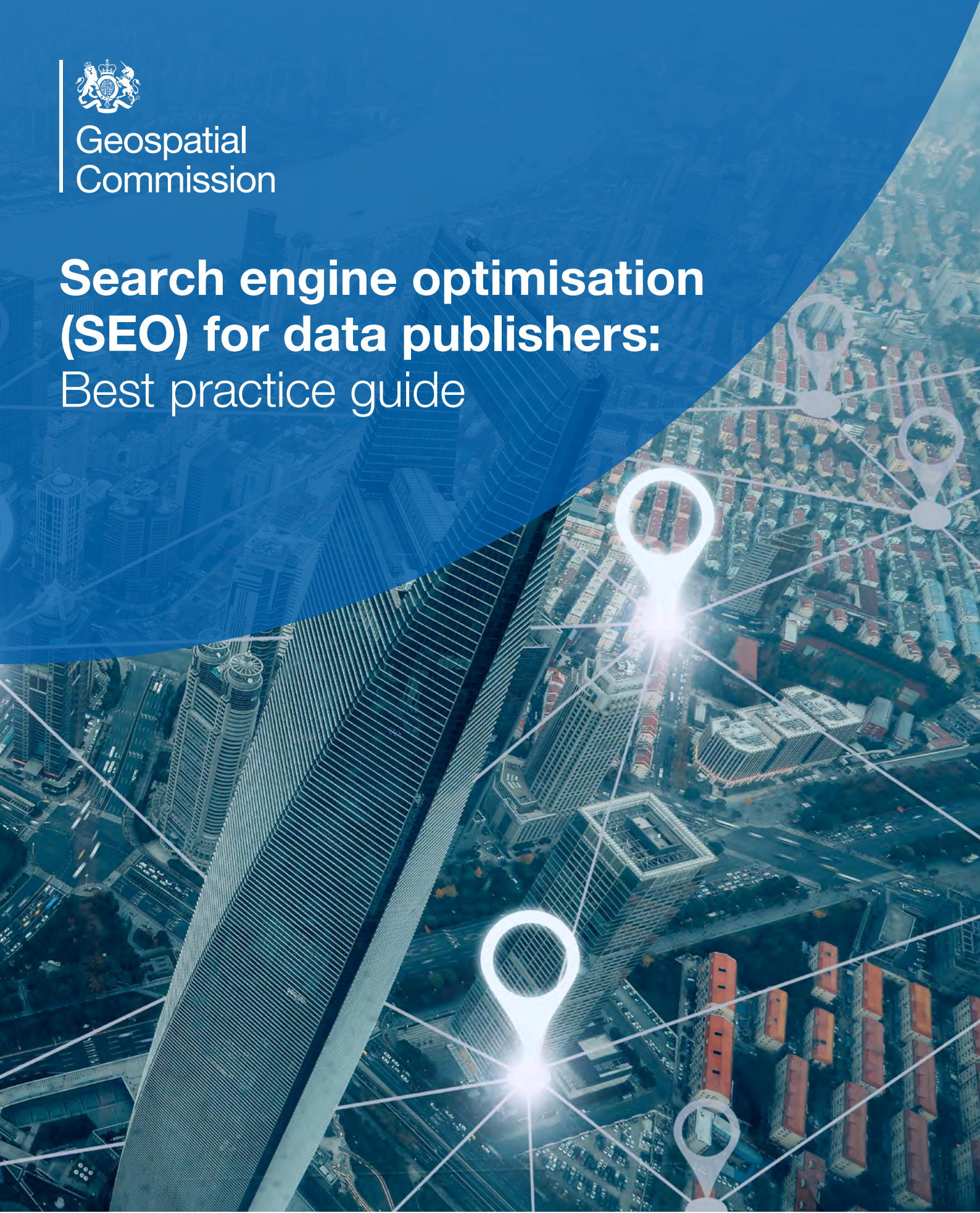




Geospatial  
Commission

# Search engine optimisation (SEO) for data publishers: Best practice guide





## Introduction

This best practice guide provides advice for data publishers wanting to improve the findability of their metadata (and therefore data) through search engines. The recommendations are based on research carried out by the Geo6<sup>1</sup> on behalf of the Geospatial Commission for the Data Discoverability project. While the project was focused on geospatial data, the principles set out below can be applied to any kind of data and data publisher.

Note that these recommendations are based exclusively on search engine optimisation (SEO) best practice and do not take into account other factors that you may need to consider (such as compliance with metadata standards, industry norms or organisation culture). You should assess the risks and benefits of each recommendation in your own context.

---

<sup>1</sup> The Geo6 consists of British Geological Survey, The Coal Authority, HM Land Registry, Ordnance Survey, UK Hydrographic Office and the Valuation Office Agency.

# Best practice guide

## 1. Fill out all metadata fields on data portals

### Why it's important

The more relevant information that exists about a webpage, the easier it is for search engines to understand what the page is about. This means that the search engine can rank the page more appropriately in search results. Where information such as a title or abstract is missing, the search engine has less information to work with and will be less likely to rank it highly.

### What it means

- Make sure all the metadata fields available to you contain accurate, relevant information
- No field should be left empty

## 2. Keep page titles no longer than 50-60 characters

### Why it's important

If they are longer than this, they will be cut short in search engine results pages. See examples below. This could mean that key information about your page is not shown. If users do not understand what your page is about from search results, they will be less likely to click through.

### What it means

- Keep titles short but make the most of the characters available by using keywords that help users find your data
- Front-load the title with the most important keywords to maximise their impact on SEO
- Carry out A/B tests and use tools such as Google Trends to identify which keywords have the biggest impact on the number of people finding and using your data

#### Page title: HM Land Registry Monthly Property Transaction Data

HM Land Registry Monthly Property Transaction ... - Data.gov.uk

<https://data.gov.uk> › [dataset](#) › [datafile](#) › [preview](#) ▼

HM Land Registry Monthly Property Transaction Data Number and types of transactions for value by all account customers. You're previewing the first 4 rows of ...

#### Page title: Bathymetric Survey - 2000-07-19 - Cromarty Deep Water Anchorage

Bathymetric Survey - 2000-07-19 - Cromarty Deep Water ...

<https://data.gov.uk> › [dataset](#) › [bathymetric-survey-2000-07-19-cromarty-d...](#) ▼

19 Jul 2000 - IPR Holder: Ministry of Defence; Purpose: Safety of navigation; IHO Sea: North Sea - 4; Survey Start: 2000-07-19; Survey End: 2000-07-25; ...

#### Page title: VOA Non domestic rating - addresses, floor areas, characteristics and attributes of properties

VOA Non domestic rating - addresses, floor areas ...

<https://data.gov.uk> › [dataset](#) › [voa-non-domestic-rating-addresses-floor-are...](#) ▼

19 Dec 2013 - VOA Non domestic rating - addresses, floor areas, characteristics and attributes of properties. Availability: Not released; Published by: ...

Examples of page titles cut short in search engine results pages

### 3. Optimise the content of abstracts

#### Why it's important

As outlined above, the more relevant information that exists about a webpage, the easier it is for search engines to understand the content and rank it appropriately. If you include confusing or inaccurate information it can negatively affect the page's position in search results. Search engines may interpret duplicated words or content within the page as an artificial attempt to improve the SEO, and could penalise the page as a result.

#### Summary

IPR Holder: Maritime and Coastguard Agency; Purpose: Safety of navigation; IHO Sea: North Sea - 4; Survey Start: 2014-06-14; Survey End: 2014-06-29; Primary Instrument Type: Echosounder - multibeam; Primary Navigation Type: DGPS;

#### Summary

A link to Ordnance Survey datasets

*Examples of metadata page content not optimised for search engines*

#### What it means

- Enhance all textual fields with suitable keywords, front-loading them with the most important ones
- Be as informative as possible but only include information relevant to the page and avoid repeating keywords
- Pay particular attention to the first 120 characters of any abstract as this will appear as a content preview in search results (alongside the title)
- Carry out A/B tests and use tools such as Google Trends to identify which keywords have the biggest impact on the number of people finding and using your data

### 4. Do not include lists of keywords

#### Why it's important

Search engines are very good at recognising artificial attempts to improve SEO and will penalise pages that list random keywords. To be effective, keywords must form part of the core content of a page and should be embedded in full sentences.

#### OS Commercial Addresses data

**Published by:** Ordnance Survey  
**Last updated:** 03 June 2019  
**Topic:** Not added  
**Licence:** Other Licence  
[View licence information](#)

#### Summary

Available for England, Scotland, Wales, N Ireland, Channel Islands and the Isle of Man as a point geometry in 5 Ordnance Survey (OS) products: AddressBase, AddressBase Plus, AddressBase Premium, AddressBase Plus Islands and AddressBase Premium Islands. The commercial nature of properties is determined by the Local Authority addressing custodian as part of their statutory role as Local Land Property Gazetteer (LLPG) custodians. This data is often used for geocoding, service delivery, address matching and verification, customer relationship management and statistical analysis. **Keywords:** Addresses, Unique Property Reference Numbers, UPRN, House Name, House Number, Postcode, Basic Land and Property Unit, BLPU, Land and Property Identifier, LPI, Commercial, Address, Gazetteer, LLPG

*Example of a metadata page that includes a list of keywords*

#### What it means

- Make sure all page content is written in natural, full sentences
- Never use non-sensical text or word strings

## 5. Check whether you can influence the URL

### Why it's important

Research shows that pages with long URLs are more likely to appear low down in Google search results. To avoid this, URLs should ideally be no longer than 50-60 characters.

### What it means

- You may not have any control over the length of a URL on your metadata pages, but should check whether you can influence any part of it – for example, it might be generated using the dataset title. If this is the case, you should aim to keep it short and optimise it using relevant keywords
- Avoid low-value words such as conjunctions and prepositions where possible.

data.gov.uk/dataset/2f33c8db-8626-4018-ba3c-36af6b7e0d93/voa-council-tax-addresses-characteristics-and-attributes-of-properties

Page title

*Example of URL generated using the dataset title*


## 6. Avoid special characters where these are not displayed correctly

### Why it's important


Search engines penalise webpages containing information that does not make sense. If you have published metadata on a portal that does not recognise your special characters (such as copyright or trademark symbols) this will have a negative impact on the page's ranking.

### What it means

- Check how special characters are displayed on websites where you publish your metadata
- If special characters are displayed incorrectly you should remove them
- If you still need to convey the sense of the character (for example, to convey copyright or trademarking) you should do this by explaining it within the text

**AddressBase<sup>®</sup>  Plus**

<b>Published by:</b>	Ordnance Survey
<b>Last updated:</b>	30 March 2019
<b>Topic:</b>	Not added
<b>Licence:</b>	Other Licence <a href="#">View licence information</a>

**Strategi<sup>®</sup>**

<b>Published by:</b>	Ordnance Survey
<b>Last updated:</b>	30 March 2019
<b>Topic:</b>	Not added
<b>Licence:</b>	Other Licence <a href="#">View licence information</a>

*Examples of unrecognised special characters*

## 7. Keep the same URL if your data is updated

### Why it's important

Pages gain authority with search engines over time. If you create a new page, it will take time for it to build up that authority with the search engine and it will usually be ranked lower in search results than a similar page that has been available for longer.

### What it means

- If you have metadata that is regularly updated (for example, if your data is refreshed every few months) you should reuse the same page (and URL) wherever possible
- Avoid creating a new page each time

## 8. Remove out-of-date pages

### Why it's important

The more data that is published, the harder it is for users to filter out the noise and find the source they really want. Search engines view recently updated pages more positively so pages that are out of date, or not actively managed, are less likely to be seen by users. If they are not seen by users they do not add value, but simply add to the long tail of results.

### What it means

- Design processes for managing the metadata you publish, ensuring it is regularly reviewed and removed if it is no longer useful or relevant to users
- If possible, set up a permanent redirect for any pages removed so users are directed to other relevant content and do not encounter broken links
- Where older metadata must be maintained (for example, where it refers to a snapshot in time or is part of a series), include a link to where users can find the most recent version of the data

## 9. Avoid duplication of metadata

### Why it's important

Duplication is another issue that makes it harder for users to find the source that best meets their needs. There are two main types of duplication:

1. similar or identical metadata available from multiple publishers
2. similar or identical metadata from the same publisher, available in more than one place

### 1. Similar metadata available from multiple publishers

## Search results

Q

**Filter by**

Publisher

Topic

Format

Open Government Licence (OGL) only

Apply filters

**13,014 results found** Best match ▼

[Ordnance Survey Code-Point Open](#)

**Published by:** Ordnance Survey  
**Last updated:** 30 November 2018

September 2017 OS Code-Point Open for Greater London and London boroughs are downloadable via the links below. OS Code-Point Open provides a National Grid coodinate for a point within each...

[Code-Point® Open](#)

**Published by:** Ordnance Survey  
**Last updated:** 30 March 2019

Keywords: Postcodes Code-Point Open is a dataset that contains postcode units, each of which have a precise geographical location. There are approximately 1.7 million postcode units in...

## 2. Identical metadata from the same publisher, available in more than one place:

### Public Health England (PHE) - British Geological Survey (BGS) Joint Radon Dataset for England and Wales

<b>Published by:</b>	British Geological Survey
<b>Last updated:</b>	04 June 2015
<b>Topic:</b>	Environment
<b>Licence:</b>	Other Licence <a href="#">View licence information</a>

#### Summary

The joint PHE-BGS digital radon potential dataset provides the current definitive map of radon Affected Areas in England and Wales. It will also allow an estimate to be made of the probability that an individual property in England and Wales is at or above the Action Level for radon. This information also provides an answer to one of the standard legal enquiries on house purchase in England and Wales, known as CON29 standard Enquiry of Local Authority; 3.13 Radon Gas: Location of the Property in a Radon Affected Area. The radon potential dataset will also provide information on the level of protection required for new buildings under as described in the latest Building Research Establishment guidance on radon protective measures for new buildings (BR 211 2007). This radon potential hazard information for England and Wales is based on Public Health England (PHE) indoor radon measurements and BGS digital geology information. This product was derived from DigMap50 V3.14 and PHE in-house radon measurement data. The indoor radon data is used with the agreement of the PHE. Confidentiality of measurement locations is maintained through data management practices. Access to the data is restricted. This dataset has been superseded by PHE-BGS Joint Radon Potential Dataset For Great Britain. Radon is a natural radioactive gas, which enters buildings from the ground. Exposure to high concentrations increases the risk of lung cancer. The Health Protection Agency recommends that radon levels should be reduced in homes where the annual average is at or above 200 becquerels per cubic metre (200 Bq m<sup>-3</sup>). This is termed the Action Level. The Health Protection Agency defines radon Affected Areas as those with 1% chance or more of a house having a radon concentration at or above the Action Level of 200 Bq m<sup>-3</sup>. The dataset was originally developed by BGS with the Health Protection Agency (HPA) which is now part of Public Health England.



## Public Health England (PHE) - British Geological Survey (BGS) Joint Radon Dataset for England and Wales

The joint PHE-BGS digital radon potential dataset provides the current definitive map of radon Affected Areas in England and Wales. It will also allow an estimate to be made of the probability that an individual property in England and Wales is at or above the Action Level for radon. This information also provides an answer to one of the standard legal enquiries on house purchase in England and Wales, known as CON29 standard Enquiry of Local Authority; 3.13 Radon Gas: Location of the Property in a Radon Affected Area. The radon potential dataset will also provide information on the level of protection required for new buildings under as described in the latest Building Research Establishment guidance on radon protective measures for new buildings (BR 211 2007). This radon potential hazard information for England and Wales is based on Public Health England (PHE) indoor radon measurements and BGS digital geology information. This product was derived from DigMap50 V3.14 and PHE in-house radon measurement data. The indoor radon data is used with the agreement of the PHE. Confidentiality of measurement locations is maintained through data management practices. Access to the data is restricted. This dataset has been superseded by PHE-BGS Joint Radon Potential Dataset For Great Britain. Radon is a natural radioactive gas, which enters buildings from the ground. Exposure to high concentrations increases the risk of lung cancer. The Health Protection Agency recommends that radon levels should be reduced in homes where the annual average is at or above 200 becquerels per cubic metre (200 Bq m<sup>-3</sup>). This is termed the Action Level. The Health Protection Agency defines radon Affected Areas as those with 1% chance or more of a house having a radon concentration at or above the Action Level of 200 Bq m<sup>-3</sup>. The dataset was originally developed by BGS with the Health Protection Agency (HPA) which is now part of Public Health England.

Source: [BGS Website](#)

### What it means

- Avoid unnecessarily duplicating metadata for data owned by other publishers
- Avoid duplicating your own metadata across multiple portals as one high quality metadata record will be ranked more highly by search engines than several similar records on different platforms

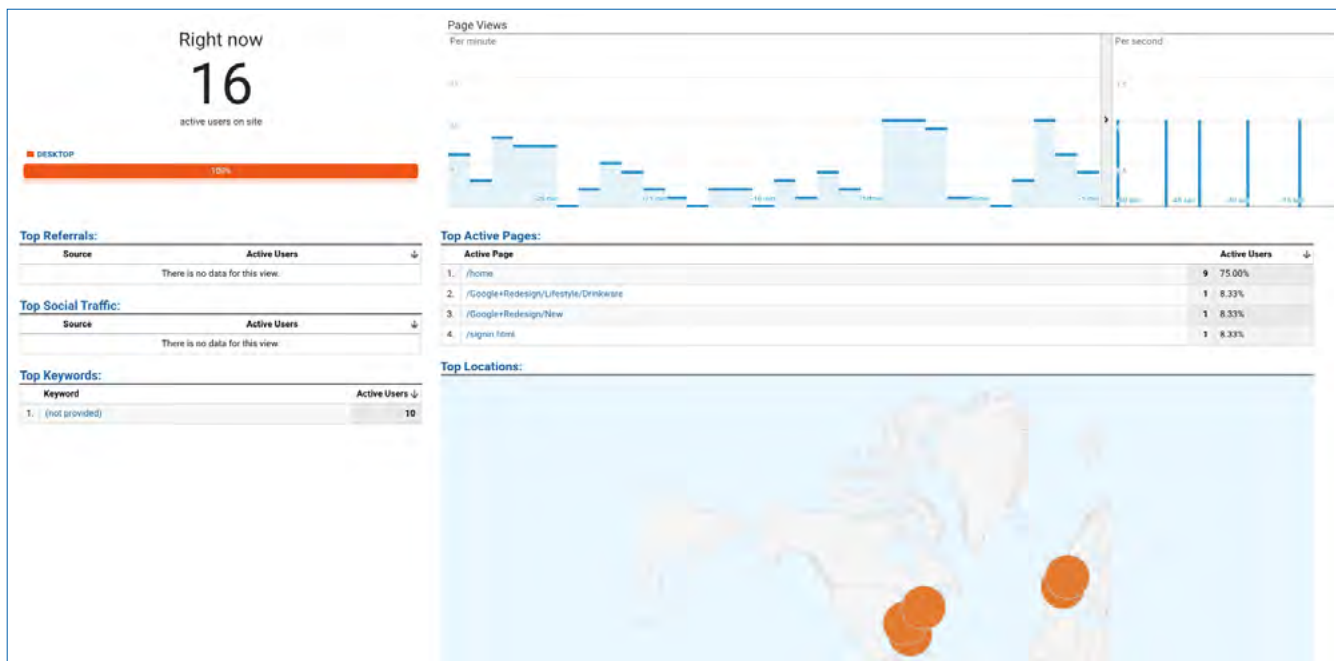
## 10. Use tools and tests to understand users

### Why it's important

While there are many common principles for SEO, the target audience for your organisation will be more nuanced than this guide. What works for one organisation may be less impactful for another so you must understand your users to know which levers to pull.

### What it means

- Use tools such as Google Analytics and Google Search Console (or equivalents) to understand the behaviour of your users and make evidence-based decisions about how to reach your target audience
- Use techniques such as A/B tests, user research interviews, user surveys and observations to inform your actions – card sorting exercises may be a good way to test whether users understand the terminology and keywords you are using



*Example of data available via Google Analytics*

## 11. Use tools to identify the best keywords for your subject

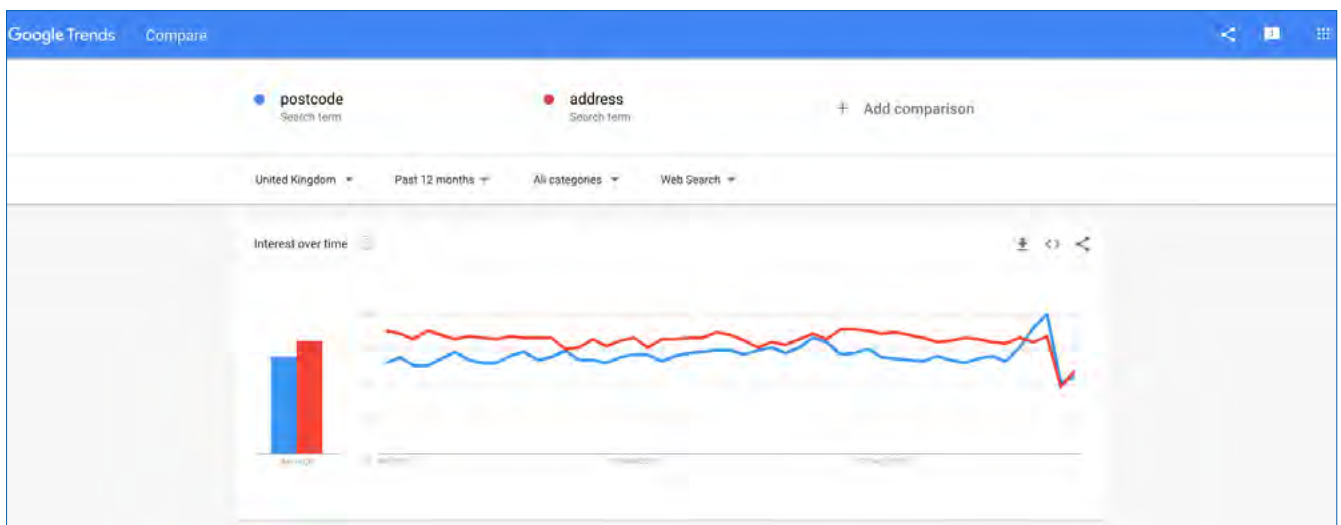
### Why it's important

To choose the best keywords for the SEO of your webpages you need to understand:

- which keywords are most popular among your target audience
- how competitive those keywords are (i.e. how many other web pages are also trying to rank highly using those words)

### What it means

- Use research tools such as Google Trends to check search volumes for different keywords
- If appropriate, use other tools (usually commercial) to identify the highest ranking synonyms for your keywords and understand how difficult it is to rank for your preferred terms



*Example of how Google Trends can help inform which keywords to use*

## 12. Apply these recommendations to all your pages

### Why it's important

Applying the SEO techniques above to one webpage should have a positive impact on its findability. However, it will be negatively impacted by pages on the same site that have a low SEO value. This is because search engines treat website domains as single entities.

### What it means

- Make sure all pages under your control are optimised for search engines using the techniques above
- If there are pages with poor SEO on the same website that are not owned by you, consider raising it with the organisation responsible or share this guide



