Improving the Usability of Geospatial Data – an Academic Perspective

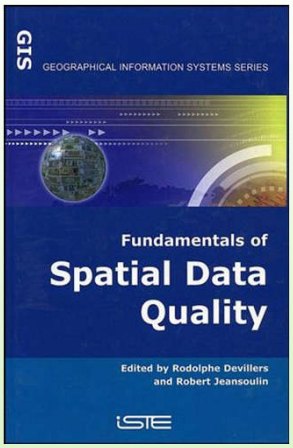
Claire Ellul
University College London

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 produsers, 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 3rd 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.

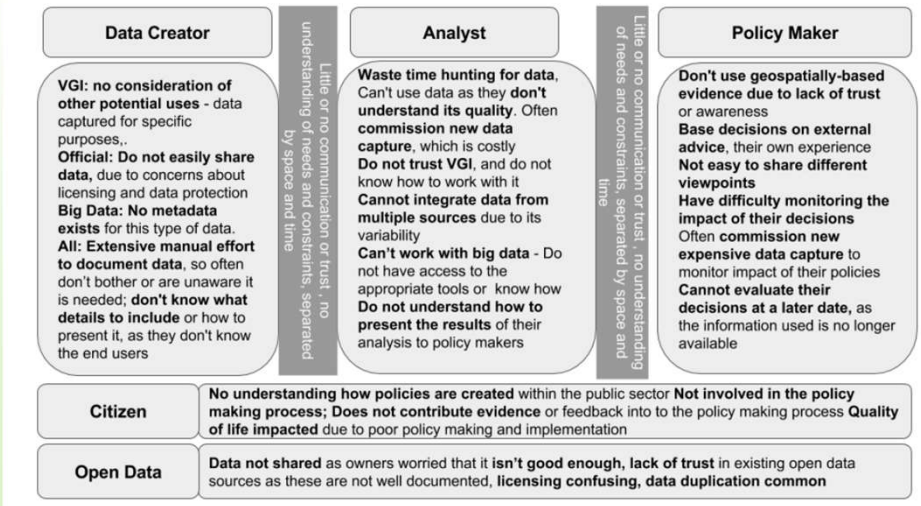
Challenges for Data Sharing

- **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 Source	Other descriptive information about the data resource
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/-QXXPRREeSqg/UAOQ4tHzZUI/AAAAAAAAAJo/6OOP29cQ90Y/s200/5403222-ice-cube-droped-in-cola-glass-and-cola-splashing.jpg

Choose One

http://2.bp.blogspot.com/-BwNd3GlSSlU/T1r5tEzFPZI/AAAAAAAAVW8/-Svyblyh5o0/s1600/Coke-and-Pepsi.jpg

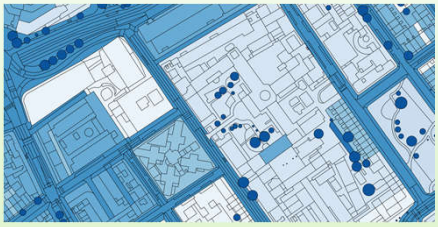
Choose One

Option A

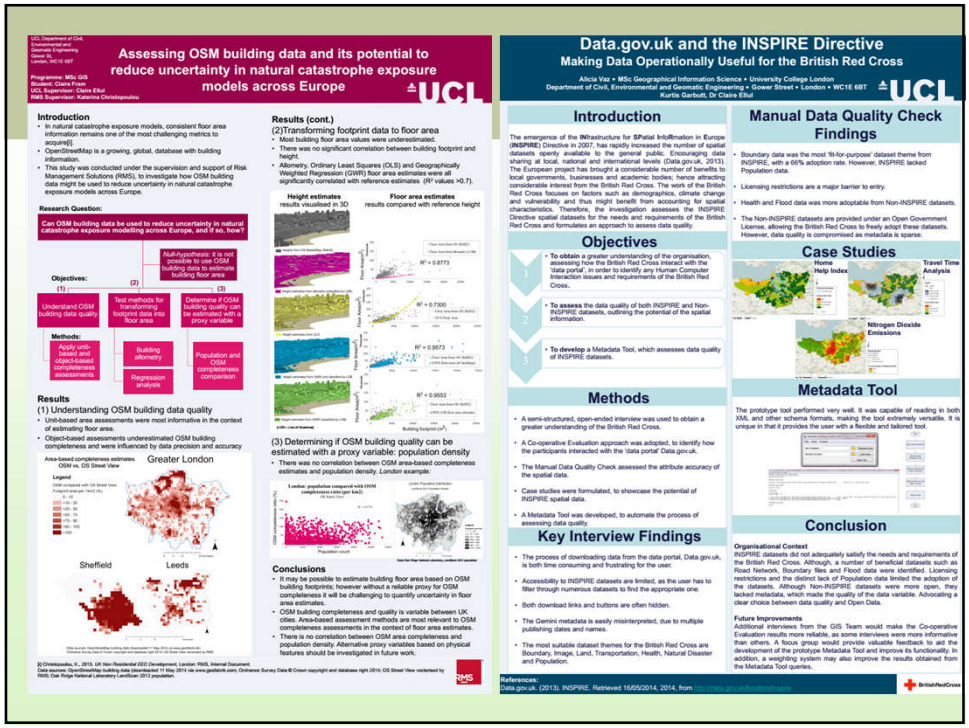
Option B

Choose One

Ordnance Survey Master Map



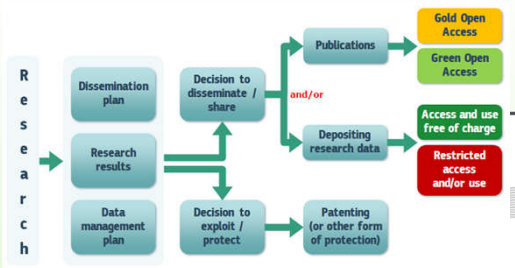
UK Map (Geo-Information Group)



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

Horizon 2020 Programme



EUROPEAN COMMISSION
Directorate-General for Research & Innovation

H2020 Programme

Guidelines on
FAIR Data Management in Horizon 2020

http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm

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 m
  research purposes upon reque
- All data will be stored at the
  protected against unauthoris
  regular off-premise backups.

WeGovNow

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

Grant Agreement number: 693514

Deliverable

D6.3

Data Management Framework

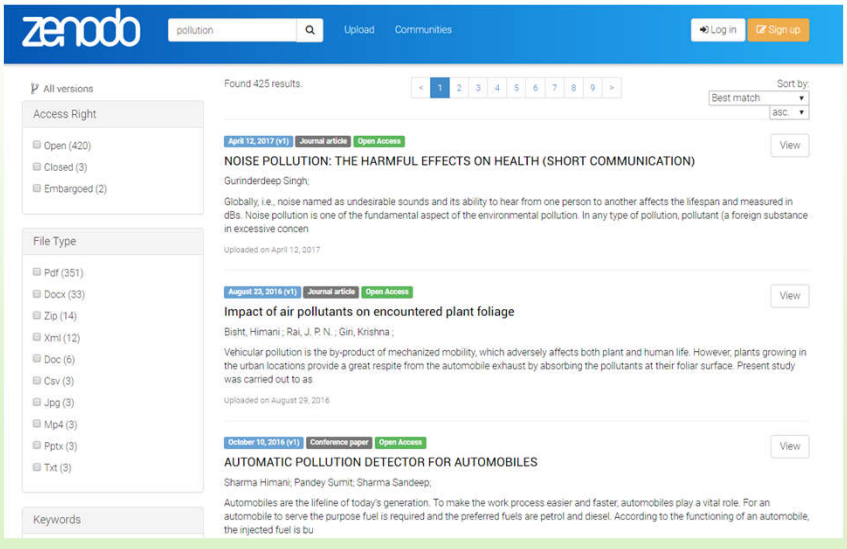
4 Standards and metadata

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.

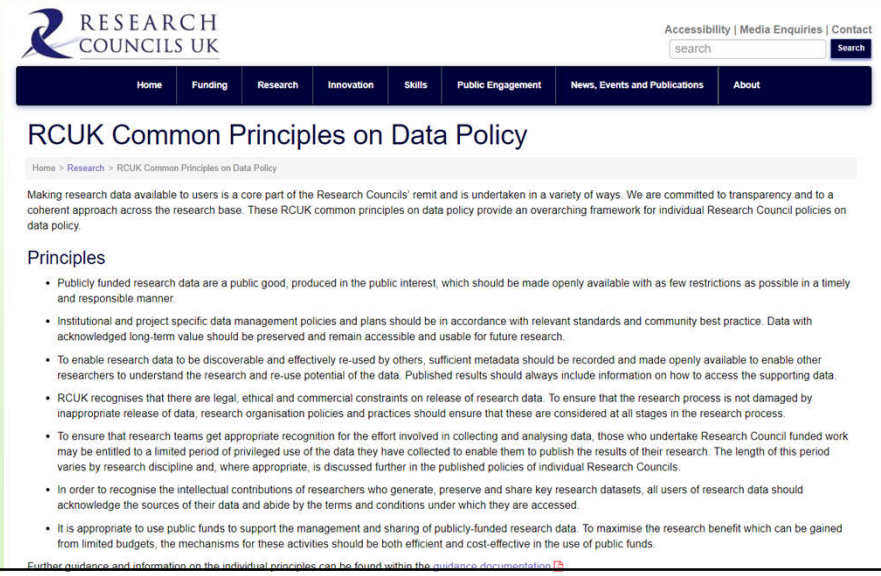
http://wegovnow.eu/fileadmin/wegovnow/images/deliverables/d6_3_final.pdf

H2020 Repository

UCL



UK Research Councils



RESEARCH COUNCILS UK

Accessibility | Media Enquiries | Contact

search 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.

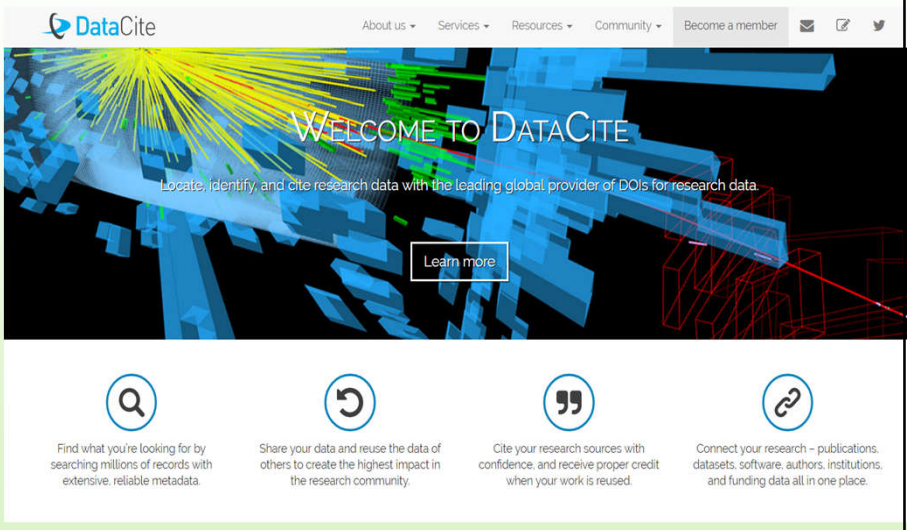
Further guidance and information on the individual principles can be found within the guidance documentation

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



https://schema.datacite.org/meta/kernel-4.0/doc/DataCite-MetadataKernel_v4.0.pdf

Table 1: DataCite Mandatory Properties

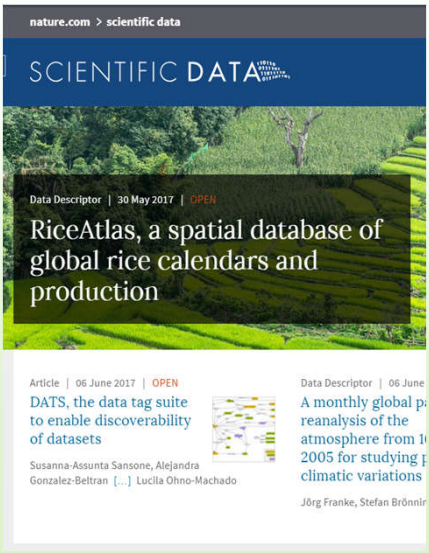
ID	Property	Obligation
1	Identifier (with mandatory type sub-property)	M
2	Creator (with optional name identifier and affiliation sub-properties)	M
3	Title (with optional type sub-properties)	M
4	Publisher	M
5	PublicationYear	M
10	ResourceType (with mandatory general type description sub-property)	M

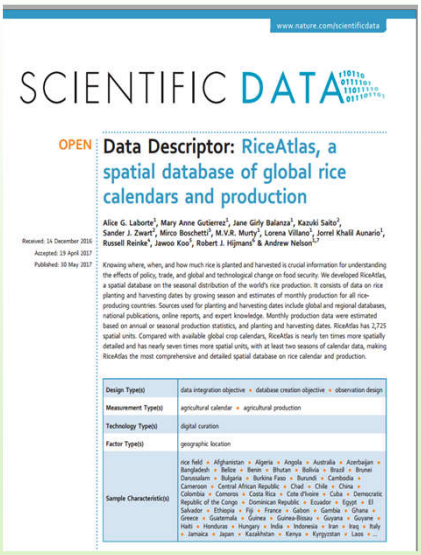
Table 2: DataCite Recommended and Optional Properties

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
[illegible]	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

Other Avenues for Publication

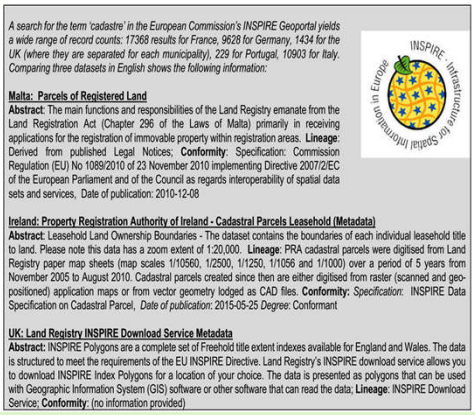




Improving the Usability of Geospatial Data – an Academic Perspective

- 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

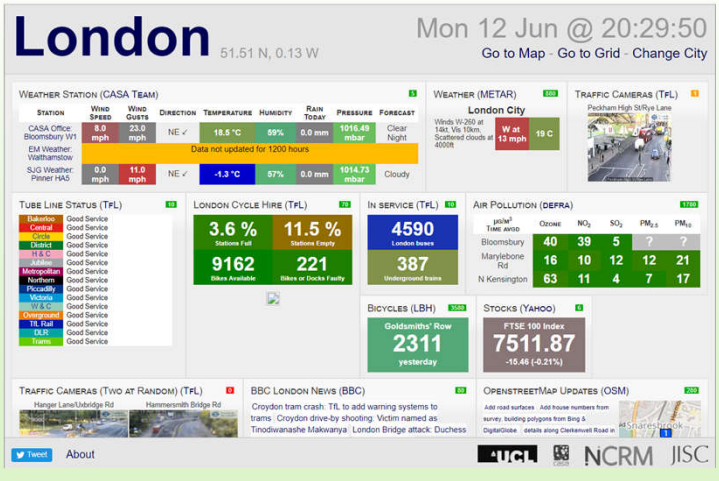
Ireland: Property Registration Authority of Ireland - Cadastral Parcels Leasehold (Metadata)

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 geo-positioned) 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

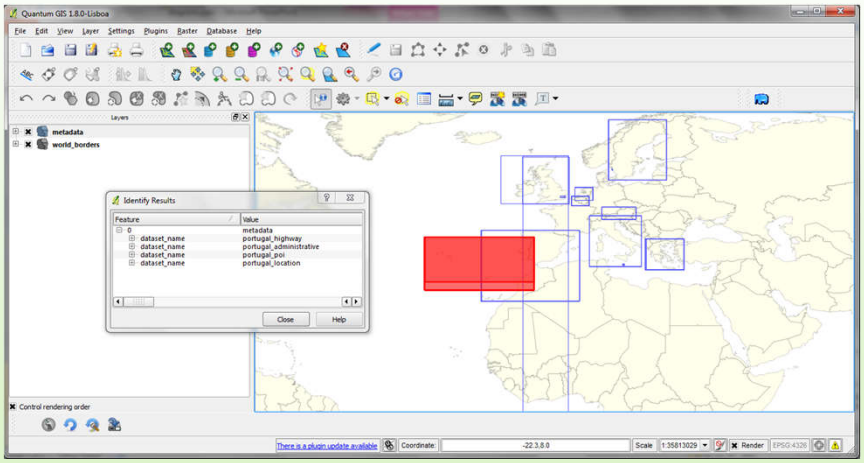
UK: Land Registry INSPIRE Download Service Metadata

Abstract: INSPIRE Polygons are a complete set of Freehold title 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)

Challenge 2 – Fitting in with how people work



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

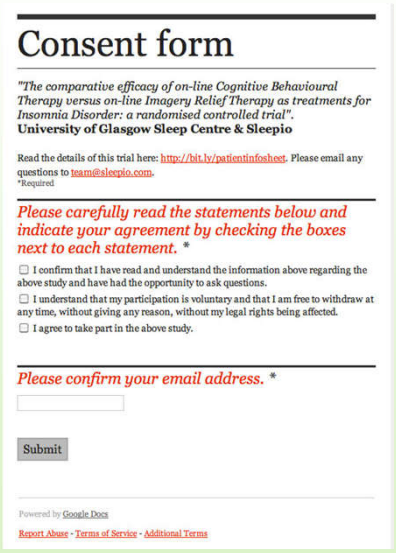
Challenge 4 – Learning from Other Disciplines

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/patientinfosheet. Please email any questions to team@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