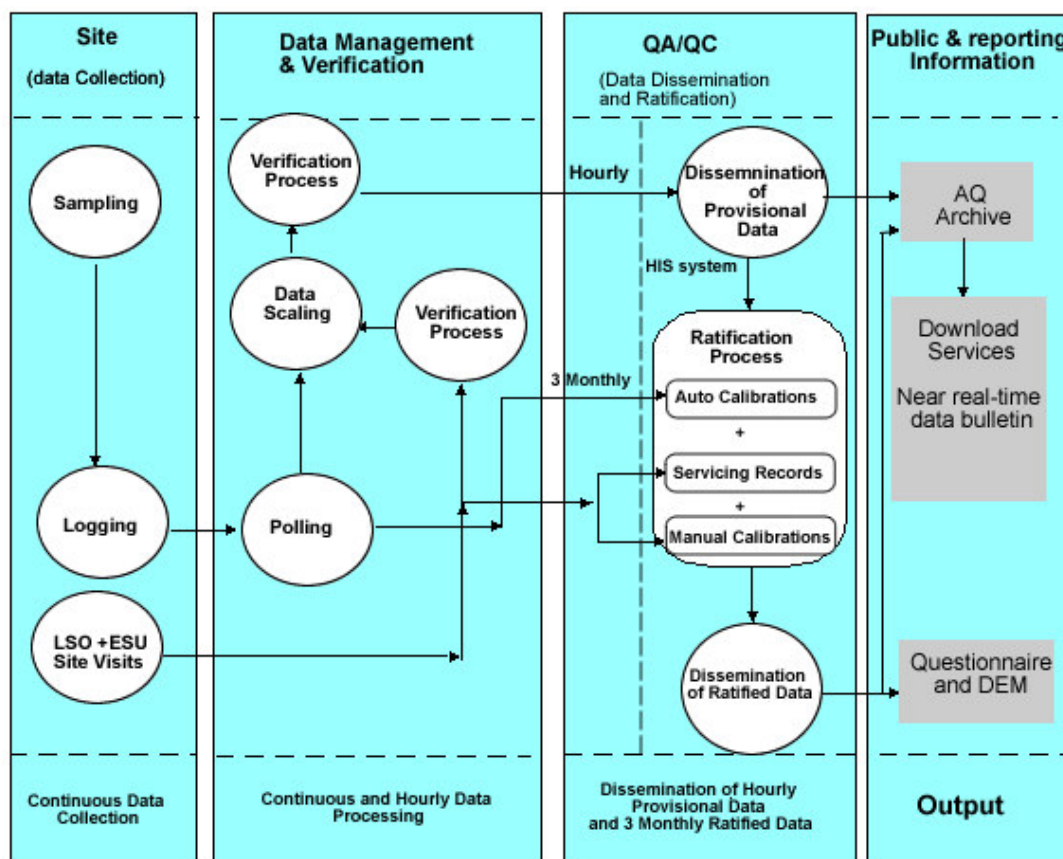
The UK Air Quality Information Archive (the Archive) on the other hand, has been developed by AEA since the mid 1990s as a public interface for air quality data (both provisional and validated). It provides a mechanism for disseminating information on air pollution to various user groups ranging from the general public to the UK Government, European Commission and the scientific research community. Air quality information for the UK is available via a range of bulletin, viewing and download services hosted by the Archive and are available for public information services relating to CAFÉ (near real-time data in the form of alert and information threshold exceedences) and for official data and aggregated statistics.

The two systems and their high-level interconnectivity are broadly summarised in **Figure 1**.

⁷ <http://www.airquality.co.uk>

⁸ Directive on ozone in ambient air (2002/3/EC) http://eur-lex.europa.eu/pri/en/oj/dat/2002/l_067/l_06720020309en00140030.pdf

Figure 1 Data and work flows in the UK validated and near real-time reporting and data exchange mechanism



Note: acronym reference list provided in Appendix 2

It should be recognised that both systems have relevance to the prototype CAFE reporting mechanism for the UK; the HIS system providing a robust, tried and tested tool for data ratification, whereas the Archive provides web enabled download services compliant (in the main) to open standards and high-level INSPIRE concepts. Continuous real-time updates from the HIS system to the Archive ensures the synchronous operation of the two systems.

The broad data and workflows for each system are summarised in the following sections.

2.1.1 Reporting official validated measurement data via the DEM data submission

Currently, all of the mandatory data reporting obligations set out by the Exchange on Information decision (Eol) on Air⁹ via the DEM, are conducted by AEA's HIS system.

Because the DEM reporting / submission systems have been used for many years, the majority of metadata relevant to network, station and measurement configuration have been stored in the EU DEM system either as flat files or are stored in the HIS system itself.

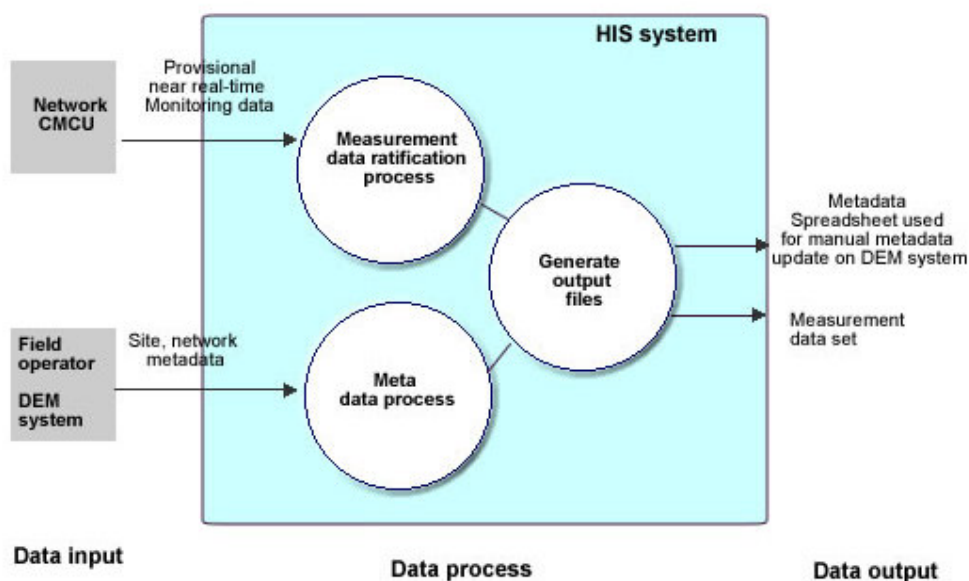
On an annual basis, by the end of September each year, meta information are obtained either by extracting data from the HIS system and outputting to a range of Excel files, or modifying the existing

⁹ Exchange on Information decision (Eol) on Air, 97/101/EC (OJ L 35, 5.2.1997, p. 14–22)
<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31997D0101:EN:HTML>

metadata files downloaded from the DEM system. These are then updated and uploaded to the DEM system using one of the preferred upload and file format options.

A summary of the work and data flows relevant to CAFE reporting under the existing DEM mechanism via the UK's HIS system is presented in Figure 2.

Figure 2 Generic data and work flows relevant to CAFE reporting under the existing DEM mechanism via the UK's HIS system



Note: acronym reference list provided in Appendix 2

2.1.2 Reporting near real-time data

At present, there are a variety of near real-time data flows for exchange of air quality information via HIS and the UK Air Quality Information Archive. These data flows extend to services for

1. Public information services; including CAFE alert and information threshold information, national air quality bulletins, indices and forecasts to media services and other subscriber communities (carried out by the UK's Archive)
2. Monthly summer ozone reporting under the 3rd Daughter Directive on air quality¹⁰ (carried out by the UK's HIS system)
3. Services to the EEA's ozone web service (carried out by the UK's Archive)
4. Data for scientific research purposes e.g. air quality forecast model validation and data assimilation (GEMS, ECWMF, PREVAIR etc.)

Only the near real-time data flows extending to CAFE alert and information thresholds are included in the scope of this pilot study.

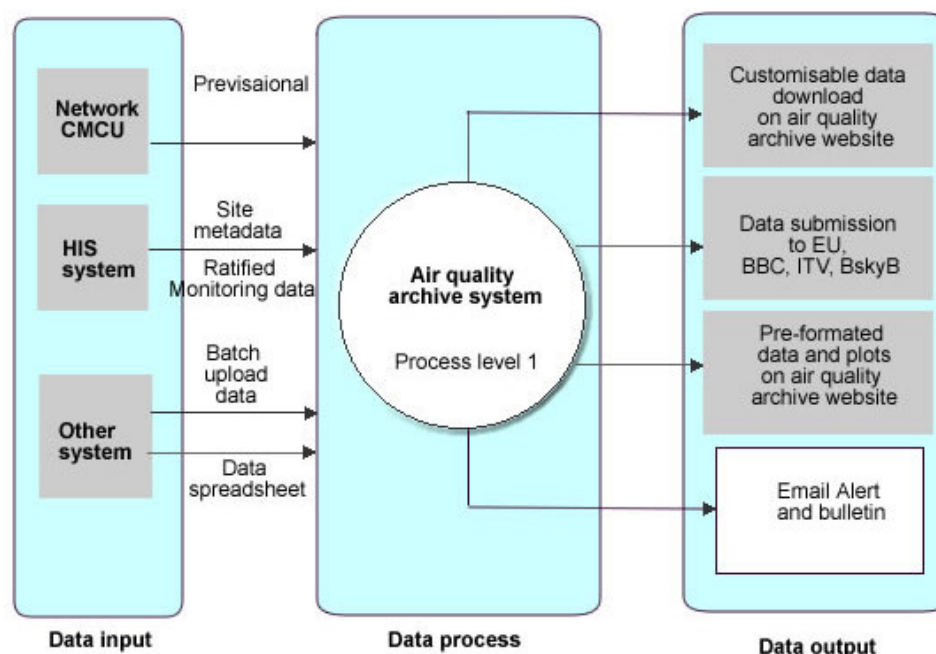
The main file formats used in these reporting systems are csv or text documents. The mechanisms of exchange use predominantly ftp or sftp (file transfer protocol or secured file transfer protocol). However, the Archive does also provide to its users a full range of near real-time services including information on the latest measurement data, exceedence data for national and EU regulatory metrics and national pollution indices for public information. The outputs of these services are available via viewing services using KML, standard http interfaces and via data download or ftp. These services

¹⁰ Directive on ozone in ambient air (2002/3/EC) http://eur-lex.europa.eu/pri/en/oj/dat/2002/l_067/l_06720020309en00140030.pdf

are in general not currently INSPIRE compliant, in their file formats (predominant use of non-xml based formats), the level of metadata support for interoperability, nor in the viewing services used.

Figure 3 provides a description of the data flows and reporting services for near real-time datasets in the UK. Further detail on the processes populating this system with information is given in the following sections.

Figure 3 Top level population of the UK Air Quality Information Archive with air quality data



Note: acronym reference list provided in Appendix 2

Network, station and sensor configuration metadata for automatic monitoring stations are managed by site operator(s) and collected by AEA's HIS system. Subsets of the metadata information stored in HIS (and also supplied in the DEM submissions to the CDR) are uploaded to Archive system by the administrator. Similar datasets for non-automatic data are collated by the Archive and periodically uploaded to the HIS system.

The near real-time monitoring data from the UK's automatic monitoring networks are uploaded to the Archive system via the managers of the monitoring networks (CMCU) in flat file format. This process performed every hour. The status of the data files is assigned a 'provisional' QA/QC flag at this stage. Comprehensive QA/QC checks are carried out over a three-month cycle using AEA's off-line HIS system tools to ratify (validate) data, and thereby progress the status of data from provisional near real-time data to official (ratified) data. This process involves scaling of data against network calibration information, identifying and removing spurious information and uses the monitoring network audit results to confirm satisfactory analyser performance. After completion of the ratification procedure, data are updated in the HIS system with appropriate QA/QC flags and transferred to the UK Archive again with new ratified status.

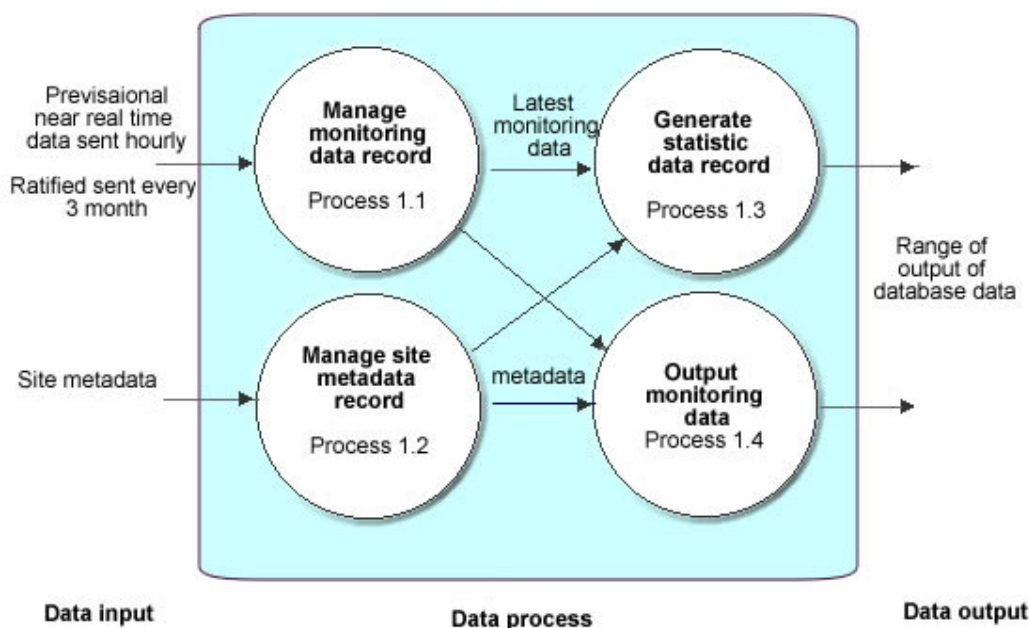
Some of the UK's air quality monitoring datasets are collected from non-automatic monitoring networks at daily or monthly intervals. These data are periodically provided to AEA by the network service providers in a ratified form. The files are formatted into the standardised upload files (batch upload files), similar to those for near real-time data. The files are manually transferred to the process server prior to statistical processing (along with the near real-time datasets) to generate aggregated metrics for compliance checking against CAFÉ and UK regulations. This data flow, represented by the 'Other system' data flow in Figure 3, does not return a temporal resolution relevant to CAFÉ alert and

information thresholds, although is relevant to PM₁₀, PM_{2.5}, lead and pollutants regulated by the 4th Daughter Directive¹¹.

An overview of the high level functionality of the data processing of the Archive system relevant to the pilot study and INSPIRE is provided in Figure 4 below. The data processing and dissemination mechanisms for near real-time and ratified data within the Archive system have been broken down into 4 generic processes with discrete functionality as presented below. Further details of the component processes are provided in the following sections.

- | | |
|---------------------|---|
| Process 1.1. | Management of measurement data records. |
| Process 1.2. | Management of network metadata records. |
| Process 1.3. | Generation of aggregated statistical information. |
| Process 1.4. | Data output services. |

Figure 4 Overview of Level 1 processes in the UK air quality archive system



Note: acronym reference list provided in Appendix 2

Process 1.1 – Management of measurement data records

Raw measurement data files are uploaded to the Archive server from the HIS system and/or the monitoring network operators in hourly cycles. The details of the component sub-processes are presented graphically in Figure 5. The highest temporal resolution at which raw measurements data are collected is as hourly averages (for all pollutants measured using automatic methods) with the exception of sulphur dioxide for which a 15- minute average is collected for UK regulatory purposes.

A sub-process (process 1.1.1 in Figure 5) identifies newly uploaded files in the directory. When new files are found, the upload process (process 1.1.2) is activated. The metadata content of supplied upload files is validated against pre-existing information held by the Archive; each field, such as monitoring station codes, pollutant, network type, sample date and time, measurement period, being validated. Where validation is not successful, the record/data file is rejected.

A check is also carried out for duplicate records in process 1.1.3, whereby if there is no existing record for the site at the monitoring date and time, a new record is created with the QA/QC status flag set as

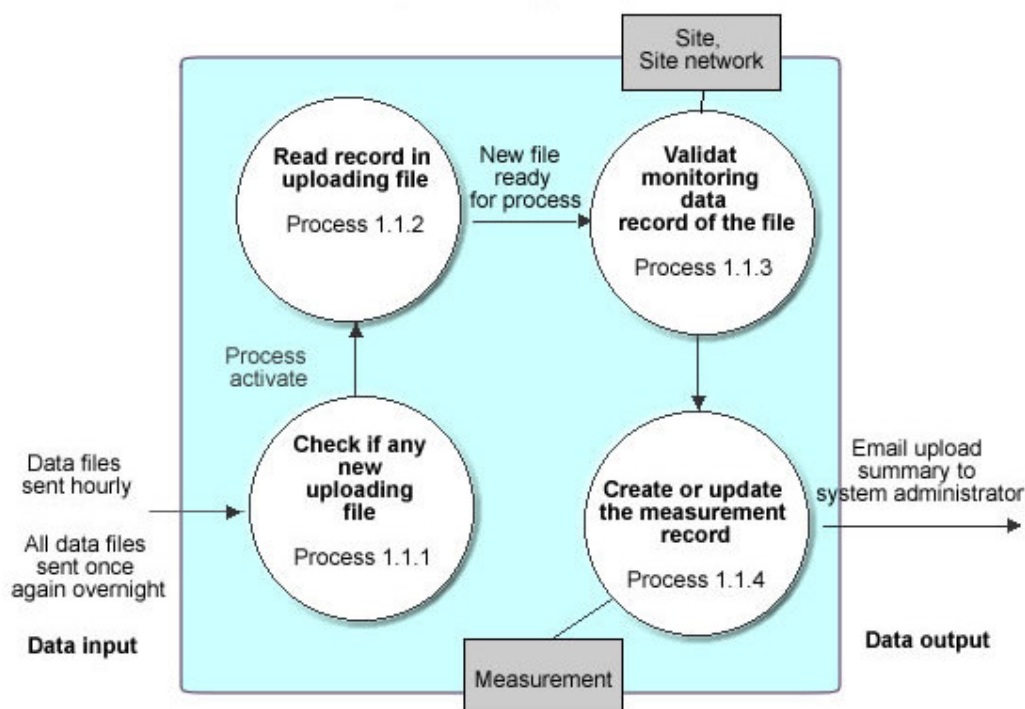
¹¹ 2004/107/EC http://eur-lex.europa.eu/LexUriServ/site/en/oj/2005/l_023/l_02320050126en00030016.pdf

provisional. If a record does exist however, then this is updated with the new values. When monitoring data are ratified within the HIS system in 3-month cycles, these updated data files with appropriate QA/QC flags are uploaded to the Archive as ratified (validated) data.

All records created, updated and rejected are logged in process 1.1.4 and a summary of each upload is sent to system administrators via email and stored in database. The uploaded files are backed up to a new directory. Rejected records and the errors are written to an errors file.

The entire process is reiterated with all upload files once a day to ensure that files which failed to upload initially are captured.

Figure 5 Sub-processes to Process 1.1 management of measurement data records



Note: acronym reference list provided in Appendix 2

Process 1.2 Management of network metadata records

The transfer of metadata files between HIS and Archive systems initiates the update of metadata within the Archive relating to monitoring network, monitoring station and measurement configuration. The details of this process are presented graphically in Figure 6.

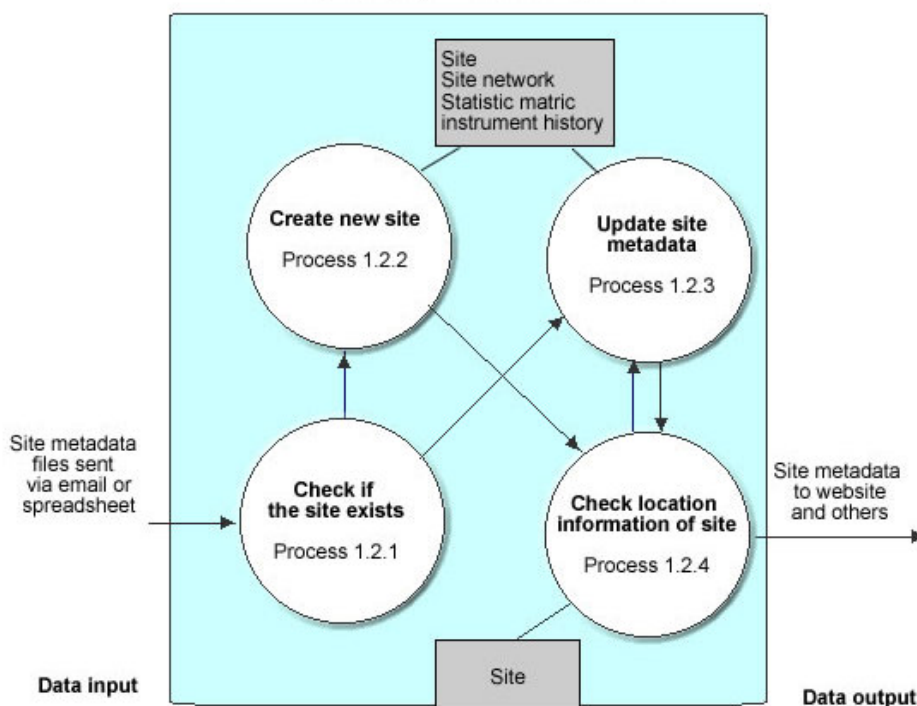
Sub-process 1.2.1 performs a check to establish if all station information exists in the Archive system. If not, sub-process 1.2.2 generates records and relevant information. The information included in this metadata transfer covers:

- Monitoring station details (Station codes, name, physical location and location type etc.).
- Network information (which monitoring network a station is operated under).
- Station configuration (parameters measured, monitoring start date, instrument used etc.).
- Statistic calculation configuration.

However, it should be noted, that the Archive system (at present) does not store all metadata information required under the DEM submission. Rather, a subset of this information is currently stored, which is adequate for effective dissemination of data to core Archive end-user communities.

Sub-process 1.2.3 then checks whether station records exist and the content of the metadata. If the content needs updating changes are made and validated. Sub-process 1.2.4 describes the follow up checks on metadata provided by user AEA's internal air quality community data users who validate metadata content such as location type, grid reference etc and provide a further reality check on metadata content.

Figure 6 Sub-processes to Process 1.2 Management of network metadata records



Note: acronym reference list provided in Appendix 2

Process 1.3 Generation of aggregated exceedance and descriptive statistical information

Process 1.1 described the maintenance of raw data within the Archive resolved to hourly averages (15-minute averages for SO₂). Aggregated exceedance and descriptive statistics based on these monitoring data are calculated by Process 1.3 as shown in the Figure 4 schematic of the system.

Aggregated metrics such as daily or annual mean concentrations provide users with analysed and interpreted information relevant to research, public information and for compliance checking against the various national and European standard levels, thresholds and targets. In order to provide the aggregated statistics efficiently the datasets are pre-calculated daily and stored in the Archive system. Calculation of aggregated statistics on-the-fly, at point of request by the end-user, would be prohibitive in terms of processing and download speeds.

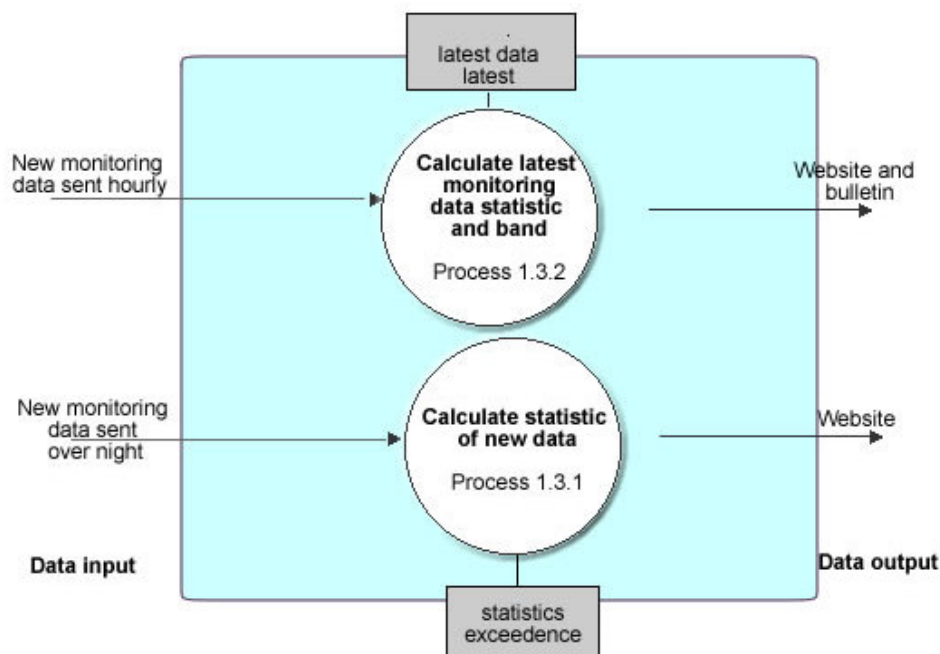
The pre-calculated aggregated statistics include all metrics covered by existing CAFE IP and regulated by the Air Quality Daughter Directives and the UK's Air Quality Regulations¹² plus a wide range of metrics providing information on air pollution trends, emissions, spatial and temporal patterns of exposure and high pollution episodes and with progress towards meeting the prescribed objectives and targets.

The calculation of aggregated statistics may be broadly sub-divided into calculation of short and long-term aggregated metrics as shown in Figure 7. The former being calculated every hour as new measurement data comes in from the monitoring stations (sub-process 1.3.2 in Figure 7) and relates to information such as CAFE information and alert threshold data, and the UK's air quality banding

¹² Supported by the UK's Air Quality Strategy (AQS) <http://www.defra.gov.uk/environment/airquality/strategy/index.htm>

and index system¹³. The latter being calculated on a daily basis by sub-process 1.3.1 and are relevant to longer-term statistics such as annual and daily mean concentrations, percentiles, running means concentrations etc.

Figure 7 Sub-processes to Process 1.3 Generation of aggregated statistical information



Note: acronym reference list provided in Appendix 2

Approximately 60 short-term descriptive statistics are calculated hourly and 30 long-term exceedance statistics are calculated at lower temporal resolutions by the Archive system¹⁴.

Process 1.4 Data output services

The main data output services and processes offered by the Archive are summarised below and in Figure 8. These include:

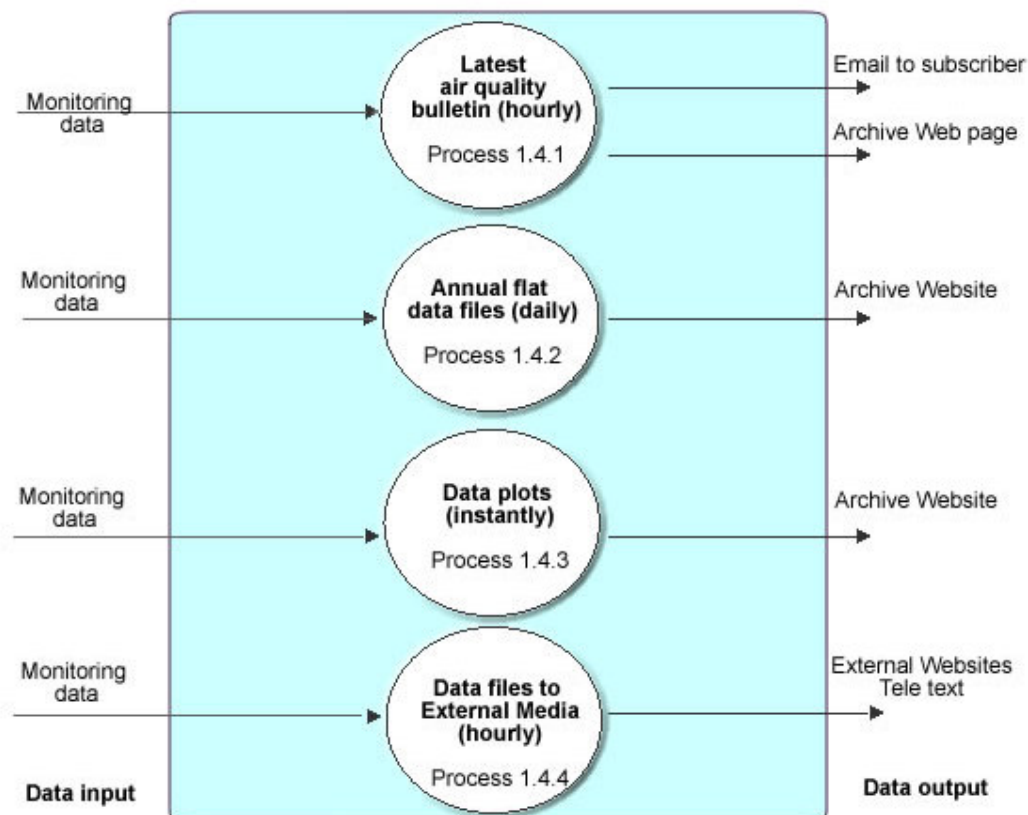
- Process 1.4.1. Short-term statistical downloads;** generates download services for short-term aggregated metrics. Outputs latest monitoring data summary as measured data and analysed results for UK regions to inform citizens about the quality of the air we all breathe. The data are listed on the air quality archive and accessible by a viewing service and links. The data are also sent to users subscribing to an email air quality bulletin service.
- Process 1.4.2. Long-term statistics downloads;** generates download services allowing users to select measurement, statistic data and exceedance data from on line forms. The selected data can be viewed on screen and sent via email. All the download queries are stored in the database for monitoring purposes.
- Process 1.4.3. Time-series downloads;** generates updated weekly monitoring data plots.

¹³ <http://www.airquality.co.uk/bulletin.php?type=Current>

¹⁴ <http://www.airquality.co.uk/standards.php#std>

Process 1.4.4. Bulletin services; generates bulletin text files for dissemination to media broadcasts.

Figure 8 Sub-processes to Process 1.4 Data output services



Note: acronym reference list provided in Appendix 2

2.2 CAFÉ assessment information and zone datasets

CAFÉ data themes (4) and (5) as presented in section 1.3 of this report relate to datasets delimiting CAFÉ zones within Member States for air quality assessment and management purposes and assessment information for compliance checking against the CAFÉ Directives various regulatory thresholds and targets.

Under existing CAFÉ IP, assessment information and supporting information such as datasets delimiting CAFÉ zones may be reported to the ReportNet Central Data Repository (CDR) in variety of file formats. For the most recently published CAFÉ reporting round (2007 at time of preparation of this report), the files presented in Table 1 were uploaded to the CDR by AEA on behalf of the UK in order to fulfil formal reporting obligations. The content and format of each being are controlled by guidelines provided by the Commission¹⁵.

¹⁵ Guideline to Commission Decision 2004/461/EC laying down a questionnaire to be used for annual reporting on ambient air quality assessment under Council Directives 96/62/EC, 1999/30/EC, 2000/69/EC and 2002/3/EC August, 2006

Table 1 CAFÉ assessment information and zones delimitation file formats reported to the CDR

File name	Format	Content
080902_New_questionnaire_monitoringandmodellingdata_final_v2.xls	MS Excel	CAFÉ assessment information
ukmaps2007.doc	MS Word	Images of UK supplementary assessment raster outputs
ukmaps200710kmdata.doc	MS Word	ASCII grid of UK supplementary assessment raster outputs, resample from 1x1km resolution to 10x10km
ukzones2007_shapefile.zip	Shapefile (zipped)	ESRI shapefile delimiting UK's CAFÉ zones for 2007

The bulk of CAFÉ assessment information (including datasets delimiting CAFÉ zones) is reported to the CDR using a Microsoft Excel spreadsheet. This spreadsheet is commonly known as 'the Questionnaire' and is supplied by the Commission; updates to content and format are being available annually¹⁶.

The UK's Questionnaire compilation process is semi-automated, utilising information from many data streams already described in previous sections in addition to others which have been developed to accommodate the UK's unique situation; the UK being heavily reliant on modelled (supplementary) assessment data for both formal CAFÉ compliance purposes and to optimise the UK monitoring networks for compliance with the CAFÉ monitoring requirements.

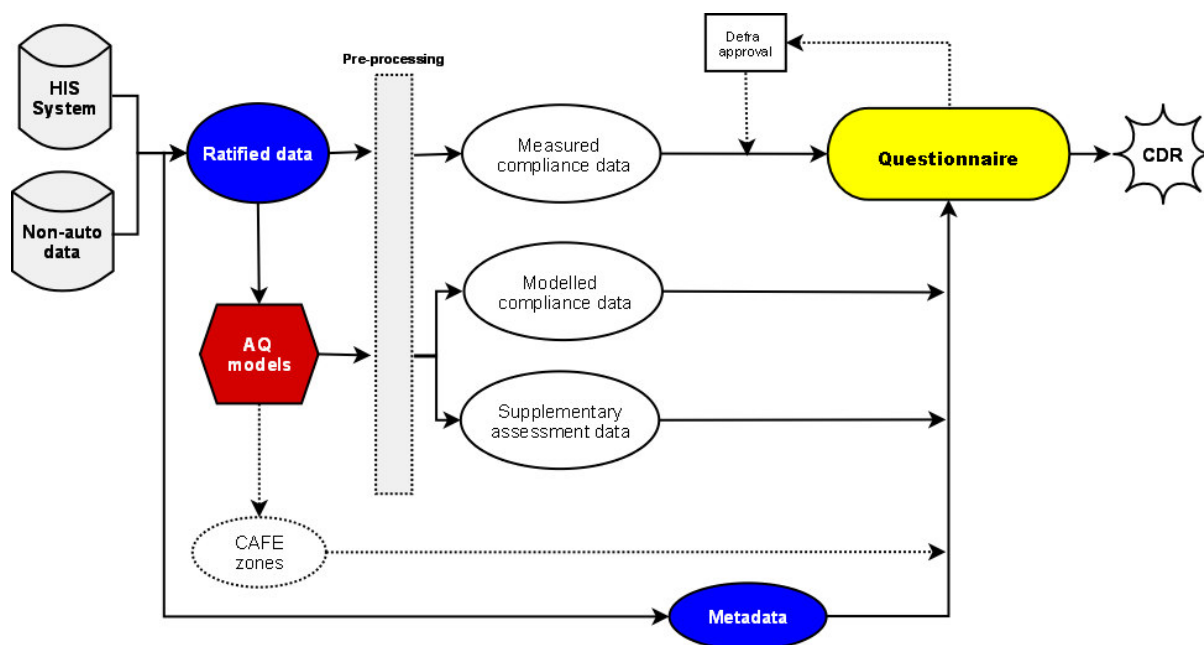
In compiling the Questionnaire, the AEA HIS system provides the ratified aggregated measurement statistics relevant to CAFÉ for all automatic monitoring stations. These are supplied as flat export files from the HIS statistical database. Ratified non-automatic data are provided directly by the network operators in flat file format. Modelled supplementary assessment and exceedence outputs from AEA's GIS enabled air quality models are provided by AEA's Pollution Climate Mapping team, again in flat file format. This team also provides supporting GIS outputs outlined Table 1 above, in the formats indicated.

The CAFÉ Questionnaire is populated with a combination of manual interactions from AEA's air quality experts (providing updates to the files supplied to AEA) and automated exports from the HIS statistical database. To a large extent the formats of non-automated datasets are beyond the control of AEA, these datasets being provided by external organisations. As a result, they necessitate manual processing and transfer of data into the Questionnaire format. However, this direct user interface does allow for a high degree of crosschecks and reality-checks to be performed using the expert judgement of the operators to input a level of QA/QC.

A schematic of the data flows and processes involved in the compilation of the Questionnaire is provided in Figure 9.

¹⁶ <http://ec.europa.eu/environment/air/quality/legislation/reporting.htm>

Figure 9 Data flows and processes involved in the compilation of the UK Questionnaire submission



Note: acronym reference list provided in Appendix 2

An important feedback loop in the Questionnaire compilation process relates to the approval by officials in UK Government. Only after formal approval by these responsible parties is the CAFÉ Questionnaire and supporting documents uploaded to the CDR. As part of this feedback loop the Questionnaire content is further scrutinised by UK Government's Department for Environment Food and Rural Affairs (Defra) air quality officials, anomalous data entries are identified and corrected by AEA. Once both Defra and AEA officials agree on the final content of the Questionnaire and its potential impacts within the European regulatory framework, the Secretary of State is briefed and the finalised Questionnaire approved for uploading to the CDR with supporting documents.

3 Implementation of INSPIRE concepts

The following section provides an appraisal of the existing systems within the UK for CAFÉ data themes presented in section 1.3. We present the systems, data, software and process gaps, which are to be closed in order to achieve an INSPIRE compliant system in terms of interoperability, services and metadata.

We have broken these down into three broad groups which require additional functionality/capability in order to support INSPIRE concepts. These groups relate to;

1. Data specifications to promote interoperability
2. Services to support discovery of data
3. Services to support download

3.1 Data specifications

3.1.1 Measurement datasets reported via the DEM

The preferred data formats used by the current DEM submissions are not INSPIRE compliant being based largely on ISO 7168-1 (amongst others). The file formats supported by the DEM are presented in Table 2.

Table 2 File formats supported by the existing DEM

Raw air quality data	Statistics	Meta information	Ozone Exceedences
ISO 7168:1985	2002/3/EC (only ozone)	DEM-Meta	2002/3/EC
ISO 7168-2:1998	ISO 7168-2:1998	ISO-7168-1:1998 (ext)	ISO-7168-1:1998 (ext)
ISO 7168-1:1998 (ext)	ISO 7168-1:1998 (ext)		
NASA-Ames 1001	DEM-AQ		
DEM-AQ	2002/3/EC (only ozone precursors)		
DEM-AQ (with Nasa Ames information)			

In order to achieve compliance with INSPIRE, we propose to migrate the existing reporting mechanism represented by the DEM submission and its metadata specification to the INSPIRE compliant harmonized Environmental Monitoring Facilities data specification¹⁷. However, we also recognise that although this specification is expected to accommodate the basic data requirements for air quality measurement stations, networks including configuration and observational data, the EMF data specification does not make provision for specific CAFÉ information elements (e.g. information on main emission sources close to the point of measurement, street canyon geometry information, CAFÉ Questionnaire codes for measurement methods etc). Although some of the specific CAFÉ data elements may be accommodated in the SupplementalInformation data element of the EMF schema, AEA's view is that some specific additions to this basic EMF schema may be needed to optimise support for CAFÉ reporting.

The current DEM data specification is provided in Appendix 1 of this report. Additional data elements have been added to this specification to support the draft CAFÉ IP¹⁸ and a provisional mapping of DEM data elements to the EMF data specification is provided.

¹⁷ Proposed Harmonized Data Specification Profile for Reporting 2008/50/EC Under 2007/2/EC, 27 March 2009, European Commission, Joint Research Centre, Institute for Sustainable Development.

¹⁸ Indicated in Appendix 1 by text in blue

Additional mandatory INSPIRE metadata elements

It is also clear from studying the INSPIRE compliant EMF data specification that there are some elements, mandatory in INSPIRE (and ISOs 19115 and 19119), which are not populated in the existing DEM datasets. A particular example being the level of information required for Responsible Party. Most of these data elements are mandatory in INSPIRE which covers parties responsible for data collection, analysis and reporting and is currently not found explicitly in the DEM data sets. We propose to populate the mandatory metadata requirements under INSPIRE and additional ISO elements where these add value to the data specifications particularly for tracking QA/QC status.

3.1.2 CAFE assessment information and zone datasets

The existing spreadsheet submission mechanism for reporting of assessment information is to be replaced with XML based documentation. The compilation of component datasets underlying the Questionnaire will continue to be carried out by AEA's HIS and dispersion modelling teams. Greater automation of this process will however be sought and the migration to XML based documentation will offer an opportunity for increased internal validation of document contents.

In order to achieve streamlining of the CAFE reporting mechanism AEA has investigated a clear demarcation of functionality in the reporting mechanisms;

- the DEM submission data files providing network and station configuration information and measurement data itself and, crucially, acting as the sole repository for this information
- the CAFE Questionnaire providing higher level summary of the exceedence situation with supporting high-level information to give context to the declaration of exceedence situations (providing data on; when, where, how much)

Our approach in this respect has been to analyse the functionality of each form in the Questionnaire and to identify if this is essential metadata for the Questionnaire and whether the content aligns best with the role of the Questionnaire or DEM. Table 3 presents a summary of this analysis.

On the basis of these observations AEA has streamlined the CAFE Questionnaire to concentrate on its core functionality as outlined below. We have developed a preliminary XML schema covering this core functionality for CAFE reporting as indicated in Table 3. Duplication of reporting effort within the Questionnaire and DEM data flows has been removed from the Questionnaire where this is perceived unnecessary as contextual information to assessment or compliance, although this information remain or will be added to the DEM prototype data specification. The XML schema will be consistent with INSPIRE concepts on mandatory metadata.

Table 3 Preliminary analysis of the component functionality of the CAFE questionnaire

Form(s)	Data type	Commentary	Action
Form 0	Metadata	MS submitting & contact details of the MS submitting, plus revision / creation details	Merge into single schema form 0-1
Form 1	Metadata	Further details on MS submitting & contact details plus notes	Merge into single schema
Form 2	Metadata	CAFE AQ zone information, including revision / creation details	Replace with reference to kml / gml / shp
Form(s) 3-6	Metadata	Standard details of monitoring stations in operation for each Directive & pollutant, plus measurement technique info. All this data should be in the DEM	Align / replace with DEM info
Form 7	Metadata	Metadata on measurement technique used could be included in DEM	Align / replace with DEM info
Form(s) 8-9	Dataset	Declaration of exceedence of LV, MOT, TV & LTOs	Merge into single schema

Form(s)	Data type	Commentary	Action
Form 10	Dataset	Declaration of exceedence of assessment thresholds	Merge into single schema
Form 11	Dataset	List of exceedence situations & reason codes	Merge into single schema
Form 12	Metadata	Optional additional reason codes, could be part of DEM	Align / replace with DEM info
Form(s) 13-16	Dataset	Not used in the UK but same function as 11,13,14,15 & 16	Align / replace with DEM info
Form 17	Dataset	Not used in the UK but same function as 11,13,14,15 & 16. Also no longer required under new Questionnaire for 2008	Align / replace with DEM info
Form 18	Dataset	Needs updating with revised CAFE IP for PM _{2.5}	Align / replace with DEM info
Form(s) 19	Dataset	Results from supplementary assessment needs updating with revised CAFE IP for PM _{2.5}	Merge into single schema
Form 20	Metadata	Reference to reports on supplementary assessment techniques	Merge into single form 19 schema
Form(s) 21, 23, 24	Dataset	Declaration of natural events (needs updating with revised CAFE IP for PM _{2.5})	Merge into single schema
Form 22	Metadata	Optional additional reason codes, could be part of DEM	Align / replace with DEM info
Form 23-24	Dataset	Declaration of natural events (needs updating with revised CAFE IP for PM _{2.5})	Merge into single schema
Form(s) 25	Metadata	Consultations on transboundary events	Merge into single schema
Form 26	Dataset	Declaration of exceedence of repealed LVs, TV & LTOs	Merge into single schema
Form 27	Metadata	Optional additional reason codes, could be part of DEM	Align / replace with DEM info

It has also been recognized that the existing reporting mechanism for CAFE zone information and supporting datasets such as supplementary assessment data (modelled outputs in the case of the UK) is poorly supported by the current Questionnaire. In the prototype XML document to replace the current Questionnaire format, provision will be made for declaration of all CAFE zone information and supporting documents/datasets via the reference to appropriate data schema for this purpose.

In order to ensure that the CAFE reporting document is not over burdened with detailed information on zone delimitation or supplementary assessment outputs we propose that references to these datasets be provided by xpath or URIs to files available and located on a public access storage area. Full metadata on the content and file formats of these supporting documents will be embedded in the CAFE reporting XML document and/or the supporting documents themselves.

3.2 Discovery metadata

The UK Archive will serve as the platform for INSPIRE compliant CAFE reporting services. Once the functionality of this online resource is extended, services will include:

1. Annual measurement data exchange and download (DEM extended to draft CAFE IP)
2. CAFE assessment information (Questionnaire extended to draft CAFE IP)
3. CAFE zone delimitation and supplementary assessment data (in appropriately marked GIS or NetCDF formats)
4. Near real-time data download exchange and download

Metadata relevant to discovery services have already been the subject of a collaborative report within this contract¹⁹. We also recognize, however, that the metadata files for discovery services will need to be updated regularly and registered with INSPIRE registries to take into account the frequent changes in metadata for near real-time data. An automated system will be required to generate up to date discovery metadata encompassing near real-time data on a routine and scheduled basis in order for the UK services to be discovered readily by users from the INSPIRE GoePortal and EEA.

The metadata files will be compliant to the INSPIRE discovery implementing rules²⁰, with each dataset being allocated an individual metadata set for discovery. Where there is commonality in data blocks, these will be stored centrally and utilized as a common resource. Other elements of the metadata files, for example the unique object identifiers, will be specific to the metadata file and generated automatically by either by internal systems or the external discovery registries.

There are extensive pre-existing services already on the UK Archive which are also important and useful to users although may not be INSPIRE compliant in their current formats. These services relate to:

- Near real-time and historical data enquiries via download services.
- Fixed format Annual measurement data files.
- View services via Google Maps and Earth.
- Raster based air quality background concentrations and forward projection details.

We propose that discovery metadata may be created for users to also discover these non-INSPIRE services as this has relevance also to promoting data sharing across user groups.

3.3 Download services

As we have seen in previous sections, reporting of measurement data via the DEM is currently the responsibility of the HIS system. This system was originally developed as an off-line QA/QC tool for the UK's centrally coordinated Air Quality Monitoring Networks. It was not developed with the aim of providing an online information system for air quality or CAFE datasets. As a result, continued use of this system in its current state will be a barrier to compliance with INSPIRE concepts particularly in relation to discovery, view and download of CAFE datasets. Our proposal, therefore, is to migrate reporting of CAFÉ measurement data themes (validated data via the DEM and near real-time data) and CAFÉ assessments data themes (including zones and supplementary assessment data) from the existing systems to the UK's Archive system to provide an INSPIRE compliant online CAFE reporting and download service.

Download services, in this context, are the mechanisms by which users can have access to CAFE reporting datasets transformed into INSPIRE compliant formats and architectures. The download services provide the capability to access the information in order to fulfil a series of possible tasks including the visualization of information in a variety of ways, integration with other information, and to allow analysis for knowledge and decision making. A CAFE download service shall be developed to give access to the dataset or a part of the data set. It will provide access to spatial objects whether representing discrete or continuous phenomena.

At a high level an INSPIRE compliant download service will allow users either directly or indirectly to access any available dataset. The services metadata elements will define if the service is a direct access download service or not.

If the service does not offer a query capability, then only a complete pre-defined spatial data set or a pre-defined part of a spatial data set can be downloaded. If the download service offers a query capability, it is a direct access download service whereby access to the service allows more control over the retrieved data than the simple download of a pre-defined dataset or pre-defined part of a dataset. It can therefore be considered to be more 'advanced' than the other download service variant.

¹⁹ Proposed Harmonized Metadata Specification Profile For Reporting 2008/50/EC Under 2007/2/EC by The 4 Lots

²⁰ INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119 (Revised edition)

In this document, we only consider the pre-defined data download services at this stage.

In further detail, INSPIRE download services are defined by five functions:

- Get Download Service Metadata
- Get Spatial Objects
- Describe Spatial Object Types
- Define Query
- Link Download Service.

This pilot study will concentrate on Get Download Service Metadata and Link Download Service, the others being beyond the scope of the current study for CAFE.

INSPIRE has recommended that metadata for download services be established in conformance with the corresponding Implementing Rule on the interoperability for spatial datasets.²¹ Detailed documentation on this subject is under preparation by the INSPIRE drafting teams (at time of writing this report). However, according to Article 11 of the INSPIRE Directive, download services require two kinds of metadata elements:

- Metadata for the INSPIRE Download Services.
- Metadata for each dataset represented in the Download Service.

²¹ D3.9 Draft Implementing Rules for Download Services

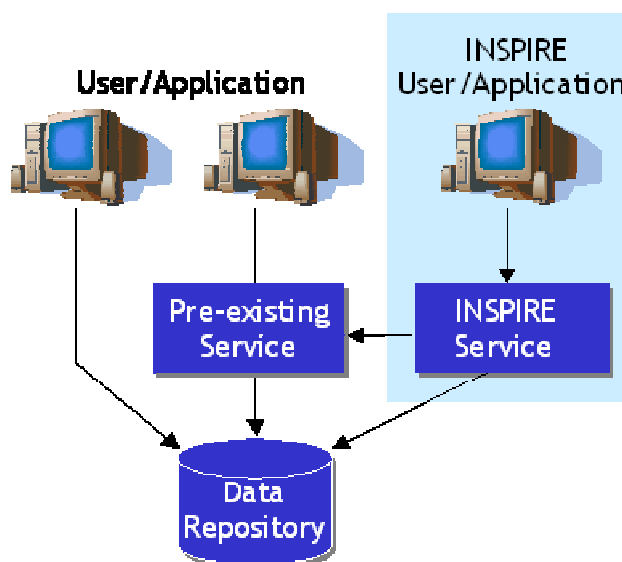
4 Architecture options for the prototype

A key component of this pilot study will relate to enabling spatial data sets to be transformed with a view to achieving interoperability. This is especially important with regard to download services. The existing UK Archive system will be augmented to provide the INSPIRE compliant services and achieve interoperability. This may be achieved by using a transformation service.

We propose that there are two core architecture options. One option, in which the INSPIRE services interact with the Archive system directly, mirroring its current non-INSPIRE functionality, as shown in Figure 10. Under this system, the XML files are generated from the existing database but constrained to the relevant Implementing Rules. The data and services are accessed via an INSPIRE service layer of the system.

A system of this architecture contains only one data repository and would necessitate a range of processes in order to support the INSPIRE services. Such a system would be a good choice if the existing data repository were fully INSPIRE compliant in terms of the database schema, metadata and data. In such an instance, data structure and processes can be shared by both INSPIRE users and native users. However, existing native data structures in the Archive system do not adhere to INSPIRE, being constructed prior to its implementation and most of the existing systems being designed to meet the particular needs of its local UK user groups.

Figure 10 Architecture for direct access INSPIRE services



Note: acronym reference list provided in Appendix 2

The second architecture option is illustrated in Figure 11. Under this architecture INSPIRE users may be provided with access to INSPIRE compliant datasets information via a separate route from the existing Archive users and applications.

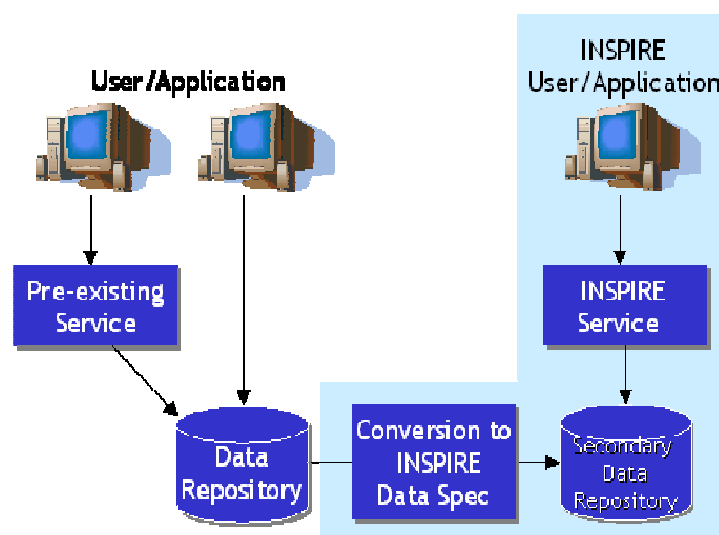
Firstly, the data in the native data repository (the Archive) are converted to a new data repository, with conformity at the service level through transformed virtual data sets. Secondly, the services are supported by separate dedicated file processes (INSPIRE services).

This architecture offers flexibility, being capable of maintaining existing non-INSPIRE compliant services required for internal MS obligations and providing an INSPIRE compliant service with least resource expenditure. Additional advantages to implementing such architecture relate to;

- Quick response to INSPIRE users or applications resulting in better performance;
- Less dependence, which will result in easier management of system changes.
- The system is more robust

For these reasons the architecture presented in Figure 11 is our proposed system for achieving INSPIRE compliance through data transformations²². Because the INSPIRE service layer is linked to the INSPIRE database schema and not to the native database, it is possible for a generic tool set to support the INSPIRE services to be deployed in the long-term. This architecture option also offers benefits to the pilot study being less resource intensive approach to supporting existing legacy functionality requirements of the Archive system, which was designed to serve UK air quality information and data needs. In their respective development phase(s) both systems must have minimal impact on the other to ensure service provision, even if it may be expected that both systems will converge in the future.

Figure 11 Architecture of indirect access INSPIRE services



Note: acronym reference list provided in Appendix 2

The UK Archive database architecture is currently based on MySQL as the data storage engine. MySQL implements spatial extensions following the specification of the OGC. However, at this moment in time MySQL does not yet implement certain GIS features and hence is currently not fully compliant with the *OpenGIS® Simple Features Specifications For SQL*.

Therefore in order to achieve the long-term goal of having a fully INSPIRE compliant database, able to offer the required web services as detailed above (with viewing services), the UK's INSPIRE data repository will use the PostGreSQL database system with the PostGIS extension. PostGIS adds support for geographic objects to the PostgreSQL object-relational database. In effect, PostGIS "spatially enables" the PostgreSQL server, allowing it to be used as a backend spatial database for geographic information systems. PostGIS follows the *OpenGIS® Simple Features Specifications For SQL* and has been certified as compliant with the "Types and Functions"

The architecture of the prototype is also an important consideration in the context of data transformation services, which in the case of the architecture presented in Figure 11, form an additional service layer between native and INSPIRE compliant content access services (Discovery/View/ Download). Transformation functionality may be either;

1. Incorporated as an integral part of the content service itself. As the functionality is then not exposed to the outside world, there are no standardization requirements involved.

²² Draft Implementing Rules for Transformation Services, Network Services Drafting Team, 10-02-2009.

2. As a layer closely coupled with the content access service. In this case the transformation module acts as an opaque proxy for the content access service and thus must be exposed to the calling application.

In this pilot study, we plan to use the first option, to allow for flexibility to implement different transformation tools.

4.1 Data transformations

Transformation Services may implemented to perform file format transformations, language translations, geometric transformations and schema translations for INSPIRE services. As part of this pilot study, schema and file format transformations will be carried out using separated processes.

The basic principle of data transformation in the context of data interchange is presented in the ISO Standard 19118:2005 Geographic information – Encoding. According to this standard data interchange may occur between different systems using a commonly agreed application schema, which is applied during the data transfer. The systems involved in the data interchange transaction also implement a mapping functionality, which translates the data instances from the internal (native) schema to the common schema (INSPIRE schema) and vice versa. After the dataset is mapped to the common schema, it is encoded into the data exchange language according to encoding rules defined in the standard above. The receiving system carries out the corresponding processes in a reverse.

4.1.1 Data schema transformation

A data mapping and transformation tool will be implemented within the LOT3 prototype to convert the native dataset to INSPIRE data repository. The resulting database shall match an agreed INSPIRE harmonized data schema. Because the scope of this pilot study is limited to direct access data as outlined earlier, we do not propose to carry out reverse transformation which would involve transforming the data query from INSPIRE request into native query functions, and output of data in INSPIRE compliant formats. The latter functions may be developed in-house at a later date. Preferred transformation tools are discussed in the following section.

INSPIRE database schema

Although our preferred database schema is not presented here, there are fundamental elements of the metadata with regard to the measurement of air pollution data which define the measurements made both spatially and temporal, and must be common between any database system, and as such will allow the transformation of data from one system to another. These fundamental elements are;

- **What** - What is being measured
- **Where** - Where are the measurements being made
- **When** - When are these measurements happening
- **How** - How are the measurements being made
- **How long** - Over what time period do these measurements refer

By associating the metadata from the UK Archive database, based upon the elements described above, with the corresponding elements of the INSPIRE compliant metadata it will be possible to transform the UK Archive database into an INSPIRE compliant database,

Xloader

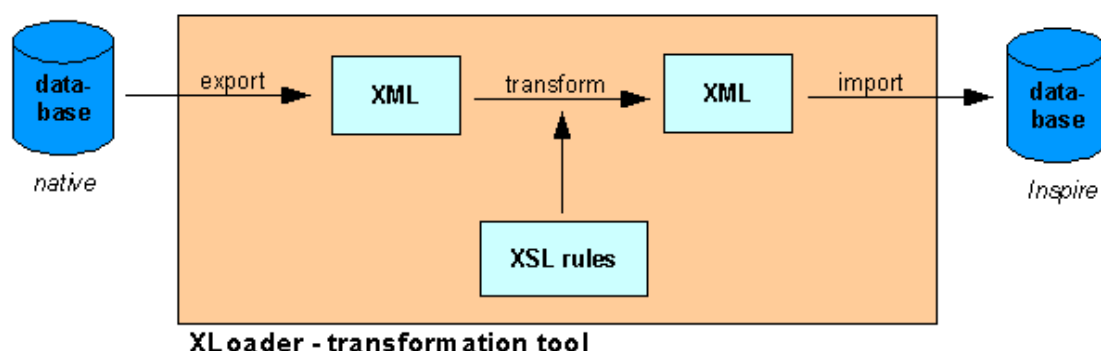
The Xloader ("cross loader") transformation tool is a Java based tool developed specifically for the INSPIRE EMF project. It fulfils core INSPIRE concept requirements being free-licensed and open-source. But it does not have all the features included in some other available products. The product is currently under trial within the UK Archive, AEA being non-contracted participants in the testing phase of this tool for the EMF project.

Xloader has been designed to work with two fundamental data sources, these are table-based and tree-based, each written in various formats. To accommodate this diversity in file formats, the tool is able to process table-based data (accessed via SQL) as well as tree-based data (accessed as XML). The UK Air Quality Archive uses a table-based database, and hence the Xloader tool will access the data via SQL.

Xloader will be used to export data from the Archive database, having first been configured with an appropriate agreed schema, output files from this export being XML documents. Subsequently, the exported XML files will be transformed using transformation rules written in XSLT language and stored in XSL files. Many rules can be defined in each XSL file and many XSL files can be used in one transformation run. The result of this transformation will be an XML file based on schema that maps directly to the destination database schema of the INSPIRE compliant EMF harmonized database schema or any other agreed schema identified. For example, the assessment information stored by the CAFE Questionnaire. The final action of transformation will be to populate the new database by importing the data from the XML files generated. A schematic of the transformation process is provided in Figure 12 below.

AEA have installed this tool and are testing it on the UK Air Quality Archive system.

Figure 12 Xloader – data schema transformation tool



Note: acronym reference list provided in Appendix 2

4.2 Additional archive processes to support INSPIRE services

Process gaps in existing UK services and systems have been identified relative to the INSPIRE implementing rules for different services, and the publishing workflow requirements. The main documents considered are given in Table 4.

Table 4 Information sources for the analysis

Dataset	Relevant Implementing Rules and Provisions	
	Metadata for Discovery Services	Data Specifications and application schema
DEM data files	<p>INSPIRE Prototype Environmental Monitoring Facilities Metadata Profile, Ref: BV/AG/GIS/10GG.0237029.</p> <p>Eol Decision, Council Decision 97/101/EC</p> <p>Implementing Provisions for reporting under Directive on ambient air quality and cleaner air for Europe Data exchange Group - Preparatory material</p> <p>INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119 (Revised edition)</p> <p>INSPIRE Metadata Regulation</p>	<p>Harmonised data specification Ref: BV/AG/GIS/AGGX0893/WP4.1</p> <p>Harmonised Data Specification profile for Reporting 2008/50/EC under 2007/2/EC</p> <p>Airbase Data Exchange Module (DEM) Manual, Version 11.0, June 2008</p> <p>Eol Decision, Council Decision 97/101/EC</p> <p>D2.7: Guidelines for the encoding of spatial data, Version 3.0</p>
CAFE questionnaire CAFE zone	<p>Proposed Harmonized Metadata Specification Profile For Reporting 2008/50/EC Under 2007/2/EC By The 4 Lots</p> <p>Implementing Provisions for reporting under Directive on ambient air quality and cleaner air for Europe Data exchange Group - Preparatory material</p> <p>Eol Decision, Council Decision 97/101/EC</p> <p>INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119 (Revised edition)</p> <p>INSPIRE Metadata Regulation</p>	<p>Proposed prototype harmonized data specification profile for reporting 2008/50/EC Under 2007/2/EC By The 4 Lots</p> <p>Implementing Provisions for reporting under Directive on ambient air quality and cleaner air for Europe Data exchange Group - Preparatory material</p>
INSPIRE download services	<p>D3.9 Draft Implementing Rules for Download Services, By Drafting Team "Network Services"</p> <p>WP 3.1 INSPIRE Prototype Environment Monitoring Facility Metadata Profile.</p> <p>Draft Technical Guidance Document for INSPIRE Discovery Services By Network Services Drafting Team</p>	In preparation

4.2.1 Processes to support INSPIRE compliant download services

Previous sections have shown how, at a conceptual level, the relevant data will be extracted from the INSPIRE data repository, output into XML file format based on the agreed bespoke data specifications and schemas. The exported datasets may then be stored on a pickup point on prototype the system. From this location users and applications will be able to discover, access and download CAFÉ datasets. The following section presents the processes required to achieve this within the prototype.

Air Quality Monitoring Data I: Near real-time data

The UK Archive already collects near real-time data and datasets relevant to CAFÉ alert and information thresholds, publishing these to UK systems for public information. These systems therefore need to be incorporated into the INSPIRE compliant CAFÉ services by:

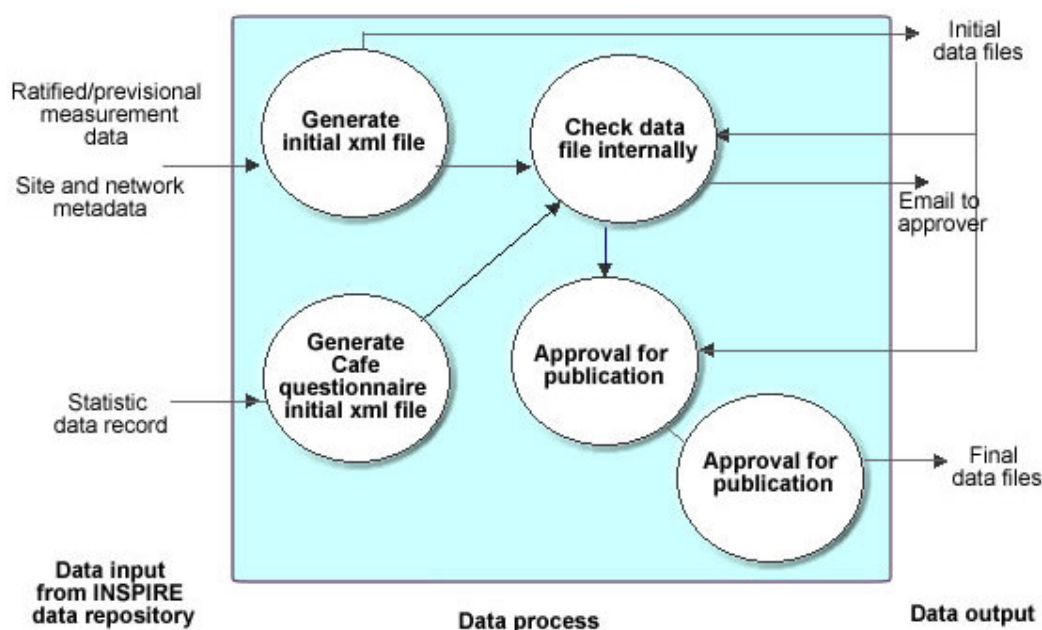
1. Transforming data into predefined XML data specifications and schema
2. Transferring data to the INSPIRE-CAFÉ data repository every hour
3. Mirroring updates to monitoring station and network metadata within the Archive to the INSPIRE-CAFÉ data repository
4. Generation and registration of metadata documents for discovery services

Essentially, this may be viewed as a synchronisation of UK's existing Archive systems and INSPIRE-CAFÉ datasets/schemas.

Air Quality Monitoring Data II (validated data) and CAFÉ assessment data themes

Owing to the official nature of the documents to be accessible via the download service, (being legally binding documents in many cases), an internal publication process and administration tool is needed to generate the files and support the process traceable way. This is particularly relevant to data services supporting the DEM and CAFÉ Questionnaire datasets, although less relevant to near real-time data which will be marked as officially 'provisional' using a QA/QC flag. The data flows under this system are illustrated in Figure 13.

Figure 13 Proposed publication system for generation of official CAFÉ documentation



Note: acronym reference list provided in Appendix 2

DEM and CAFE Questionnaire datasets are officially published by the end of September each year. Our proposal is to evaluate the use of an administrative front-end to facilitate this publication task allowing compilers to:

1. Define the data set type (e.g. measurement or assessments datasets)
2. Define the dataset content (data to be processed, monitoring sites, parameters, annual/exceedence statistics, etc.).
3. Populate provisional official data files within the secure area of INSPIRE data repository.
4. Validate file content based on CAFE validation rules.
5. Obtain both internal and external approvals to proceed to publication

As with existing UK systems, both internal and external crosschecks and reality-checks on the datasets will be required with input from air quality experts within AEA and UK Government. Provisional data files for these purposes may be viewed via the administrative tool, either on screen or via downloadable spreadsheet formats or forms. XMLT may be deployed for presenting the XML files in web pages.

After this checking process is complete at all levels, the responsible UK Government officials at Defra may be informed either by an email sent from the system automatically or offline. Nominated staff at Defra may then login to the system and view, download and approve the datasets via the secure web site. (We note that a formal hard copy of the datasets may be required for official ministerial sign-off).

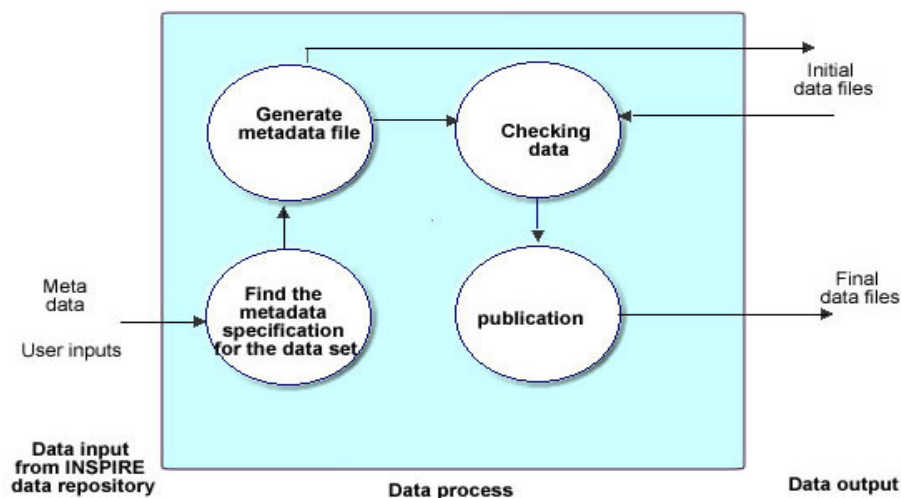
After official approval of the datasets by UK Government, the system can then perform three tasks;

1. Move the files to a public access area of the download service for pick up or access by users and applications.
2. Update the content of the data files to reflect the official status of the datasets, publication date, version numbering etc.
3. Generate metadata documents relevant to the new datasets for registration with discovery services.

4.2.2 Processes to support discovery metadata services

The content of metadata files for discovery services will be relatively consistent over the lifetime of the datasets and services. They will be generated according to the INSPIRE IR by using relevant processes. Figure 14 provides a high level illustration of the processes involved. The resulting metadata files will be validated and may be made available to INSPIRE registries either via a pickup point on the UK INSPIRE–CAFE download service or sent via email or ftp to the INSPIRE Geoportal for the registration.

Figure 14 Processes involved in the generation of INSPIRE-CAFE metadata files for discovery services



Note: acronym reference list provided in Appendix 2

4.2.3 UK INSPIRE-CAFE data repository file structure

Non-real time CAFE data themes will routinely be updated on an annual basis and contain data relevant to air pollution levels and assessment information for a full calendar year. This administrative structure is reflected in the file structure on the EIONET CDR for both DEM and CAFE data themes. We propose to mirror this file structure within the UK prototype. A generic example of this file structure and file nomenclature is provided in Table 5.

File structures for near real-time data are potentially more diverse depending on the level of functionality offered by the prototype. Table 5 presents a simple option for near real-time data files where a single file is made available by pollutant and threshold. This file would theoretically be generated on the 1 January each year or for the 1st hour in the year for which an exceedance is measured and subsequently incrementally updated for each additional exceedance as data becomes available.

An alternative to this option would be to provide users with a queryable interface, allowing for user defined downloads at higher temporal resolution for specified threshold metrics or pollutants. To accommodate this potentially large variety of query conditions, it would be necessary to record the query parameters entered when a user selects the data and create a unique identifier associated with the query. Subsequently, when the extracted data are placed in a data file it would then be necessary to associate this unique identifier with the data requested and as such this unique identifier will be used as part of the download file name. We will evaluate the preferred options for provision of download services for near real-time datasets.

Table 5 Generic UK INSPIRE-CAFE repository file structure and nomenclature

Dataset and shared metadata		Directory and file name structure	
DEM	Data files	INSPIRE_pickup/DEM/data/2008/	UK_DEM_2008_ReposibleParties.xml UK_DEM_2008_network.xml UK_DEM_2008_station.xml UK_DEM_2008_measurement.xml UK_DEM_2008_sensor.xml
	Discovery metadata	INSPIRE_pickup/DEM/	UK_DEM_metadata
CAFE questionnaire dataset	Data files	INSPIRE_pickup/CAFE/data/2008/	UK_cafe_Questionnaire_ID1_2008.xml UK_cafe_AQ_zones_2008.gml UK_cafe_roadside_modelled_data_2008.gml UK_cafe_modelled_data_2008_no2.nc UK_cafe_modelled_data_2008_pm10.nc UK_cafe_modelled_data_2008_pm2_5.nc UK_cafe_modelled_data_2008_so2.nc UK_cafe_modelled_data_2008_pb.nc UK_cafe_modelled_data_2008_co.nc UK_cafe_modelled_data_2008_bz.nc UK_cafe_modelled_data_2008_o3.nc UK_cafe_roadside_modelled_data_2008.gml
	Discovery metadata	INSPIRE_pickup/CAFE/	UK_cafe_metadata
Near real-time download services	Data files	INSPIRE_pickup/CAFE/	UK_cafe_metadata
		INSPIRE_pickup/real_time/data/	UK_cafe_alert_exceedence_2008_so2.xml UK_cafe_alert_exceedence_2008_no2.xml UK_cafe_alert_exceedence_2008_o3.xml UK_cafe_info_exceedence_2008_o3.xml UK_DEM_2008_ReposibleParties.xml UK_DEM_2008_network.xml UK_DEM_2008_station.xml UK_DEM_2008_measurement.xml UK_DEM_2008_sensor.xml
	Discovery metadata	INSPIRE_pickup/real_time/	UK_air_quality_metadata

5 Summary and conclusions

5.1 Key gaps in the existing UK systems relative to INSPIRE IR

The main gaps in the UK's existing reporting and data exchange systems under the CAFE Directive have been assessed relative to the INSPIRE IR, concepts and draft CAFE IP. Gaps have been identified in three key areas:

1. Data specifications and formats to promote interoperability
2. Services to support download
3. Services to support discovery of data

The key findings for data provision are summarised below.

In order to improve data interoperability, the prototype harmonized EMF data specification shall be applied for the reporting of all measurement datasets under CAFE (currently the DEM submission of validated data and near real-time data reporting). It has been recognised however, that in some instances this data specification schema may require extending to accommodate CAFE specific information instances.

A preliminary mapping of existing DEM and draft CAFE IP data specifications has been made against the EMF data specification. This information is provided in Appendix 1 and shows that to a large extent the CAFE specific information instances not explicitly existing in the EMF data specification may be accommodated via the use of the EMF's SupplementalInformation data element. However, we recommend further evaluation of data requirements is needed to ensure data specifications are optimised for CAFE reporting.

An XML based preliminary data specification and schema for reporting of CAFE assessment information, delimitation of CAFE zone's and supporting documentation (supplementary assessment data in the form of modelled outputs) is required and under development by LOT3.

In the majority of CAFÉ data themes within the scope of this pilot study, mandatory INSPIRE metadata elements are missing. In particular, metadata elements such as responsible parties, data lineage, versions controlling and other QA/QC related INSPIRE metadata elements are missing. Proposals have been made to populate these mandatory elements and to provide additional ISO 19115 and 19119 elements where these add value to the data specifications.

Existing UK download services are compliant, at a high-level, with INSPIRE concepts for such services. However, existing viewing services are neither INSPIRE compliant nor directly focused on CAFE datasets. The main gap relative to INSPIRE for existing data download services relates to file formats used. Therefore, the key gap relative to supporting INSPIRE concepts in the UK's CAFE data themes systems, relates to the transformation of data and the processes supporting this activity to achieve interoperability.

Metadata to support discovery services are missing from the UK's existing systems. A CAFE specific metadata profile relevant to such services has already been the subject of a collaborative report within this contract²³. This profile will be used within an automated system to generate up to date metadata to support discovery services for all UK CAFE data themes. Metadata will be registered with INSPIRE registries.

²³ Proposed Harmonized Metadata Specification Profile For Reporting 2008/50/EC Under 2007/2/EC by The 4 Lots

5.2 Summary of steps to close the gaps

The following bullets summarise the steps to be implemented in order to close the gaps between existing UK CAFE reporting systems and the draft CAFE IP and INSPIRE IR.

1. Migrate all CAFE datasets and metadata relating to air pollution measurements (validated and near real-time) and CAFE assessment information (including zones and supplementary assessment data) to the UK Archive system using the PostGreSQL database system with the PostGIS extension
2. Develop and implement appropriate data specifications and XML schema for CAFE data themes encompassing
 - a. the CAFE Questionnaire
 - b. CAFE Questionnaire supplementary assessment datasets (modelled data)
 - c. CAFE zones delimitation
 - d. measurement datasets and metadata (based on an extended harmonized EMF data specification)
3. Develop INSPIRE data repository including:
 - a. Secure area for data testing / approval
 - b. Public access pickup point / download service for validated measurements and CAFE assessment information
 - c. Public access pickup point / download service for near real-time data
 - d. Public access pickup point / download service for metadata for discovery
4. Develop CAFE focused transformation tools to migrate CAFE datasets to INSPIRE compliant data specifications and data formats
5. Implement transformation in an administration tool to generate and maintain datasets and metadata for discovery services.

We also recommend the development of a CAFE code list or vocabulary resource to support the CAFE data specifications and validation of CAFE dataset content. We would welcome input from ReportNet, JRC and other air quality experts within the LOTs on the development of such a resource.

Appendices

Appendix 1: Extended DEM data specifications with preliminary mappings to the harmonised EMF data specification

Appendix 2 List of acronyms

Appendix 1

Extended DEM data specifications incorporating draft CAFÉ IP and with preliminary mappings to the harmonised EMF data specification

Contents

Appendix Table 1	Monitoring network metadata specifications
Appendix Table 2	Monitoring station metadata specifications
Appendix Table 3	Measurement instrument configuration specifications

Appendix Table 1 Monitoring network metadata specifications
 Additional data elements not part of the DEM v11 data specification and relevant to draft CAFÉ IP shown in blue text

CAFE Requirement				EMF Mapping			
Keyword	Description	Type of parameter	Domain	EMF FeatureType	EMF attribute	Domain	Multiplicity
Dem_status	type of mutation		A = add network M = mutate network D = delete network "" = no mutation	EnvironmentalMonitoringNetwork	status	StatusCodes: Active Inactive deprecated	0..1
Network_code	Identification network	Key	CCNNXX, where: CC=ISO-country code NNN=serial number within Country X=A (id. general AirBase Network)	EnvironmentalMonitoringNetwork	NetworkID	ScopedName	0..1
Network_name	Name of network	Mandatory	1-255 characters	EnvironmentalMonitoringNetwork	name	CharacterString	1
Network_short_name	Abbreviation network	Optional	1-10 characters	-	-	-	-
Network_start_date	Date network operational	Essential	YYYYMMDD, Valid date, > 19000101 <= current date	EnvironmentalMonitoringNetwork	operationPeriod	TM_Object	1
Network_end_date	Date network shut down	Optional	YYYYMMDD, Valid date, > network_start_date				
Network_time_reference_basis	Time reference results reported for network	Mandatory	UTC or LOCAL (time reference basis defined for country)	-	-	-	-
Network_type	Type of network	Desirable	Local, urban, regional, national	EnvironmentalMonitoringNetwork	Type	GenericName	
Network_qaqc_classes	Information about status QA/QC defined for network	Desirable	1, 2a, 2b, 3, 4, 5	-	-	-	-
Organization_name	Organization name network-manager and data supplier	Mandatory	1-80 characters	EnvironmentalMonitoringNetwork	ResponsibleParty Role needs to be defined: owner sponsor manager dataSupplier	EMF_ResponsibleParty → organisationName	
Organization_address	Organization address network-manager and data supplier	Mandatory	1-160 characters	EnvironmentalMonitoringNetwork	ResponsibleParty	EMF_ResponsibleParty → address	

CAFE Requirement				EMF Mapping			
Keyword	Description	Type of parameter	Domain	EMF FeatureType	EMF attribute	Domain	Multiplicity
Organization_city	Organization city network-manager and data supplier	Mandatory	1-30 characters	EnvironmentalMonitoringNetwork	ResponsibleParty	EMF_ResponsibleParty → address	
Organization_website_address	Organization website address network-manager and data supplier	Optional	7-80 characters, format: X://X.X where X = 1 or more character(s)	EnvironmentalMonitoringNetwork	ResponsibleParty	EMF_ResponsibleParty → URL	
Organization_phone_number	Organization full phone-nmbr (incl. nat. code) network-manager and data supplier	Mandatory	2-16 characters, format: +NNN where N=1 or more figures[0-9]	EnvironmentalMonitoringNetwork	ResponsibleParty	EMF_ResponsibleParty → telephone	
Organization_fax_number	Organization fax-nmbr network-manager and data supplier	Mandatory	2-16 characters, format: +NNN where N=1 or more figures[0-9]	EnvironmentalMonitoringNetwork	ResponsibleParty	EMF_ResponsibleParty → faxNumber	
Manager_organization_name	Organization name network-manager	Mandatory	1-80 characters	EnvironmentalMonitoringNetwork	ResponsibleParty Role = manager	EMF_ResponsibleParty → organisationName	
Manager_organization_address	Organization address network-manager	Mandatory	1-160 characters	EnvironmentalMonitoringNetwork	ResponsibleParty Role = manager	EMF_ResponsibleParty → address	
Manager_organization_city	Organization city network-manager	Mandatory	1-30 characters	EnvironmentalMonitoringNetwork	ResponsibleParty Role = manager	EMF_ResponsibleParty → address	
Manager_organization_website_address	Organization website address network-manager	Optional	7-80 characters, format: X://X.X where X = 1 or more character(s)	EnvironmentalMonitoringNetwork	ResponsibleParty Role = manager	EMF_ResponsibleParty → URL	
Manager_organization_phone_number	Organization full phone-nmbr (incl. nat. code) network-manager	Mandatory	2-16 characters, format: +NNN where N=1 or more figures[0-9]	EnvironmentalMonitoringNetwork	ResponsibleParty Role = manager	EMF_ResponsibleParty → telephone	

CAFE Requirement				EMF Mapping			
Keyword	Description	Type of parameter	Domain	EMF FeatureType	EMF attribute	Domain	Multiplicity
Manager_organization_fax_number	Organization fax-nmbr network-manager	Mandatory	2-16 characters, format: +NNN where N=1 or more figures[0-9]	EnvironmentalMonitoringNetwork	ResponsibleParty Role = manager	EMF_ResponsibleParty → faxNumber	
Manager_person_last_name	Lastname network-manager	Mandatory	1-20 characters	EnvironmentalMonitoringNetwork	ResponsibleParty Role = manager	EMF_ResponsibleParty → individualName	
Manager_person_first_name	Firstname network-manager	Optional	1-20 characters	EnvironmentalMonitoringNetwork	ResponsibleParty Role = manager	EMF_ResponsibleParty → individualName	
Manager_person_title	Title network-manager	Optional	1-20 characters	EnvironmentalMonitoringNetwork	ResponsibleParty Role = manager	EMF_ResponsibleParty → individualName	
Manager_person_email_address	Email-address network-manager	Desirable	5-80 characters, Format:X@X.X Where X = 1 or more character(s)	EnvironmentalMonitoringNetwork	ResponsibleParty Role = manager	EMF_ResponsibleParty → emailAddress	
Data_supplier_organization_name	Organization name data supplier	Desirable	1-80 characters	EnvironmentalMonitoringNetwork	ResponsibleParty Role = dataSupplier	EMF_ResponsibleParty → organisationName	
Data_supplier_organization_address	Organization address data supplier	Optional	1-160 characters	EnvironmentalMonitoringNetwork	ResponsibleParty Role = dataSupplier	EMF_ResponsibleParty → address	
Data_supplier_organization_city	Organization city data supplier	Optional	1-30 characters	EnvironmentalMonitoringNetwork	ResponsibleParty Role = dataSupplier	EMF_ResponsibleParty → address	
Data_supplier_organization_website_address	Organization website address data supplier	Optional	10-80 characters, format: http://X.X where X = 1 or more character(s)	EnvironmentalMonitoringNetwork	ResponsibleParty Role = dataSupplier	EMF_ResponsibleParty → URL	
Data_supplier_organization_phone_number	Organization full phone-nmbr (incl. nat. code) data supplier	Optional	2-16 characters, format: +NNN where N=1 or more figures[0-9]	EnvironmentalMonitoringNetwork	ResponsibleParty Role = dataSupplier	EMF_ResponsibleParty → telephone	
Data_supplier_organization_fax_number	Organization fax-nmbr data supplier	Optional	2-16 characters, format: +NNN where N=1 or more figures[0-9]	EnvironmentalMonitoringNetwork	ResponsibleParty Role = dataSupplier	EMF_ResponsibleParty → faxNumber	

CAFE Requirement				EMF Mapping			
Keyword	Description	Type of parameter	Domain	EMF FeatureType	EMF attribute	Domain	Multiplicity
Data_supplier_person_last_name	Lastname data-supplier	Desirable	1-20 characters	EnvironmentalMonitoringNetwork	ResponsibleParty Role = dataSupplier	EMF_ResponsibleParty → individualName	
Data_supplier_person_first_name	Firstname data-supplier	Optional	1-20 characters	EnvironmentalMonitoringNetwork	ResponsibleParty Role = dataSupplier	EMF_ResponsibleParty → individualName	
Data_supplier_person_title	Title data-supplier	Optional	1-20 characters	EnvironmentalMonitoringNetwork	ResponsibleParty Role = dataSupplier	EMF_ResponsibleParty → individualName	
Data_supplier_person_email_address	Email-address data-supplier	Desirable	5-80 characters, Format:X@X.X Where X = 1 or more character(s)	EnvironmentalMonitoringNetwork	ResponsibleParty Role = dataSupplier	EMF_ResponsibleParty → emailAddress	

Appendix Table 2 Monitoring station metadata specifications

Additional data elements not part of the DEM v11 data specification and relevant to draft CAFÉ IP shown in blue text

CAFE Requirement				EMF Mapping			
Keyword	Description	Type of parameter	Domain	EMF FeatureType	EMF attribute	Domain	Multiplicity
Dem_status	type of mutation		A = add station M = mutate station D = delete station "" = no mutation	Environmental MonitoringFacility	Status	StatusCode: active inactive deprecated	0..1
Station_code	Identification station	Key	CCNNNNX, where: CC=ISO-country code NNNN=serial number within Country X=A : id. general AirBase Station G,U,R,I: id. Station type defined by EMEP	Environmental MonitoringFacility	InternationalIDCode	ScopedName	0..1
Network_code	Identification network	Key	CCNNNNX, where: CC=ISO-country code NNN=serial number within Country X=A (id. general AirBase Network)	-			
Station_local_code	Local identification code	Desirable	1-10 characters	Environmental MonitoringFacility	identificationCode	ScopedName	1
Station_name	Name of station	Mandatory	1-60 characters, no double quote (")	Environmental MonitoringFacility	Name	CharacterString	0..1
station_description	Additional information station	Optional	1-240 characters	Environmental MonitoringFacility	Description	CharacterString	0..1
Station_nuts_level1	NUTS-information (level 1)	Desirable	1 character	EnvironmentalMonitoringFacility → Location	geographicDescriptor	Value	0..*
Station_nuts_level2	NUTS-information (level 2)	Desirable	1 character	EnvironmentalMonitoringFacility → Location	geographicDescriptor	Value	0..*
Station_nuts_level3	NUTS-information (level 3)	Desirable	1 character	EnvironmentalMonitoringFacility → Location	geographicDescriptor	Value	0..*
Station_lau_level1	LAU1-information	Optional	2 characters (If the LAU1 code starts with 0, a single quote (') has to be placed before the 0)	EnvironmentalMonitoringFacility → Location	geographicDescriptor	Value	0..*

CAFE Requirement				EMF Mapping			
Keyword	Description	Type of parameter	Domain	EMF FeatureType	EMF attribute	Domain	Multiplicity
Station_lau_level2	LAU2-information	Optional	3 characters (If the LAU2 code starts with 0, a single quote (') has to be placed before the 0)	EnvironmentalMonitoringFacility → Location	geographicDescriptor	Value	0..*
Station_lau_name2	LAU2 name	Optional	1-255 characters	EnvironmentalMonitoringFacility → Location	geographicDescriptor	Value	0..*
Station_sabe_icc	Country ID	Optional	1-2 characters	EnvironmentalMonitoringFacility → Location	geographicDescriptor	Value	0..*
Station_sabe_shn	Sabe ID	Optional	1-20 characters	EnvironmentalMonitoringFacility → Location	geographicDescriptor	Value	0..*
Station_sabe_gen	Sabe name	Optional	1-80 characters	EnvironmentalMonitoringFacility → Location	geographicDescriptor	Value	0..*
Station_start_date	Date station operational	Essential	YYYYMMDD, Valid date, >= network_start_date	EnvironmentalMonitoringFacility	operationPeriod	TM_Object	0..1
Station_end_date	Date station shutdown		YYYYMMDD, Valid date, > station_start_date				
Station_latitude_d	Co-ordinate station: latitude defined as decimal value	Mandatory	[<SIGN>] <DEG>.<MIN>.<SEC> where: SIGN= [-,+], default SIGN=+ DEG = [0-90] MIN_SEC=[0-999999]	Specialised Environmental Monitoring Facility → MonitoringPoint	Position	GM_Point GM_Curve GM_Surface GM_Solid (includes CRS)	1
Station_latitude_dms	Co-ordinate station: latitude defined as text value	Mandatory	[<SIGN>]<DEG>.<MIN>.<SEC> where: SIGN=[-,+], default SIGN=+ DEG=[0-90] MIN=[0-59] SEC=[0-59]				
Station_longitude_d	Co-ordinate station: longitude defined as decimal value	Mandatory	[<SIGN>] <DEG>.<MIN>.<SEC> where: SIGN= [-,+], default SIGN=+ DEG = [0-180] MIN_SEC=[0-999999]				
Station_longitude_dms	Co-ordinate station: longitude defined as text value	Mandatory	[<SIGN>]<DEG>.<MIN>.<SEC> where: SIGN=[-,+], default SIGN=+ DEG= [0-180] MIN= [0-59] SEC= [0-59]				
Station_altitude	Altitude station in meters	Mandatory	Valid integer value	SpecialisedMonitoringFacility	elevation	Measure (include units)	0..1
Type_of_station	Type of station	Essential	"traffic", "industrial", "background", "unknown"	EnvironmentalMonitoringFacility	Purpose	GenericName	0..*

CAFE Requirement				EMF Mapping			
Keyword	Description	Type of parameter	Domain	EMF FeatureType	EMF attribute	Domain	Multiplicity
Station_type_of_area	Type of area in which the station is located	Mandatory	"urban", "suburban", "rural", "unknown"	EnvironmentalMonitoringFacility	Type	GenericName	0..*
Station_characteristic_of_zone, list	One or more characteristics of the zone in which the station is located	Optional	"unknown", "residential", "commercial", "industrial", "agricultural", "natural", "airport", "park", "mountain", "valley", "seaside", "lakeside"	EnvironmentalMonitoringFacility	Type	GenericName	0..*
station_ozone_classification	type of area in which the station is located for ozone stations (Ozone classification)	Optional	"urban", "suburban", "rural", "rural background"	EnvironmentalMonitoringFacility	Type	GenericName	0..*
Station_main_emission_source list	One or more main emissions in area in which station is located	Optional	"Public power, co-generation and district heating", "Agriculture", "Nature", "Other", "Commercial, institutional and residential combustion", "Industrial combustion", "Production processes", "Extraction and distribution of fossil fuels", "Use of solvents", "Road transport", "Other mobile sources and machinery (specify)", "Waste treatment and disposal"				
Station_type_emission	Station type in relation to main emission source	Mandatory	Traffic, industrial, background	Repeats type of station			
Main_emission_type		Mandatory for industrial emission	NFR code <i>not</i> SNAP ?	-	-	-	-
Emission_distance	Distance to main emission source	Mandatory for industrial emission	Metres	EnvironmentalMonitoringFacilities	SupplementalInformation	SupplementalInformation	0..*
Emission_rt	RT emission	Voluntary	tonnes	EnvironmentalMonitoringFacilities	SupplementalInformation	SupplementalInformation	0..*

CAFE Requirement				EMF Mapping			
Keyword	Description	Type of parameter	Domain	EMF FeatureType	EMF attribute	Domain	Multiplicity
Emission_dom	Domestic emission	Voluntary	tonnes	EnvironmentalMonitoringFacilities	SupplementalInformation	SupplementalInformation	0..*
Emission_ind	Industrial emission	Voluntary	tonnes	EnvironmentalMonitoringFacilities	SupplementalInformation	SupplementalInformation	0..*
Station_area_of_representativ	Area (defined by radius in km) for which the monitoring results are representative	Optional	>= 0	-	-	-	-
Station_distance_from_kerb	Distance (in km) between station and kerbstone	Optional	>= 0	EnvironmentalMonitoringFacilities	SupplementalInformation	SupplementalInformation	0..*
Station_distance_to_source	Distance (in km) to emission source	Optional	>= 0	EnvironmentalMonitoringFacilities	SupplementalInformation	SupplementalInformation	0..*
Station_distance_to_junction	Distance (in m) to traffic junction	Optional	>= 0	EnvironmentalMonitoringFacilities	SupplementalInformation	SupplementalInformation	0..*
Station_snap, list	One of more SNAP-codes. A SNAP-code is a combination of level1-level2-level3 code	Optional	Format "<L1><L2><L3>" where format <L1>, <L2>, <L3> = "NN" N = [0,9]	EnvironmentalMonitoringFacility	Type	GenericName	0..*
Station_subcat_rural_background	Sub-category defined for rural background stations	Optional	"near city", "regional", "remote", "unknown"	EnvironmentalMonitoringFacility	Type	GenericName	0..*
Station_city	City-name in which station is located	Essential/ Desirable	1-30 characters	EnvironmentalMonitoringFacility → Location	geographicDescriptor or address	Value CharacterString	0..*
Population	Number (/ 1000) of inhabitants in Station_city	Desirable	>= 0	EnvironmentalMonitoringFacilities	SupplementalInformation	SupplementalInformation	0..*
Street_name	Streename in which station is located	Desirable	1-40 characters	EnvironmentalMonitoringFacility → Location	geographicDescriptor or address	Value CharacterString	0..*

CAFE Requirement				EMF Mapping			
Keyword	Description	Type of parameter	Domain	EMF FeatureType	EMF attribute	Domain	Multiplicity
Street_type	Type of street	Desirable	"canyon street", "highway", "wide street", "unknown"	EnvironmentalMonitoringFacility	Type	GenericName	0..*
Number_of_vehicles	Number (/1000) of vehicles in 24 hours passing station	Optional	>= 0	EnvironmentalMonitoringFacilities	SupplementalInformation	SupplementalInformation	0..*
Station_lorry_percentage	Heavy-duty fraction of traffic	Optional	>= 0 and <= 100	EnvironmentalMonitoringFacilities	SupplementalInformation	SupplementalInformation	0..*
Station_traffic_speed	Traffic speed (in km/hour)	Optional	>=0 and < 200	EnvironmentalMonitoringFacilities	SupplementalInformation	SupplementalInformation	0..*
Station_ratio_distance	Ratio distance/heigh building facades	Optional	>= 0	EnvironmentalMonitoringFacilities	SupplementalInformation	SupplementalInformation	0..*
Station_street_width	Street-width in m.	Optional	>= 0	EnvironmentalMonitoringFacilities	SupplementalInformation	SupplementalInformation	0..*
Monitoring_objective, list	Objective(s) for monitoring	Optional	>= 0 and <=255	ReportingInformation	LegislativeObjective → obligation	CharacterString	0..*
Other_monitoring_objective	Objective other than defined in AirBase	Optional	1-160 characters	ReportingInformation	LegislativeObjective → obligation	CharacterString	0..*
Exposure	Kind of exposure	Optional	"population", "ecosystem", "material"	EnvironmentalMonitoringFacilities	SupplementalInformation	SupplementalInformation	0..*
Meteorological_parameters list	List of meteorological parameters	Mandatory	99 or list of ISO-codes ISO-code: >= 51 and < 99	EnvironmentalMonitoringFacilities	SupplementalInformation	SupplementalInformation	0..*
Local_dispersion_situation	Describes impact of buildings on dispersion	Voluntary	CAFE IP code list – Jan 2009	EnvironmentalMonitoringFacilities	SupplementalInformation	SupplementalInformation	0..*
Regional_dispersion_situation	Describes impact of topographic situation	Voluntary	CAFE IP code list – Jan 2009	EnvironmentalMonitoringFacilities	SupplementalInformation	SupplementalInformation	0..*

Appendix Table 3 Measurement instrument configuration specifications

Additional data elements not part of the DEM v11 data specification and relevant to draft CAFÉ IP shown in blue text

CAFE requirement				EMF Mapping			
Keyword	Description	Type of parameter	Domain	EMF FeatureType	Attribute	Domain	Multiplicity
Dem_status	type of mutation		A = add measurement configuration M = mutate measurement configuration D = delete measurement configuration "" = no mutation	Currently don't have a status value on observedProperty			
station_code	Identification station	Key	CCNNNNX, where: CC=ISO-country code NNNN=serial number within Country X=A : id. general AirBase Station G,U,R,I: id. Station type defined by EMEP	Repeated attribute as this is encoded within EnvironmentalMonitoringFacility	identificationCode	ScopedName	1
CAFE_zone_code	Name of the CAFE zone	Mandatory		Repeated attribute as should be encoded within EnvironmentalMonitoringFacility Feature Type	Location	Location → geographicDescription	0..*
station_name	name of the station	Optional	1-60 characters, no double quote (")	Repeated attribute as this is encoded within EnvironmentalMonitoringFacility	name	CharacterString	0..1
component_code	Identification component	Mandatory	> 0	ObservedProperty → PropertyValue	propertyCode	GenericName	0..1
component_name	name of the component	Optional	1-40 characters	ObservedProperty → PropertyValue	propertyName	GenericName	1
component_short_name	component short name	Optional	1-30 characters	ObservedProperty → PropertyValue	alternatePropertyName	GenericName	0..1
zone_objective_code1	first component zone objective code		"human", "eco", "human/eco", "ozone"	- This information will be encoded in the Annex III theme Area Management/Restriction/Regulation and Reporting Units			

CAFE requirement				EMF Mapping			
Keyword	Description	Type of parameter	Domain	EMF FeatureType	Attribute	Domain	Multiplicity
zone_code1	first component zone code		1-20 characters	Repeated attribute as should be encoded within EnvironmentalMonitoringFacility Feature Type	Location	Location → geographicDescription (this should be encoded using Xlink href to the zone identifier in the Annex III theme Area Management/Restriction/Regulation and Reporting Units)	0..*
zone_name1	first component zone name	Optional	1-40 characters	Repeated attribute as should be encoded within EnvironmentalMonitoringFacility Feature Type	Location	Location → geographicDescription (this should be encoded using Xlink href to the zone identifier in the Annex III theme Area Management/Restriction/Regulation and Reporting Units)	0..*
zone_objective_code2	second component zone objective code		"human", "eco", "human/eco", "ozone"	- This information will be encoded in the Annex III theme Area Management/Restriction/Regulation and Reporting Units			
zone_code2	Second component zone code		1-20 characters	Repeated attribute as should be encoded within EnvironmentalMonitoringFacility Feature Type	Location	Location → geographicDescription (this should be encoded using Xlink href to the zone identifier in the Annex III theme Area Management/Restriction/Regulation and Reporting Units)	0..*
zone_name2	second component zone name	Optional	1-40 characters	Repeated attribute as should be encoded within EnvironmentalMonitoringFacility Feature Type	Location	Location → geographicDescription (this should be encoded using Xlink href to the zone identifier in the Annex III theme Area Management/Restriction/Regulation and Reporting Units)	0..*

CAFE requirement				EMF Mapping			
Keyword	Description	Type of parameter	Domain	EMF FeatureType	Attribute	Domain	Multiplicity
						Regulation and Reporting Units	
measurement_code	Identification measurement as part of a measurement configuration group	Key	>= 1 and <= 999	-			
meas_conf_group	Identification measurement configuration group	Key	>= 1 and <= 999	-			
measurement_start_date	Date monitoring measurement operational	Essential	YYYYMMDD, > station_start_date	ObservedProperty	observationPeriod	TM_Object	1
measurement_end_date	Date monitoring measurement stopped	Optional	YYYYMMDD, > measurement_start_date				
measurement_method_type	Type of measurement method (automatic analyzer/remote sensor) or principle (active sampler)	Essential	"automatic analyzer", "remote sensor", "active sampler"	Procedure	procedureType	GenericName (this should be derived from a dictionary of procedures)	1
measurement_technique_principle	Measurement technique or analytical principle	Essential	1-60 characters	MeasurementProcedure	procedureMethodType	GenericName (this should be derived from a dictionary of procedures)	0..*
Analytical_CEN_std	CEN std no. relevant to the analysis method	Mandatory		MeasurementProcedure	procedureMethodType	GenericName (this should be derived from a dictionary of procedures)	0..*
measurement_equipment	Monitoring- or sampling equipment	Essential	1-80 characters	MeasurementProcedure	procedureMethod	GenericName	1
Equivalence_demonstration	Has equivalence with reference method been demonstrated	Mandatory	Yes/No/R (?)	-			
Equivalence_demonstration	url to Equivalence demonstration report	Mandatory	url	-			
Detection_limit		Mandatory	Float or n/a	This is linked to a more full description of the sensor which			

CAFE requirement				EMF Mapping			
Keyword	Description	Type of parameter	Domain	EMF FeatureType	Attribute	Domain	Multiplicity
				has not been included in the current EMF data specification. Though this could be provided by extending the EMF data specification to include sensorML elements			
Sampling_time	Replaces sampling frequency differentiation essential for discontinuous sampling. Duration of one sample (expressed in unit sampling_time_unit)	Mandatory	?	Procedure	procedureDate	TM_Object	0..*
Sampling_interval	Replaces sampling frequency differentiation essential for discontinuous sampling	Mandatory	?	Procedure	frequency	Measure (includes units)	0..1
sampling_time_unit		Optional	"second", "minute", "hour", "natural day", "week", "month", "year"				
integration_time_frequency		Optional	>=0	Procedure	IntegrationFrequency	Measure (includes Units)	0..*
integration_time_unit		Optional	"second", "minute", "hour", "day", "week", "month", "year"				
height_sampling_point		Mandatory	Metres	These are linked to a more full description of the sensor which has not been included in the current EMF data specification. Though this could be provided by extending the EMF data specification to include sensorML elements			
horiz_sampling_point	Horizontal distance of the sampling point inlet to building facade	Mandatory	Metres	-			

CAFE requirement				EMF Mapping			
Keyword	Description	Type of parameter	Domain	EMF FeatureType	Attribute	Domain	Multiplicity
Kerb_dist_sampling_point	distance of the sampling point inlet to kerb of nearest road		Metres	-			
Sampling_point_coords	Physical cords of inlet if accuracy of cords exceed the dimension of station	Voluntary		-			
length_sampling_line		Optional		-			
location_sampling_point	Description location sampling point	Optional	"facade of building", "pavement", "kerbside", "courtyard", "free air flow", "roof", "park", "unknown"	-			
sampling_time	Duration of one sample (expressed in unit sampling_time_unit)	Optional	>= 0	-			
calibration_frequency	Frequency of callibration (expressed in unit calibration_unit)	Optional	>= 0	Calibration Procedure	procedureFrequency	Measure (includes units)	1
calibration_unit		Optional	"second", "minute", "hour", "day", "week", "month", "year"				
calibration_method	Calibration method used	Optional	"calibration cell", "diffusion cell", "permuation tube", "compressed gases", "calibration filter", "uv lamp (ozone generator)", "standard solution", "reference measurement meth.", "exterior air purified", "blank filter", "blank solution", "unknown", "not	CalibrationProcedure	procedureType or procedureMethodType	Generic Name GenericName	1 0..*

CAFE requirement				EMF Mapping			
Keyword	Description	Type of parameter	Domain	EMF FeatureType	Attribute	Domain	Multiplicity
			applicable", "electric current", "flow rate check", "freon electrode", "standard mass", "O3 generator stopped", "black filter (reflectometry)"				
calibration_description		Optional	1-160 characters	CalibrationProcedure	procedureDescription	CharacterString	0..1
Measurement_uncertainty		Voluntary		-			
Quality_system_flag		Mandatory	Accredited network & method, Accredited network & documented method, Documented QA/QC	QualityAssurance → AccreditationInformation	accreditationStatus accreditationType	Boolean GenericName	1 1
QA_QC_resource	url to QA/QC report	Mandatory for CAFE compliance	url	QualityAssurance	Report	SupplementalInformation → URL	0..*
Monitoring_objective	Continuous or indicative etc for CAFE			MonitoringProgrammeInformation Or EnvironmentalMonitoringFacility	monitoringObjective or purpose	MonitoringObjective Or purpose	0..* Or purpose
body_or_programme, list		Desirable	"EU Eol", "EEA EuroAinet", "EMEP", "OECD", "OSPARCOM", "HELCOM", "GEMS", "WHO"	MonitoringProgrammeInformation	monitoringProgramme	GenericName	1

Appendix 2

List of Acronyms

AIRBASE	AirBase is the public air quality database system (of the EEA)
AQ	air quality
CAFÉ	Clean Air For Europe initiative
CAFÉ IP	Clean Air For Europe implementation provisions applicable the CAFÉ Directive
CDR	Central Data Repository (of Reportnet)
CMCU	Central Management and Control Unit
csv	comma separate value
Defra	Department of Environment Food and Rural Affairs (UK Government)
DEM	Data Exchange Module (of EIONET)
EEA	European Environment Agency
EIONET	European Topic Centre on Air and Climate Change
EMF	Environment Monitoring Facilities
Eol	Exchange of Information Decision on Air (ref. 97/101/EC (OJ L 35, 5.2.1997, p. 14–22)
ESU	Equipment Support Unit
ftp	file transfer protocol
GIS	Geographical Information System
GML	geographic mark-up language
HIS system	the UK's HIS QA/QC tool
INSPIRE	Infrastructure for Spatial Information in the European Community (INSPIRE Directive 2007/03/14)
INSPIRE IR	INSPIRE implementation rules
INSPIRE-CAFÉ	Development and demonstration of technical IT solutions for data exchange and reporting under the CAFE Directive using INSPIRE services. Contract Number - 384306
ISO	International Organisation for Standardisation
JRC	Joint Research Centre
KML	keyhole mark-up language
LSO	Local Site Operator
MS	Member State
MySQL	an open source database http://mysql.com
NetCDF	network common data form
PAH	Poly Aromatic Hydrocarbons
QA/QC	Quality Assurance and Quality Control
ReportNet	the Eionet's infrastructure for supporting and improving data and information flows
sftp	secure file transfer protocol
SQL	<i>Structured Query Language</i>
the Archive	the UK Air Quality Information Archive http://www.airquality.co.uk
XML	Extensible mark-up language
XMLT	Extensible Markup Language Translator
XSL	Xml Stylesheet Language
XSLT	Extensible Stylesheet Language Transformation



AEA group
329 Harwell
Didcot
Oxfordshire
OX11 0QJ

Tel: 0870 190 xxxx
Fax: 0870 190 xxxx