m(App) My Data! : The Power of Location as a Data Integrator

Paul Reynolds paul.reynolds.22@ucl.ac.uk

Supervised by Dr Claire Ellul

UCL

PROBLEM

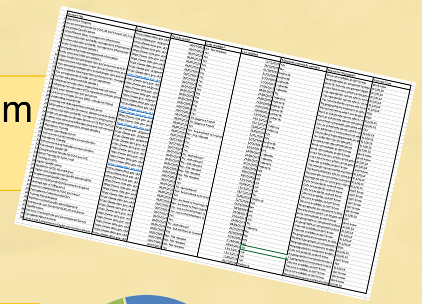
- Linking data together (**data integration**) via location can offer comprehensive insights and aid in evidence-based decision-making.
- Data integration helps tackle pressing challenges, including urbanisation, resource management, education and environmental sustainability.
- **Mapping** is a key method to visualise location-integrated data.
- However**,** integrating data can be **difficult** due to the varying nature of the data. Data often comes in formats that lack the geographical coordinates necessary for direct mapping.
- This means that researchers must spend **significant time** converting non-map-ready data.

CONTEXT – *the current state of data mapping*

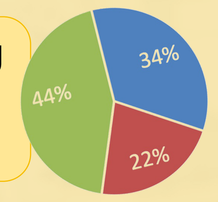
- Current use cases demonstrate the **transformative power** of geospatial mapping, such as tracking progress towards **Net Zero** carbon emissions by overlaying environmental data with industrial outputs, urban development patterns, and natural resource consumption.[1]
- Modern geospatial tools and platforms have **broadened** the range of datasets that can be visualised, with methods continually evolving to map **diverse** sources.[2,3]
- While tools like GIS and **Google Maps** excel at handling data with inherent geographical coordinates, many datasets, especially those related to socio-economic indicators or narrative reports, remain **challenging** to map directly.[4]
- Platforms like OpenStreetMap are pioneering the integration of varied data types and from **unconventional** data sources, suggesting a trend towards more inclusive mapping ecosystems. This validates the significance of mapping in realms like disaster response, tourism, and infrastructure planning.[5,6]
- The concept of the **'real-time city'** captures a significant shift towards immediate data interpretation and visualisation, emphasising the **dynamic** nature of today's mapping processes. This can be crucial for traffic management, pollution monitoring, and crime analysis.[7,8]
- Despite advancements, the mappability of datasets remains a core challenge, pressing the need for innovative solutions and strategies to streamline and **optimise** mapping workflows.

METHOD

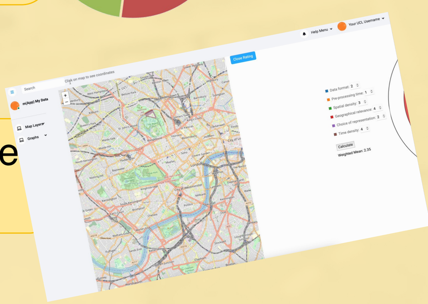
step 1
Collect and attempt mapping a sample of datasets from the 'Education' filter on data.gov.uk.



step 2
Record common factors and challenges of the mapping process, using findings to iteratively improve the rating system.



step 3
Implement the rating system in a web app, enabling the map visualisation of the data and its rating.

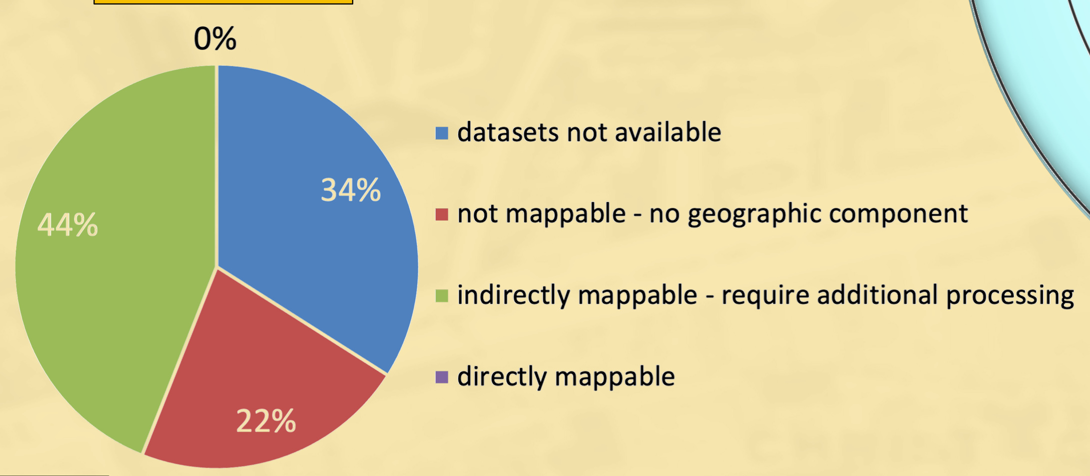


dataset FINDINGS

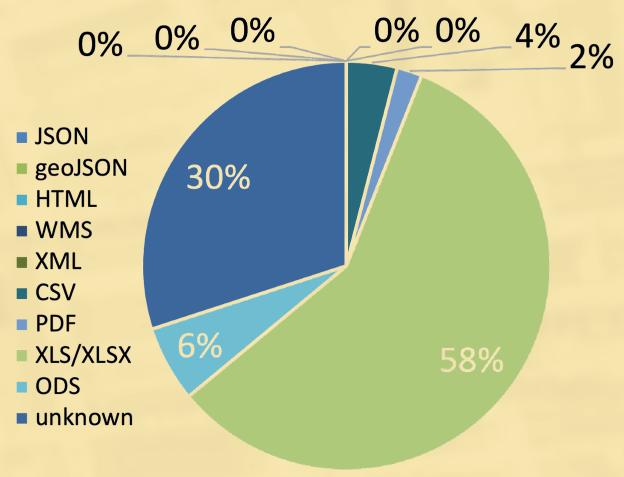
Filters in the search tool generally worked well. However, they sometimes displayed **illogical** results or **hid** key datasets. For example, 'Average age of rollingstock', a dataset about trainlines, was include under the 'Education' filter.

Out of the sample of **50** education datasets:

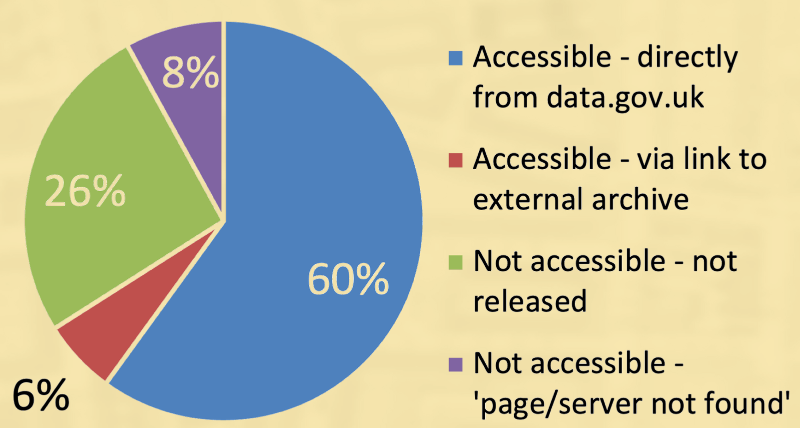
accessibility



primary format



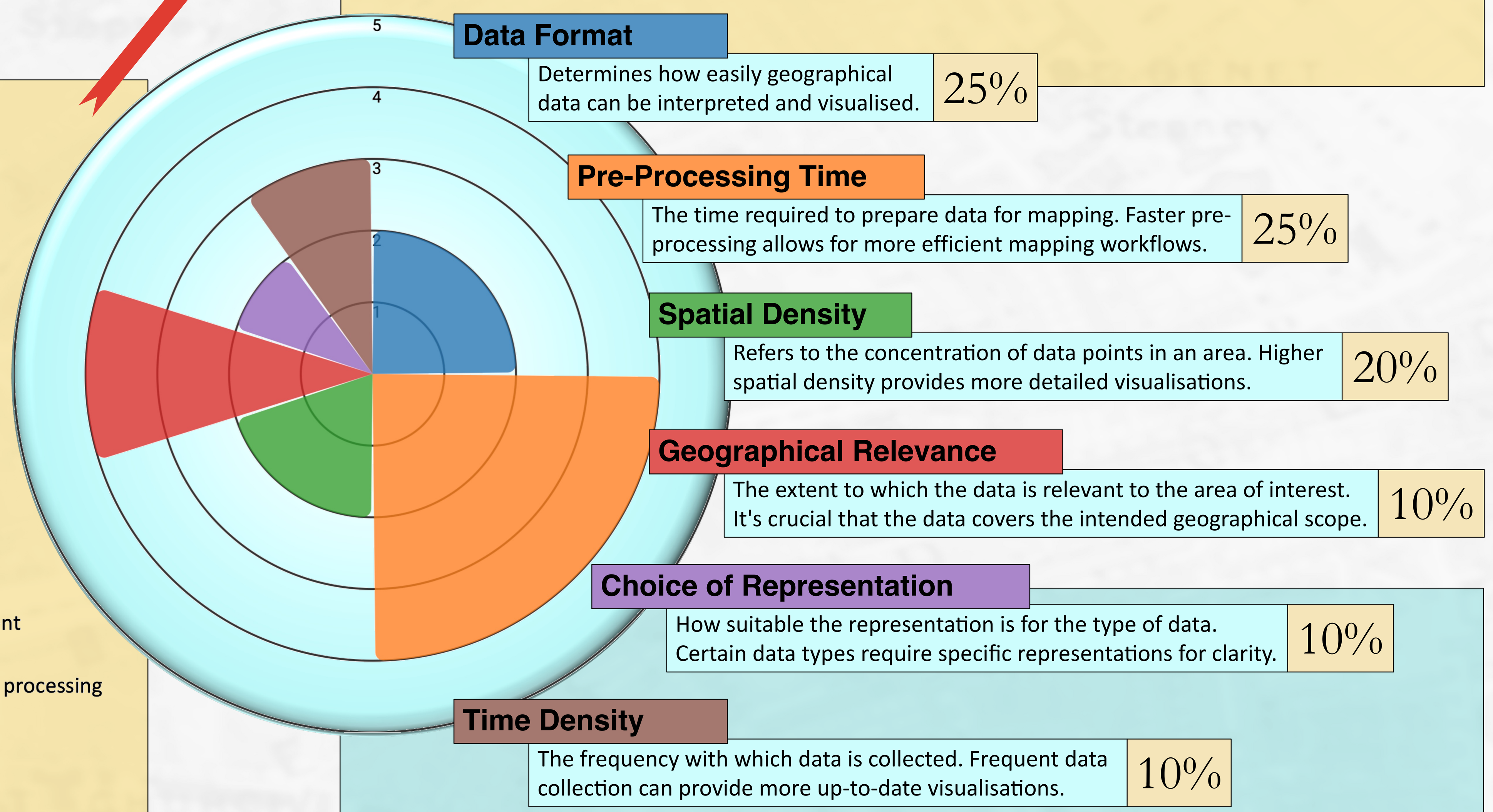
mappability



SOLUTION

- The **m(Appability) Rating System** evaluates datasets based on a weighted, six-factor scale.
- Each factor is graded on a **scale** from 0 to 5, allowing for a clear assessment of its suitability for mapping:

  0 indicates the non-existent suitability
  5 Indicates optimal suitability
- The cumulative score for a dataset is calculated using a **weighted mean**, factoring in each criterion's grade and its respective weighting.
- This concise grading system equips users with a structured approach to evaluate the aptness of datasets for mapping, promoting more **informed** decisions and **efficient** workflows.



- This study has developed a **multi-factor rating system** to assess mapping suitability.
- This should **reduce** the time that researchers spend on tedious data pre-processing, allowing them to focus more on actual data science and extracting insights.
- Thus, the expectation is to empower researchers to better **select and understand** the data that suits their needs.

1. British Government, 2023; UN Geospatial Strategy
2. Wang et al., 2015; A comparative study of landslide susceptibility maps using logistic regression, frequency ratio, decision tree, weights of evidence and artificial neural network. Geosciences Journal, 2015.
3. Soliman et al., 2022; Assessment of implementing land use/land cover LULC 2020-ESRI global maps in 2D flood modeling application. Water, 2022.
4. Tretiakov & Hunter, 2021; User Experiences of the NZ COVID Tracer App in New Zealand: Thematic Analysis of Interviews. JMIR mHealth and uHealth, 2021.
5. Neis & Zipf, 2012; Analyzing the Contributor Activity Of A Volunteered Geographic Information Project — The Case Of OpenStreetMap. International Journal of Geo-Information, 2012.
6. Haklay & Weber, 2008; OpenStreetMap: User-generated Street Maps. IEEE Pervasive Computing, 2008.
7. Kitchin, 2013; The Real-time City? Big Data and Smart Urbanism. GeoJournal, 2013.
8. Núñez-Andrés et al., 2022; Spatial Data Infrastructure (SDI) For Inventory Rockfalls With Fragmentation Information. Natural Hazards, 2022.

LAIDLAW SCHOLARS