

# Improving the Usability of Geospatial Data – an Academic Perspective

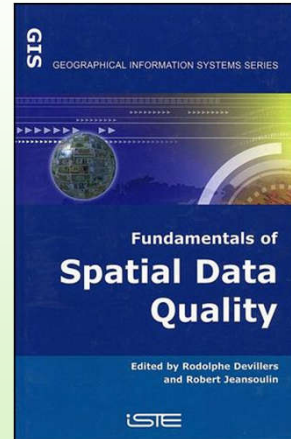
Claire Ellul  
University College London

## Improving the Usability of Geospatial Data – an Academic Perspective

- Overview
  - **Research Perspective**
    - Usability and investigating user needs
  - Teaching Perspective
    - Making data usability interesting
  - Research Data Curation Perspective
    - Drivers and approaches
  - Research Challenges

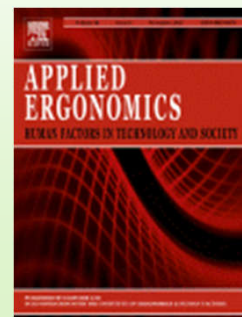
## History

- Data used by experts in isolation
  - Some work on quality issues e.g. fuzzy boundary representation
- Data shared/sold
  - Data Quality = “metadata”
  - Created by producers for expert users
- Web 2.0, Open Data, Open Software, VGI
  - Concept of producers, anyone can capture and use data
  - More focus on non-expert users, downstream data use, data integration
  - Data Quality = “Fitness for purpose” - > evolves into usability of data



## General Research Themes

- Data Quality evaluation (in particular VGI)
  - Intrinsic, Extrinsic
  - Automated, Manual
- Data Quality description
  - Metadata standards and beyond
  - Metadata automation
- Spatial Data Infrastructures



## Usability



- “The degree to which something is able or fit to be used” (Oxford English Dictionary)
- From Software Engineering
  - Usability is a **quality attribute** that assesses how easy user interfaces are to use.
  - Usability is defined by **5 quality components**:
    - **Learnability**: How easy is it for users to accomplish basic tasks the first time they encounter the design?
    - **Efficiency**: Once users have learned the design, how quickly can they perform tasks?
    - **Memorability**: When users return to the design after a period of not using it, how easily can they re-establish proficiency?
    - **Errors**: How many errors do users make, how severe are these errors, and how easily can they recover from the errors?
    - **Satisfaction**: How pleasant is it to use the design?

• From: <https://www.nngroup.com/articles/usability-101-introduction-to-usability/>

### Preparing for the Event

Before you attend the event on the 3<sup>rd</sup> May, we'd like you to think about some geospatial data management challenges that you are currently facing in your organisation, and more specifically to identify **the top challenge** that you face when sharing data with each of the following user groups (where applicable in your organisation):

*Note: for the purpose of this exercise, imagine that you are the 'geospatial data manager', responsible for managing and sharing spatial data across your organisation – although we realise that in reality it is most likely that your role may also encompass at least part of the others listed here.*

- **Specialist/expert colleagues** – “GIS Analysts”. These are the people who have good knowledge of Geographical Information Systems and may be using the data your organisation holds to answer questions posed by colleagues across the organisation or from elsewhere. They will be familiar with desktop GIS tools (e.g. MapInfo, ArcMap, QGIS) and perhaps also of other analytical tools such as R.
- **Non-expert colleagues** – “Decision Makers”. These are the people who use the outputs of your colleagues analysis - e.g. the maps, the charts, the paragraphs of text describing the results – to make decisions for the organisation. Normally they would not have any GIS training.
- **External Users** – “Non-Experts” – these are people outside your organisation, including the general public – who will have access to any open data you provide, and may combine this data with other sources of information. They may or may not have GIS training.

Be prepared to write your three answers on post-it notes at the start of the session – we'll collate them and share them with the group in order to help us understand the most pressing data management challenges we face.

## Challenges for Data Sharing



- **Specialist/expert colleagues**

- Ownership and provenance seems to be the top issue when discussing sharing GI data.
- As data is passed around and transferred, the information about the origin of the data may be lost.
- Where data is being updated, there is not enough coordination to ensure that other dependent datasets are updated accordingly.

## Challenges for Data Sharing



- **Non-expert colleagues**

- Lack of understanding of how to use different GI datasets “correctly”.
- May not understand the limitations of the dataset leading to incorrect applications.
- Some decision makers have a tendency to have absolute faith in the GIS products and lack an understanding of potential sources of errors.
- The lack of understanding may be a result of lack of experience, but more often than not may also be **because of the lack of time to investigate and consider data quality issues.**

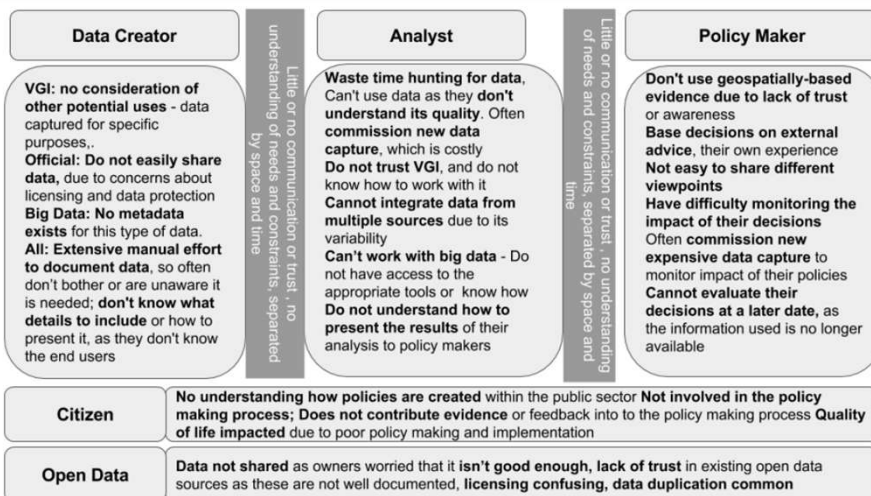
## Challenges for Data Sharing



- **External Users**

- Similar challenges to non-expert colleagues
- Main issue is lack of “understanding” of the data.
- The inappropriate use of data in the wrong contexts and drawing large assumptions from incomplete information is a large concern.

## User Needs – Policy Making



## Improving the Usability of Geospatial Data – an Academic Perspective

- Overview

- Research Perspective

- Usability and investigating user needs

- **Teaching Perspective**

- Making data usability interesting

- Research Data Curation Perspective

- Drivers and approaches

- Research Challenges



### The Problem with Metadata

#### Metadata

- Is “data about data”
- Gives you important information such as
  - When the data was created
  - Who by
  - For what purpose
  - When it was updated
  - How to obtain the data

## The Problem with Metadata

Metadata is:

- Complex and time consuming to create
- Requires expertise about the data
- Requires expertise about how to create useful metadata
  - How much detail should be included?
  - Who are the end users of the metadata?
- Requires MAINTENANCE when data changes!

Field	Description
Spatial reference system	Name or description of the system of spatial referencing, whether by coordinates or geographic identifiers, used in the data resource
Spatial resolution	Measure of the granularity of the data (in metres) 1. For data capture in the field, it is the precision at which the data is captured. This may be the accuracy for topographic surveys, or the average sampling distance in an environmental survey. 2. For data taken from maps, it is the positional accuracy of the map. 3. For image data, it is the resolution of the image. In many cases, only approximate values can be given.
Resource locator	Location (address) for on-line access using a Uniform Resource Locator (URL) address or similar addressing scheme
Data format	Format in which the digital data can be provided
Responsible organisation	Details of the organisation(s) responsible for the establishment, management, maintenance and distribution of the data resource 1. contact position 2. organisation name 3. full postal address 4. telephone number 5. facsimile number 6. email address 7. web address 8. responsible party role
Frequency of update	Frequency with which modifications and deletions are made to the data resource after it is first produced
Limitations on public access	Restrictions imposed on the data resource for security and other reasons
Use constraints	Restrictions and legal restraints on using the data resource
Additional information	Other descriptive information about the data resource
Source	
Metadata date	Date on which the metadata was last updated, or was confirmed as being up-to-date, or if not updated, then the date it was created
Metadata language	Language used for documenting the metadata
Metadata point of contact	Party responsible for the creation and maintenance of the metadata
Unique resource identifier	Value uniquely identifying the data resource
Coupled resource	Identifier of datasets that the service operates on (if the metadata describes a discovery download etc. service rather than relating to a specific dataset)
Originating controlled vocabulary	Name of the formally registered thesaurus or a similar authoritative source of keywords

### The Challenge



### Choose One



Option A



Option B

<http://www.colourbox.com/preview/2713275-557218-refreshing-glass-of-coke-with-ice-cubes.jpg>  
<http://4.bp.blogspot.com/-QXXPRREeSaa/UAQG4HhZU/AAAAAAAAAJc/6OOP29cQ80Y/s200/5403222-ice-cube-dropped-in-cola-glass-and-cola-splashing.jpg>



Choose One



<http://2.bp.blogspot.com/-BwNd3GjSSiU/T1r5IEzFPZI/AAAAAAAAAW8/-Svyblyh5o0/s1600/Coke-and-Pepsi.jpg>

Choose One



Option A

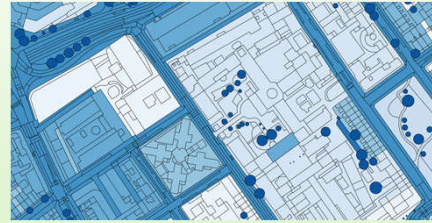


Option B

# Choose One



Ordnance Survey Master Map



UK Map (Geo-Information Group)

### Assessing OSM building data and its potential to reduce uncertainty in natural catastrophe exposure models across Europe

**UCL**

**Introduction**

- In natural catastrophe exposure models, consistent floor area information remains one of the most challenging metrics to acquire.
- OpenStreetMap is a growing, global, database with building information.
- This study was conducted under the supervision and support of Risk Management Solutions (RMS), to investigate how OSM building data might be used to reduce uncertainty in natural catastrophe exposure models across Europe.

**Research Question:**

Can OSM building data be used to reduce uncertainty in natural catastrophe exposure modelling across Europe, and if so, how?

**Objectives:**

- Understand OSM building data quality.
- Apply and assess OSM building data quality to estimate floor area.
- Determine if OSM building data quality can be estimated using a proxy variable.

**Methods:**

- Apply and assess OSM building data quality to estimate floor area.
- Building density regression analysis.
- Population and OSM completeness correlation.

**Results**

(1) Understanding OSM building data quality

- Area-based assessments were most informative in the context of estimating floor area.
- Open-based assessments underestimated OSM building completeness and were influenced by data precision and accuracy.

Area-based completeness estimates: OSM vs. OS Street View

Greater London

Legend: OSM completeness vs. OS Street View

- 0-10%
- 10-20%
- 20-30%
- 30-40%
- 40-50%
- 50-60%
- 60-70%
- 70-80%
- 80-90%
- 90-100%

Sheffield Leeds

(2) Transforming footprint data to floor area

- Most building floor area values were underestimated.
- There was no significant correlation between building footprint and height.
- Altimetry, Ordinary Least Squares (OLS) and Geographically Weighted Regression (GWR) floor area estimates were all significantly correlated with reference estimates. (R<sup>2</sup> values >0.7).

**Height estimates results visualised in 3D**

**Floor area estimates results compared with reference height**

W<sup>2</sup> = 0.8772

W<sup>2</sup> = 0.7200

W<sup>2</sup> = 0.8273

W<sup>2</sup> = 0.8583

(3) Determining if OSM building quality can be estimated with a proxy variable: population density

- There was no correlation between OSM area-based completeness estimates and population density: London example.

London population compared with OSM

**Conclusions**

- It may be possible to estimate building floor area based on OSM building footprints; however without a reliable proxy for OSM completeness it will be challenging to quantify uncertainty in floor area estimates.
- OSM building completeness and quality is variable between UK cities. Area-based assessment methods are most relevant to OSM completeness assessments in the context of floor area estimates.
- There is no correlation between OSM area completeness and population density. Alternative proxy variables based on physical features should be investigated in future work.

**References:** Data.gov.uk (2013). INSPIRE. Retrieved 16/05/2014, from [www.data.gov.uk](http://www.data.gov.uk)

### Data.gov.uk and the INSPIRE Directive Making Data Operationally Useful for the British Red Cross

**UCL**

**Introduction**

The emergence of the Infrastructure for Spatial Information in Europe (INSPIRE) Directive in 2007, has rapidly increased the number of spatial datasets openly available to the general public. Encouraging data sharing at local, national and international levels (Data.gov.uk, 2013). The European project has brought a considerable number of benefits to local governments, businesses and academic bodies, hence attracting considerable interest from the British Red Cross. The work of the British Red Cross focuses on factors such as demographics, climate change and vulnerability and thus might benefit from accessing for spatial characteristics. Therefore, the investigation assesses the INSPIRE Directive spatial datasets for the needs and requirements of the British Red Cross and formulates an approach to assess data quality.

**Objectives**

- To obtain a greater understanding of the organisation, assessing how the British Red Cross interact with the data portal, in order to identify any Human Computer Interaction issues and requirements of the British Red Cross.
- To assess the data quality of both INSPIRE and Non-INSPIRE datasets, outlining the potential of the spatial information.
- To develop a Metadata Tool, which assesses data quality of INSPIRE datasets.

**Methods**

- A semi-structured, open-ended interview was used to obtain a greater understanding of the British Red Cross.
- A Co-operative Evaluation approach was adopted, to identify how the participants interacted with the data portal Data.gov.uk.
- The Manual Data Quality Check assessed the attribute accuracy of the spatial data.
- Case studies were formulated, to showcase the potential of INSPIRE spatial data.
- A Metadata Tool was developed, to automate the process of assessing data quality.

**Key Interview Findings**

- The process of downloading data from the data portal, Data.gov.uk, is both time consuming and frustrating for the user.
- Accessibility to INSPIRE datasets are limited, as the user has to filter through numerous datasets to find the appropriate one.
- Both download links and buttons are often hidden.
- The Generic metadata is easily misinterpreted, due to multiple publishing dates and names.
- The most suitable dataset themes for the British Red Cross are Boundary, Image, Land, Transportation, Health, Natural Disaster and Population.

**Manual Data Quality Check Findings**

Boundary data was the most fit-for-purpose dataset theme from INSPIRE, with a 66% adoption rate. However, INSPIRE lacked Population data.

- Licensing restrictions are a major barrier to entry.
- Health and Flood data was more applicable from Non-INSPIRE datasets.
- The Non-INSPIRE datasets are provided under an Open Government License, allowing the British Red Cross to freely adopt these datasets. However, data quality is compromised as metadata is sparse.

**Case Studies**

Home Help Index

Travel Time Analysis

Nitrogen Dioxide Emissions

**Metadata Tool**

The prototype tool performed very well. It was capable of reading in both XML and other schema formats, making the tool extremely versatile. It is unique in that it provides the user with a flexible tailored tool.

**Conclusion**

**Organisational Context**

INSPIRE datasets did not adequately satisfy the needs and requirements of the British Red Cross. Although, a number of beneficial datasets such as Road Network, Boundary Area and Flood data were identified. Licensing restrictions and the direct lack of Population data limited the adoption of the datasets. Although Non-INSPIRE datasets were more open, they lacked metadata, which made the quality of the data variable. Advocating a clear choice between data quality and Open Data.

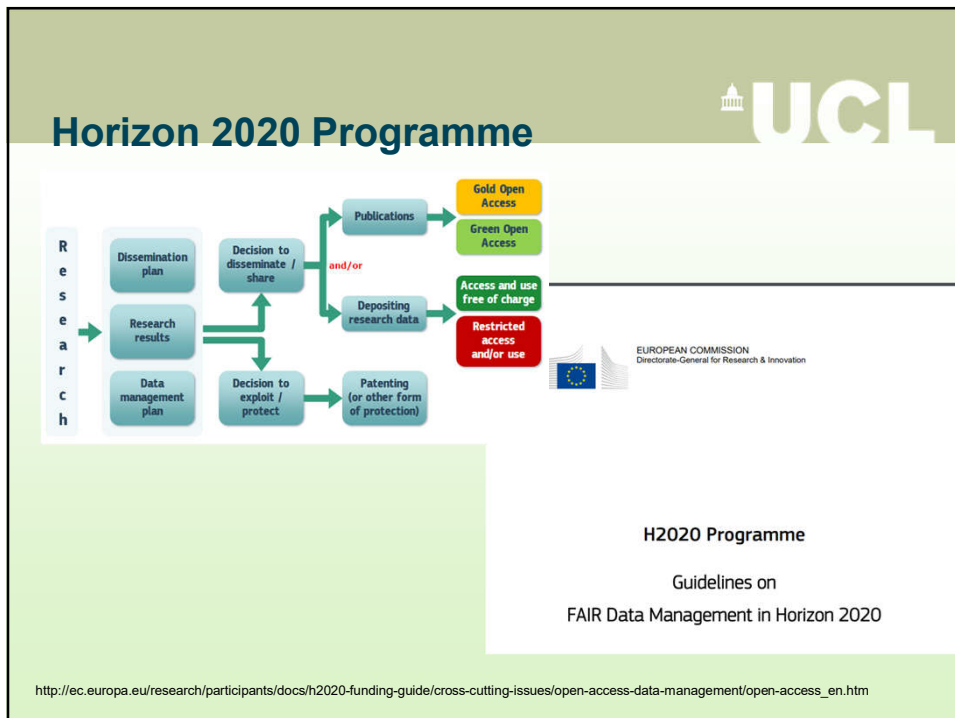
**Future Improvements**

Additional interviews from the GIS team would make the Co-operative Evaluation results more reliable, as some reviewers were more informative than others. A focus group would provide valuable feedback to aid the development of the prototype Metadata Tool and improve its functionality. In addition, a weighting system may also improve the results obtained from the Metadata Tool queries.

**References:** Data.gov.uk (2013). INSPIRE. Retrieved 16/05/2014, from [www.data.gov.uk](http://www.data.gov.uk)

# Improving the Usability of Geospatial Data – an Academic Perspective

- Overview
  - Research Perspective
    - Usability and investigating user needs
  - Teaching Perspective
    - Making data usability interesting
  - **Research Data Curation Perspective**
    - Drivers and approaches
  - Research Challenges



## Executive Summary

H2020 projects are to provide a first version of the Data Management Plan (DMP) within the first six months of the project, with a view to being updated during the project lifetime. The present document presents the initial DMP for the WeGovNow project, thereby describing the project's current view on the data management life cycle for the datasets to be collected, processed or generated for the purposes of the WeGovNow pilot evaluation. This refers to the handling of evaluation data during and after the project. According to the workplan, the current version of the DMP will be updated on the basis of a dedicated evaluation framework to be developed until project month 12 (D4.1). The final DMP will be presented in a dedicated deliverable (D6.3). The current view can be summarized as follows:

- Data set references and names will be specified on the basis of the evaluation framework to become available by project month 12 (D4.1).
- Qualitative evaluation data on positive and/or negative impacts of utilising the WeGovNow platform and services, as perceived at the part of the participating municipalities, will be generated. These will be augmented by quantitative data to be collated (e.g. time spent on utilising the WeGovNow platforms by municipal staff).
- Qualitative evaluation data on positive and/or negative impacts of utilising the WeGovNow platform and services, as perceived at the part of civil society stakeholders (e.g. representatives local NGOs participating at a given pilot site) and citizens, will be generated.
- Quantitative data on WeGovNow platform and service utilisation which can be automatically derived from the technical infrastructure to be piloted (e.g. platform utilisation statistics) will be aggregated.
- Currently it is envisaged that available in an anonymised manner for research purposes upon request.
- All data will be stored at the protected against unauthorised regular off-premise backups.

## 4 Standards and metadata

During the evaluation plan development phase lasting until project month 12, metadata, procedures and file formats for note-taking, recording, transcribing, and anonymising semi-structured interview and focus group discussion data will be developed and agreed. This will also be achieved for any quantitative data to be generated throughout the WeGovNow pilot duration.



Towards We-Government: Collective and participative approaches for addressing local policy challenges

Grant Agreement number: 693514

Deliverable

D6.3

Data Management Framework

[http://wegovnow.eu/fileadmin/wegovnow/images/deliverables/d6\\_3\\_final.pdf](http://wegovnow.eu/fileadmin/wegovnow/images/deliverables/d6_3_final.pdf)

## H2020 Repository



The screenshot shows the Zenodo H2020 Repository search results for the query 'pollution'. The interface includes a search bar with the query 'pollution', a search button, and navigation links for 'Upload' and 'Communities'. The search results are displayed in a list format, showing the date of publication, the title of the document, the author(s), and a brief description. The results are sorted by 'Best match' and 'asc'. The first result is 'NOISE POLLUTION: THE HARMFUL EFFECTS ON HEALTH (SHORT COMMUNICATION)' by Gurinderdeep Singh, published on April 12, 2017. The second result is 'Impact of air pollutants on encountered plant foliage' by Bhisht, Himani; Raj, J. P. N.; Giri, Krishna, published on August 23, 2016. The third result is 'AUTOMATIC POLLUTION DETECTOR FOR AUTOMOBILES' by Sharma Himani; Pandey Sumit; Sharma Sandeep, published on October 10, 2016. The left sidebar shows filters for 'Access Right' (Open, Closed, Embargoed) and 'File Type' (Pdf, Docx, Zip, Xmi, Doc, Csv, Jpg, Mp4, Pptx, Txt). The bottom of the page shows the 'Keywords' field.

UK Research Councils 

 RESEARCH COUNCILS UK

Accessibility | Media Enquiries | Contact

search

Home Funding Research Innovation Skills Public Engagement News, Events and Publications About

## RCUK Common Principles on Data Policy


Home > Research > RCUK Common Principles on Data Policy

Making research data available to users is a core part of the Research Councils' remit and is undertaken in a variety of ways. We are committed to transparency and to a coherent approach across the research base. These RCUK common principles on data policy provide an overarching framework for individual Research Council policies on data policy.

### Principles


- Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner.
- Institutional and project specific data management policies and plans should be in accordance with relevant standards and community best practice. Data with acknowledged long-term value should be preserved and remain accessible and usable for future research.
- To enable research data to be discoverable and effectively re-used by others, sufficient metadata should be recorded and made openly available to enable other researchers to understand the research and re-use potential of the data. Published results should always include information on how to access the supporting data.
- RCUK recognises that there are legal, ethical and commercial constraints on release of research data. To ensure that the research process is not damaged by inappropriate release of data, research organisation policies and practices should ensure that these are considered at all stages in the research process.
- To ensure that research teams get appropriate recognition for the effort involved in collecting and analysing data, those who undertake Research Council funded work may be entitled to a limited period of privileged use of the data they have collected to enable them to publish the results of their research. The length of this period varies by research discipline and, where appropriate, is discussed further in the published policies of individual Research Councils.
- In order to recognise the intellectual contributions of researchers who generate, preserve and share key research datasets, all users of research data should acknowledge the sources of their data and abide by the terms and conditions under which they are accessed.
- It is appropriate to use public funds to support the management and sharing of publicly-funded research data. To maximise the research benefit which can be gained from limited budgets, the mechanisms for these activities should be both efficient and cost-effective in the use of public funds.





Further guidance and information on the individual principles can be found within the [guidance documentation](#).

UK Research Councils 

- Research organisations will ensure that **appropriately structured metadata describing the research data they hold** is published (normally within 12 months of the data being generated) and made freely accessible on the internet;
- in each case the metadata must be sufficient to allow others to understand:
  - what research data exists,
  - why, when and how it was generated,
  - and how to access it.
- Where the research data referred to in the metadata is a digital object it is expected that the metadata will include use of a robust digital object identifier (For example as available through the [DataCite organisation](#)).


<https://www.epsrc.ac.uk/about/standards/researchdata/expectations/>

**UKRC Recommended Repository** 


 About us ▾ Services ▾ Resources ▾ Community ▾ Become a member   

**WELCOME TO DATACITE**  
 Locate, identify, and cite research data with the leading global provider of DOIs for research data.


[Learn more](#)




Find what you're looking for by searching millions of records with extensive, reliable metadata.



Share your data and reuse the data of others to create the highest impact in the research community.




Cite your research sources with confidence, and receive proper credit when your work is reused.



Connect your research – publications, datasets, software, authors, institutions, and funding data all in one place.

[https://schema.datacite.org/meta/kernel-4.0/doc/DataCite-MetadataKernel\\_v4.0.pdf](https://schema.datacite.org/meta/kernel-4.0/doc/DataCite-MetadataKernel_v4.0.pdf)



**Table 1: DataCite Mandatory Properties**

ID	Property	Obligation
1	Identifier (with mandatory type sub-property)	M
2	Creator (with optional name identifier and affiliation sub-properties)	M
3	Title (with optional type sub-properties)	M
4	Publisher	M
5	PublicationYear	M
10	ResourceType (with mandatory general type description sub-property)	M

**Table 2: DataCite Recommended and Optional Properties**

ID	Property	Obligation
6	Subject (with scheme sub-property)	R
7	Contributor (with type, name identifier, and affiliation sub-properties)	R
8	Date (with type sub-property)	R
9	Language	O
11	AlternateIdentifier (with type sub-property)	O
12	RelatedIdentifier (with type and relation type sub-properties)	R
13	Size	O
14	Format	O
15	Version	O
16	Rights	O
17	Description (with type sub-property)	R
18	GeoLocation (with point, box and polygon sub-properties)	R
19	FundingReference (with name, identifier, and award related sub-properties)	O

[https://schema.datacite.org/meta/kernel-4.0/doc/DataCite-MetadataKernel\\_v4.0.pdf](https://schema.datacite.org/meta/kernel-4.0/doc/DataCite-MetadataKernel_v4.0.pdf)

## Other Avenues for Publication



nature.com > scientific data

SCIENTIFIC DATA

Data Descriptor | 30 May 2017 | OPEN

RiceAtlas, a spatial database of global rice calendars and production

Article | 06 June 2017 | OPEN

DATS, the data tag suite to enable discoverability of datasets

Susanna-Assunta Sansone, Alejandra Gonzalez-Beltran [...] Lucila Ohno-Machado

Data Descriptor | 06 June 2017

A monthly global reanalysis of the atmosphere from 1950 to 2005 for studying climatic variations

Jörg Franke, Stefan Brönnig

www.nature.com/scientificdata

SCIENTIFIC DATA

OPEN Data Descriptor: RiceAtlas, a spatial database of global rice calendars and production

Alice G. Laborte<sup>1</sup>, Mary Anne Gutierrez<sup>2</sup>, Jane Gily Balanza<sup>3</sup>, Kazuki Saito<sup>4</sup>, Sander J. Zwart<sup>5</sup>, Mirco Boschetti<sup>6</sup>, M.V.M. Murty<sup>7</sup>, Lorena Villano<sup>8</sup>, Jonel Khali Aunario<sup>9</sup>, Russell Roberts<sup>10</sup>, Jeevan Koor<sup>11</sup>, Imbert J. Hignani<sup>12</sup> & Andrew Nelson<sup>13</sup>

Received: 14 December 2016  
Accepted: 19 April 2017  
Published: 30 May 2017

Knowing when, where, and how much rice is planted and harvested is crucial information for understanding the effects of policy, trade, and global and technological change on food security. We developed RiceAtlas, a spatial database on the seasonal distribution of the world's rice production. It consists of data on rice planting and harvesting dates by growing season and estimates of monthly production for all rice-producing countries. Sources used for planting and harvesting dates include global and regional databases, national publications, online reports, and expert knowledge. Monthly production data were estimated based on annual or seasonal production statistics, and planting and harvesting dates. RiceAtlas has 2725 spatial units. Compared with available global crop calendars, RiceAtlas is nearly ten times more spatially detailed and has nearly seven times more spatial units, with at least two seasons of calendar data, making RiceAtlas the most comprehensive and detailed spatial database on rice calendar and production.

Design Type(s)	data integration objective • database creation objective • observation design
Measurement Type(s)	agricultural calendar • agricultural production
Technology Type(s)	digital curation
Factor Type(s)	geographic location
Sample Characteristics	rice field • Afghanistan • Algeria • Angola • Australia • Azerbaijan • Bangladesh • Bolivia • Brazil • Bhutan • Burkina Faso • Burundi • Cambodia • Cameroon • Central African Republic • Chad • China • Chile • Colombia • Comoros • Costa Rica • Cote d'Ivoire • Cuba • Democratic Republic of the Congo • Dominican Republic • Ecuador • Egypt • El Salvador • Ethiopia • Fiji • France • Gabon • Gambia • Ghana • Guinea • Guatemala • Guinea • Guinea-Bissau • Guyana • Guyana • Haiti • Honduras • Hungary • India • Indonesia • Iran • Iraq • Italy • Jamaica • Japan • Kazakhstan • Kenya • Kyrgyzstan • Laos •

## Improving the Usability of Geospatial Data – an Academic Perspective



- Overview
  - Research Perspective
    - Usability and investigating user needs
  - Teaching Perspective
    - Making data usability interesting
  - Research Data Curation Perspective
    - Drivers and approaches
  - Research Challenges

# Challenge 1 – What is useful metadata? Who creates/uses metadata and how?

A search for the term 'cadastre' in the European Commission's INSPIRE Geoportal yields a wide range of record counts: 17368 results for France, 9628 for Germany, 1434 for the UK (where they are separated for each municipality), 229 for Portugal, 10903 for Italy. Comparing three datasets in English shows the following information:

**Malta: Parcels of Registered Land**  
**Abstract:** The main functions and responsibilities of the Land Registry emanate from the Land Registration Act (Chapter 296 of the Laws of Malta) primarily in receiving applications for the registration of immovable property within registration areas. **Lineage:** Derived from published Legal Notices. **Conformity:** Specification: Commission Regulation (EU) No 1089/2010 of 23 November 2010 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards interoperability of spatial data sets and services. Date of publication: 2010-12-08

**Ireland: Property Registration Authority of Ireland - Cadastral Parcels Leasehold (Metadata)**  
**Abstract:** Leasehold Land Ownership Boundaries - The dataset contains the boundaries of each individual leasehold title to land. Please note this data has a zoom extent of 1:20,000. **Lineage:** PRA cadastral parcels were digitised from Land Registry paper map sheets (map scales 1:10560, 1:2500, 1:1250, 1:1056 and 1:1000) over a period of 5 years from November 2005 to August 2010. Cadastral parcels created since then are either digitised from raster (scanned and georectified) application maps or from vector geometry lodged as CAD files. **Conformity:** Specification: INSPIRE Data Specification on Cadastral Parcel. Date of publication: 2015-05-25 Degree: Conformant

**UK: Land Registry INSPIRE Download Service Metadata**  
**Abstract:** INSPIRE Polygons are a complete set of Freehold life extent indexes available for England and Wales. The data is structured to meet the requirements of the EU INSPIRE Directive. Land Registry's INSPIRE download service allows you to download INSPIRE Index Polygons for a location of your choice. The data is presented as polygons that can be used with Geographic Information System (GIS) software or other software that can read the data. **Lineage:** INSPIRE Download Service. **Conformity:** (no information provided)



**Please fill in the form**  
 All fields marked with \* must be filled in

Metadata ID\*: 4568

Title\*:

Abstract\*:

Type of Data\*: Unknown

Work Package(s)\*: WP1

Time Period Covered by Dataset:

Dataset Creation Process:

Contact Email: c.ellul@ucl.ac.uk

Contact: London Metropolitan University

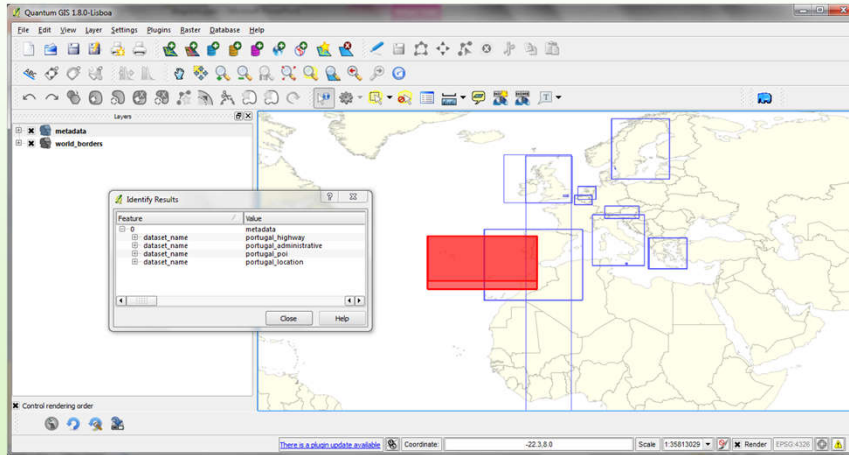
# Challenge 2 – Fitting in with how people work

**London** 51.51 N, 0.13 W Mon 12 Jun @ 20:29:50  
Go to Map - Go to Grid - Change City

<b>WEATHER STATION (CASA TEAM)</b> <table border="1"> <thead> <tr> <th>STATION</th> <th>WIND SPEED</th> <th>WIND GUSTS</th> <th>DIRECTION</th> <th>TEMPERATURE</th> <th>HUMIDITY</th> <th>RAIN TOTAL</th> <th>PRESSURE</th> <th>FORECAST</th> </tr> </thead> <tbody> <tr> <td>CASA Office, Bloomsbury 1911</td> <td>8.0 mph</td> <td>23.0 mph</td> <td>NE ✓</td> <td>18.5 °C</td> <td>59%</td> <td>0.0 mm</td> <td>1016.48 mbar</td> <td>Clear Night</td> </tr> <tr> <td>EM Weather, Watlington</td> <td colspan="8">Data not updated for 1200 hours</td> </tr> <tr> <td>SJO Weather, Pinner HAS</td> <td>9.0 mph</td> <td>11.0 mph</td> <td>NE ✓</td> <td>-1.3 °C</td> <td>57%</td> <td>0.0 mm</td> <td>1014.72 mbar</td> <td>Cloudy</td> </tr> </tbody> </table>	STATION	WIND SPEED	WIND GUSTS	DIRECTION	TEMPERATURE	HUMIDITY	RAIN TOTAL	PRESSURE	FORECAST	CASA Office, Bloomsbury 1911	8.0 mph	23.0 mph	NE ✓	18.5 °C	59%	0.0 mm	1016.48 mbar	Clear Night	EM Weather, Watlington	Data not updated for 1200 hours								SJO Weather, Pinner HAS	9.0 mph	11.0 mph	NE ✓	-1.3 °C	57%	0.0 mm	1014.72 mbar	Cloudy	<b>WEATHER (METAR)</b> <b>London City</b> Winds 1100 at 15kt Vis 10km Scattered clouds at 4500ft W at 13 mph 19 C	<b>TRAFFIC CAMERAS (TFL)</b> Podium High St/Rye Lane 													
STATION	WIND SPEED	WIND GUSTS	DIRECTION	TEMPERATURE	HUMIDITY	RAIN TOTAL	PRESSURE	FORECAST																																											
CASA Office, Bloomsbury 1911	8.0 mph	23.0 mph	NE ✓	18.5 °C	59%	0.0 mm	1016.48 mbar	Clear Night																																											
EM Weather, Watlington	Data not updated for 1200 hours																																																		
SJO Weather, Pinner HAS	9.0 mph	11.0 mph	NE ✓	-1.3 °C	57%	0.0 mm	1014.72 mbar	Cloudy																																											
<b>TUBE LINE STATUS (TFL)</b> <table border="1"> <tr><td>Central</td><td>Good Service</td></tr> <tr><td>District</td><td>Good Service</td></tr> <tr><td>N&amp;C</td><td>Good Service</td></tr> <tr><td>Metropolitan</td><td>Good Service</td></tr> <tr><td>North</td><td>Good Service</td></tr> <tr><td>Piccadilly</td><td>Good Service</td></tr> <tr><td>Victoria</td><td>Good Service</td></tr> <tr><td>W&amp;C</td><td>Good Service</td></tr> <tr><td>Charing Cross</td><td>Good Service</td></tr> <tr><td>TL Rail</td><td>Good Service</td></tr> <tr><td>DLR</td><td>Good Service</td></tr> <tr><td>Tram</td><td>Good Service</td></tr> </table>	Central	Good Service	District	Good Service	N&C	Good Service	Metropolitan	Good Service	North	Good Service	Piccadilly	Good Service	Victoria	Good Service	W&C	Good Service	Charing Cross	Good Service	TL Rail	Good Service	DLR	Good Service	Tram	Good Service	<b>LONDON CYCLE HIRE (TFL)</b> 3.6 % Stations Full 11.5 % Stations Empty 9162 Bikes Available 221 Bikes or Docks Faulty	<b>IN SERVICE (TFL)</b> 4590 London buses 387 Underground Buses	<b>AIR POLLUTION (DEPRA)</b> <table border="1"> <thead> <tr> <th>Station</th> <th>Ozone</th> <th>NO<sub>2</sub></th> <th>SO<sub>2</sub></th> <th>PM<sub>10</sub></th> <th>PM<sub>2.5</sub></th> </tr> </thead> <tbody> <tr> <td>Bloomsbury</td> <td>40</td> <td>39</td> <td>5</td> <td>?</td> <td>?</td> </tr> <tr> <td>Manlybone Rd</td> <td>16</td> <td>10</td> <td>12</td> <td>12</td> <td>21</td> </tr> <tr> <td>N Kensington</td> <td>63</td> <td>11</td> <td>4</td> <td>7</td> <td>17</td> </tr> </tbody> </table>	Station	Ozone	NO <sub>2</sub>	SO <sub>2</sub>	PM <sub>10</sub>	PM <sub>2.5</sub>	Bloomsbury	40	39	5	?	?	Manlybone Rd	16	10	12	12	21	N Kensington	63	11	4	7	17
Central	Good Service																																																		
District	Good Service																																																		
N&C	Good Service																																																		
Metropolitan	Good Service																																																		
North	Good Service																																																		
Piccadilly	Good Service																																																		
Victoria	Good Service																																																		
W&C	Good Service																																																		
Charing Cross	Good Service																																																		
TL Rail	Good Service																																																		
DLR	Good Service																																																		
Tram	Good Service																																																		
Station	Ozone	NO <sub>2</sub>	SO <sub>2</sub>	PM <sub>10</sub>	PM <sub>2.5</sub>																																														
Bloomsbury	40	39	5	?	?																																														
Manlybone Rd	16	10	12	12	21																																														
N Kensington	63	11	4	7	17																																														
<b>TRAFFIC CAMERAS (TWO at RANDOM) (TFL)</b> Hanger Lane/Ubbridge Rd Hammersmith Bridge Rd	<b>BBC LONDON NEWS (BBC)</b> Croydon tram crash: TFL to add warning systems to trams - Croydon drive-by shooting: Victim named as Timodiwanase Makwanyia - London Bridge attack: Duchess	<b>STOCKS (YAHOO)</b> Goldsmiths' Row 2311 yesterday FTSE 100 Index 7511.87 -16.46 (-0.21%)	<b>OPENSTREETMAP UPDATES (OSM)</b> Add road surfaces Add house numbers from survey, building polygons from Bing & DigitalGlobe - create along Cheshamwell Road in																																																



## Challenge 2 – Fitting in with how people work



## Challenge 3 – Theory versus reality

- Students are taught about metadata and data management best practices in their courses
- However, in real life there is perhaps more pressure to produce outputs to deadline and also to delivery a very summarised version of information without the end user understanding its strengths and limitations

## Challenge 4 – Learning from Other Disciplines



Data Protection  
Act 1998

Information is:

- used fairly and lawfully
- used for limited, specifically stated purposes
- used in a way that is adequate, relevant and not excessive
- accurate
- kept for no longer than is absolutely necessary
- handled according to people's data protection rights
- kept safe and secure
- not transferred outside the [European Economic Area](#) without adequate protection

<https://www.griffinhouseconsultancy.co.uk/wp-content/uploads/dpa-crown.png>

### Consent form

*"The comparative efficacy of on-line Cognitive Behavioural Therapy versus on-line Imagery Relief Therapy as treatments for Insomnia Disorder: a randomised controlled trial".*  
University of Glasgow Sleep Centre & Sleepio

Read the details of this trial here: <http://bit.ly/patientinfo sheet>. Please email any questions to [iam@sleepio.com](mailto:iam@sleepio.com).

\*Required

**Please carefully read the statements below and indicate your agreement by checking the boxes next to each statement. \***

- I confirm that I have read and understand the information above regarding the above study and have had the opportunity to ask questions.
- I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my legal rights being affected.
- I agree to take part in the above study.

**Please confirm your email address. \***

Submit

Powered by [Google Docs](#)

[Report Abuse](#) - [Terms of Service](#) - [Additional Terms](#)

<https://www.sleepio.com/img/research/consent-form.jpg>

## Challenge 5 - Finding Research Funding



- Interdisciplinary approach needed!
- Funding Options
  - H2020
  - EPSRC
  - Wellcome Trust
  - Other sources?
- But
  - Topic is not 'blue-sky research'
  - No calls on data curation
  - Difficult to fit into other proposals
    - Data science currently focussing on big data analytics

# Improving the Usability of Geospatial Data – an Academic Perspective

UCL

**Any Questions?**



[c.ellul@ucl.ac.uk](mailto:c.ellul@ucl.ac.uk)