RICARDO-AEA

Defra Air Quality: Open Data Roadmap

Report for Defra

Customer:

Defra

Customer reference:

RMP5856
ED39099002

Confidentiality, copyright & reproduction:

This report is the Copyright of Defra and has been prepared by Ricardo-AEA Ltd under contract to Defra dated 31 July 2013. 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 Defra. Ricardo-AEA Ltd 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.

Contact:

Tony Bush
Ricardo-AEA Ltd
Gemini Building, Harwell, Didcot, OX11 0QR
t: 01235 75 3022
e: tony.bush@ricardo-aea.com
Ricardo-AEA is certificated to ISO9001 and ISO14001

Author:

Tony Bush and Rachel Yardley

Approved By:

Paul Willis

Date:

22 April 2014

Ricardo-AEA reference:

Ref: Ricardo-AEA/R/3418
ED39099002- Issue Number 1

Executive summary

Aims of the study

Ricardo-AEA has been commissioned by Defra and the Devolved Administrations to develop an Open Data Roadmap for the UK Air Information Resource (UK-AIR), to extend existing information sharing services to achieve three key goals:

1. Meet the legal requirements of the **In**frastructure for **Sp**atial **I**nfo**R**mation in **E**urope (INSPIRE) Directive (European Directive 2007/2/EC)
2. Meet new and existing user needs for air quality data
3. Improve access to Defra air quality data

The Policy and evidence drivers for this study are clear. In addition to meeting legislative requirements, there is a need to inform people about environmental and human health risks from poor air quality and to provide public information services. Improving air quality through decreased emissions or reducing human exposure to air pollution requires behavioural change and this can be supported and driven by knowledge and information. Additional Cabinet Office policy drivers on Open Data and Transparency are also central to this study and this report sets out how the Open Data "score" of Defra air quality data can be improved. Information is an asset to Government and opening up data for the emerging technology communities helps support the creation of an information marketplace, and in turn, jobs and growth.

This study also considers how Defra air quality data should be managed, published and shared to enable greater reuse of data by both traditional "human" users and automated machines and technologies. The study considers user needs for access to data and the benefits to these users and government policy from the Open Data initiative, the regulatory requirements of INSPIRE and the UK Government Digital Service (GDS) guidelines.

INSPIRE and Web Services

The INSPIRE Directive requires public bodies to establish a range of web services (internet systems with a particular function) to publish spatial environment datasets. The services have three aims, to:

- Allow datasets to be discoverable (Discovery Service) i.e. finding the data
- Allow datasets to be visualised (View Service) i.e. seeing the data
- Provide access to data content (Download Service) i.e. getting data

This work builds upon previous studies in 2010 that investigated (i) the scope for integration of air quality and related datasets into a single information system and (ii) the specifications for Defra's air quality communication systems. Defra has already made considerable steps on the path to open air quality data and in the development of web services, recognising work on INSPIRE as a medium term priority in the Atmosphere and Local Environment Evidence Plan[1]. Recent work has included:

- Making clear for users the licencing terms of the data under the Open Government Licence (OGL)
- Cataloguing the data on UK-AIR with descriptive metadata and linking this to data.gov.uk to make data discoverable
- Development of Web Mapping Services (WMS) to allow users to view and interrogate measurement stations and air quality data online, using Google Maps and Environmental Systems Research Institute's (ESRI) ArcGIS Server technologies

---

[1] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/221080/pb13906-evidenceplan-atmosphere-local-environment.pdf

In developing the roadmap, both lessons learned and the maturation in data and web-service standards since 2010 have been considered.

In December 2013 Defra air quality data met the initial INSPIRE obligations. However, further work remains to develop the web services to provide access to the data and meet future deadlines in 2015 and 2020. This present study makes recommendations for the period 2014-2015 based on current knowledge and state of technology. It is also clear however, that standards and technology will continue to evolve. In recognition of this it is recommended that all systems and services be periodically reviewed and feedback sought from users and the expert community on an ongoing basis in order to drive continuous improvement.

User Needs Analysis

This study has used GDS design principles to support its recommendations. UK-AIR services fall within the GDS scope as they are services that only Government can provide, they have a clear demand from users and they constitute public information systems and services for which there are legal obligations.

A cross section of core air quality data users and stakeholders were engaged in an informal consultation via questionnaire and a workshop held in October 2013. Subsequently, user needs analysis was undertaken and the following key user needs were identified:

- Services optimised to the needs of the air quality community were essential
- Data feeds (downloads) that are robust, traceable and available in the highest time resolution possible
- Unrestricted and simple access to pre-defined data downloads within the scope of available update frequency(ies) and fair use policies for service usage
- Scalable web services for data sharing across all user groups including human and computer users
- The continued availability of the services supporting a Comma Separated Variable (CSV) formatted data download, was at least as important as the development of automated downloads in machine readable formats going forward

On the back of these findings, developments that are recommended to the existing Defra air quality data services have been made to meet, not only immediate user needs, but also to comply with the requirements of the INSPIRE Directive, whilst addressing the guidelines and polices from the GDS and the UK Open Data Strategy.

Summary of recommendations

A series of technical recommendations have been made in this report that have been grouped by themes to assist with understanding.

Data Standards and Formats	Web Services	Access to Data
Recommendation 1. Adopt the Air Quality e-Reporting data model and schema for encoding of downloads of observational data and metadata for air quality spatial objects	**Recommendation 2.** Adopt download web services that are conformant with INSPIRE guidance on network services for download services for all datasets already identified as within scope of INSPIRE i.e. those datasets with discovery metadata and viewing services registered on the UK-AIR data catalogue and data.gov.uk.	**Recommendation 4**. Adopt an improved system for sign-posting to metadata in the CSV downloads to promote openness and transparency. Implement based on the identifier and metadata resources developed for Air Quality e-Reporting to align metadata standards for human and machine readable downloads
Recommendation 11. Implement open source technologies where possible and where proprietary solutions do not present significant financial or technology benefits.	**Recommendation 3.** Maintenance of the current implementation(s) for viewing services, noting the current use of proprietary software, and to undertake additional work to evaluate the conformance of existing viewing services with INSPIRE requirements and to inform future INSPIRE conformance options when these become available.	**Recommendation 5**. Extension of download services to 15-minutes for all pollutants where this is stored in the UK-AIR database
Recommendation 12. UK-AIR should continue to offer existing downloads in CSV formats. XML encodings should conform to the Air Quality e-Reporting IPR data model schema implementation.	**Recommendation 6** Register all web services for viewing and download with the appropriate discovery services registries	**Recommendation 7**. Develop a UK-AIR air quality object registry to support the use of HTTP URIs from the UK Air Quality e-Reporting data model in future UK-AIR CSV and XML downloads, allowing them to be resolved to HTML and XML documents.
	Recommendation 8. Implement an INSPIRE RSS Atom feed for historical data, statistics, air quality spatial object attribute data and other stable spatial datasets e.g. AQMAs because existing Web Feature Service (WFS) options are not optimised for observational data and cannot be connected directly (at present) to the UK-AIR data management systems without costly re-engineering of UK-AIR content.	

Data Standards and Formats	Web Services	Access to Data
	Recommendation 9. Implement the SOS standard for the download of small quantities of near-real-time observational data with some degree of spatial sub-setting.	
	Recommendation 10. Migration to alternative Relational Data Management Systems in order to support a Web Feature Service based service should be avoided in the interest of cost versus the core capabilities required. A watching brief for the development of WFS implementations with MySQL connectivity should be kept to support spatial sub-setting of air quality spatial object attribute and Air Quality e-Reporting via WFS in the future.	
	Recommendation 13. UK-AIR should keep informed on the potential extension of INSPIRE Technical Guidance to cover WCS and, if appropriate, indicate a user need that would support European Air Quality policy implementation.	
	Recommendation 14. UK-AIR should as appropriate and subject to availability implement INSPIRE compliant WCS download services to support sharing of gridded datasets and coverages and to support Air Quality e-Reporting	

Based on the analysis of the software technology suitable for delivering the core capability requirements for air quality, the software solutions identified and recommended are as follows:

1. Implementation of an Rich Site Summary (RSS) Atom feed for download of all pre-defined static historical data, statistics and air quality spatial object attribute data
2. Implementation of a Sensor Observation Service (SOS) for near-real-time data feeds
3. Development of a UK-AIR air quality object registry for spatial and non-spatial air quality objects within the UK Air Quality e-Reporting data model enabling dereferencing of HTTP Uniform Resource Identifiers (URIs), in HTML and XML
4. Continued use of GeoNetwork as a data catalogue registered to data.gov.uk
5. Continued use of the ArcGIS Server WMS, (extended appropriately and where licensing allows) to deliver INSPIRE compliant WMS

In addition to these recommendations on specific technologies, further over-arching philosophies for continuous improvement of UK-AIR services area were advised. These include (i) the use of open source technologies where possible and where proprietary solutions do not present significant financial or technology benefits, (ii) the provision of practical advice to the air quality community on roles, responsibilities and Technical Guidance on how to apply and reuse these new services to get value for money.

Table of contents

Glossary of terms and abbreviations

Terminology / abbreviation	Definition
Atom	Atom Syndication Format, an XML language used for web feeds, an alternative to RSS
AQD	The Air quality directive 2008/50/EC
API	Application Programming Interfaces
AQMA	Air Quality Management Areas
AURN	Automatic Urban and Rural Network
BADC	British Atmospheric Data Centre
CEH	Centre for Ecology and Hydrology
CRS	Coordinate Reference System
CSV	Comma Separated Variable
CSW	Catalogue Service for the Web, a web service standard for retrieving and access to catalogue information on datasets and services
Data Object (data type)	An abstract representation of a real-world phenomenon, attribute, entity without location
DG ENV	European Commission's Directorate-General for the Environment
ESRI	Environmental Systems Research Institute, an international supplier of Geographic Information System (GIS) software
EEA	European Environment Agency
EF	The INSPIRE Environmental Monitoring Facilities theme
ELA	Enterprise Level Agreement
EoI	Exchange of Information Decision
EIONET	European Environment Information and Observation Network
ETC	European Topic Centre
Feature (feature type)	An abstract representation real world phenomena related to a specific location or geographical area [ISO 19101]
GDS	Government Digital Service

Terminology / abbreviation	Definition
GML	Geography Mark-up Language, an XML-format for geographical information
HTTP	Hypertext Transfer Protocol
ICT	Information and communications technology
INSPIRE	Infrastructure for Spatial Information in Europe
IPR	Implementing Provisions for Reporting (in relation to Air Quality e-Reporting)
ISO	International Organisation for Standardisation
JRC	Joint Research Centre
KML	Keyhole Mark-up Language, an XML-based language schema for expressing geographic annotation and visualization on existing (or future) Web-based, two-dimensional maps and three-dimensional Earth browsers
LAMP	Linux, Apache, MySQL and PHP
LAQM	Local Air Quality Management
LGA	Local Government Association
Linked Data	A method of publishing structured data so that it can be interlinked and become more useful. It builds upon standard Web technologies such as HTTP, RDF and URIs, but rather than using them to serve web pages for human readers, it extends them to share information in a way that can be read automatically by computers.
MySQL	An open-source relational database management system
NAEI	National Atmospheric Emissions Inventory
NERC	Natural Environment Research Council
NGO	Non-governmental organisation
NII	National Information Infrastructure
O&M	Observation and Measurements, a general model and XML encoding for observations and measurements.
OGC	Open Geospatial Consortium
OGL	Open Government Licence
QA/QC	Quality assurance/quality control
PSI	Public Sector Information

Terminology / abbreviation	Definition
PSMA	Public Sector Mapping Agreement
Register	A set of files containing identifiers assigned to items with descriptions of the associated items [ISO 19135]
Registry	An information system on which a register is maintained [ISO 19135]
RDMS	Relational data base management system
RSS	Rich Site Summary, an XML language used for web feeds
Schema	An (XML) schema is a description of a type of XML document, typically expressed in terms of constraints on the structure and content of documents of that type, above and beyond the basic syntactical constraints imposed by XML itself.
SEIS	Shared Environmental Information System
SOA	Service Orientated Architecture
SOS	Sensor Observation Service, a web service standard to provide an open interface to obtain observations and sensor and platform descriptions from one or more sensors.
SQL	Structured Query Language - a programming language designed for managing data held in a relational database management system.
SWE	Sensor Web Enablement, a standards framework enabling developers to make all types of sensors, transducers and sensor data repositories discoverable, accessible and useable via the Web.
TSV	Tab separated values
UKEOF	UK Earth Observation Framework
UML	Unified Modelling Language, a general-purpose modelling language in the field of software engineering. The basic level provides a set of graphic notation techniques to create visual models of object-oriented software-intensive systems.
URI	Uniform Resource Identifier, a compact sequence of characters that identifies an abstract or physical resource [IETF RFC 3986]
URI dereferencing	The act of making use of a URI in order to retrieve a representation of its associated resource [IETF RFC 3986]. NOTE Dereferencing a URI which identifies a spatial object returns that object encoded, for example, in GML, KML, RDF, JSON, PDF or HTML
WCS	Web Coverage Service
Web services	A Web service is a method of communication between two electronic devices over World Wide Web. A web service is a software function provided at a network address over the web or the cloud; it is a service that is "always on" as in the concept of utility computing.

Terminology / abbreviation	Definition
WFS	Web Feature Service, a web service standard for retrieving or altering feature descriptions via a web service
WMS	Web Map Service, a web service standard for retrieving map images
WRF	Weather Research and Forecasting modelling format
XML	(eXtensible Markup Language) A mark-up language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable

1 Introduction

Ricardo-AEA has been commissioned by Defra to scope and develop an Open Data roadmap for the development of a UK Air Quality web-Portal, to extend existing information sharing services to achieve three key goals:

1. To meet the legal requirements of the **In**frastructure for **Sp**atial **I**nfo**R**mation in **E**urope (INSPIRE) Directive for Defra air quality data[2]
2. To meet new and existing user needs for air quality data
3. To improve access to Defra air quality data

This work builds upon previous studies in 2010 that investigated (i) the scope for integration of air quality and related datasets into a single information system, (ii) the specifications of the service contract supporting Defra's existing air quality communication systems (see Section 1.3). In developing the roadmap, the current study recognises both the lessons learned from previous studies and the maturation in data and web-service standards since 2010. In addition a cross section of core air quality data users and stakeholders were consulted via questionnaire and a workshop held in October 2013.

The roadmap proposed in Section 6 covers the period between 2014 and the end of 2015. The options for data encoding and web services for data sharing have been considered and recommendations have been made to support Defra's approach to centralised data provision. Specifically, options for download services for data have been evaluated and recommendations are provided for web services to support these, data formats and standards to be supported and mechanisms for improved access to data, all of which help to meet the three stated aims above.

The roadmap has been heavily influenced by a number of external drivers that have evolved since 2010. The drivers are discussed in detail in section 3 of this report but may be summarised as:

- EU legislative requirements for reporting air quality data electronically: "Air Quality e-Reporting"
- The regulatory demands of the INSPIRE Directive
- Alignment with the Cabinet Office's and Defra's Open Data Strategy[3] and UK Location Strategy[4], regulatory reporting data flows, and data covered by the Freedom of Information Act[5] and Environmental Information Regulations[6]
- The UK Government Digital Service (GDS) Design Principles Meeting the needs of traditional end-users and emerging consumers of air quality data e.g. the London Data Store[7] and Cleaner Air for London[8] websites

In response to these drivers and increased demand for machine accessible and readable data formats hosted on the web, this study will inform on how future Defra's air quality web data services should be implemented in a prioritised and efficient way.

The roadmap detailed in this report includes a consideration of the minimum technical requirements necessary to achieve conformance with the INSPIRE Directive, and includes a timeline for the roadmap setting out key stages along the journey to 2020.

Users of existing UK air quality data services have been consulted to identify their current and future requirements. Air quality data providers have been consulted to understand their approaches and plans for publishing data, in particular, other public sector organisations with

---

[2] Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32007L0002:EN:NOT
[3] http://data.gov.uk/sites/default/files/Cabinet%20Office%20Open%20Data%20Strategy.pdf
[4] http://data.gov.uk/sites/default/files/uk-location-strategy_10.pdf
[5] http://www.legislation.gov.uk/ukpga/2000/36/introduction
[6] http://www.legislation.gov.uk/uksi/2004/3391/contents/made
[7] http://data.london.gov.uk/
[8] http://www.cleanerairforlondon.org.uk/

obligations under INSPIRE. Defra's experience of developing INSPIRE conformant data models and vocabulary for air quality was shared with members of the air quality community to support others who have similar obligations.

This new study draws out the changes which need to be implemented to allow Defra's air quality data to be published to meet the requirements of the regulatory drivers as well as the needs of current and future users. In contrast to the 2010 studies the focus now is on accessibility and management of data through web-based machine readable services rather than integration. Applications to support accessibility and data management will concentrate on Defra datasets only, rather than considering health, meteorological or transport data. In addition the emphasis of tools is on improving visualisation and accessibility, enabling others to build tools and applications from an air quality information marketplace, instead of designing tools for them, which was part of the original 2010 studies. Overall the 2013 roadmap now places greater weight on managing and publishing data at a single source and using data.gov.uk to facilitate discovery, redirecting users to web-services that can be read by humans and computers, to ensure all users can access the data they require, in the formats they require.

1.1 Objectives

The objectives of this study are to:

- Identify user requirements
- Identify policy framework requirements including European Air Quality e-Reporting[9] INSPIRE and the Government Digital Service[10]
- Recommend system architecture for Defra's air quality systems
- Recommend consistent and interoperable approaches for publishing air quality data, in particular for download services which have thus far not been implemented using a data standards based approach for UK-AIR
- Produce a timeline for the recommended approaches

1.2 Structure of this report

In Sections 1 and 2 of this report we describe the conclusions and recommendations from previous studies and the existing data services that have evolved as a result of these in order to open up data. Section 3 describes and discusses the external initiatives that are driving the changes in the publication and sharing of data, and the barriers and opportunities they have created. Section 4 addresses the needs of the users, both human and machine. An important input into this section has been a stakeholder consultation exercise and analysis of how existing data services are used. The design principles required in order to meet the needs of the users and various external drivers are defined in Section 5, and the resulting recommendations and roadmap for infrastructure and architecture are outlined in Section 6. Finally, practical recommendations for collaborative working within the air quality community are made in Section 7.

---

[9] http://www.eionet.europa.eu/aqportal

1.3 Previous studies

In 2010, Defra commissioned two complementary studies to evaluate the future of Defra Air quality data and information services. Since the studies were reported, the air quality data landscape has continued to evolve. This has largely been driven by regulatory obligations to the Ambient Air Quality Directive 2008/50/EC[11] and the INSPIRE Directive 2007/2/EC, but strategic initiatives such as the Cabinet Office's Transparency and Open Data agenda have become increasingly important.

1.3.1 Data Management and Integration Study

In 2010, AEA Technology plc. carried out a scoping study[12] to investigate the feasibility of an integration process to maximise the overall availability and use of Defra air quality data. The study included user engagement workshops and found that many data users did not have access to all of the monitoring, modelling, emissions and other non-air quality data they needed. This was partly due to many of Defra's air quality service contracts (and in some cases each dataset) having bespoke systems to capture and manage the data, making data sharing and analysis difficult and manual.

The objectives of the scoping study were to:

- Make recommendations to improve the accessibility and reusability of UK air quality data
- Examine data presentation issues in respect of spatial presentation and online tools
- Propose how the UK could move towards a common data infrastructure
- Consider both air quality datasets and others including meteorological data, traffic and health statistics

Two workshops were held with stakeholders to present and discuss the datasets, data management practices and barriers to integration. The stakeholders discussed common problems with availability and analysis and potential user tools that could be developed with an integrated dataset.

Datasets were assessed for their suitability for integration and three key problems were identified:

- Lack of structure in the architecture of the overall system
- Disparity of datasets in terms of standard formats
- Lack of or inconsistent metadata

After a comprehensive review of the datasets, the authors proposed an integration process that would:

- Maximise the overall availability and reuse of the data to support stakeholder objectives
- Standardise definitions and data formats
- Standardise data updates
- Catalogue datasets and disseminate information about the data availability
- Allow different datasets and different systems to be interrogated as one standard system
- Increase data processing efficiency and reduce operating costs through automation
- Simplify statutory reporting procedures
- Make it possible to develop useful tools for policy makers

The proposed solution was a new data approach and architecture incorporating a spatial data infrastructure, complying with INSPIRE requirements. The goal was to have data collected and stored locally but with wider access through using a common approach and data standardisation. The proposed implementation route for sharing data across contractors and systems was to make extensive use of Service Orientated Architecture.

In the shorter term a number of tasks were identified, some of which have been implemented, including:

- Creation of a comprehensive catalogue of air quality datasets

---

[11] Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe
[12] Air Quality Data Management and Integration System Scoping Study, Issue 1.2 AEAT/ENV/R/3005, 2010

- Improving access to meteorological data
- Standardisation of data and metadata

1.3.2 Air Quality Communications Review

Also in 2010, Aether carried out a review[13] of the Air Quality Communications Contract for Defra and evaluated Defra's arrangements on the provision of air quality information to the public. At the time of writing the report, the majority of this information was provided via the Air Quality Archive web pages, which were the predecessor to the UK-Air Information Resource web pages at uk-air.defra.gov.uk.

The specific objectives of the review were to assess:

- How well the Communications Contract delivered information to stakeholders
- How effective the Archive was in terms of usability, functionality and the extent to which the UK fulfilled its requirements for public information under the Ambient Air Quality Directive and the 4th Daughter Directive (2004/107/EC)[14]
- Whether there was best practice being used by other websites etc., which could be drawn upon

Aether performed a thorough review of the website content, structure and accessibility, considering the variety of key users. A stakeholder engagement survey was carried out; including an online questionnaire and telephone interviews with users to obtain feedback on the existing services and needs for the future. With consideration to international best practice, the study proposed tools, platforms and technologies that could be used by Defra in the future to improve the user experience, including approaches used for data handling and information dissemination. Finally, consideration was given to the way in which Defra provided information to the public, other Government departments, and the way it worked with Local Authorities.

Following the study, Aether identified gaps and made recommendations for improved website structure, data selection tools, metadata, Twitter feeds, OpenAir tools and interactive maps, which have since all been actioned. Aether highlighted the value of having a single portal to access air quality data, although the demand for non-air quality datasets were less consistent among end-users, apart from the provision of meteorological data, which was strongly supported (and has since been made available through the Weather Research and Forecasting (WRF) Model).

As in the Data Management and Integration Study, the Aether review recommended that Defra develop a Service Oriented Architecture, containing existing databases of data with registries and data dictionaries to support registration, discovery and extraction of data from third party services, to comply with Shared Environmental Information System (SEIS)[15] and INSPIRE principles.

---

[13] Review of the Air Quality Communications Contract, Defra/04F10CD/4295, 2010
[14] Directive 2004/107/EC of the European Parliament and of the Council of 15 December 2004 relating to arsenic, cadmium, mercury, nickel and polycyclic aromatic hydrocarbons in ambient air
[15] Shared Environmental Information System, http://ec.europa.eu/environment/seis/

2 Existing Defra data services

Defra and the Devolved Administrations of Northern Ireland, Scotland and Wales provide accessible and easy to use information on air quality in the UK for members of the public, vulnerable individuals and the air quality research community. Since 2010, a number of developments to the services and their delivery mechanisms have been made, designed to meet the needs of a wider variety of users and in response to the recommendations made in the 2010 reports.

This section of the report outlines the status of the web-services and other information services provided Defra and the Devolved Administrations as of March 2014.

2.1 UK Air Information Resource

In 2011, Defra launched a new UK Air Information Resource online. The UK-AIR website is available at uk-air.defra.gov.uk and was explicitly developed to address many of the recommendations made by Aether during the 2010 review of the Air Quality Communications Contract. UK-AIR contains textual background information on history of air quality, legislation and pollutants, alongside a library of reports. It also serves as a gateway to data services, the specification for which have been guided by the end user needs for data discovery and downloads.

In 2012, all owned data and published by Defra and the Devolved Administrations and hosted on the UK-AIR and National Atmospheric Emissions Inventory (NAEI) websites were licensed under the Open Government Licence[16] (OGL). Attribution of an open licence to data is a fundamental requirement for Open Data and the release of UK-AIR and NAEI datasets under the OGL was a significant step forward. Licencing via the OGL encourages third parties to use and re-use the published datasets in free and flexible way, with only a few conditions. This in turn promotes the value of Defra's air quality datasets and evidence programme within the context of scientific, research and business communities looking to use the data in an applied way.

The UK-AIR website map including live URLs is provided in Appendix 1. The following sections briefly describe the range of services developed and hosted by UK-AIR in response to the recommendations made in the 2010 reports and highlights the gaps relative the INSPIRE obligations.

2.1.1 UK-AIR services to support data discovery

Metadata Catalogue http://uk-air.defra.gov.uk/data/data-catalogue

A searchable catalogue of UK air quality monitoring, modelling and emissions datasets has been developed and is available on the UK-AIR website. This resource supports searches based on keywords, location, web link, responsible organisation and contact details. The metadata profile used is the UK GEMINI2.1 standard, which is conformant with the INSPIRE profile.

During 2013, Defra worked closely with data architects and the team behind data.gov.uk to ensure that the UK-AIR catalogue is harvested by the data.gov.uk thereby publishing the datasets to a wider audience and fulfilling the INSPIRE requirement for metadata discovery services. The catalogue has been developed adhering to the INSPIRE principles that data should be collected once, stored where it can be maintained most effectively, and that information from different sources should be made available in a form to share with many users and applications.

---

[16] http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/

2.1.2 UK-AIR viewing services to support data visualisation

A range of data visualisation tools have been implemented to support a variety of data formats, update frequencies and user needs.

Interactive monitoring networks map *http://uk-air.defra.gov.uk/interactive-map*

The UK-AIR web site hosts a Web Mapping Service (WMS) that acts as a front end to an interactive information retrieval system for data at all levels, and provides:

- A map of stations in each monitoring network
- The latest pollution index for the automatic monitoring stations
- Hyperlinks to maps, photographs and other information about each station
- Hyperlinks to data selectors

The monitoring networks map WMS is not INSPIRE conformant and has been designed with the specific user needs in mind within the UK-AIR's tailored host environment. The WMS layers exist to support the users in finding information they want rather than to publish map images for direct consumption.

Interactive GIS maps *http://uk-air.defra.gov.uk/data/gis-mapping*

The UK-AIR web site hosts a map viewer and geo-processing tool that allows users to browse, visualise, interrogate and download data from modelled air quality datasets including:

- Maps for 14 pollutants
- A choice of background or roadside model predictions
- Predictions for all years 2001-2012
- A toolset to select, interrogate and download data for specific areas as CSV and shape files
- Layers representing local authority boundaries, monitoring stations etc.

A GIS toolkit is also available on Defra's NAEI website at http://naei.defra.gov.uk/data/gis-mapping, allowing users to retrieve emissions data from specific sectors in user-defined areas. The GIS toolkits also hosts WMS layers for INSPIRE conformance for each pollutant layer included in the tools.

OpenAir Toolset *http://uk-air.defra.gov.uk/data/openair*

UK-AIR hosts the OpenAir toolset to allow users to perform complex and innovative visual analysis of current and archived air pollutant data from the Automatic Urban and Rural Network (AURN). Recent developments have added tools that use meteorological inputs from the WRF weather forecasting model. The website makes it straightforward for users to select data, visualise and download the graphical outputs (plus underlying datasets and code) from OpenAir in document-ready formats.
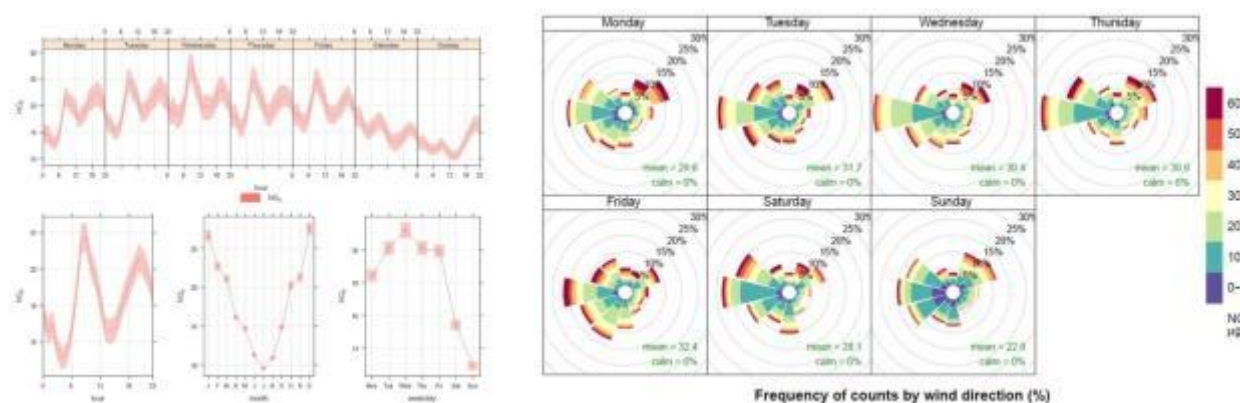
The OpenAir toolset has been designed with the specific user analysis needs in mind within the UK-AIR's tailored host environment. It is not a web service targeted for INSPIRE compliance.

2.1.3 UK-AIR data download services to support data visualisation

A range of web services to support data download have been implemented. These have been developed with specific user needs and data update frequencies in mind and to support a variety of data formats. Because of the application-specific way in which many of these web pages and services have been developed, there are some crossovers in the service types, i.e. many of the download services include viewing tools.

Data selector *http://uk-air.defra.gov.uk/data/data_selector*

The UK-AIR database is searchable through the data selector tool on the UK-AIR website. Users can search automatic and non-automatic monitoring networks using several variables to generate a customised dataset, which can be viewed graphically, on screen and is also available to download as a CSV file. Both raw data and calculated statistics are available through the data selector. Historical data are also available separately.

Data are uploaded to the database from numerous organisations through a specially developed portal. The data selector was not designed with INSPIRE conformance in mind and owing to the data encoding formats (CSV) and vocabularies used, this web service does not follow the prescribed standards for an INSPIRE conformant download service. However, it does provide an essential, accessible and easy to use service for human UK-AIR users.

Latest measurement summaries *http://uk-air.defra.gov.uk/latest/*

UK-AIR's latest measurement summary web pages host maps of the latest pollutant air quality index for zones and agglomerations around the UK including:

- Time series graphs of latest pollutant concentrations
- Tables of latest pollutant concentrations and air quality index for each automatic site
- RSS (Rich Site Summary) feeds for latest measurements. RSS feeds are commonly used to automatically publish frequently updated information on the Web. The XML file format allows the information to be published once and viewed by many different programs. This benefits the user who is able to subscribe to updates from UK-AIR or to aggregate the feed along with those from other sites into one place
- The latest measurement summary is posted daily on Twitter @DefraUKAIR

The latest measurement summaries web pages were not designed with INSPIRE conformance in mind and owing to the data encoding formats (CSV) and vocabularies used, this web service does not follow the prescribed standards for an INSPIRE conformant download service.

Forecast of pollution *http://uk-air.defra.gov.uk/forecasting/*

UK-AIR's pollution forecast web pages host a clickable map showing the forecast of pollutant air quality index in each region for the present and following day, and an outlook for the next few days, in addition to:

- Modelled outputs for forecasted weather, air quality and air mass trajectories
- RSS feeds of XML data for pollution forecasts
- Plug-ins to allow users to easily embed the latest UK air pollution forecast into their own webpages by automatically generating the necessary HTML code, which can simply be copied and added to the user's web page
- A forecast summary that is posted on Twitter @DefraUKAIR daily

The pollution forecast web pages were not designed with INSPIRE conformance in mind and owing to the data encoding formats (CSV) and vocabularies used, this web service does not follow the prescribed standards for an INSPIRE conformant download service. Air pollution forecasting vocabularies are not currently supported in the UK or at a European level. These will need to be developed to support migrating existing services to an INSPIRE conformant service.

Air quality bulletins subscription service *http://uk-air.defra.gov.uk/subscribe*

UK-AIR's air quality bulletins provide users with an email subscription service for:

- Bulletins on the latest pollution levels
- 5-day air quality forecasts
- Other news items by email

The air quality bulletins subscription service was not designed with INSPIRE conformance in mind and owing to the data encoding formats (CSV) and vocabularies used, this web service does not follow the prescribed standards for an INSPIRE conformant download service. Air pollution forecasting vocabularies are not currently supported in the UK or at a European level. These will need to be developed to support migrating existing services to an INSPIRE conformant service.

2.1.4 UK-AIR web analytics statistics

In the 12 months from November 2012 to October 2013, the UK-AIR website was visited more than 250,000 times by 164,418 individuals, visiting an average of five different web pages on each visit. Table 1**Error! Reference source not found.** shows that the five most popular pages include two of the tools that have been recently upgraded, such as the data selector and interactive map.

Table 1   Most popular UK-AIR web pages

Page	% of Page Views
Home page	8.9%

Effects of Air Pollution	6.5%
Data selector	3.5%
Interactive Map	3.4%
Data homepage	2.1%

Extended user analysis indicates that visitors to the UK-AIR can be categorised into five broad groups, shown in Table 2, listed in order of group size[17].

Table 2   ***UK-AIR user types***

User Type Identified	% of Visits
Schools and Universities	6.7%
Local Authorities	3.9%
Air or Environmental Consultants	3.4%
Government and Government Agencies	2.1%
Media, Banking, Industry and Charity	1.4%

2.2 Wider air quality data services

Defra's NAEI website at naei.defra.gov.uk was redesigned in 2013 to give it the same look and feel as other Defra websites using the Defra design templates in use at the time. The NAEI web pages provide information on atmospheric emissions of air quality pollutants and greenhouse gases from UK sources such as cars, trucks, power stations and industrial plant. These emissions estimates are used to find ways of reducing the impact of human activities on the environment and human health. The website hosts both static information about air quality pollutants emissions, method statements on how the inventory is compiled and the following data services:

- Static emissions maps for the latest inventory year
- An interactive GIS based emissions map viewer
- Spatial data from point sources provided as a MS Excel file
- A searchable database of emission factors
- Local and regional interactive end-user CO_2 maps, showing CO_2 by sector, local authority and year
- A searchable database of emissions data, viewable on screen and downloadable as CSV or MS Excel files

Defra also provides more local air quality services on its smoke control web pages[18] and Local Air Quality Management (LAQM) web pages[19]. Defra is considering ways to streamline and improve these websites to meet user needs, considering the content in terms of citizen, policy and technical information requirements. With the databases and supporting spatial data tools behind both these websites, there is an opportunity to align them more closely with the mapping information and data portal on UK-AIR in the future.

Defra is also working with Local Authorities to reduce the level of the air quality-INSPIRE burden at a variety of levels.

---

[17] Analysis of the top 1000 visitors
[18] http://smokecontrol.defra.gov.uk/
[19] http://laqm.defra.gov.uk/

- In 2012, local authorities were invited to submit metadata on their air quality data holdings to UK-AIR's data catalogue. By agreement with these participants[20], these data can, in turn, be registered with data.gov.uk to meet the INSPIRE requirement for discovery level metadata This offer does not extend to the responsibility for air quality download services
- As part of the LAQM process, Defra has recently agreed an exemption with Ordnance Survey with respect to a Derived Data Exemption request for the UK Air Quality Management Areas (AQMAs). Defra and the Devolved Administration collect, compile and combine the AQMA datasets from Local Authorities annually. The exemption now allows Defra to prepare INSPIRE-conformant discovery level metadata and viewing services for the national (UK) AQMA dataset, thereby relieving Local Authorities of this burden. It is one of the recommendations of this report that all datasets with INSPIRE-conformant discovery level metadata and viewing services will be available for download via an INSPIRE compliant download, see section 6

---

[20] Non-affiliate monitoring stations of Air Quality in Wales, Air Quality in Scotland, Northern Ireland Air, Herts and Beds Network, Air Quality Monitoring in London and South East England, Norfolk Air Quality Network, Kent and Medway Air Quality Monitoring Network, Greater Manchester Districts network, other local authorities represented by UKAirQuality.net.

3 Influencing external initiatives

Section 1 of this report indicated that the roadmap has been heavily influenced by a number of external drivers and initiatives that have evolved since 2010. The core drivers were summarised as:

- EU legislative requirements for reporting air quality data electronically: "Air quality e-Reporting"
- Conformance obligations with the regulatory demands of the INSPIRE Directive
- Alignment with the Open Data Agenda within Government
- Design standards from the UK GDS
- Meeting the changing needs of end-users and emerging consumers of air quality data

In this section, we explore the scope and impacts of these drivers and other lesser actors on air quality and recommendations made in this roadmap.

3.1 Air Quality e-Reporting

Air Quality e-Reporting is the response of the European Commission to the need to adapt procedures for data provision, regulatory assessments and reporting of air quality information. Changes in European Union legislation in recent years has provided an opportunity to streamline European air quality data reporting and exchange. As a result of the organic growth of air quality reporting over time, there has been a strong business case in recent years for a move towards a more automated assessment of air quality by organising and harmonising data and information, and by following standardised approaches.

The European Commission Implementing Decision 2011/850/EC was introduced on 12 December 2011 and laid down rules for the Air Quality Directive in force as regards the reciprocal exchange of information and reporting on ambient air quality[21]. This Decision provided an opportunity to modernise data reporting, improve data quality, facilitate information sharing and reduce the administrative burden of reporting. In adapting the procedures, Air Quality e-Reporting has embraced digital formats for reporting and the internet as the core media for reporting. The new procedures have also been designed and built in line with the requirements for INSPIRE where possible, extending (and adapting) the scope of INSPIRE to specific air quality needs as required in support of regulatory and informative air quality reporting.

Assisted by the European Topic Centre (ETC)[22] and the European Environment Agency (EEA), the new reporting and exchange of information system has been developed in close liaison with the European Environment Information and Observation Network (EIONET) air quality community. Moreover, the European Commission's Directorate-General for the Environment (DG ENV) also invited countries to participate in the activity of designing the Air Quality e-Reporting systems and testing the data flows for each type of reporting covered in the implementing Decision. Testing of e-Reporting systems commenced in 2011 as part of a pilot study programme. This programme was aimed at optimising file formats for e-Reporting, data management systems, quality assurance/quality control (QA/QC) routines, aggregation routines, and evaluating and developing support for all eventual EIONET participants. The pilot programme was operational throughout 2012–2013, during which time operational systems, guidance and support were prepared in advance of the first tangible e-Reporting deadline on 1 January 2014. The pilot has on-going responsibilities from 2014 for continued optimisation of systems.

---

[21] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:335:0086:0106:EN:PDF
[22] European Topic Centre on Air and Climate Change Mitigation (ETC/ACM)

The scope of the new reporting system covers all regulatory and informative reporting agreements set out by the Exchange of Information Decision[23] (EoI) (Council Decision 97/101/EC), the Air Quality Directive and the 4th Daughter Directive. By adopting data modelling approaches prescribed by INSPIRE, redundancies and duplications that had arisen in the old reporting data model(s) over time were removed in order to deliver a merged and internally consistent data model that has been upgraded in line with modern standards for data encoding and data sharing. The data model is now organised into eight broad air quality data themes that service all reporting and information sharing needs of the air quality community. A simplified diagram of the Air Quality e-Reporting data model is presented in Figure 1   S.

Figure 1   Simplified schematic of the Air Quality e-Reporting data model

The new model has been designed to reinforce internal consistency in regulatory data flows. This has been achieved by the migration of the reporting formats to XML and widespread user of internal and external cross references using standard XML functionality in preference to duplicating information. The machine readability that XML brings also supports automated checking and QA/QC processing.

The new data model and framework for reporting has been designed to be reusable at all levels (local, regional, national and Europe wide).

The development of the Air Quality e-Reporting system has been instrumental in engaging the air quality community across Europe with the INSPIRE initiative. The INSPIRE goals of improved accessibility and interoperability have, in turn, had wider benefits allowing the air quality community to readily align with the UK's broader Open Data Strategy.

As well as an early INSPIRE adopter, air quality is also an interesting test case for INSPIRE as it draws upon so many INSPIRE data specifications to support the range of air quality assessments activities: from monitoring activities, computational prediction of air quality using models, management practices to improve air quality in special areas and assessment practices against regulatory thresholds. There are five core INSPIRE themes related to air quality including human health and safety, area management, atmospheric conditions, meteorological features and environmental monitoring facilities. These specifications are supplemented with the Open Geospatial Consortium (OGC) observations and measurements and sensor web enablement standards for encoding observational data.

3.2 INSPIRE

European Directive 2007/2/EC is known as the 'INSPIRE' directive or initiative, which aims to establish an **In**frastructure for **Sp**atial **I**nfo**R**mation in **E**urope to support Community environmental policies or activities that may have an impact on the environment. It establishes an infrastructure for sharing public sector spatial information in the European Union and it was transposed into UK law in December 2009.

Better environmental policy across the EU will be facilitated by:

- Improving the connectivity of and access to existing spatial data across the European Union at a local, regional, national and international level

---

[23] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:1997:035:0014:0022:EN:PDF

- Improvements in the sharing mechanisms for spatial data between public authorities
- Improving public access to spatial data

In support of INSPIRE, Member States must make spatial datasets that come within the scope of the Directive available in a consistent format and create network services for accessing the datasets via the internet. Datasets in scope of INSPIRE are ones that come under one or more of the 34 environmental themes set out in the Directive Annexes[24]. Milestones are set for when metadata, data, and network services for dataset in each Annex are to be made available. Technical Implementing Rules have subsequently been made by the EC to support implementation of INSPIRE.

To implement INSPIRE in the UK, separate but broadly consistent regulations covering England, Northern Ireland and Wales; (SI 2009 No 3157) and Scotland (SSI 2009 No 440) came into effect on 31 December 2009. The INSPIRE (Amendment) Regulations 2012 continue the transposition into UK law of INSPIRE. The Amendment Regulations came into effect on 1 August 2012. The INSPIRE (Scotland) Amendment Regulations 2012 came into effect on 23 November 2012.

The implementation of INSPIRE in the UK is led in Defra by the UK INSPIRE Project. The project provides standards, guidance, technology and assistance to organisations with a duty to publish data under INSPIRE. The work of the UK INSPIRE Project is directed by the UK INSPIRE Compliance Board, which is the decision making body that supports Defra and sets the implementation approach, taking account of infraction risk. Technical assurance is provided by the Architecture and Interoperability Board. Defra's obligations to the INSPIRE Directive for establishing an infrastructure for spatial information have been recognised as a medium term priority within the Atmosphere and Local Environment Evidence Plan,[25] which will require changes in air quality data holdings and data management systems.

The Local Government Association (LGA) is leading on INSPIRE implementation in Local Government and provides advice and guidance to Local Authorities. See also the LGA's INSPIRE Knowledge hub[26] and LGA INSPIRE team contact points at inspire@local.gov.uk. The devolved administrations have their own arrangements for the coordination of INSPIRE[27].

3.2.1.1   *Standards and services supporting INSPIRE objectives*

INSPIRE requires public bodies to establish a range of web services to publish spatial environment datasets. However, it is not sufficient to just publish the HTML document, PDF or similar describing the datasets, nor just provide an interactive map. INSPIRE requires a specific network (web) service to allow data to be streamed (downloaded) into other applications on request. The services will:

- Allow datasets to be searchable / discoverable (Discovery Service)
- Allow datasets to be visualised (View Service)
- Provide access to data content (Download Service)

Discovery Services enable searches to be made on metadata in order to determine the existence and location of data. The searches are made via a metadata catalogue and are supported by metadata or descriptive data on the datasets being published under INSPIRE. There are two types of discovery metadata:

- Metadata for the dataset being published
- Metadata for the Network Services being used to publish the data

Both need to be managed together, as they must reference each other.

View Services or WMS are publically-accessible network services that provide an image of a spatial data set in response to user-defined requests defined by IT standards. These services

---

[24] http://inspire.ec.europa.eu/index.cfm/pageid/2/list/7
[25] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/221080/pb13906-evidenceplan-atmosphere-local-environment.pdf
[26] https://knowledgehub.local.gov.uk/group/inspiredirectiveanduklocationprogramme
[27] http://www.gistrategyni.gov.uk/contact.htm

present a snapshot of spatial data as an image (GIF, JPEG or PNG). INSPIRE View Services must be compliant with the INSPIRE Regulations on Network Services[28].

Download Services or web feature services (WFS) are publically-accessible network services that provide access to the underlying environment data relating to the spatial object. There are three types of download service, two providing pre-defined datasets and one providing direct access:

1. The RSS Atom feed delivering pre-defined datasets
2. WFS delivering pre-defined datasets
3. WFS providing direct access to features

INSPIRE Download Services must be compliant with the INSPIRE Regulations on Network Services[29]. Figure 2**Error! Reference source not found.** describes the confirmation of web services supporting INSPIRE.

Figure 2 INSPIRE network services architecture
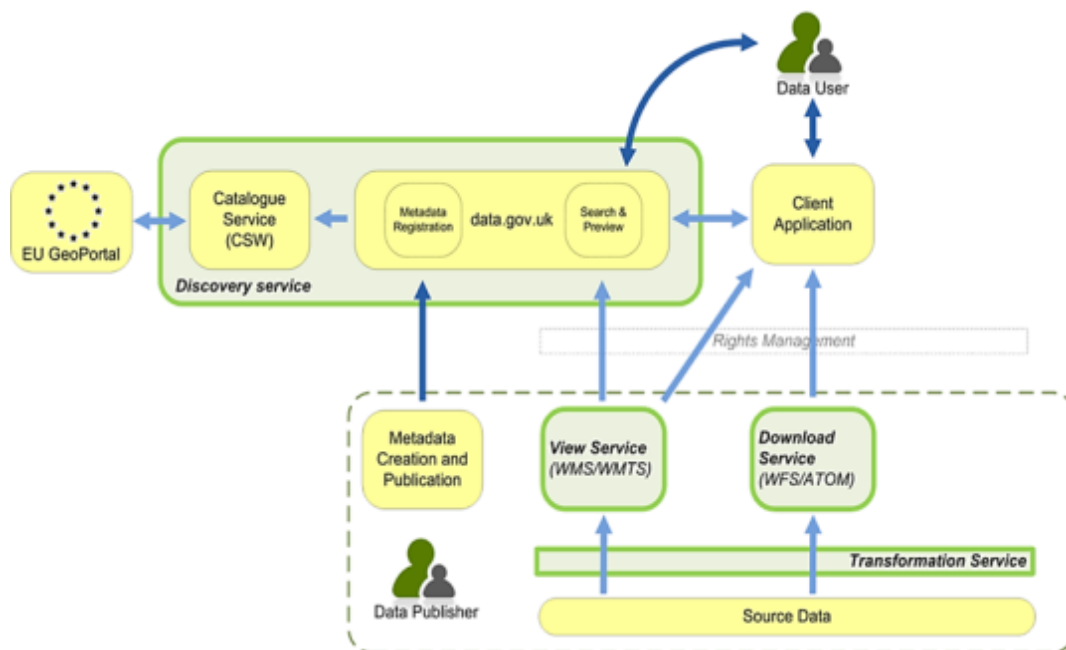
To support INSPIRE network services and promote interoperability of the published data, INSPIRE also prescribes the use of IT standards to encode the data. INSPIRE specifies that all data shall be encoded in XML to support machine readability. A range of International Organisation for Standardisation (ISO) and OGC standards in addition to those developed specifically by the INSPIRE initiative are used to encode attribute data of environmental data objects. These standards are generic, being designed to be applicable to a broad environmental data theme e.g. the Environmental Monitoring Facilities (EF) theme covers all types of environmental monitoring activity. The data specifications may subsequently be extended by domain communities for specific application needs. INSPIRE requires that its core data specifications and any extensions are modelled in Unified Modelling Language (UML).

3.2.2 INSPIRE implementation timetable

The implementation of INSPIRE is divided into two distinct phases:

Phase 1 publishing your data as it is now

Phase 2 publishing INSPIRE Annex I, II and III compliant data

Milestones for complying with the INSPIRE Regulations are as follows (see also the data.gov.uk guidance[30]).

---

[28] http://inspire.jrc.ec.europa.eu/index.cfm/pageid/3
[29] http://inspire.jrc.ec.europa.eu/documents/Network_Services/Technical_Guidance_Download_Services_v3.1.pdf

Phase 1	
Date	**Milestone**
December 2010	Compliant Discovery Metadata created (and available) for INSPIRE Annex I & II datasets
May 2011	View Services established for INSPIRE Annex I & II datasets (start of initial operating capability)
November 2011	Full View Services established for INSPIRE Annex I & II datasets (end of initial operating capability)
June 2012	Download and Coordinate Transformation Services established (start of initial operating capability)
December 2012	Full Download & Coordinate Transformation Services capability (end of initial operating capability)
December 2013	Full Discovery Metadata, View and Download Services made available for INSPIRE Annex III datasets

Phase 2	
Date	**Milestone**
December 2012	Newly collected and extensively restructured INSPIRE Annex I compliant data available
Expected Oct/Dec 2015	Newly collected and extensively restructured INSPIRE Annex II and III compliant data available
June 2017	Publish INSPIRE Annex I compliant Data
Expected Oct/Dec 2020	Publish INSPIRE Annex III comp

As of 3 December 2013, UK-AIR was aligned with this timetable, with services in place for discovery, view and download in existing formats.

3.2.3 Steps to INSPIRE conformance

The LGA has broken down the journey to conformance with the INSPIRE Regulations into a series of 12 steps, which have been replicated here as an easy reference guide. Steps 1 to 6 are the most urgent for public sector bodies such as Local Authorities or non-government organisations (NGOs) having a deadline of 3 December 2013 for Annex III datasets, followed by steps 7 to 10. LGA advise Local Authorities to follow these steps to ensure that they are compliant with the INSPIRE legislation.

3.2.3.1 Steps 1-4 on collating and publishing metadata

Steps 1-4 focus on collating and publishing discovery metadata on the spatial datasets held by public bodies (or their contractors). This discovery metadata is designed to enable user groups to find and locate the datasets. Further information about metadata and how to create it is provided by data.gov.uk[31]:

> Step 1. Each public body should evaluate its data catalogue/inventory to identify data within scope of INSPIRE

Step 2. All data that falls under INSPIRE should be published, with view and download services, by 3 December 2013 (or as soon as practical). Existing data do not need to conform to the INSPIRE data specifications until 2018/2020. However, we advised to make data available in these formats as soon as possible

Step 3. Create INSPIRE compliant metadata for each dataset in the UK Location profile of GEMINI2.1 must be used to ensure the data.gov.uk discovery service is INSPIRE compliant

Step 4. Register the metadata on data.gov.uk so that your authority's data can be searched externally and in-turn registered with INSPIRE

3.2.3.2 *Steps 5-9 on publishing the data through network services*

There are a number of publication options available to public bodies including locally-hosted own web services, a partnership site, a hosted site or through a national solution. The LGA is looking into how commissioning publishing services for local authorities can potentially be provided and funded.

Step 5. Some data can be published through existing national portals, the indicative topic list at Annex B includes a column which shows the data topics which can be published in this way[32]

Step 6. Discuss your publishing options with your business/IT or finance services department to identify the best way to provide INSPIRE compliant view and download services

Step 7. Ensure you comply with licensing arrangements before publishing the data. If data are not bound by any third party licensing restrictions such as the Public Sector Mapping Agreement (PSMA)[33], it should be published under the OGL

Step 8. Create INSPIRE compliant metadata for each service and update the dataset metadata to reference these services. In the UK the UK Location profile of GEMINI2 must be used to ensure the data.gov.uk discovery service is INSPIRE compliant

Step 9. Register the metadata on data.gov.uk so that your authority data and services can be searched externally and in-turn registered with INSPIRE

3.2.3.3 *Steps 11 to 13 on publishing the data through network services using INSPIRE data specifications*

Step 10. INSPIRE datasets need to be transformed to new formats for interoperability. The INSPIRE data specifications describe the requirements for each of the 34 data themes. Guidance and data models are available at http://inspire.jrc.ec.europa.eu/index.cfm/pageid/2/list/2. The data models although appearing complicated are generic, being designed to be applicable to broad environmental data themes and extended by domain communities for specific application needs

Step 11. Data can be mapped (transformed) to the new INSPIRE data schema

Step 12. Once the data has been transformed, update your metadata entry and publish it

---

[32] http://www.local.gov.uk/c/document_library/get_file?uuid=ba19b779-eb8d-404c-af07-25ecd921aed9&groupId=10180
[33] http://www.ordnancesurvey.co.uk/business-and-government/public-sector/mapping-agreements/public-sector-mapping-agreement.html

3.3 The Open Data initiative

The Open Data initiative is a movement promoting and supporting Open Data and open access to data at all levels across the public and private sectors. Open Data itself is data that can be freely used, shared and built on by anyone, anywhere, for any purpose. It is preferable that data meet the following criteria:

- Accessible (ideally via the internet) at no more than the cost of reproduction
- Unencumbered by limitations based on user's identity or intent
- In a digital, machine readable format for interoperation with other data
- Free of restriction on use or redistribution in its licensing conditions

The UK Government published the Open Data White Paper[34] in June 2010, which for the first time set out a commitment to a 'right to data', releasing Public Sector Information (PSI) as machine readable Open Data wherever possible. Open data is now at the heart of UK Government's Transparency Agenda to promote greater clarity across government. It underpins the government's commitment to enable the public to hold politicians and public bodies to account, to reduce the national economic deficit and deliver better value for money in public spending, to drive growth built on data as a raw material and to inform choice over the public services. This was formally set out in June 2012, in the Cabinet Office Open Data Strategy[35].

The Open Data White Paper committed government (public bodies) to identifying and publishing priority datasets and over time removing barriers to the publication of more datasets. These were more directly identified within the Departmental Open Data Strategies[36] with a clear timeline for publication and a route for users to request further data to be made open.

Defra's strategic vision of Open Data is aligned with the PSI Principles[37] developed under the Government information and communications technology (ICT) Strategy[38] and informed as a whole by two key principles:

- Information is valued as an asset
- Information is managed

An on-going programme of work is in place to assure the confidentiality, integrity and availability of information, and to ensure that it is appropriately managed. As part of this vision, Defra recognises the importance of making scientific and research information and data as readily and easily available as possible. The Evidence Investment Strategy: 2010-2013 and beyond[39] highlights Defra's commitment to this where feasible, recognising that widening access will help gain best value from investments, encouraging the reuse of data, and improving knowledge exchange, innovation and economic growth.

Defra's work to adopt open standards for air quality and plans for future work are expected to place air quality on a 4-star rating as set out in the Open Data scoring system described in Appendix 2. UK-AIR has made steady progress in improving its Open Data score since 2010. Progress is summarised by the information presented in   , which shows that in 2010, although UK-AIR downloads were generally available in structured (Excel) machine readable formats and non-proprietary (CSV), the absence of open licence attribution meant UK-AIR datasets did not even score a 1-star rating. In 2012 the OGL was adopted for all datasets that were available for public download. This immediately lifted the UK-AIR Open Data score to a 3-star rating and, with the adoption of the recommendations made in this report, it is anticipated that Defra's data holding served via UK-AIR will achieve a 4-star rating. To date, Defra air quality teams have focused efforts on achieving regulatory compliance and Open Data rather than Linked Data, having focused upon establishing a robust and stable core reference dataset using a controlled vocabulary for describing air quality objects. With these in place following the recommendations made in this

---

[34] http://data.gov.uk/sites/default/files/Open_data_White_Paper.pdf
[35] http://data.gov.uk/sites/default/files/Cabinet%20Office%20Open%20Data%20Strategy.pdf
[36] http://data.gov.uk/sites/default/files/Defra%20Open%20Data%20Strategy_10.pdf
[37] https://www.gov.uk/government/publications/information-principles-for-the-uk-public-sector
[38] https://www.gov.uk/government/publications/uk-government-ict-strategy-resources
[39] https://www.gov.uk/government/publications/defra-s-evidence-investment-strategy-2010-to-2013-and-beyond-2011-update

report, there is ample scope for the Open Data community to establish linkages with air quality data.

Figure 3 Summary of improving in UK-AIR Open Data score since 2010
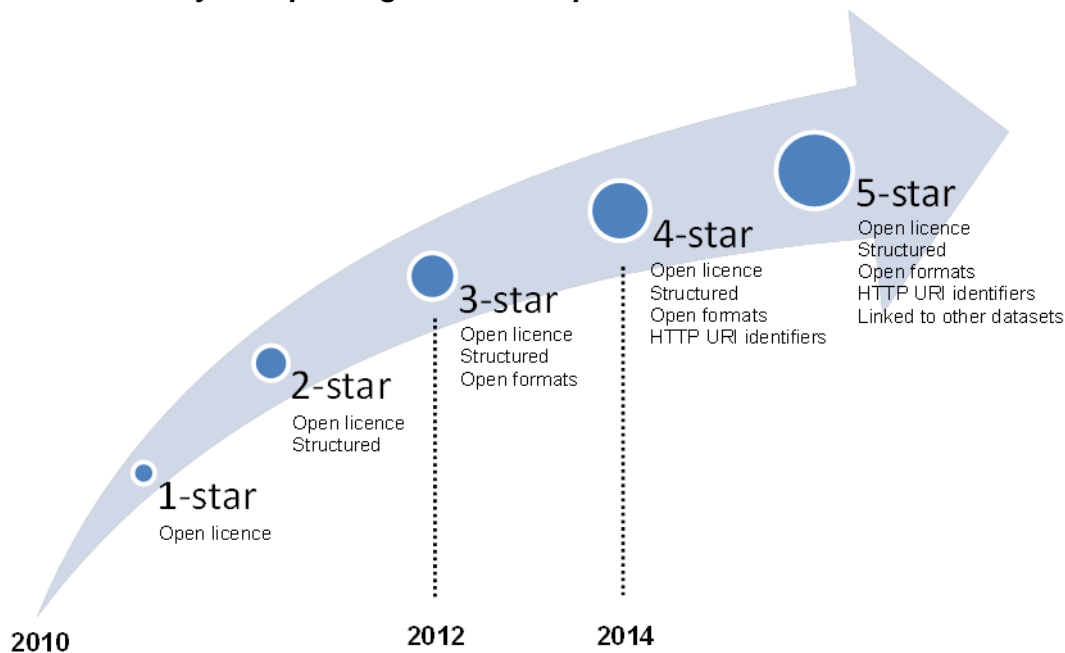
3.3.1 Core outcomes of the Shakespeare Review for Air Quality

A key commitment in the Open Data White Paper was the commissioning of an independent review of PSI. Launched in October 2012, the Shakespeare Review[40] looked at progress since the White Paper on opening up public data, and assessed how the Government should best use PSI to support economic growth.

In response to the Shakespeare Review the Government has recognised the need to present a coherent strategy on the government's policies on opening up PSI, and the resultant National Data Strategy will be taken forward through the Information Economy Strategy[41].

The review also called for the publishing of PSI to follow a twin-track policy for data-release, which recognises that the perfect should not be the enemy of the good: a simultaneous 'publish early even if imperfect' imperative and a commitment to a 'high quality core'. This twin-track policy aims to maximise the benefit within practical constraints, reducing the barriers for poor or slow delivery; it says 'get it all out and then improve'. The review also called for a high quality core dataset to be defined as 'National Core Reference Data'.

The Government is supportive of the call to set out a clear, predictable top-down process for the release of the remaining major datasets across Government departments, prioritised to drive economic growth. This release should complement rather than replace the existing, and much valued, bottom-up process that allows data businesses to highlight individual datasets that are valuable for their enterprise, but which might struggle to be defined as 'core'. Given the task of identifying these data and the breadth of stakeholders involved, the Cabinet Office Transparency Team has set out a collaborative process for identifying those datasets, which should be part of the National Information Infrastructure[42] (NII). In addition to Open Data, published and unpublished air quality datasets already have a presence on the NII via data.gov.uk and UK-AIR.

---

[40] https://www.gov.uk/government/publications/shakespeare-review-of-public-sector-information
[41] https://www.gov.uk/government/publications/information-economy-strategy
[42] https://www.gov.uk/government/publications/national-information-infrastructure

There are clear similarities in the proposed twin-track publishing model and existing practices for air quality, where it has long been common place to publish 'provisional' data in near-real-time (every hour), followed by high quality 'ratified' datasets in a quarterly timeframe.

3.4 The Government Digital Service

The GDS is a new team within the Cabinet Office Efficiency and Reform Group, tasked with transforming government digital services. GDS was established in 2010 to ensure that government offers world-class digital products that meet people's needs.

GDS works in three core areas:

- Transforming 25 high volume exemplars from across government into quality digital services
- Building and maintaining the consolidated GOV.UK website, which brings government services together in one place
- Changing the way government procures IT services

The aim of GDS is to be the owner of a high quality user experience between people and government by being the architect and the engine room of GDS provision. To do this GDS is developing different devices and channels to offer services to people across a wide range of demographics to promote digital inclusion. GDS is working collaboratively across government to:

- Make accessing government services easier for users
- Drive cost efficiencies and help save government money
- Open up government to the public

A core part of the GDS strategy to achieve these goals is to manage the transition of existing government digital services to a single government domain, GOV.UK, thereby replacing all the individual websites of government departments and agencies. This will provide, for the first time, a single destination for government information designed with users in mind, making it quick and easy for them to find the information or service they need. The ambition behind this transition is to remove duplication, save money and present content in a much more joined up and clear way, recognising that people should not have to wade through the complex structure of government to find what they need.

GDS plans to migrate all mainstream content to GOV.UK by April 2014, although this deadline will be kept under review. Even so, the transition and timescales have a potentially significant impact on the future of Defra air quality data services.

Transition to GOV.UK is prioritised, phased and process driven, being supported by guidance at many levels. In its initial roll out, GOV.UK will accommodate many content types, including guidance, announcements, publications and corporate information. In the planning process, content analysis is used to identify the scope of content from existing websites that should be transferred and where it should go on GOV.UK. However, owing to the large amount of web content to be transferred and the variety of content types, GDS recognises that in its initial roll-out, tools and transactions that current GOV.UK formats cannot accommodate are out of scope for the transition. Preliminary discussions with GDS in relation to the NAEI website has identified this service as outside of the scope of the current transition priorities, owing to the level of specialist tools, fast refresh content and web services; a viewpoint that is expected to be shared in relation to UK-AIR, which has a higher degree of specialist content, tools and web services.

However, in order to align the redesign of Defra's data services with the GDS design principles[43] this report has taken on board GDS best practice, in particular in the analysis of user needs[44] and in choosing technology[45].

Section 5 sets out an approach to identifying user needs using data from UK-AIR analytics logs and a user engagement programme aimed at capturing core user needs. Section 6 outlines a

---

[43] https://www.gov.uk/design-principles
[44] https://www.gov.uk/service-manual/user-centered-design/user-needs.html
[45] https://www.gov.uk/service-manual/making-software/choosing-technology.html

recommended approach, assessing technologies for implementation based on GDS best practice where possible and taking into account the other external drivers set out in this section.

3.5 Open Government Licence

Public sector information, including all air quality datasets owned by Defra, is subject to the OGL[46] . The OGL licensing terms allow people to 'use and re-use the Information that is available under the licence freely and flexibly, with only a few conditions.' Under this licence, data users are free to:

- Copy, publish, distribute and transmit the Information
- Adapt the Information
- Exploit the Information commercially and non-commercially for example, by combining it with other Information, or by including it in your own product or application

The licence is expected to encourage wider use of Defra's air quality data, which will be further promoted by the improved accessibility to data brought about by the proposals in this report.

3.6 Benefits and barriers

Section 3 has shown that there a range of initiatives, regulatory and policy drivers that are influencing the digital data services at a number of levels. This section explores the benefits and barriers of future Defra air quality digital data services, which have been shaped by the initiatives presented and careful consideration of user needs.

3.6.1 Policy benefits

The Policy benefits of aligning Defra's air quality data services are clear and are summarised below:

1. Alignment of Defra's air quality data services and Policy with the regulatory requirements of the Air Quality Directive (AQD) and its Air Quality Implementing Rules for Reporting (IPR) for e-Reporting
2. Alignment of Defra's air quality data services and Policy with the regulatory requirements of the INSPIRE Directive
3. Alignment of Defra's air quality data services and Policy with the Cabinet Office and Defra Open Data Strategies
4. Improved evidential value of Defra's air quality evidence programme
5. Improved access to air quality data for all in order to inform, empower, promote behavioural change and serve Policy
6. Improved access to air quality data for all to create the maximum value.
7. Reduced burden of Freedom of Information and Environmental Information Regulations requests
8. An improved governance model for UK air quality data
9. Facilitating evaluation of whether Defra's air quality data are being used to maximum effect in developing, evaluating and adapting policy

3.6.2 Benefits to the air quality community

The benefits to the air quality user community arising from greater openness in air quality data through Air Quality e-Reporting, the INSPIRE initiative, government Open Data strategy and digital design reform are expected to be many. The key benefits are summarised below:

10. Alignment with the regulatory obligations, data model and vocabulary of the Air Quality e-Reporting initiative and with comparable datasets across Europe
11. Alignment with the regulatory obligations, data specifications and service standards of the INSPIRE Directive to enable comparable datasets across Europe and cross environmental theme linkages

---

[46] http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/

12. Consistency with Defra and wider Government strategies supporting Open Data
13. Improved data discovery systems for air quality
14. An improved governance model for UK air quality data
15. Reusable, centrally published web maps and download services that can be consumed by third party services
16. Improved transparency in the air quality data models arising from controlled vocabularies and improved air quality metadata on monitoring stations, sampling points, zones, measurement techniques etc.
17. Improved traceability supported by adoption of internet standards allowing hyperlinking of downloads to cross-reference and cite metadata e.g. providing hyperlinks to UK-AIR's station information archive which contains attribute data on air quality monitoring stations
18. Machine readable Application Programming Interfaces (API)-type interfaces to Defra's air quality databases facilitating automated processing and post-processing of data including QA/QC checks, consistency and compliance/conformance checks etc.
19. A scalable and extensible data model that is reusable at local, regional and national levels.

3.6.3 Benefits to wider communities

The benefits to other non-air quality communities will include:

20. Improved accessibility via new discovery systems for air quality data
21. Improved openness via the use of standard based data models and data formats
22. Improved accessibility via implementation of machine readable API-type interfaces to Defra's air quality databases
23. Open Data to adapt and improve public services and businesses and enhance our whole way of life, bringing economic growth and wide-ranging social benefits.
24. Reusable, centrally published web maps and download services that can be consumed by third party services
25. Placing air quality on the threshold of the semantic web by supporting Open Data that can be linked to and support its re-use at all-levels

3.6.4 Barriers to adoption

An important potential barrier to wider uptake of the INSPIRE conformant Air Quality IPR e-Reporting data model is its complexity. It is imperative that this is recognised. At a European level within the Air Quality e-Reporting pilot group, inertia to the uptake of the data model has been observed. This has arisen as a result of the new approaches to data modelling, unfamiliar terminology and new best practices that INSPIRE has brought to the air quality community that are often alien to traditional air quality data users. Documentation and guidance supporting the uptake of the new technologies has also proved to be problematic, interestingly not because of the lack of guidance, but more because the amount of it and the style in which it has been written. The documentation has been heavily influenced by the IT standards community and is not intuitive for air quality practitioners. These barriers are important to recognise and support will be required to help air quality users navigate the inevitable learning curve. Valuable experience in this respect has been gained from the Air Quality IPR e-Reporting pilot and these lessons learned can be brought to support future Defra data services and service users. To a degree the harder work to support and promote uptake has already been done at a European level, although application to UK specific needs will require some additional work in some areas.

Of particular importance is that traditional CSV downloads must be maintained for human users and improved where possible via the inclusion of metadata links.

4 Design principles for digital services

The GDS Design Standards will apply, from April 2014, to government digital services that meet the following criteria:

- Services that are completely new and/or being redesigned
- Services that are processing (or are likely to process) more than 100,000 transactions every year
- Services that are the responsibility of a central government department, agency or non-departmental public body

Based on these criteria this roadmap study falls partially in scope of the GDS Design Standards. Clearly the UK-AIR web pages are the responsibility of a central government department. In addition, UK-AIR handles 251,835 visits per year although these are not within the definition of service transactions. However, the work recommended by this project is not strictly developing a new service or redesign of existing services. Rather it seeks to develop particular aspects of the current service to meet emerging user needs and legislative drivers. Even so, the GDS design standards provide some important lessons and guiding principles[47] which are relevant to this study.

The principles can be summarised as follows:

1. Start with needs
2. Do less
3. Design with data
4. Do the hard work to make it simple
5. Iterate. Then iterate again.
6. Build for inclusion
7. Understand context
8. Build digital services, not websites
9. Be consistent, not uniform
10. Make things open: it makes things better

Of these, points 1, 2 and 10 are instantly applicable to the roadmap but all have relevance and merit consideration as may impact on the approach for future service provisions relating to air quality data. Points 1, 2 and 10 are evaluated in more detail in the remainder of this section in conjunction with the service standards criteria set by the Digital by Default Design Standards team[48].

4.1 User needs

GDS guidance recommends that "we need to be clear about the need the *user* has of government and not the need of government to impart information to the user." In response to this requirement this project has endeavored to update an already good understanding (as a result of previous studies) of user needs, in particular those needs that are not currently met by UK-AIR but could be captured by improved services. Section 5 presents an analysis of the user needs conducted by this project, including a stakeholder engagement workshop focussing on gathering user stories (and needs) based on guidance provided by GDS[49].

Reflecting on the user stories guidance provided by GDS, there are a number of criteria that resonate and validate the level of service (and proposed services) provided by UK-AIR. Clearly it is

---

[47] https://www.gov.uk/design-principles
[48] https://www.gov.uk/service-manual/digital-by-default
[49] https://www.gov.uk/service-manual/users/user-needs.html

important to set the boundaries for services that government can and should offer, to prevent a bottomless list of user needs. The most important services to be provided to users are those which:

- Are *only* provided by government
- Have a clear demand from users (i.e. through search and traffic logs)
- The government is legally obliged to provide

In the context of these criteria and UK-AIR, providing user friendly access to Defra and the Devolved Administration's air quality data is something that only government can do and is also something the government is legally obliged to provide, to meet the requirements of INSPIRE and the public information requirements of the AQD. There are also other legislative drivers specific to particular datasets. Section Identifying user needs5 also evidences a clear demand from users for the information as indicated by the website traffic and customer insight work carried out under this project. In addition, providing access to the information and research that government pays for is strongly supported by the Open Data and transparency agendas.

It is important to consider the user types, including 'human users' wishing to manually access data through user defined queries for desktop analysis, and the emerging 'machine users' requiring automated access to data in machine readable digital formats. User needs and the imparting of information to the user are also important in the context of service design, and should not be seen as contradictory to the aims of Open Data. Data are an asset, should be made available under the Open Government License and the services should meet user needs in terms of the methods of accessing, visualising and retrieving those data.

4.2 Doing Less

As indicated above, boundaries need to be set to define what a service can and should offer, and in this context government should concentrate on the irreducible core needs. Reuse of external resources should be promoted where practicable and appropriate, with links to these services from government websites. If resources, like APIs, are available that will help users access data, then these should be utilised.

For the roadmap project this means concentrating on the datasets that are government owned and for which government are responsible for managing and publishing. However, this ethos should not detract from the expected benefits of Defra taking a lead to encourage, support and promote consistent approaches across the air quality community, particularly where this helps to reduce burdens on local authorities.

4.3 Make things open: it makes things better

GDS design standards advocate sharing what we doing whenever we can 'with colleagues, with users, with the world. Share code, share designs, share ideas, share intentions, share failures. The more eyes there are on a service the better it gets.'

For the roadmap this means sharing our experiences with others in the air quality domain, learning from the work of the UK INSPIRE Project team, using existing infrastructure and using open source methods where we can.

4.4 Further resources

GDS also provides resources for Service Managers[50] for example, when considering APIs, or for choosing technology or appropriate formats. The roadmap will consider the principles and considerations set out in this manual.

---

[50] https://www.gov.uk/service-manual/service-managers

4.4.1 Build an API by building with the API

GDS guidance[51] recognises that when building an API there is always a danger of building the wrong thing in the wrong way for the wrong people. This is especially a risk in the absence of a developer community driving the needs behind the API.

For the roadmap this means making sure that the API is appropriate for a wide range of needs, uses open standards, formats and vocabulary and is consumable by users at many levels.

4.4.2 Technology

GDS guidance also recognises the difficulties of selecting the right technology[52] and accepts that an educated guess may be the best or only way of selecting technology when little is known about the domain and the user need. After implementation, it is important to challenge the choice of technology as lessons are learnt about the domain, the user needs and the required capabilities of the technology.

For the roadmap this means thinking broadly about what the technology is to deliver and considering what is fit for purpose, provides best value, and being open to new ways of working.

4.4.3 Formats

Choice of data/file format to fit the nature of the information they contain, and the uses to which they will likely be put, is an important consideration[53]. For data, use of CSV for human users is an important consideration whereas other 'structured data' formats (for example JSON and XML) will be the preference for automated data sharing. Publishing structured data in unstructured formats such as PDF is discouraged. In summary, it is necessary to consider the users, and the uses to which published data and content will be applied. The assumption that users can read proprietary formats is not appropriate and, wherever possible, publication in accessible, patent-free, open formats, for which software is widely available on a variety of platforms, is to be promoted. If publishing in proprietary formats, a non-proprietary alternative should always be available. For textual reports, GDS recommend that government provides HTML, plain text (.txt), or PDF rather than formats that require proprietary software to view, such as Word documents (.doc/.docx). For tabular data, provide CSV or tab separated values (TSV) rather than Excel spreadsheets (.xls/.xlsx). If in doubt, the native format of the web, HTML, should be treated as the default option.

For the roadmap this means considering the storage format of the data as well as the access options, which should be available through queries. This report addresses the needs of the users, the legislative drivers and considers the role of proprietary formats, in particular Environmental Systems Research Institute (ESRI) products, where these can add value to the data.

---

[51] https://www.gov.uk/service-manual/making-software/apis.html
[52] https://www.gov.uk/service-manual/making-software/choosing-technology.html
[53] https://www.gov.uk/service-manual/design-and-content/choosing-appropriate-formats.html

5 Identifying user needs

This section evaluates the needs of communities consuming air quality data with a view to identifying the services required. The evaluation makes use of website usage statistics to inform the design of future data services analysing which pages, tools and apps are most often used and the most frequent entry points to UK-AIR. The outcomes of the stakeholder engagement workshop, focussing on gathering user stories and needs, are also presented.

The evidence generated updates our understanding of user needs, in particular those needs which are not currently met by UK-AIR but could be captured by improved services.

5.1 Google Analytics

Analysis of recent use (12 months from November 2012 to October 2013) of the UK-AIR website using Google Analytics shows that:

- 164,402 individuals visited the website over 251,835 visits
- Users view an average of 5 different pages during their visit
- 36% of the people visiting the site have visited previously, while the remaining 64% were first–time visitors
- Only 63% accessed the website from the UK, with a significant number of visitors from India and the United States
- 21% of the visits during the year were 3 minutes or longer in duration. However, the majority (58%) of the visits during the year were less than 10 seconds in duration
- 87% of users accessed the website from a desktop, although access via mobiles and tablets was also popular

The most popular pages are listed in Table 3:

Table 3   Top Ten Most Visited Pages UK-AIR

Number	Page	% Page views
1	Home Page	8.9%
2	/air-pollution/effects	6.1%
3	/interactive-map	3.6%
4	/data/data_selector	3.5%
5	/data/	2.1%
6	/latest/	2.1%
7	/forecasting/	1.9%
8	/air-pollution/	1.7%
9	/networks/	1.5%
10	/data/data_selector?q=36286	1.4%

Number 10 is an example of an innovative, although inefficient, automated data scraping process that has been set up by a particular user to re-use the same database query to extract the latest data for all monitoring stations every hour.

The top 1000 users during the year have been reviewed and from these, several groups have been successfully identified from their IP address, shown in Table 4:

Table 4   *UK-AIR User Types identified by IP address*

User Type Identified	% of Visits
Schools and Universities	6.7%
Local Authorities	3.9%
Air or Environmental Consultants	3.4%
Government and Government Agencies	2.1%
Media, Banking, Industry and Charity	1.4%
Other, unidentifiable within top 1000	64.6%
Others not in top 1000	17.6%

5.2 Email Queries

Defra provides an email address for users to contact the UK-AIR team with specific queries. In the 12 months from November 2012 to October 2013, 398 genuine queries were received and dealt with. The majority of these were from members of the public with most questions falling into three categories:

1. Data requests for monitoring data before 1961, species that are not measured or for data from analysers that are closed or out of service
2. Requests for links to other websites to be included on UK-AIR
3. Questions about the health impacts of specific local pollution sources

5.3 Stakeholder consultation

A workshop was held in London during October 2013 for key stakeholders, including the Defra Evidence team, Defra UK Project representative, the Greater London Authority, air quality data providers and researchers and Local Authority representatives. The objectives of the workshop were to:

- Update attendees on policy drivers, including Defra's work on Air Quality e-Reporting, INSPIRE and Government Open Data projects
- Outline proposals for future air quality data services
- Discuss user needs for access to Defra air quality data and to elicit feedback on the proposals for future air quality data services
- Obtain updates from stakeholders on their air quality data management activities
- Discuss how the air quality data community can work together to meet user needs and support the Open Data agenda

The agenda, minutes of the meeting and a list of attendees can be found in Appendix 3.

As part of the Defra INSPIRE monitoring and reporting programme, the Defra INSPIRE project team have asked Ricardo-AEA to contact the UK air quality community to gather information on progress towards INSPIRE Directive obligations. The 3 December 2013 marked the first deadline

for Annex III data themes (including air quality) under INSPIRE. As of this date, public bodies responsible for air quality data were obliged to have set in place discovery metadata and services, viewing services and download services, the latter in existing formats. Ricardo-AEA contacted all air quality data holders in England using the UK Earth Observation Framework (UKEOF) catalogue to identify non-Defra initiatives in place to meet INSPIRE obligations. Ricardo-AEA is in the process of collating questionnaire responses and will report back to the Defra INSPIRE project team in due course.

5.3.1 Key messages

1. The benefits of ensuring ease of access to data for users and compatibility/interoperability of datasets were supported by attendees.
2. Data providers presented a large range of interesting developments in their services, with many common themes and requests. Web based services and APIs have already been developed by some providers.
3. There was interest in and support for a future flow down of the EU air quality e-Reporting vocabulary and data models. It is likely that using these in practice will present many technical queries and challenges as they are adapted and extended for different user needs and the UK domain.
4. Risks and issues were discussed and included a growing use of air quality data by non-experts with scope for misinterpretation and misleading results. Technical issues included demand and pressure on servers from web service access (particularly at high time resolutions), data management overheads related to serving spatial datasets, the verbosity of the XML data model and the need for fast lightweight data feeds, up-skilling of user communities to maximise use of new XML technologies, as well as ensuring data is correctly licensed and clearly attributable to a source.
5. Other data providers also have catalogues including the Centre for Ecology and Hydrology (CEH), the British Atmospheric Data Centre (BADC), the Natural Environment Research Council (NERC), UKEOF and the Met Office. Linking these electronically for automatic harvesting where records relate to the same dataset will be important. Catalogue content will need to be compliant with the UK INSPIRE Project implementation of the Gemini2 metadata profile in order for them to be accepted by data.gov.uk.
6. Actions and points for further discussion were raised and are documented in the minutes.

5.4 Stakeholder Questionnaire

In addition to the workshop the key stakeholders were asked to complete a questionnaire to capture user needs for Defra air quality data and the approaches of data providers to publishing data. The questionnaire is attached in Appendix 4.

5.4.1 Key messages

1. 75% of data users felt they did not have access to all the Defra air quality data and metadata they needed
2. Half of data users felt they did have access arrangements to meet their needs
3. Half of data users wanted options to download data in both CSV and XML formats. A further 33% had no interest in XML
4. Half said they were likely to use other format options WMS, WFS, SOS
5. Around half thought that the new data catalogue & INSPIRE metadata tools met their needs. The rest thought that things had improved but were "not quite there yet"
6. Everyone who expressed an opinion wanted automated access to data
7. Four organisations identified themselves as having obligations under INSPIRE (CEH, Department of the Environment Northern Ireland, King's College London and Ricardo-AEA)
8. The INSPIRE Directive requirements and obligations are not clear to everybody
9. Multiple formats are being used for different user needs and services: Excel, PDF, CSV, Word, HTML, Text, XML/GML, phone apps, SMS, bulletins, email, voicemail, website widgets, WMS services, JSON, GeoJSON, NetCDF, ASCII grid

10. The commonly recognised opportunities were for standardised formats, vocabulary and aggregation rules, improve and link metadata, LAQM and open standards based web services

11. Challenges identified by data providers included funding, up-skilling, linked datasets access, for example, meteorological and traffic data, and systems which scale for different levels of complexity and involvement

User stories, describing user requirements and goals, can be found in Appendix 5. A summary of the key user needs is given in Table 5 below.

Table 5 Key User Needs

Actors	Narratives	Goals
I am the Defra Air Quality team and I am responsible for implementing INSPIRE for air quality.	I need services to respond to my legal / policy obligations.	Delivery of services that meet the legal/policy obligations and supporting opportunities for businesses/service providers.
I am the GLA and I am responsible for coordinating response to domestic and regional policy.	I need services to provide me with access to central and local data sources.	Delivery of high quality and timely information and advice on AQ to the public.
I am a consultant delivering client needs with AQ data.	I need services to deliver data in time resolutions appropriate to my client needs.	Delivery according to the needs of my clients.
I am a data provider: Network owner (funding authority), Network operator, Data publisher/ manager	I need services that allow me to check the consistency of the data that I have provided to UK-AIR systems.	To ensure that AQ information in UK-AIR is accurate and of high quality.
I am a researcher publishing scientific papers.	I need services to deliver data in time resolutions that support my research needs for modelling and monitoring.	Delivery of quality scientific research to advance understanding, policy & improve the environment.
I am a computer programmer (hacker)	I need data feeds that allow free unrestricted access to data in easy to use standard formats to support the development of web and mobile apps.	Delivery of data and information social media channels to enhance well-being and knowledge
I am a contingency planner	I need instant access to real time data in formats compatible with my systems	To ensure evidence based decisions are taken and emergency situations can be monitored.

5.5 Conclusions of Stakeholder Analysis

There are some core themes coming through from the user story feedback. User needs can be summarised as follows:

1. Robust and reliable data and access to services that allow the integrity of data to be checked or verified
2. Robust metadata
3. Flexibility in accessing aggregated statistics based on user driven interfaces
4. Standardised CSV downloads are as important as computer readable formats
5. Users would look to harvesting data via XML feeds in the future, particularly for an API type interface with UK-AIR
6. Access to data in the highest time resolution possible. Traditionally 1-hour averages were the highest time resolution possible. Requests for 15-minute average data for all automatic datasets were made.

Some of these needs align with the outcomes of the previous study in 2010 but with subtle differences; the majority of the user needs focus on access to data in highest time resolution in a

consistent/standard format(s) which is traceable to metadata. Previously the emphasis was on data in consistent formats that was discoverable. These needs will be address in the proposed solutions presented in Section 6.

It should be noted that data services in support of Local Authority air quality Action Planning were not considered due to the scope restraints on the project. However, it is recognised that there are continued user needs for improved access to information on air quality measures by local authorities seeking to understand which measures are effective at tackling air quality and by Defra who require improved management of this information to enable data-sharing and reporting to the EU.

The missing service components required to improve UK-AIR's Open Data scores, achieve INSPIRE conformance and meet emergent user needs have been summarised in Table 6.

Table 6 Essential UK-AIR components and proposed solutions

	Requirement	Proposed implementation
Services	INSPIRE conformant discover service	• GeoNetwork / data.gov.uk
	INSPIRE conformant viewing service	• ArcGIS Server WMS
	INSPIRE conformant download services	• Either RSS Atom, WFS or SOS. To be determined.
Formats	Human & machine readability	• Air Quality e-Reporting XML schema to support machine readability hosted on an INSPIRE conformant download service, type to be determined. • CSV formats to support human users.
Accessibility	Improved traceability	• To be determined.

Via these new services, UK-AIR could deliver :

- Services optimised for the specific needs of the intended air quality community user groups i.e. service oriented
- Data feeds (downloads) that are robust, traceable and available in the highest time resolution possible
- Unrestricted and simple access to pre-defined data downloads within the scope of agreed update frequency(s) of the Defra network data ratification contracts and fair use policies for service usage
- Scalable web services for data sharing across all user groups including human and computer users
- An improved CSV formatted data download, the persistence of which would be equally important as automated downloads going forward

6 Infrastructure and architecture supporting Defra AQ data services

Section 2 of this report summarised the information services supported by UK-AIR. This section investigates the infrastructure and systems used to support and deliver these and evaluates the options for extending the services to meet INSPIRE obligations and to improve Defra air quality data's Open Data score.

The UK-AIR website is hosted on a Linux, Apache, MySQL and PHP (LAMP) framework which is a powerful and flexible web application platform and is open source. LAMP platforms were the de facto development standard at the time when the UK-AIR was being commissioned. When reviewing the existing systems with a view to improving UK-AIR's Open Data score and conformance with INSPIRE it is helpful to distinguish the different types of information published on UK-AIR:

1. Static pages (HTML pages, PDF documents) that provide textual information and static images, graphics or reports
2. Proprietary and non-proprietary information visualisation tools e.g. Google Maps, Google Earth, Google Charts or ESRI's ArcGIS Server that can deliver rich data viewing, analysis and interrogation capabilities for specific applications
3. Communication channels e.g. Twitter feeds and email bulletins that deliver information
4. Observational data (measurements, predictions, forecasts etc.) and metadata (attribute data on the configuration of the measurement, model etc.)

This distinction helps to separate services that publish data or spatial data from information, from tools that facilitate knowledge sharing and policy decision-making. Hence, when considering the content of UK-AIR in this way it is possible to further focus on the core services supplying data; evaluate the data formats, user needs, update frequencies and the payload (file size) of downloads in order to explore the best options for extended services to support Open Data and INSPIRE.

Further considerations include the options available to improve or support transparency and machine readability. Both of these are explored in this section for UK-AIR data services.

6.1 Closing the gaps between existing UK-AIR systems, INSPIRE and improved Open Data scores

Previous sections have shown that UK-AIR systems host a variety of data download services, which can be summarised as:

1. Predefined downloads of flat files for bulk download of historical data
2. User defined downloads from the UK-AIR data selector via a Hypertext Transfer Protocol (HTTP) query string (form)
3. RSS feeds with content
4. Data viewing tools for target and guide information sharing and download of spatial datasets in proprietary GIS formats other open formats

This combination of services ensures that the UK-AIR's data can be easily accessed by users and also automated via scripts from client machines. As a result UK-AIR's data scores well on the simple Open Data scoring systems (see Appendix 2). Even so, there are key gaps which must be addressed to achieve INSPIRE conformance and improved Open Data scores. These gaps are principally related to non-conformance of UK-AIR data encoding specifications with those mandated by INSPIRE and the non-conformance of UK-AIR's download services with INSPIREs guidance on network services (web services). The following section(s) make a series of technical

recommendations aimed at closing the gap with INSPIRE and achieving Open Data goals. The recommendations are presented in call out boxes and framed by a description to provide context with regard to the requirement / need of UK-AIR, INSPIRE or Open Data.

In the main, data downloads from UK-AIR are available as CSV, although some use of Keyhole Mark-up Language (KML) is made for a limited number of downloads via Google Maps / Earth. INSPIRE requires downloads to conform with the data specifications for INSPIRE, which are XML based, and use specific vocabularies to encode the attributes / content of downloads and associated metadata. In addition the download services must conform with download standards for web services. None of these requirements are currently supported by UK-AIR download services. Adoption of INSPIRE conformant formats for encoding downloads from UK-AIR and INSPIRE conformant web services for the download process would also greatly improve UK-AIR's Open Data scores.

Recommendation 1. Adopt the Air Quality IPR e-Reporting data model and schema for encoding of downloads of observational data and metadata for air quality spatial objects.

Recommendation 2. Adopt download web services that are conformant with INSPIRE guidance on network services for download services for all datasets already identified as within scope of INSPIRE i.e. those datasets with discovery metadata and viewing services registered on the UK-AIR data catalogue and data.gov.uk.

Adoption of the Air Quality IPR e-Reporting data model and schema for encoding of downloads represents a quick win as the hard work in the development of this data model in line with the INSPIRE data specification has already been done at a European level. It would also be relatively straightforward to extend this data model to the UK's specific needs in support of the LAQM process and Air Quality Strategy policy needs.

Recommendation 3. Maintenance of the current implementation(s) for viewing services, noting the current use of proprietary software, and to undertake additional work to evaluate the conformance of existing viewing services with INSPIRE requirements and to inform future INSPIRE conformance options when these become available.

The user story analysis and stakeholder engagement presented in section 5.3 has highlighted a further consideration for an improved UK-AIR Open Data score. This observation relates to the accessibility of supporting metadata for CSV downloads. An example is given in Figure 4E below, which presents the content of the header for a download from UK-AIR's data selector[54]. The observation raised here by users is that it is up to the user to locate and download additional information on the station called "London Marylebone Road" and the configuration of its measurement systems observing levels of "Nitrogen Dioxide", even though all of the required information is available on other UK-AIR resources.

Figure 4   Example of metadata header associated with existing downloads from the UK-

```
Hourly measurement data supplied by UK-air on 6/2/2014
All Data are GMT hour ending
Status: R=Ratified P=Provisional A=As supplied V=Validated S=Suspect
,,London Marylebone Road,,
Date,time,"Nitrogen dioxide",status/unit

01-01-2013,01:00,73,R ugm-3
```

The current configuration of the CSV downloads requires users to know, in advance, where and how to access important metadata associated with the content of the download. A method

---

[54] http://uk-air.defra.gov.uk/data/data_selector

supporting improved sign-posting to supporting metadata would promote openness and transparency in CSV downloads.

Recommendation 4. *Adopt an improved system for sign-posting to metadata in the CSV downloads to promote openness and transparency. Implement based on the identifier and metadata resources developed for Air Quality e-Reporting to align metadata standards for human and machine readable downloads.*

6.1.1 Additional user needs

Section 5.5 summarised six key messages from the user needs analysis conducted during this study. Of these, user needs 1-3 are supported by existing UK-AIR services and reaffirmed the need for quality web services delivering data and information in human readable, desktop friendly formats. The web services supporting these user needs will also be improved in particular through the implementation of Recommendation 4 and Recommendation 7.

User need 5 indicates that users would make use of an API-type interface to harvest data from UK-AIR in machine readable format. This user need will be supported by Recommendation 1 and Recommendation 2.

User need 6, however, for access to data in the highest time resolutions available remains outstanding, in particular, for data from the AURN. UK-AIR has historically made SO_2 data as 15-minute average in support of the UK Air Quality Objective for SO_2, which uses this averaging period. However, as similar temporal resolutions are not made available for all pollutants, it is therefore recommended that 15-minute averages are supported for all pollutants.

Recommendation 5. *Extension of download services to 15-minutes for all pollutants where this is stored in the UK-AIR database.*

### 6.1.2 Other recommendations from 2010 studies

The previous studies in 2010 on Defra web-services made a number of recommendations. Many of these have already been achieved, as outlined in Section 2. However, some core recommendations remain outstanding. These include:

1. Implementation of reusable services via Service Orientated Architecture (SOA) that can be combined to create more complex composite services and flexible end user tools
2. Use of standardised approach to (and with) linking data to other datasets

Both of these recommendations are still relevant to this study and will form the basis of any response to Defra's future data services, Open Data aspirations and INSPIRE conformance. The way in which UK-AIR has and will respond to these requirements is set out in the following sections.

6.1.2.1 Service Orientated Architecture

Figure 5 presents the high-level description of the core components of SOA within the context of air quality and a suite of web services such as UK-AIR. The SOA design principle being described is that of a Service Provider, in this case UK-AIR, publishing data services to the web for a Service Consumer (a user) to access, download and reuse. In order for the Service Consumer to find the Service Provider's data services, the Provider must also register the existence of its service(s) with a Service Register. Any potential Service Consumer can then search for air quality services, discover them, locate them from metadata in the Service Register and then consume them.

Figure 5   High level components of SOA for air quality
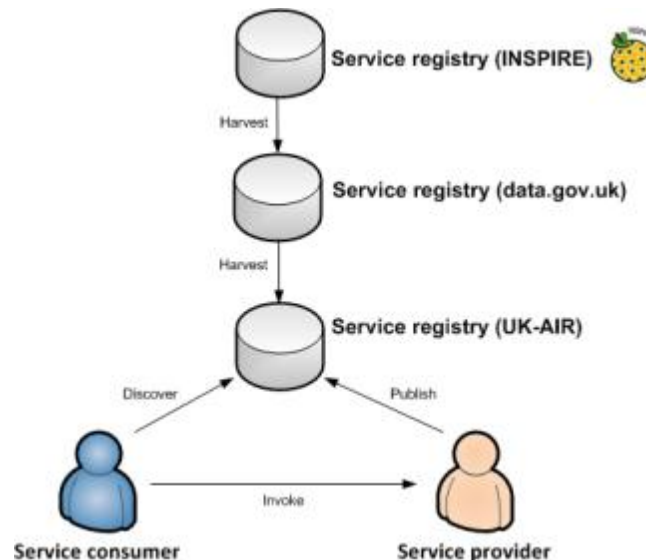
This design principle is identical to that required by INSPIRE services for discovery and, at a UK level, is supported by data.gov.uk, which acts as both the UK INSPIRE service registry for spatial datasets and as a service registry for non-spatial datasets in support of the UK Open Data agenda. As a result of the work carried out since the 2010 studies, UK-AIR is aligned with the SOA design principles. UK-AIR has a data catalogue that collects details of all the datasets and data services hosted by UK-AIR. This catalogue is available in HTML format at uk-air.defra.gov.uk/data/data-catalogue . The content of this catalogue is also hosted on a Catalogue Service for the Web (CSW). In turn, via an arrangement with data.gov.uk, the CWS is harvested, thereby registering all data and viewing services hosted by UK-AIR As the recognised INSPIRE registry for UK, data.gov.uk is also responsible for registering UK-AIR datasets with the INSPIRE registry GeoPortal. The synchronisation of the discovery data for services across the UK-AIR data catalogue, data.gov.uk and INSPRE registries in this way ensures that any services hosted by UK-AIR are consistent and accessible at community, national and European (international) levels. In summary, within UK-AIR, viewing and data services are:

1. Designed at service application level with specific application requirements in mind informed by user needs and stories
2. Services are exposed to the web for consumption
3. Discovery data for accessing published services are published at a number of levels to community and national / international registries
4. Service Consumers can request and consume UK-AIR's exposed data services via standard HTTP requests and query forms exposed within the discovery services.

As outlined in recommendations 1 to 3, a standards- based approach to the design and implementation of download services and encoding for these download content will promote improved Open Data scores. These must, in turn, be registered with the appropriate registries to achieve SOA design principles that support INSPIRE and the UK Open Data initiative.

> **Recommendation 6** *Register all web services for viewing and download with the appropriate discovery services registries.*

6.1.2.2 A standardised approach to linking data with other datasets

A 5-star Open Data rating on Sir Time Berners-Lee's scoring matrix (Appendix 2), whereby data exposed to the web is linked to other data on the web, can only be achieved by first implementing object (data) identifiers which conform with web standards. This involves implementing HTTP Uniform Resource Identifiers (URIs) to identify spatial objects (e.g. monitoring stations) within the context of the internet. The HTTP URL string can then be used as a unique handle for these

reference data objects and their attributes. For example, the HTTP URL string http://environment.data.gov.uk/air-quality/so/Station_GB0682A has been assigned to the Marylebone Road monitoring station as a unique and persistent identifier. This can be used as a means for exclusive referencing Defra's monitoring station at Marylebone Road. In addition within the context of the internet because the identifier is also a web address, the URL can also be resolved to an HTML document (or XML document) containing information on Defra's monitoring station at Marylebone Road.

Fundamentally this means implementing HTTP URIs to identify data objects within the context of the internet in order for these to be used as the unique handle for reference data objects. Use of HTTP URIs to identify spatial objects within the INSPIRE framework is also mandated[55].

To date, Defra's approach to improved Open Data ratings have been driven by regulatory obligations to Air Quality e-Reporting IPR which, as we have seen in previous sections, is underpinned by INSPIRE. In order to align Defra's air quality e-Reporting response with Open Data practices promoted by data.gov.uk, HTTP URIs have been implemented for spatial and non-spatial air quality objects within the UK Air Quality e-Reporting data model. This in itself delivers a 4-star Open Data rating for air quality. To date, Defra has focused efforts on achieving regulatory compliance and Open Data rather than Linked Data, having focused in the recent years upon defining a robust and stable core reference dataset using a controlled vocabulary for describing air quality objects. Even so, with this core infrastructure in place there is ample scope for the Open Data community to establish linkages with air quality data. A UK-AIR object registry is recommended to support linked air quality data and recommendations 1 and 3. This web page(s) would allow the HTTP URIs to be resolved as a web address, allowing objects and their attributes to be downloaded and displayed, object, rendering content in human (HTML) and machine readable (XML) formats.

> **Recommendation 7**. *Develop a UK-AIR air quality object registry to support the use of HTTP URIs from the UK Air Quality e-Reporting data model in future UK-AIR CSV and XML downloads, allowing them to be resolved to HTML and XML documents.*

In addition to these core improvements, it is also an essential requirement that:

1. Existing CSV downloads are maintained in future systems for traditional users familiar with this format, although these downloads will be improved with greater internal transparency and traceability
2. Modifications to legacy systems are kept to a minimum
3. Open source software options are taken up where possible

6.2 Choosing technology to support the Defra services

The technology options for delivering the service recommendations are varied. Following GDS design principles in selecting proposed technologies the following have been considered:

1. The desired capabilities including minimum level of service
2. Interfaces at a service level with existing UK-AIR data management software systems
3. Legacy software and the cost implication of migrating away from this software
4. Avoiding technology 'lock-in'
5. Cost

6.2.1 Capabilities required

Section 6.1 has outlined the service recommendations or capabilities required to achieve INSPIRE compliance whilst supporting regulatory reporting and promoting an improved Open Data score. These capabilities should be considered to be the minimum requirements.

---

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

At a service level technology solutions are dominated by the INSPIRE requirements. As a result any service layer must conform to:

1. INSPIRE Technical Guidance for the implementation of Discovery Services[56]
2. INSPIRE Technical Guidance for the implementation of Viewing Services[57]
3. INSPIRE Technical Guidance for the implementation of Download Services[58]

INSPIRE and these guidance documents bring the benefit of both further reducing the allowable technology options and at the same time ensuring that open source standards based implementations of each INSPIRE service type is available. The following sub-sections look at the capabilities in more detail.

6.2.1.1 Discovery services

At a minimum UK-AIR Discovery Services must publish discovery metadata for air quality data services to data.gov.uk in the correct formats. UK-AIR currently hosts the UK INSPIRE Project implementation of the GeoNetwork CSW. This implementation is open source, platform agnostic, supported by the UK INSPIRE Project and conforms to the metadata profile for data.gov.uk (Gemini2.1), the INSPIRE metadata profile and INSPIRE Technical Guidance for the implementation of Discovery Services.

The UK-AIR GeoNetwork instance is registered to data.gov.uk and is harvested routinely. No benefit is foreseen from changing this system at present. Implementation costs, on-going costs and lock-in risks are low.

6.2.1.2 Viewing services

The minimum service level requirement for Viewing Services is for publishing images of spatial air quality datasets as an INSPIRE compliant WMS. UK-AIR currently hosts WMS images served by ESRI's ArcGIS Server technology (version 10.0). The server is licensed to Ricardo-AEA as a third party authorised contractor under Defra's Enterprise Level Agreement (ELA) with ESRI. The technology has been used to host a range of air quality related GIS tools[59], which provide access to data, geoprocessing and download of native GIS datasets for targeted policy support purposes. The server implementation under Defra's ELA is cost effective, efficient and on-going costs to Defra are low. Even so, there known non-conformances of ESRI's ArcGIS Server WMS with INSPIRE guidance that require an additional software to resolve. One option would be to extend licensing via a corresponding third party authorised contractor agreement. Alternatively, an INSPIRE compliant WMS could be achieved via the range of open source WMS[60]. The latter is not the recommended option at it would incur additional data migration and management costs and duplication of data to support the ArcGIS Server for Defra GIS tools and the WMS.

Viewing Services for air quality via an OGC compliant WMS are already in place for publically available spatial air quality datasets on UK-AIR. The conformance of these services with detailed INSPIRE requirements should be evaluated to determine its current and potential future conformance and suitability for Defra's use.

6.2.1.3 Download Services

We have seen that UK-AIR download services will largely fall into services to support human users and services to support machine readability. The former, being focused at desktop human users, will primarily continue to use the CSV download services; the latter will use XML to support machine-readable INSPIRE compliant service users.

The detailed capabilities required of machine-readable INSPIRE compliant services within UK-AIR may be further subdivided into:

i. Services downloading large amounts of historical observational data for post processing by the user

---

[56] http://inspire.jrc.ec.europa.eu/documents/Network_Services/TechnicalGuidance_DiscoveryServices_v3.1.pdf
[57] http://inspire.jrc.ec.europa.eu/documents/Network_Services/TechnicalGuidance_ViewServices_v3.11.pdf
[58] http://inspire.jrc.ec.europa.eu/documents/Network_Services/Technical_Guidance_Download_Services_v3.1.pdf
[59] http://uk-air.defra.gov.uk/data/gis-mapping and http://naei.defra.gov.uk/data/gis-mapping
[60] http://mapserver.org/, http://geoserver.org/, http://www.deegree.org/

ii.   Services for downloading small quantities of near-real-time observational data

iii.   Services for downloading air quality spatial object attribute data for combining with (i) and (ii) by the user

The expectation is that in (i) there will be little or no server side spatial sub-setting of the dataset required i.e. the service will provide relatively static pre-defined downloads specifically aimed at client side post processing. In (ii) the datasets are more dynamic, the refresh rate being hourly and requiring a fast, lightweight and efficient service optimised for observational data, supporting some spatial sub-setting based upon predefined temporal envelopes and aggregates. In (iii) the air quality spatial object attribute data is static, with an annual refresh rate (perhaps bi-annual), download file size is small, typically 5-20 Mbytes. In each case the XML downloads will conform to the INSPIRE conformant Air Quality e-Reporting schema.

The characteristics of these download capabilities further inform the INSPIRE conformant service options applicable. The WFS standard is optimised for download of spatial objects and their attribute data. However, it is not well configured for observational data and, in particular, fast refresh / near-real-time data. As a result, WFS is not a strong candidate for download services (i) or (ii), which are observational datasets. The datasets served by service (i) are by nature, relatively static, and currently configured as a pre-defined download in CSV formats. As a result, the RSS Atom feed is a good candidate for service (i). For dynamic observational datasets such as the near-real-time download service (ii), the INSPIRE Sensor Observation Service (SOS) is strong candidate. This service standard, which is a being modified for conformance with INSPIRE service standards, is optimised for sharing sensor (observational) data and is configurable to interface with existing UK-AIR systems. Finally, the air quality spatial object attribute data service (iii) is well matched to RSS Atom feed as the dataset is relatively static and, is configured as a pre-defined download for regulatory reporting under Air Quality e-Reporting. In addition, the Air Quality e-Reporting application schema used is complex compared with standard simple WFS XML schema would necessitate complex mapping of database content to the application scheme. The burden of development and management of the mapping systems outweighs the benefits of a relatively small and simple pre-defined dataset, which makes WFS less attractive.

A semi-quantitative summary of the relative suitability of the different INSPIRE service options for providing the service characteristics / capabilities required is provided in

Table 7.

Table 7 Summary of the relative suitability of INSPIRE service options against service characteristics

Services required		Service characteristics	WFS	RSS Atom	SOS
(i)	**Historical observational data**	*Static, pre-defined, bulk download (file size)*	✖	✔✔✔	✔✔
(ii)	**Near-real-time observational data**	*Dynamic, fast refresh rate, some spatial sub-setting, pre-definition of download problematic*	✖	✔	✔✔✔
(iii)	**Air quality spatial object attribute data**	*Static, pre-defined, bulk download*	✔	✔✔✔	✖

6.2.2 Interfaces with UK-AIR

UK-AIR has a highly developed and sophisticated data management system hosted on an open source LAMP framework. The system utilises a range of pre-defined SQL statements and algorithms to select database objects, perform aggregations and statistics for download by users. The statements and algorithms have been thoroughly quality assured over the years.

For optimal deployment of INSPIRE services the chosen technologies should connect with existing MySQL database systems with minimal re-engineering. This would ensure that existing database systems, statements and algorithms are reused, thereby promoting deployment efficiency and rigour in the downloaded data. It has been observed that both proprietary and open source INSPIRE conformant WFS implementations do not support connectivity with MySQL at present. It is assumed that the less advanced support for spatial data types in MySQL did not adequately support the spatial features required of a WFS implementation. It is expected that this position will change going forward as support for spatial data types in MySQL improves. Taking this lack of connectivity with MySQL into account, considering that static historical observational datasets will require little spatial sub-setting and taking into account that WFS is not optimised for observational data, WFS is not the recommended option for service types (i), (ii) or (iii).

There are open source SOS implementations available that support connectivity with MySQL. As a result this option is a strong candidate for delivery near-real-time observational data over restricted time frames, (ii) in Table 7 above. The time frame, statistics and aggregates should be restricted, at least initially, to safeguard other UK-AIR service levels.

> ***Recommendation 8***. *Implement an INSPIRE RSS Atom feed for historical data, statistics, air quality spatial object attribute data and other stable spatial datasets e.g. AQMAs because existing Web Feature Service (WFS) options are not optimised for observational data and cannot be connected directly (at present) to the UK-AIR data management systems without costly re-engineering of UK-AIR content.*

> ***Recommendation 9***. *Implement the SOS standard for the download of small quantities of near-real-time observational data with some degree of spatial sub-setting.*

6.2.3 Legacy software and the cost of migration

As noted in section 6.2.2 the cost of migrating UK-AIR database systems, statements and algorithms to alternative relational data base management system (RDMS) in order to serve data via a WFS, is expected to be prohibitively expensive. The marginal benefit of the spatial sub-setting capability brought by the WFS is considered to be outweighed by the cost to migrate and maintain comparable levels of rigour in the data. A WFS also does not deliver an optimised service for observational datasets and the benefits of spatial sub-setting for air quality object attribute data is not clear at present.

> ***Recommendation 10**. Migration to alternative RDMS in order to support a WFS based service is recommended to be avoided in the interest of cost versus the core capabilities required. A watching brief for the development of WFS implementations with MySQL connectivity should be kept to support spatial sub-setting of air quality spatial object attribute and Air Quality e-Reporting via WFS in the future.*

6.2.4 Avoiding technology 'lock-in'

The software technologies currently implemented by UK-AIR are open source in the main and, as such, present a low software lock-in risk. The technology options to support Defra's INSPIRE and open source goals should, wherever possible, be based on open source implementations or open standards, except where proprietary software offers additional benefits. This approach makes the RSS Atom and SOS technology options strong candidates.

As noted in section 6.2.1.2 UK-AIR uses ESRI's ArcGIS Server technology (version 10.0) to deliver WMS and geoprocessing services for air quality datasets. The server is licensed to Ricardo-AEA as a third party authorised contractor under Defra's ELA with ESRI. Defra Resource, Atmosphere & Sustainability team are not subject to the licensing fees for this software. The recommendation is, if possible, for this software to be extended as required for INSPIRE, in order to realise benefits of data management and visualisation tools developed for Defra's geoprocessing and visualisation tools. Although a software lock-in risk, this is perceived as acceptable as the benefits of the visualisation tools requiring ESRI's ArcGIS Server technology are significant. Open source alternatives for the INSPIRE WMS are available, should licencing agreements with ESRI and Defra cease or the required extensions be unavailable, and remain a potential option. Migration costs would be incurred by moving to any alternative technologies.

> ***Recommendation 11**. Implement open source technologies where possible and where proprietary solutions do not present significant financial or technology benefits.*

6.2.5 Costs

The technology costs of the recommended solution include development costs, licensing fees and on-going management costs. These must be kept to a minimum in delivering the required capabilities. This principle has been implemented in each of the recommendations presented in Sections 6.2.1 to 6.2.4 by:

- Reusing or extending existing systems where possible
- Utilising open source technologies and standards where these are available and meet capabilities requirements
- Drawing down existing licensing agreements where these are available, meeting capabilities requirements and add value

6.3 Choosing data formats to support the Defra services

The data formats to be supported by UK-AIR going forward are clear cut from the user story analysis and from an INSPIRE perspective. Traditional CSV downloads must continue for human users to analyse and reuse UK-AIR data on their desktops. INSPIRE and the Open Data initiatives require standards based XML encodings to support machine readability. The Air Quality e-Reporting IPR data model schema has established a European air quality community data model, which is recommended to be used in conjunction with the associated controlled vocabulary for encoding XML formats.

> ***Recommendation 12**. UK-AIR should continue to offer existing downloads in CSV formats. XML encodings should conform to the Air Quality e-Reporting IPR data model schema implementation.*

6.4 Horizon scanning

Defra's air quality data holdings on UK-AIR are rich in gridded (raster or coverage) datasets. Although this type of dataset features highly in many of the data specifications for INSPIRE Annex III data themes, INSPIRE guidance on download service has concentrated on the needs of Annex I data themes. This has been recognised by INSPIRE and, while guidance currently meets the download service requirements for Annex I themes, it may not be so suitable for some of the data in Annex II and III themes. INSPIRE has further indicated that if later data specifications relating to Annex II or Annex III themes should require additional functionality, like those covered by the OGC Web Coverage Service (WCS) or the OGC SOS, the Technical Guidance document will be extended accordingly. Likewise, other implementations (e.g. Linked Data) may be considered in future extensions of this Technical Guidance. Work is already underway to extend the guidance for a SOS implementation. A similar extension (and implementation) for a WCS would be highly beneficial as a download service for the gridded datasets and other coverage data types that UK-AIR host. It is recommended that a watching brief on the extension of INSPIRE Technical Guidance to WCS and an INSPIRE compliant implementation thereof is kept. It is anticipated that extended guidance will support the future of Air Quality e-Reporting of modelled datasets for regulatory compliance and, as a result of the UK's reliance upon modelled data as a method to supplement measurements, it will be an important development area going forward.

> **Recommendation 14**. UK-AIR should keep informed on the potential extension of INSPIRE Technical Guidance to cover WCS and, if appropriate, indicate a user need that would support European Air Quality policy implementation.

> **Recommendation 13**. UK-AIR should as appropriate and subject to availability implement INSPIRE compliant WCS download services to support sharing of gridded datasets and coverages and to support Air Quality e-Reporting

7 Recommendations

This section summarises the recommended software technologies to support the future of Defra's air quality data services including INSPIRE and Open Data ambitions. A proposed architecture is summarised and a timeline of implementation is also set out.

As a general recommendation, in addition to the development of the new services, there remains a need for practical advice to the air quality community on roles, responsibilities and Technical Guidance on how to apply these new services. This should be a secondary aim of any developments to UK-AIR to support the air quality community in working together in a consistent way towards INSPIRE conformance and Open Data.

Further to these aims the scope of services considered in this report have excluded the next generation of LAQM information which is likely to emerge to support user needs and is outside the scope of this study. This will be a separate exercise once this policy has been reviewed with regard to services to support data collection aggregation, and reporting at UK and European levels.

7.1 Recommended software technologies

Based on the analysis of the software technology suitable for delivering the core capability requirements for air quality, the software solutions identified and recommended by this report to support UK-AIR goals are as follows:

Discovery Services

- Continued use of GeoNetwork as a data catalogue registered to data.gov.uk for data discovery
- Continued use of the UK-AIR HTML data catalogue interface for traditional air quality community users

View Services

- Continued use of the ArcGIS Server WMS, (extended appropriately) to deliver WMS for viewing

Download Services

- Implementation of an RSS Atom feed for download of all pre-defined static historical data and statistics service (i)
- Implementation of a SOS for near-real-time data feeds service (ii)
- Implementation of an RSS Atom feed for download of air quality spatial object attribute data service (iii)
- Development of a UK-AIR air quality object registry for spatial and non-spatial air quality objects within the UK Air Quality e-Reporting data model enabling dereferencing of HTTP URIs , in HTML and XML

RSS Atom is an XML syndication format for web feeds. It is a recommended service type for pre-defined data download services under INSPIRE. The Atom syndication format provides a simple, widely understood mechanism for publishing information on the web in the form of feeds, in a way that is compatible with existing web architecture and many tools. The Atom standard specifies an XML-based document format to describe web feeds. These feeds in turn allow software programs to check for updates published on a website and the feed can then be downloaded by programs. The INSPIRE recommended feed redirects or points the consumer to pre-defined downloads which, in the case of UK-AIR, will be a package of air quality data in the AQ e-Reporting XML format.

The SOS is an OGC Web Service interface specification that provides the necessary functionality to share such sensor observation data e.g. air quality measurements, predictions or forecasts. The SOS standard has been proposed as the basis for extending the INSPIRE Technical Guidelines for Download Services for observation data. In its current implementation, there are a number of aspects that need to be addressed to make the SOS fully INSPIRE compliant: INSPIRE requires additional functionality, which goes beyond the SOS standard. This includes, for example, the support of natural language selection or coordinate reference system (CRS) transformation. These are currently being addressed by 52°North and the European Commission's Joint Research Centre in order to propose an update of the Technical Guidelines for INSPIRE Download Services including the OGC SOS in parallel with the enhancement of the 52° North 4.x SOS implementation.

A number of WMS and visualisation tools are already operating on UK-AIR. As for the SOS standard, some of these have minor non-conformances with INSPIRE specifications. It is the expectation that these will migrate to compliant services for INSPIRE compliance for viewing services.

The air quality object registry will provide a web service, allowing dereferencing of the HTTP URI pattern used for identifying all UK air quality objects e.g. networks, stations, sampling points, management zones, models etc. The air quality objects are already fully described in the UK submissions to the Commission compiled annually. Upon dereferencing of an UK air quality HTTP URI, the registry will render an HTML or XML file for download containing attribute data of the object. The HTML file will provide support for human users to read the attribute data and the XML file will provide support for machine readability.

Table 8Table 8 summarises the software upgrades recommended to deliver support Defra INSPIRE and Open Data goals.

Table 8   Software upgrades recommended to deliver support Defra INSPIRE and Open Data goals

|	Requirement	Proposed solution
Services	INSPIRE conformant discover service	• GeoNetwork / data.gov.uk
	INSPIRE conformant viewing service	• ArcGIS Server WMS via INSPIRE GeoPortal extension
	INSPIRE conformant download services	• Predefined downloads via RSS Atom feed for historical data, statistics and air quality spatial object attribute data. • Near-real-time data via 52° North 4.x SOS implementation • Development of a UK-AIR air quality object registry in HTML and XML formats
Formats	Human & machine readability	• Air Quality e-Reporting XML schema to support machine readability • CSV formats to support human users linked via HTTP URIs to air quality object registry
Accessibility	Improved traceability	• CSV download service option • XML download service option • Via a HTTP URI object identifiers system • Air quality object registry in HTML & XML formats

7.2 Proposed service architecture

An overview of the architecture of the proposed services to support Defra air quality is presented in this section. The architecture has been grouped into four key service levels: Discovery Services,

Download Services for observational data, Download Services for modelled data / Viewing Services for spatial objects and the UK-AIR air quality object registry. Schematics describing the architecture and the actions required to invoke the services are presented below.
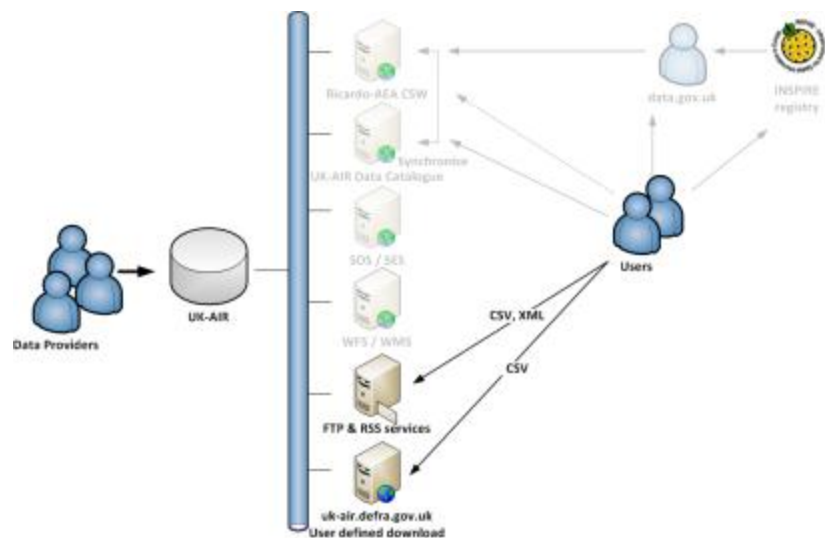
Discovery Services

- Data providers compile metadata for datasets they are responsible for
- UK-AIR adds URLs to data services on UK-AIR
- Metadata loaded to UK-AIR CSW
- UK-AIR CSW synchronised with data.gov.uk
- Users discover datasets, view and download services from UK-AIR, data.gov.uk or INSPIRE registry
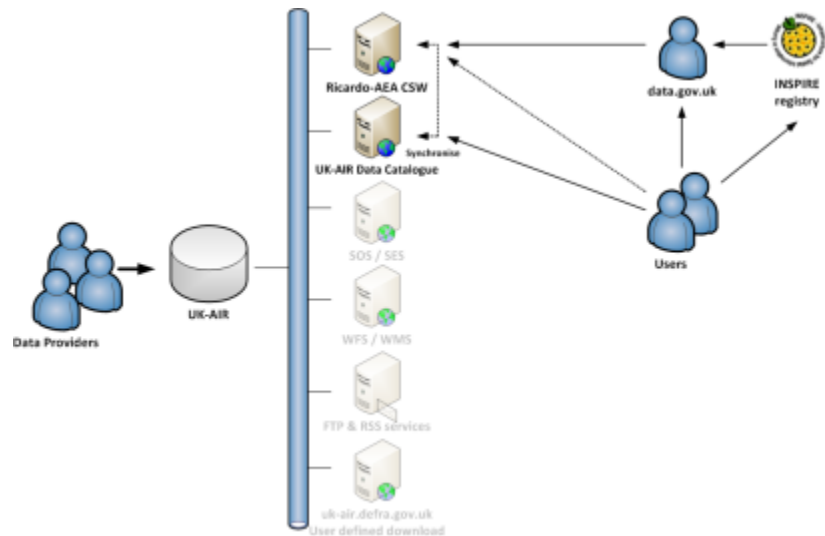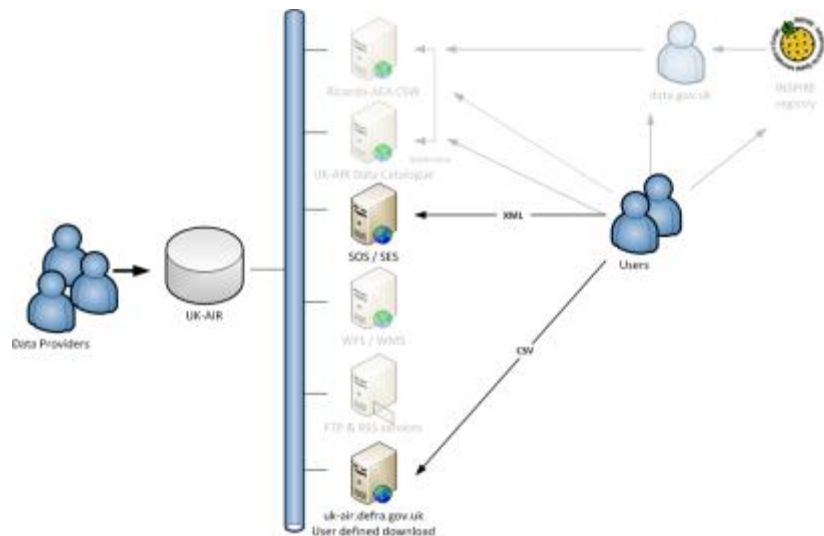

Download Services (measurements)

- Users discover datasets and download services from UK-AIR, data.gov.uk or INSPIRE registry
- Users redirected to download services
- Users access downloads in CSV via the UK-AIR data selector
- Users access download in XML UK-AIR SOS & RSS Atom feeds

Viewing Services
(stations location & models)

- Users discover datasets and download services from UK-AIR, data.gov.uk or INSPIRE registry
- Users redirected to view services for spatial datasets via WMS

Air quality object registry

- Users discover datasets and services from UK-AIR, data.gov.uk or INSPIRE registry
- Users redirected to download services via RSS Atom
- Download content contains object identifiers in the form of HTTP URIs
- Air quality object registry enables HTTP URIs to be resolved rendering (downloading) attribute data in HTML or XML e.g. the attributes of an air quality zone or monitoring stations
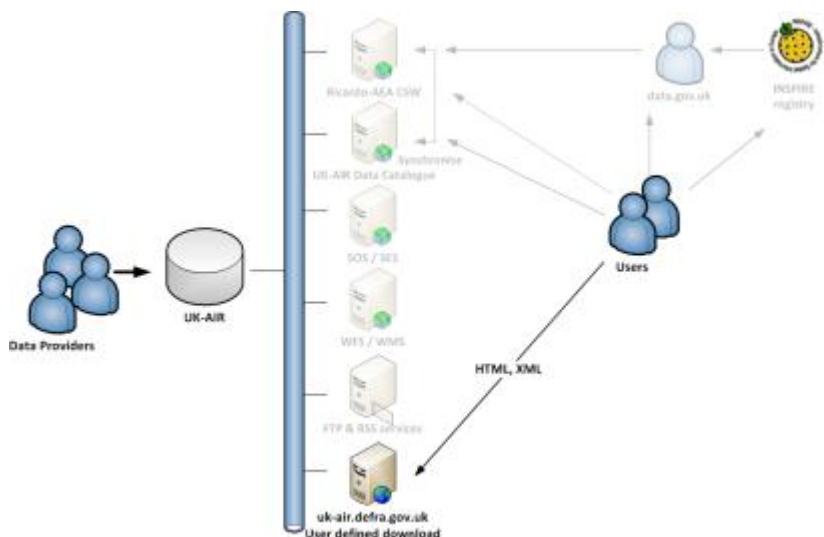
The UK-AIR site map is presented in Appendix 1 and at uk-air.defra.gov.uk/site-map. These pages provide a useful summary of the overall capability hosted by UK-AIR. It can also be further refined to summarise the services on UK-AIR delivering data to end users that may be supported by viewing and download services and the future services that are recommended to be deployed to each going forward. In this way it can be ensured that services are assigned to relevant data themes and that all INSPIRE obligations are captured.

Table 9 summarises the existing UK-AIR data discovery, download and view themes and assigns services applicable to them in addition to the HTML pages themselves.

Table 9 Summary of UK-AIR data themes recommended service types applicable

| UK-AIR Data theme	Services applicable
Interactive monitoring networks map	WMS
Latest measurement summary	RSS Atom
Latest measured levels	RSS Atom
24 Hour Summary	RSS Atom
Pollution alerts	SOS, RSS Atom
Data Selector	XML, Linked CSV
Annual and Exceedence Statistics	XML, Linked CSV
UK Particle Monitoring Programme	XML, Linked CSV
Marylebone Road, London Non-Automatic Data	XML, Linked CSV
Polycyclic Aromatic Hydrocarbons (PAH) data	XML, Linked CSV
Modelled air quality data	GML/XML, WMS/ WCS
Toxic Organic Micro Pollutants (TOMPs) data	XML, Linked CSV
Lead, Trace Elements and Industrial Metals data	XML, Linked CSV
UK Air Quality Data Catalogue	CSW
UK Ambient Air Quality Interactive Map	WMS, WCS
Site information search	XML, Linked CSV, AQ object registry (HTML/XML)
Search for monitoring sites	XML, Linked CSV, AQ object registry (HTML/XML)

7.3 Roadmap timeline

INSPIRE has set a series of phased deadlines for the implementation of services conformant with its Technical Guidance and data specifications. These have been set out in section 3.2.2 of this report and are summarised below. These deadlines are the main regulatory driver influencing when improved systems and services should be in place. There is flexibility in the dates for the implementation of INSPIRE compliant downloads, with restructured datasets being the recipient of an early implementation date, recognising that restructured / reorganised datasets can embark mapping of legacy datasets to the INSPIRE data specifications as part of the reorganisation. At a European level, the air quality e-Reporting dataset is viewed by the Commission as a restructured dataset, and therefore the expectation is that INSPIRE compliant downloads shall be in place by the end 2015. Although this is highly achievable for the air quality datasets covered by the Air Quality IPR e-Reporting data model, it should be recognised that not all of UK air quality data holdings fall within the scope of air quality e-Reporting and therefore for these data types e.g. LAQM data, AQMAs, smoke control datasets, air quality forecast information and emissions datasets, the December 2020 deadline may be more applicable. A simplified timeline for a phased implementation of INSPIRE compliant datasets and services hosted by UK-AIR is presented in Figure 6.

For datasets covered by the Air Quality IPR e-Reporting data model, it is recommended that Defra commission the services outlined over the period March 2014 to December 2015.This will include research monitoring datasets covered by former "Exchange of Information" Decision. It is recommended that AQMA data are also covered off alongside e-Reporting data as a quick win.

For emissions related datasets, Defra, and where appropriate DECC, should look to engage with similar e-Reporting exercises for emissions datasets to align initiatives and to conduct transformation of the remainder of UK datasets by the end of 2020.

Figure 6 Simplified timeline for fully compliant INSPIRE services
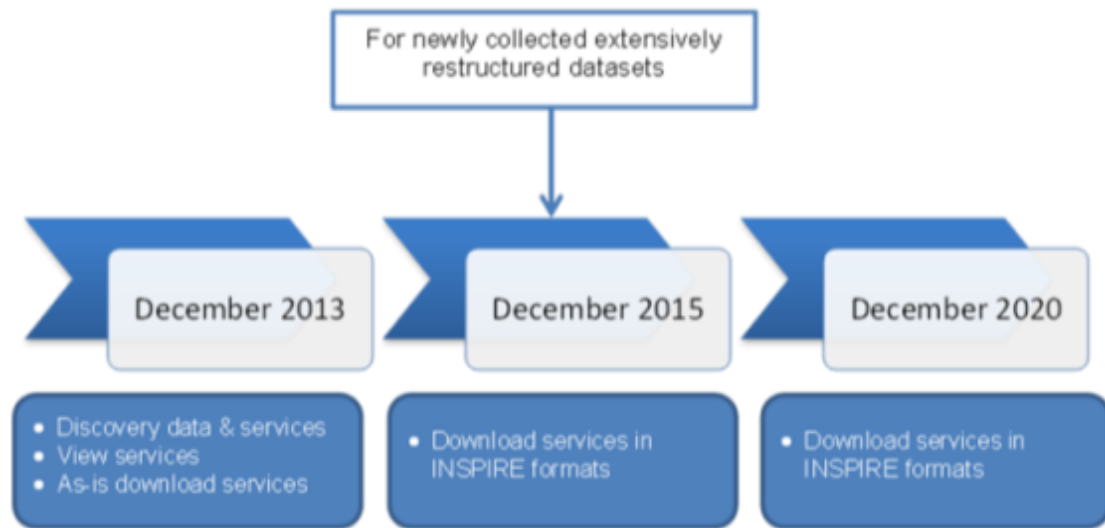
Appendices

Appendix 1   The UK-AIR website map http://uk-air.defra.gov.uk/site-map

Appendix 2   Open Data scoring metrics

Appendix 3   Agenda and minutes of the Defra web service web stakeholder consultation workshop

Appendix 4   Defra data services stakeholder consultation questionnaire

Appendix 5   Defra data services stakeholder user stories

Appendix 1   The UK-AIR website map
http://uk-air.defra.gov.uk/site-map

- ## Site map

These air quality pages contain a large amount of information and data, so this site map page can help you quickly find the area of the website you require.

- **Home**
  - Interactive monitoring networks map
  - Latest measurement summary
    - Latest measured levels
    - 24 Hour Summary
    - European Pollution Levels
    - View with Google Earth
    - Pollution alerts
  - Pollution forecast
    - Forecast Summary
    - What do the forecasts mean?
    - How are the forecasts produced?
    - WRF Forecast Summary
    - Embed UK-AIR forecasts in your website (beta)
  - UK-AIR news
  - About UK-AIR
    - UK-AIR on Twitter
    - Site map
  - Subscribe to mailing list and bulletins
    - Mailing list
    - Email Bulletins
  - Related links
    - Pages from the UK
    - UK Local Authority web pages
    - Pages from Europe
    - Pages from the USA
    - Pages from around the world
    - Meteorological organisations
    - Environment organisations
    - Campaign groups
    - Health
    - Submit a link

- **Data Archive**
  - Data Selector
  - Annual and Exceedence Statistics
  - Data availability
  - UK Particle Monitoring Programme
  - Marylebone Road, London - Non-Automatic Data
  - Polycyclic Aromatic Hydrocarbons (PAH) data
  - Important Information: Comparing data from different monitoring sites
  - OpenAir - data analysis tool
    - OpenAir data definitions
    - Summary Plot
    - Smooth Trend
    - Time Variation
    - TheilSen
    - Time Plot
    - Trend Level
    - Polar Plot
    - Wind/Pollution Rose
    - Polar Annulus
    - Scatter Plot
    - Polar Frequency
    - Calendar Plot
  - Modelled air quality data
  - Toxic Organic Micro Pollutants (TOMPs) data
  - Lead, Trace Elements and Industrial Metals data

- o   UK Air Quality Data Catalogue
- o   UK Ambient Air Quality Interactive Map
- ▪   Interactive Map FAQs
- o   DAQI regional data

- •   **About Air Pollution**
- o   UK and EU Air Quality Policy Context
- o   Daily Air Quality Index
- ▪   What is the Daily Air Quality Index?
- o   Glossary
- o   UK and EU Air Quality Limits
- o   Effects of air pollution
- ▪   Short-term effects of air pollution on health
- o   Causes of air pollution
- o   Frequently Asked Questions
- ▪   General FAQs
- ▪   Data FAQs
- ▪   Health FAQs

- •   **Monitoring Networks**
- o   Site information search
- o   Search for monitoring sites
- o   Brief history
- o   Site environment types
- o   Automatic Networks
- ▪   Automatic Urban and Rural Network (AURN)
- ▪   Automatic Hydrocarbon Network
- ▪   Automatic London Network
- o   Non-Automatic Networks
- ▪   United Kingdom Eutrophying & Acidifying Network (UKEAP)
- ▪   Acid Waters Monitoring Network
- ▪   Urban Heavy Metals Network
- ▪   Rural Heavy Metals Network
- ▪   Nitrogen Dioxide Diffusion Tube (1993 to 2005)
- ▪   Smoke and Sulphur Dioxide Monitoring Network, 1961-2005
- ▪   Black Carbon Network
- ▪   Polycyclic Aromatic Hydrocarbons (PAH)
- ▪   Toxic Organic Micro Pollutants (TOMPs) Networks
- ▪   Non-Automatic Hydrocarbon Network
- ▪   Particle Numbers and Concentrations Network
- o   Air Quality Monitoring Methods
- ▪   EU Standard Methods for monitoring and UK Approach
- ▪   Demonstration of Equivalence of PM Monitoring Methods
- ▪   Certification - MCERTS for UK Particulate Matter

- •   **Air Quality Library**
- o   Search Library
- o   Air Pollution in the UK report
- o   Acid Rain
- o   Air Pollution Forecasting
- o   Air Quality Communications
- o   Atmospheric Emissions
- o   Economic Analysis
- o   Effects of Air Pollution on Natural Ecosystems
- o   Local Authority Guidance
- o   Monitoring Networks
- o   Review Groups and Expert Groups
- o   Strategic Policy Analysis
- o   Urban Air Quality
- o   Transport
- o   Air Quality Strategy 2007 - supporting documents
- o   Air Quality Modelling
- o   Air Quality Plans for the achievement of EU air quality limit values for nitrogen dioxide (NO_2) in the UK
- o   Air Quality Expert Group
- ▪   About the Air Quality Expert Group
- ▪   Publications

- •   **Science & Research**
- o   Air Quality Modelling
- ▪   Air modelling for Defra

- ▪ Air Quality Modelling Review
- ▪ Defra's Model Intercomparison Exercise
- ▪ CMAQ-UK
- ○ Research Monitoring
- ○ Ecosystem Research

Back to top

- ▪ Air Quality Modelling Review
- ▪ Defra's Model Intercomparison Exercise
- ▪ CMAQ-UK

Appendix 2   Open Data scoring metrics

Sir Tim Berners Lee's 5-star scoring system

Level	Format		Licence
★	Make your data available on the web (in any format)	**and**	Open Licence
★★	Make your data available on the web as structured data (for example Microsoft Excel instead of an image scan of a table)	**and**	Open Licence
★★★	Make your data available on the web in an open non-proprietary format (for example CSV or XML instead of Microsoft Excel)	**and**	Open Licence
★★★★	In addition to open formats use URLs to identify things based on open standards and recommendations from the W3C, so that other people can point at your stuff	**and**	Open Licence
★★★★★	In addition to open formats and URLs to identify things, link your data to other data to provide context	**and**	Open Licence

Open Data Certificates

https://certificates.theodi.org/about

Administered by the Open Data institute (ODI), Open Data Certificates provide a seal of approval that helps deliver knowledge for everyone. An Open Data certificate gives data users a head start. It tells them what the data is about and how to get hold of it. It shares information like availability, privacy, and licensing so they can decide how much to rely on it.

Appendix 3   Agenda and minutes of the Defra web service web stakeholder consultation workshop

Workshop outcomes

Subject: Designing the future of Defra Air Quality Data Services

Date: 7th October 2013

Location: Department of Business, Innovation and Skills, Innovation Space

Time: 09:30-16:00

Capturing User Needs

Evidence from a range of sources was gathered to build a picture of user requirements for Defra AQ data in accordance with Government Digital Service (GDS) guidance. The GDS guidance indicates that any thinking about a service, whether online or offline, must start with the question: What is the user need? Evidence that supports the existence of the user need(s) is required and for this study the following information themes have been utilised or are planned for inclusion;

- Web analytics for existing content to show that content that already exists is being regularly accessed.

- Search terms via web logs to show that people are expressing a need for content or functionality by searching for it on the website.

- Customer facing staff and helplines regularly engaging with users and are able to identify what the most pressing and real user needs are.

- User research to understand what the user needs are.

- User stories and user-led development.

For the purpose of this workshop we have endeavoured to capture user needs by;

- correspondence with policy officials;

- email requests to UKAIR from UK and international users;

- previous customer insight work;

- web logs showing web site activity and access routes; and

- a specially designed pre-workshop questionnaire for all consultees and 24 workshop attendees.

The questionnaire user stories were reviewed and condensed into 5 core cases which were presented back to attendees. These are shown in Annex 1. Workshop attendees are presented in Annex 2.

Defra proposals and policy drivers

The policy drivers for new ways to exchange and publish data were presented including the Air Quality Directive[61], the Air Quality e-Reporting initiative[62], INSPIRE Directive[63] and Cabinet Office transparency and Open Data policy[64]. The role of data.gov.uk in the UK's INSPIRE response and the relevance of INSPIRE to service providers with local authority clients was discussed in detail.

The Open Government Licence (OGL)[65] was introduced and data providers should now ensure they use the attribution statement where they are using Defra air quality and emissions data[66]. The regulatory and policy drivers discussed, together with the evidence on user needs formed the basis of proposals for service orientated architecture.

Proposals for Defra's services for discovery, view and download were presented to attendees and are included in a summary diagram in Annex 3. Updates on Air Quality e-Reporting and INSPIRE preparation were also given as background information. Discussions focused on the discovery and download services as these are most relevant to users and providers. In summary it was explained that Defra will have compliant data discovery and view services in place before end 2013. Download services already exist in the form of user defined queries. Web-serviced based implementations are planned and under evaluation.

Attendees confirmed they were happy with their summarised user stories and were given the opportunity to question and comment on how the service proposals met their needs. No objections or significant issues were identified by consultees. Two additional edge cases user stories were identified and will be added to the evidence base to support hackers/the non-expert tech community, and emergency planning/civil contingencies.

Data provider community responses

The benefits of ensuring ease of access to data for users and compatibility/interoperability of datasets were supported by attendees. The relevance of the proposals for data providers and for example, to the LAQN[67], Cleaner Air for London[68] and PDP[69] websites was acknowledged. The proposed services would in future be available for use by these and any other data provider.

Data providers presented a large range of interesting developments in their services, with many common themes and requests. CEH, Met Office and BADC have been most active in preparing for INSPIRE. Other service providers working on behalf of local authorities may also become involved to present a practical implementation solution for local authorities.

There was interest in and support for of a future flow down of the EU air quality e-Reporting vocabulary and data models. Links to these are provided in Annex 4. It is likely that using these in practice will present many technical queries and challenges as they are adapted and extended for different user needs and the UK domain.

---

[61] http://ec.europa.eu/environment/air/quality/legislation/existing_leg.htm
[62] http://www.eionet.europa.eu/aqportal
[63] http://inspire.jrc.ec.europa.eu/

[64] https://www.gov.uk/government/policies/improving-the-transparency-and-accountability-of-government-and-its-services

[65] http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/

[66] http://uk-air.defra.gov.uk/about-these-pages

[67] http://www.londonair.org.uk/LondonAir/Default.aspx

[68] http://cleanairinlondon.org/

[69] http://pollutantdeposition.defra.gov.uk/

Web based services and APIs have already been developed by some providers. The extent to which these may need redefining to accommodate the EU air quality e-Reporting vocabulary and standards based methods is not clear but likely to need addressing to achieve the overall aims on a community level. Risks and issues were discussed and included a growing use of AQ data by non-experts with scope for misinterpretation and misleading results. Technical issues included demand and pressure on servers from web service access (particularly at high time resolutions), data management overheads related to serving spatial datasets, the verbosity of the XML data model and the need for fast lightweight data feeds, up-skilling of user communities to maximise use of new XML technologies, as well as ensuring data is correctly licenced and clearly attributable to a source.

Actions and points for further discussion

Discovery services

1. **Data providers are advised to contact their local authority clients to ascertain their approach to INSPIRE 2013 deadlines**, highlighting the importance of registering metadata with the UK INSPIRE node data.gov.uk before the end of the year. Publication of discovery metadata on data.gov.uk before the end of 2013 will ensure that datasets are not logged as a new or newly restructured and consequently more lenient INSPIRE obligations for fully compliant download services by 2020 (as opposed to 2015) will apply[70]. Conversations between local authority air quality officers and GIS leads will ensure that approaches to INSPIRE at local authority level are aligned where needed. Defra also plan to send a communication out to alert air quality officers.

2. The UK-AIR catalogue contains records for many non-Defra datasets, as listed in Annex 3. **Data providers were asked to confirm if they wish for their metadata records to be published to** data.gov.uk **directly from UK-AIR as part of a Defra response to INSPIRE for air quality**. **Updates required to metadata records should also be communicated to Defra.** Additional entries to the catalogue can be made, though Defra needs to consider how costs will be covered for ongoing metadata management where these are non-Defra datasets. **Further information is needed from the Devolved Administrations to ensure a joined up UK approach is taken for INSPIRE.**

3. Other data providers also have catalogues including CEH, BADC, NERC, UKEOF and Met Office. **Linking these electronically for automatic harvesting where records relate to the same dataset will be important**. **Catalogue content will need to be compliant with the UKLP implementation of the Gemini2 metadata profile in order for them to be accepted by** data.gov.uk. Compliance with the Gemini2 metadata profile will ensure INSPIRE compliance although reconciliation of local catalogues with the Gemini2 metadata standard may be needed - further discussion is needed here.

Download services

1. **Further comments and questions on the services are welcome by COB 15 November 2013.**

2. **Defra will provide advice to the GLA on recommended wording for a general contract for London** local authorities to use for future data management to overcome current issues on data interoperability, standards data communication and to support plug-and-play data connectivity.

---

[70] Register as public sector data provider here http://data.gov.uk/publisher?page=H

3. GLA also confirmed that commercial barriers to Open Data were not real barriers from the point of view of the procurer.

4. There was general encouragement for using lightweight XML format (JSON etc.) where possible.

General INSPIRE

1. There was some concern expressed on understanding in detail the scope of data providers obligations to INSPIRE both in terms of local authorities and 3rd party data providers/commercial organisations supporting local authorities. It was recognised that this was a common theme. Clarification was made that any spatial datasets generated by a public task or activity of a public authority and falling broadly in scope of the 34 INSPIRE environmental data themes would be captured by an INSPIRE obligation, even when provided by a 3rd party. As a result, encouragement for 3rd party data providers/commercial organisations supporting local authorities to contact their local authority clients to ascertain their approach to INSPIRE was reiterated.

2. The LGA confirmed that there would be financial support for local authorities in England in order for them to respond appropriately to their INSPIRE obligations. However mechanisms to distribute funding were delayed and negotiations on how funding would be distributed were on-going. The LGA also pointed to extensive guidance for local authorities see Annex 5 and encouraged local authority data managers and GIS professionals to understand what they are doing collectively in relation to INSPIRE. LGA confirmed that INSPIRE only applied to data held in digital formats as of 2013 and no re-engineer of historical datasets was required.

3. The opportunity for Defra to publish AQMA records and link to data.gov.uk and the GeoPortal was discussed. Support was given by the local authority present for such an approach which would reduce burdens on local authorities and be the most cost effective approach.

4. The importance and value of improved observation metadata (i.e. station information and data flags/quality) was discussed and Defra will be looking at how to improve availability of this information which has largely been gathered for air quality e-Reporting and regulatory purposes but is useful for aiding understanding and re-analysis. It was observed that this will be an important issue in the future as air quality data reaches wider less expert user groups.

5. The vocabulary and data dictionary developed at EU was discussed and there was general support in adopting this for the UK. **Data providers are requested to have a look at the data model and provide comments to Defra / Ricardo-AEA on what additions may be needed for a UK applications and why these would be needed**. Observations and comments on the data model, schema and controlled vocabulary supporting it may be made via http://projects.eionet.europa.eu/implementation-2011-850-eu/library. A notable addition to the model was suggested to support EMEP data quality flags which are not currently supported explicitly.

6. It was explained that the UK air quality approach to INSPIRE has already tackled data models, schema and data model data specifications for air quality as a result of the air quality e-Reporting initiative. These are not needed for Annex III until 2015 for new datasets and 2020 for existing datasets for INSPIRE compliance. **However it is recommended that data providers could start working with these, simplified or tailored to their needs and the UK applications in order to implement a consistent standards based approach to data sharing through discovery services. Comments on these are invited by data providers**.

7. Separate schemas for environmental monitoring facilities have been developed under UKEOF. The UKEOF schema is also INSPIRE EF compliant, so whilst there would be no issues with using UKEOF over e-Reporting for UK applications to describe observations, this would only provide part of the solution for encoding the air quality data particularly within a regulatory framework which falls outside of the scope of INSPIRE. **Further consideration needed by data providers and Defra.**

Annex 1: Summary of user stories

Common Actors, Narratives & Goals
Designing the future of Defra AQ data services workshop

RICARDO-AEA

Actors

- I am responsible for implementing INSPIRE for air quality

- I am responsible for coordinating response to domestic and regional policy

- I am a consultant delivering client needs with AQ data

- I am a data provider
  - Network owner (funding authority)
  - Network operator
  - Data publisher / manager

- I am a researcher publishing scientific papers

Narratives

- I need services to respond to my legal / policy obligations

- I need services to provide me with access to central and local data sources

- I need services to deliver data in time resolutions appropriate to my client needs

- I need services that allow me to check the consistency of the data that I have provided to UK-AIR systems

- I need services to deliver data in time resolutions that support my research needs for modelling and monitoring.

Goals

- Delivery of services that meet the legal/policy obligations and supporting opportunities for businesses/service providers

- Delivery of high quality and timely information and advice on AQ to the public

- Delivery according to the needs of my clients

- To ensure that AQ information in UK-AIR is accurate and of high quality

- Delivery of quality scientific research to advance understanding, policy & improve the environment

Ricardo-AEA in Confidence   © Ricardo-AEA Ltd

Annex 2: A list of workshop attendees

Emily Connolly	Defra
Dan Waterman	Defra
Sarah Honour	Defra
Jason King	Defra - UKLP
Gesche Schmid	Local Government Association
Simon Baldwin	Welsh Government
Chris Burns	DOENI
Tony Bush	Ricardo-AEA
Rachel Yardley	Ricardo-AEA
Dan Brookes	Ricardo-AEA
Andrew Grieve	King's College London
Elliot Treharne	GLA
Richard Maggs	Bureau Veritas
David Butterfield	NPL
Jayn Newman	LB Hillingdon
David Carruthers	CERC
Mark Jackson	CERC
Mike Brown	CEH
David Leaver	CEH
Geoff Broughton	AQDM
James Penman	Met Office
Ag Stephens	BADC

Annex 3: Proposed architecture for Defra's web-services for discovery, view and download
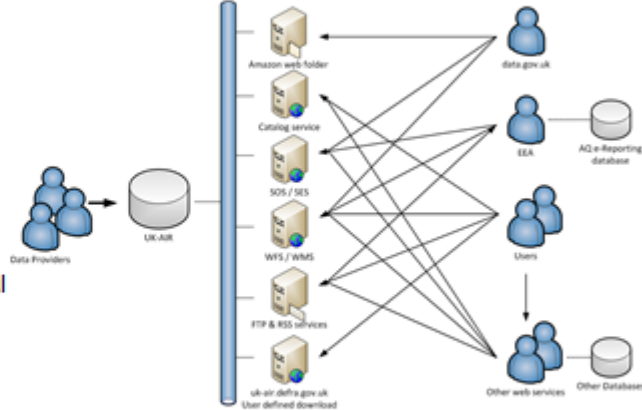
Annex 4: A list of non-Defra datasets in the UK-AIR catalogue

Air Quality in Wales, Website of the Welsh Air Quality Forum - automatic and non-automatic air quality monitoring

Air Quality in Scotland - automatic and non-automatic air quality monitoring

Herts and Beds Air Quality Network

Norfolk Air Quality Network

Greater Manchester Districts Met Stations Annual statistics, hourly data

Greater Manchester Districts Met Stations (implex)

Northern Ireland Air - automatic air quality monitoring

Air Quality Monitoring in London and South East England

Kent and Medway Air Quality Monitoring Network

Devolved Administration Activity Data and emissions by source

Devolved Administration emissions by end user category

England Nitrogen Dioxide Diffusion Tube Data Centre

Annex 5: Useful links

- The EU Air Quality e-Reporting
  http://www.eionet.europa.eu/aqportal

- EU Air Quality e-Reporting data dictionary
  http://www.eionet.europa.eu/aqportal/codelists

- EU Air Quality e-Reporting schema
  http://dd.eionet.europa.eu/schema/id2011850eu/AirQualityReporting.xsd/view

- EU Air Quality e-Reporting schema & guidance documents
  http://www.eionet.europa.eu/aqportal/toolbox

- LGA guidance on inspire
  http://www.local.gov.uk/local-government-intelligence/-/journal_content/56/10180/2833786/ARTICLE

- INSPIRE Guidance
  http://inspire.jrc.ec.europa.eu/index.cfm

- data.gov.uk
  http://data.gov.uk/

- UK-AIR data catalogue
  http://uk-air.defra.gov.uk/data/data-catalogue

Appendix 4   Defra data services stakeholder consultation questionnaire

Department for Environment, Food and Rural Affairs

Defra Air Quality Data Services

Project Background and Questionnaire

September 2013

Contents

•   **Background to the Project**

In 2010 Defra commissioned two complementary studies to evaluate the future of Defra Air Quality data and information services:

- A scoping study on Defra's Air Quality Data Management and Integration Systems[71] investigated the feasibility of a data integration process to optimise access to air quality data and other supporting datasets. The study sought to identify the datasets and tools to support air quality users at all levels and aimed to align these with emergent policy initiatives relating to the SEIS[72], INSPIRE[73] and Transparency agenda[74].

- A review of Defra's arrangements on the provision of air quality information to the public[75]. This review focused largely on the Air Quality Archive but also extended to a basic assessment of Defra's approach to air quality communications in general.

Since the studies reported the air quality data landscape has continued to evolve. This has largely been driven by regulatory obligations to the Ambient Air Quality Directive 2008/50/EC[76] and the INSPIRE Directive 2007/2/EC, although the Cabinet Office's Transparency and Open Data agenda has become increasingly important.

In response to these drivers and increased demand for machine accessible and readable data formats hosted on the web, Defra has commissioned a new study to inform on how its future web data services should be implemented in a prioritised way. The objectives of the new study are to:

- Identify user requirements.

- Identify policy framework requirements including European Air Quality e-Reporting[77], INSPIRE and the Government Digital Service[78].

- Recommend system architecture for Defra's systems.

- Recommend consistent and interoperable approaches for publishing air quality data.

As part of this work, users of existing UK-AIR data services will be consulted to identify their current and future requirements. Air quality data providers will be consulted to understand their approaches and plans for publishing data, in particular, other public sector organisations with obligations under INSPIRE. Defra's experience of developing

---

[71] http://uk-air.defra.gov.uk/reports/cat09/1102161123_Data_Integration_Report_v1-2.pdf
[72] http://ec.europa.eu/environment/seis/index.htm
[73] http://inspire.jrc.ec.europa.eu/
[74] https://www.gov.uk/government/policies/improving-the-transparency-and-accountability-of-government-and-its-services
[75] http://uk-air.defra.gov.uk/reports/cat14/1112190909_100614_Comms_Contract_Review_Report_FINAL1.pdf
[76] http://ec.europa.eu/environment/air/quality/legislation/directive.htm
[77] http://www.eionet.europa.eu/aqportal
[78] http://digital.cabinetoffice.gov.uk/about/

INSPIRE conformant data models and vocabulary for air quality will be shared with the air quality community to support others who have similar obligations.

This new study will therefore draw out the changes which need to be implemented to allow Defra's air quality data to be published to meet the requirements of the regulatory drivers as well as the needs of current and future users.

This document provides a questionnaire for data users and providers to complete as part of the consultation process.

• **About this Questionnaire**

As a user or provider of air quality data, you are invited through this questionnaire and our workshop on 7th October 2013 to help shape the future of Defra's air quality data services. The questionnaire is divided into three sections, the first "About You", the second for Data Users and the third for Data Providers. Please complete which ever sections are relevant to you and send the whole document back to us.

Changing policy requirements, user needs and technologies have created a need for Defra to consider development of web services to allow standardised and open access to Defra air quality data. These services would complement Defra's existing data access channels such as email bulletins, Twitter and user defined downloads as detailed in the Guide to UK Air Pollution Information Resources[79].

Ensuring Defra captures and considers user stories is a central part of the service development process. Defra needs to implement services which are optimised for the core and common user needs but also understand the "edge cases". The Government Service Design Principles provide a clear framework for user-led development[80].

In this questionnaire we want to capture your needs for Defra air quality data and your approaches as data providers to publishing data, before further discussion at the workshop. With many recent developments in Open Data and with rapidly approaching deadlines under the INSPIRE directive for many public sector data managers, this a timely opportunity for the UK air quality community to review its progress and work together to improve the consistency, usability and accessibility of the air quality information marketplace.

Please return your questionnaire by COP 2nd October 2013 to
Emily.connolly@defra.gsi.gov.uk and tony.bush@ricardo-aea.com

---

[79] http://uk-air.defra.gov.uk/reports/cat14/1307241318_Guide_to_UK_Air_Pollution_Information_Resources.pdf
[80] https://www.gov.uk/service-manual/user-centered-design/user-needs.html

1.   <u>About You</u>

- **Name:**

- **Company/Organisation and Job Title:**

- **Air Quality Involvement:**

  ☐ Data provider/publisher   ☐ Data user

  Please supply details:

- **Telephone number:**

- **Email Address:**

1.   <u>About You</u>

2. Air Quality Data Users

Please complete this section if you use or require access to Defra air quality data.

1. Please fill in this use case story table to provide an overview of who you are, why you need access to Defra air quality data and what you seek to do with it.

About You:	As a	*Complete as appropriate: (researcher/consultant/member of the public/policy maker/etc.)*
Narrative:	I want to	*(describe in here your need(s) for Defra air quality data)*
Goal:	So that	*(describe your goal(s) through using Defra air quality data)*

2. What Defra air quality data do you use to inform research, analysis and decision making (e.g. automatic or non-automatic monitoring data for particular pollutants, emissions inventories, maps and forecasts, etc.)?

3. How do you currently access these data from Defra? (User defined download, website, RSS feed, email bulletin etc.)

4. If you currently use or would like to use and publish Defra data in your own website, tool or application please provide details of your use requirements. Please include details of the data you use or would seek to use.

5. Please indicate whether you agree or disagree with the following statements. Please make additional comments if you wish**:**

   a. I have access to all the Defra **air quality** data and metadata I need

   ☐ Agree   ☐ No opinion   ☐ Disagree

   Comments: If disagree: what Defra air quality datasets or metadata would you like access to but don't have?

   b. The access arrangements for the Defra air quality data and metadata meet my needs.

   ☐ Agree   ☐ No opinion   ☐ Disagree

Comments: if agree, what do you like about the current access arrangements? If disagree, what access arrangements, formats, units or other changes would you request?

6. INSPIRE and the Open Data initiative are driving the implementation of platform independent, machine readable formats for publishing data via the web. The formats will be based upon XML. To what extent do you feel these formats can replace traditional CSV download formats?

☐ I still want CSV   ☐ I'd prefer XML   ☐ I'd prefer both   ☐ No opinion

7. UK-AIR now has a new data catalogue[81] allowing users to search for Defra and non-Defra AQ data via a keyword search. These metadata will also be registered with data.gov.uk and from there the INSPIRE GeopPortal[82]. Eventually this metadata will enable users to discover historical and up-to-date resources and services for viewing and downloading data. To what extent would this solve your data access issues?

☐ Fully meets my needs   ☐ Better, but not quite there   ☐ Not at all

If you have additional requirements please describe them below

8. Which of the following data formats are you likely to make use of now and in the future?

CSV format	☐ Likely	☐ Not likely	Any Comments:
XML format via a web service call e.g. a Sensor Observation Service (SOS) or RSS, or other APIs	☐ Likely	☐ Not likely	
Web Map Service (WMS) format georeferenced map images	☐ Likely	☐ Not likely	
Web Feature Service (WFS) format geographical data	☐ Likely	☐ Not likely	
GIS (shapefile) format geographical data	☐ Likely	☐ Not likely	

---

[81] http://uk-air.defra.gov.uk/data/data-catalogue
[82] http://inspire-geoportal.ec.europa.eu/discovery/

9. I would like to be able to automate access to UK-AIR data and obtain data in machine readable formats.

☐ Agree   ☐ No opinion   ☐ Disagree

3.   Air Quality Data Providers

Please complete this section of the questionnaire if you are a provider or publisher of air quality data.

1.   Please fill in this use case story to describe your work as an air quality data provider or publisher.

About You:	As a	*(Air quality data provider/publisher/other)*
Narrative:	I want to	*(describe your data provider/publishing activities/services)*
Goal:	So that	*(describe your goal(s) through publishing/providing air quality data)*

Additional Info: Please provide details of the air quality data you provide/publish and its location**.**

2.   Do you have obligations under INSPIRE to publish air quality data, or (for data service providers) do you manage or publish data which falls under INSPIRE?

☐ Yes   ☐ No   ☐ Don't Know

If yes, please describe the work done to date, the data you hold which falls under INSPIRE and what will you have in place to meet the 2013 legal deadlines for discovery, view and download services.

3.   Please describe your access arrangements for the air quality data you provide or publish including details of any machine-readable web access channels which you have implemented. Please describe any plans you have to extend or change these access options in future.

4.   Please provide details for your data services of the data formats available and any standards adopted, including any plans for future developments.

5. Please provide your thoughts on how the air quality community could work together to provide standardised, compatible air quality data for all users and the technical barriers and challenges that will be faced. Provide details of approaches you recommend should be taken to maximise interoperability and data re-use.

6. Please provide any other comments you have in relation to this study and air quality data management

Thank you for completing this questionnaire. Please send your response to Emily.connolly@defra.gsi.gov.uk and tony.bush@ricardo-aea.com before COP 2nd October 2013.

Appendix 5   Defra data services stakeholder consultation user stories

Defra Air & Noise team

Actor
- As Defra AN team I am responsible for complying with the INSPIRE Directive for datasets falling in scope that are published in the public domain

Narrative
- I need web based metadata, viewing & download services to achieve this

Goal
- To comply with European regulations, & open AQ data in interoperable, standards compliant formats

Regional authority data user

Actor
- As a strategic regional authority for AQ & transport & regional coordinator for the LAQM process

Narrative
- I want: to be able to access high quality real-time information about air pollution monitoring in API format; an improved national forecast capacity; improved alerts information

Goal
- So that we can make the data more widely available, we have the forecast & alert information needed to inform members of the public & Defra fulfils its statutory duties.

Regional authority data provider

Actor
- As a strategic regional authority for AQ & transport & regional coordinator for the LAQM process

Narrative
- I want to publish: a London specific inventory in a use friendly format which makes the key outputs accessible to ordinary people; to support accessible real-time air quality monitoring data, nowcasting & forecasting as well as other tools to enable people to make active choices about how, when & where they travel & stay.

Goal
- So that people are informed about air pollution & can take action to minimise their exposure; boroughs & developers have access to the technical information required to do reporting, assessment & planning applications.

Consultant data user

Actor
- As a consultant I use data in provision of consultancy services as experts in the field of air quality to a wide variety of clients across all sectors. This forms part of requirements on public dissemination of information but also provides for validation of modelling work, & comparison between data sets for reporting the context of air quality at any single location, or within an area. Analysis of longer-term data sets is occasionally undertaken.

Narrative
- I want to have confidence in the use of quality assured data that is transparent & auditable to industry standards

Goal
- So that I meet client requirements & provide confidence in our deliverables

Consultant data provider

Actor
- As a data provider. Data is provided for compliance measurements (AURN) & requirements on public information (AURN & local authority websites). We also publish data for construction related monitoring projects in adherence to planning conditions imposed on planning & development works.

Narrative
- I want to meet the obligations for quality assured & transparent data

Goal
- So that I meet client needs on public information & compliance assurance requirements.

AQ researcher

Actor
- As researcher

Narrative
- I want to have easy retrieval of multi pollutant datasets

Goal
- So that I can publish peer review papers on UK air quality

AQ network operator

Actor
• As network operator

Narrative
• I want to provide data in a standard format year on year

Goal
• So that data is produced as efficiently as possible

Consultant data user 2

Actor
• As consultant

Narrative
• I want to use AURN measurements to compare with our measurements during quality control & enhance public dissemination by displaying our measurements alongside the nearby multi-pollutant AURN stations on our air quality websites.

Goal
• So that I can maximise the quality of our air quality measurements & to enhance public dissemination by displaying our measurements alongside the multi-pollutant measurements from nearby AURN stations.

Consultant data provider 2

Actor
• As a data provider / publisher

Narrative
• I want to provider low cost high quality air quality measurements & statistics (to the public, Local Authorities industry & consultants), display real-time & historic air quality measurements, background information & reports on attractive websites & build on the hourly updating of the Northern Irel& website to include the Welsh, Scottish & AURN.

Goal
• So that our air quality monitoring networks provide the best value for money & highest quality in the UK – except for the AURN, Scottish & Welsh networks

Defra network operator

Actor
- As a data provider for Defra

Narrative
- I want to ensure that data QC is carried through to UKAir, to facilitate the link between the Data provider & the data user & ensure that metadata records are up-to-date

Goal
- So that UKAir provides up-to-date & efficient dissemination of data & metadata for Defra networks & the contracts which UKAir provides the service for

AQ researcher

Actor
- As a researcher

Narrative
- I want to be able to easily access AQ measurements & direct others (public, other scientists & policy makers) to data sources. I want the data to be accessible in different template formats

Goal
- So that I can use the dataset efficiently & maximise the data availability to the global research & policy community

As an environmental physicist

Actor
- As an environmental physicist

Narrative
- I want to compare my field measurements with UK-AIR data, examine trends in trace-gas concentration, map concentrations of ozone & effects indices across the UK

Goal
- So that I can quantify & underst& the processes leading to the exposure of ecosystems & people to harmful concentration of trace-gases

As an environmental physicist

Actor
- As an Air Pollution Modeller

Narrative
- I want access to all data for all sites at once

Goal
- So that I can perform automatic ACTM model evaluation

Researcher

Actor
- As a Environmental Modeller

Narrative
- I want to access air quality data to analyse potential human & ecosystem health effects

Goal
- So that I can contribute to the development of efficient air pollution control strategies

Researcher

Actor
- As an Air Pollution Modeller

Narrative
- I want use air quality data to validate atmospheric chemical transport models

Goal
- So that I can contribute to the development of efficient air pollution control strategies

Consultant

Actor
- As a consultant & Model Developer

Narrative
- I want to obtain Defra air quality data

Goal
- So that I can carry out modelling studies for local authorities, developers, & other customers; enable model users to carry out modelling studies; produce air quality forecasts for local authorities

Consultant / researcher

Actor
- As an AQ monitoring network manager

Narrative
- I want to continue being provided with hourly FDMS data from the AURN network & hourly FDMS + gas data from AURN sites & also 6 monthly ratified data from the same sites from Ricardo-AEA

Goal
- So that I can continue to provide the VCM service to researchers & local authorities in the UK & also display real-time pollution levels at AURN sites on our web sites & & our iPhone, Android & Chrome apps.

Policy Maker

Actor
- As a policy maker

Narrative
- I want to make good use of the UK-AIR resources for the support of policy in my jurisdiction

Goal
- Transparency & alignment of policy across UK where possible with the evidence to support it

Policy maker & data provider

Actor
- As an Air quality data provider

Narrative
- I want to make sure that the data we are providing on our website is useful & EU compliant

Goal
- So that the people/organisations that need air quality data have the best service possible.

Consultant

Actor
- As a consultant

Narrative
- I want to access to UK-AIR data in my own web sites & websites I develop for others in a machibne readable format

Goal
- So that I can reuse it in apps, tools & web sites that I develop for my clients

![RICARDO-AEA]

The Gemini Building
Fermi Avenue
Harwell
Didcot
Oxfordshire
OX11 0QR

Tel: 01235 75 3000
Web: www.ricardo-aea.com