Finding Geospatial Data

INTRODUCTION

Based on our research in our first Data Discoverability project, around 75% of users turn to a search engine first when searching for geospatial data. The remaining 25% eventually turn to a search engine when their first attempts are unsuccessful. It is important to understand how users go about finding geospatial data so that we can ensure that our data gets 'found'. Who are they? Where do they go to find data (besides the obvious search engines)? What terms do they search for? These were the questions explored and the insights derived have laid the foundation for improving user experience.

WHAT WE DID

Search Engine Optimisation (SEO):
SEO focusses on optimising your website to rank higher in search engine results and attract users to your content. Geospatial data publishers and portals must apply SEO best practices to compete for users, especially where the data the publishers hold is perceived as being the most authoritative. For guidance, a ***Search engine optimisation (SEO) for data publishers: Best practice guide*** *(https://bit.ly/2vpH2fX)* was published.

Dataset Granularity:
Dataset granularity looks at search terms users type into search engines when looking for geospatial data. Analysing the data catalogues for the Geo6 highlighted that entries were listed at different levels of granularity. Ordnance Survey, for example, published their catalogue at a very detailed feature level, whereas HM Land Registry published at a product level. The terms used to describe the dataset need to align with what people are searching for in order for it to be found.

Geospatial Data Portals:
Data portals are web-based interfaces designed to help users find and access datasets. Significant time and cost savings for businesses, research institutes, organisations, charities and the general public could be achieved when searching for data, if the provision of data is improved through consistent and well-designed data portals.

KEY FINDINGS

The project team identified that metadata was key to making data more discoverable and had the greatest impact on user experience. Quality metadata meant improved chances of search engines finding and ranking a dataset highly and offered sufficient information to users to make decisions about whether the data is suitable for their purpose.

1. **Search Engine Optimisation**: Search is performed on metadata rather than the data itself. Good quality metadata will be rewarded by search engines. Using structured data such as Schema.org to describe page

contents will ensure webpages can be 'crawled' and indexed by search engines. Finally, applying standard SEO best practice to webpages will improve ranking and position in search engines and they will be discovered by more users.

2. **Granularity:** For all user archetypes and all scenarios, 'Chunky Middle' terms (see below) were used far more often than any other type. Chunky Middle terms target relevant and refined keywords that attract the greatest number of people with the correct intent whilst minimising the effort required to attain a top position on a search engine result page.

   To put Chunky Middle keywords into perspective, 'Fat Head' keywords typically consist of one or two words and cover a broad area (example, 'post offices'). 'Long Tail' keywords are on the other end of the spectrum providing very specific detail and usually contain more than 3 words (example, 'post offices that are open on Sundays in Knightsbridge). 'Chunky Middle' terms fall between the two (example, 'post offices in Knightsbridge').

3. **Geospatial Portals Landscape:** Users emphasised easy navigation of portals and the ability to visualise the data before carrying on with their user journey. Many portals overwhelm users with results making it difficult to identify the most appropriate ones for their purpose. A more useful results page would, for example, have authoritative, relevant and up-to-date datasets highlighted.
   Further findings indicate that quality metadata is considered important as it informs users as to whether a dataset is suitable for their purpose. Users also require that licence information is stated clearly and simply.

RECOMMENDATIONS

1. Get webpages indexed as soon as possible and improve their SEO to perform better in search engine results. It is recommended that webpages have Schema.org embedded, use sitemaps and ensure pages have their own URLs. With these in place it will be easier for search engines to find, crawl, navigate and index websites. When the webpage is indexed it will be listed in search results.
2. Further SEO techniques will improve the position of the webpage in search results. This is particularly important as around three-quarters of users never click beyond the first page of results. Improving position is therefore essential to being more discoverable.
3. Levels of granularity used in metadata titles and content should incorporate chunky middle terms and reflect real-world concepts. It should be written in plain English to reflect what users search for. Organisation name or acronym should be used in titles or be prominent in the abstract as this would boost brand recognition.
4. Develop SEO recommendations, design guides and data quality rating methodology for spatial data portals.
5. Develop symbiotic relationship between portal developers and metadata creators to improve the quality of metadata content and its presentation to users within a data portal.
6. Allocate unique and persistent global identifiers to all Geo6 datasets and third party datasets and publish externally. This should be considered in the context of the data publisher creating an ID and possibly also the creation of a related web ID. It should be implemented in a way to support a many-to-many relationship between datasets as there will be cases where one dataset is used in the creation or derivation of multiple other datasets and vice versus.

BENEFITS

1. Applying SEO best practices and better quality metadata will make geospatial data more discoverable.
2. Future readiness for increased focus on dataset search by search engine providers.
3. Data portals will be more consistent, interlinked, intuitive, user friendly, reliable and accessible.
4. Increase in confidence in relevance and quality of discovered data.

The Coal Authority   HM Land Registry   Valuation Office Agency   UK Hydrographic Office   British Geological Survey — Expert | Impartial | Innovative   Ordnance Survey