Linking Geospatial Data

5th - 6th March 2014, Campus London, Shoreditch

Position Paper

Air Quality: the INSPIRE Annex III Early Bird

Authors: Tony Bush (Ricardo-AEA), Emily Connolly (Defra)

Overview

The UK is at the forefront of advances in air quality data management and digital air quality data publication practices. UK air quality projects are pushing the boundaries for publishing geospatial environmental data using web-standards. Since 2010 Ricardo-AEA (on behalf of Defra) has been leading work to make air quality data "open" and interoperable. There are multiple drivers for this work including legal requirements for EU e-Reporting[1] and INSPIRE[2] and the Cabinet Office transparency agenda[3]. Governmental policy needs for the open air quality data are clear - to inform, protect and serve the public interest. This paper outlines recent developments in UK air quality data management practices, the challenges of being one of the first INSPIRE Annex III data themes creating and implementing a domain-specific data model, including the use of http URI patterns for object identifiers.

Background

The quality of the air we breathe is a health concern for many. In the UK air pollution has an effect equivalent to 29,000 deaths each year and an annual economic cost of £16 billion. It can be a major factor in how much we enjoy life - whether we go for a run, ride a bike, walk the dog or indeed whether we go out at all.

There is a broad community of scientific practice in public and private sectors, academia, central and local governments making air quality observations, predictions and deriving high quality structured

[1] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:152:0001:0044:EN:PDF
[2] http://inspire.jrc.ec.europa.eu/
[3] https://www.gov.uk/government/policies/improving-the-transparency-and-accountability-of-government-and-its-services

data to help protect human and environmental health. These data can be used to inform the actions of individuals: changing behaviour, empowering decisions and reducing health effects – providing science as a service.

In its infancy, the INSPIRE initiative was all about location and geography, but to realise its full benefits the focus must be on the data. This new combination of IT, geography and science has created its own brand of technical tensions, but it's clear that INSPIRE must work to serve policy at all levels to be successful.

Within the EU's air quality community, use of INSPIRE based formats were enshrined into European law in 2008. In addition, regulatory deadlines requiring the use of these new formats were set for the end of December 2013, significantly in advance of reciprocal deadlines under INSPIRE's regulatory timetable. In this sense, air quality is the fledgling INSPIRE Annex III early bird: adopting new standards and services to support open air quality data sharing and regulatory reporting. Ricardo-AEA has led a 3-year programme of work, on behalf of Defra and the European Topic Centre[4], to prepare the UK, European Member States and the European Environment Agency for the December 2013 deadlines. This work has included:

- Defining a European air quality controlled vocabulary
- Modelling of the air quality community data model in UML and generation of an application schema for air quality e-Reporting, based on and extending INSPIRE data specifications. This application schema provides coverage for encoding metadata descriptions of all types of air quality measurement, prediction activities and assessments, as well as observational and assessment data
- Preparation of guidance documentation
- Generation of a http URI set for UK object identifiers to align INSPIRE with wider web-standards.

Many of these topics are yet to be tackled by the other Annex III communities.

[4] http://acm.eionet.europa.eu/

Challenges

The main challenges of the adoption of INSPIRE concepts and principles break down into three core groups:

1. 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 a range of air quality related activities: from monitoring activities, computational prediction of air quality using air quality models, management practices to improve air quality in special areas and assessment practices to check against regulatory thresholds. There are five core INSPIRE themes related to air quality: human health and safety, area management, atmospheric conditions, meteorological features and environmental monitoring facilities. Adding to these the Open Geospatial Consortium (OGC) observations and measurements and sensor web enablement standards for encoding observational data, it becomes clear that the air quality data model is not un-complicated. It is this complexity that has been a major barrier to adoption or perhaps more importantly comprehension of the new INSPIRE data model by the air quality community and many scientific experts. Whereas the adoption of INSPIRE in European air quality regulations was sold to its community as a delivery mechanism for streamlining, burden reduction and automated reporting in an efficient way, the verbosity of XML and INSPIRE standards have at times challenged this idea. Considerable energy has been directed at on-boarding the air quality community, bug fixing and optimising. At this early stage of operational e-Reporting, the real benefits of machine readable data supporting automated checks, cross checks and promoting rigour in data feeds are in development and remain at arm's length.
2. There has been misalignment of regulatory deadlines within the air quality and INSPIRE domains whereby the former has been reliant on the latter for the definition of standards but the latter has no obligation to deliver the required standards according to the timescales of air quality. This has caused some tension and reiteration of testing with air quality work in an implementation phase, in advance of mature and stable standards from INSPIRE and wider stakeholders.
3. Documentation and guidance at an INSPIRE level was (and is) an issue, not necessarily the lack of it, more the amount of it. There is now a common mantra within the air quality community which goes along the lines of "I need documentation on the documentation". Frequently, non-INSPIRE experts simply do not know where to start and a simple non-technical set of pointers and FAQs is essential, going forward, to support non INSPIRE

experts and enable the widespread up take of INSPIRE and the expected downstream benefits.

Benefits

Challenges aside, the benefits of adopting INSPIRE-based reporting by the UK air quality community are expected to be significant. Immediate benefits include;

- An improved governance model for UK air quality data
- Improved transparency in the air quality data model arising from controlled vocabularies and air quality metadata on monitoring stations, sampling points, zones, measurement techniques etc.
- Alignment of air quality with the Cabinet Office transparency agenda
- Higher profile for air quality and improved traceability in the evidential value (re-use) of air quality data through the INSPIRE discovery mechanisms
- Greater openness in air quality data via adoption of the Open Government Licence (OGL)[5] for all Government funded data

Downstream benefits which are expected to deliver real value and savings include;

- Improved and automated access to data via web services in machine readable formats
- Automated processing and post-processing of data including QA/QC checks and compliance checks
- Improved sharing / harvesting via web services
- Air quality on the semantic web to support its re-use at all-levels

Linking Geospatial Air Quality Data

Persistent object identifiers are important to INSPIRE and the semantic web. INSPIRE's approach to object identifiers has changed since its anticipated early implementation. Then, as now, the expectation is that INSPIRE data and services will be delivered over the web. However, the further expectation now is that web-standards and practises for identifier management will allow INSPIRE and Linked Data to successfully crossover. In the web, http URIs have become the primary way to reference information resources. In order to align with this, the UK air quality INSPIRE implementation has adopted a http URI pattern approach to object identifiers within its air quality data model. The governance model for the URI set has been designed with reference to Cabinet Office and UK Government Linked Data Working Group guidance which advocates the construction of URIs using the syntactic structure defined in RFC3986 and HTTP URIs in RFC2616. In a simplified form we have defined a http URI set using the following constructs:

http://{sector}.data.gov.uk{/collection*}{/type}/[/{concept}/{reference}]*

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

Hence for a spatial object representing an air quality monitoring station, its URI might be defined as;

e.g. http://environment.data.gov.uk/air-quality/so/Station_GB0002R

Retrieving this URI will redirect – using standard HTTP redirect mechanisms– to a "doc" URI and download service providing representations of the spatial object in a variety of formats, e.g. in: GML, JSON, HTML, RDF, etc. In this way we aim to bridge the gap between INSPIRE and the semantic web.

Wider air quality data sharing

Extensive, user-led improvements to air quality information sharing have been carried out in response to the Open Data policy agenda and UK Government Digital Service guidelines. Traditional communications media – such as telephone lines and HTML pages, have been supplemented with Twitter and RSS feeds for air quality, online data visualisation tools and web based services to share information in an open but controlled way. In 2012, a UK air quality data catalogue[6] was launched and this is now harvested by data.gov.uk. To open up modelled air quality data, online mapping applications[7] using server technologies have been developed to allow users to interact with, download data and support Web Map Service (WMS) for INSPIRE compliance . All data are provided under the OGL.

Next steps

Ricardo-AEA is currently developing web services for improved access to data and metadata in human and machine readable formats whilst continuing to support "traditional" csv file downloads of data - user story analysis evidences that this continues to be the most popular format. These planned services for early 2014 will include;

- A Sensor Observation Service (SOS) that will be deployed for real time data feeds
- RSS Atom feeds for metadata and historical data archives.
- De-referencable http URI for documentation in html, XML and RSS Atom.
- Extension of the air quality data model for UK specific applications and further engagement with the air quality community to share learning.

Summary

Innovative projects linking location, geography and air quality data are being undertaken with the aim of turning science into a service. Domain specific data models, vocabularies and http URI sets have been created for air quality at a much earlier stage than other INSPIRE Annex III datasets for regulatory obligations - others can benefit from this learning process too. Publishing air quality data

[6] http://uk-air.defra.gov.uk/data/data-catalogue
[7] http://uk-air.defra.gov.uk/data/gis-mapping

in open formats will support the creation of an information marketplace, allowing a wider community to develop innovative services and meet emerging user needs.

Acknowledgements

The following colleagues and organisations deserve recognition and thanks for their contributions in delivering the European air quality e-Reporting system and supporting this work.

ETC ACM - Jaume Targa (4sfera), Wim Mol, Patrick Van Hooydonk, Frank de Leeuw, Paul Ruyssenaars (RIVM), Kathi Schleidt, Barbara Magagna (Umweltbundesamt, Vienna).

European Environment Agency - Valentin Foltescu, Sheila Cryan, Artur Gsella, Michel Houssiau, Soren Roug, Peter Kjeld, Jan Bliki (EEA).

European Commissions (DG-ENV) - Giorgio Arduino, Daniela Buzica.

Ricardo-AEA - Xingyu Xiao

Defra - Emily Connolly