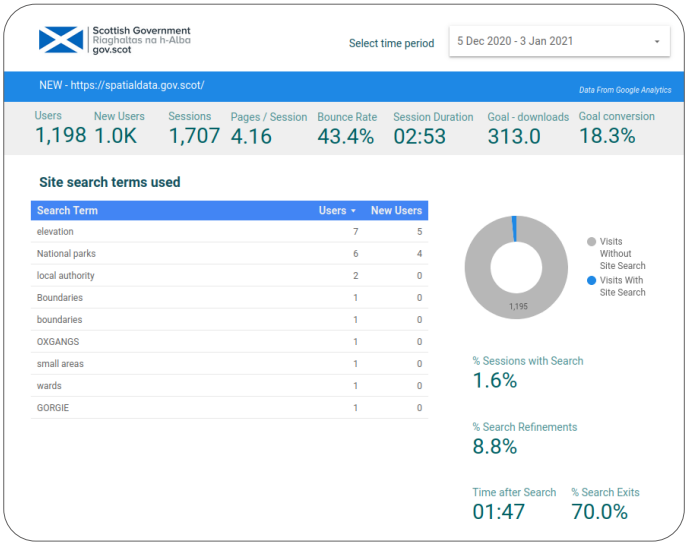
ASTUN TECHNOLOGY

We've recently completed a project on Data Discoverability for the Scottish Government, which meshes very nicely with a best practice guide on search engine optimisation that the Geospatial Commission have just published. There are three facets to the work we've done:

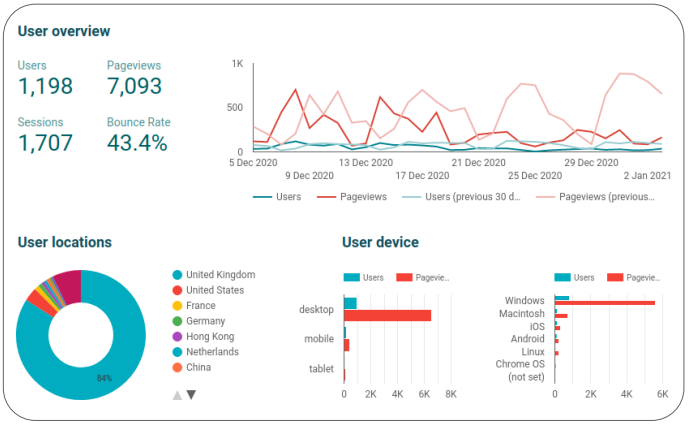
Search Engine Optimisation

The basic idea is that data needs to be discoverable via search engines, because that is where the majority of searches start, not really by visits to metadata portals. The portals are, however, the way in which search engines find and display that data, so they can be subject to the same sort of Search-Engine Optimisation techniques that any other website can use.

For the Scottish Government project we worked with experts from JNCC to improve the "SEO" of their metadata portal spatialdata.gov.scot. This uses Geonetwork Open Source, with a number of changes to make it easier to analyse how people are using the portal, including things like being able to report on the number of people clicking on WMS or WFS URLs, or other data downloads, as well as recording overall site visitor metrics. Consequently Scottish Government can find out useful information about which datasets are actually being used, and what search terms led people to the site.



Search terms used on spatialdata.gov.scot



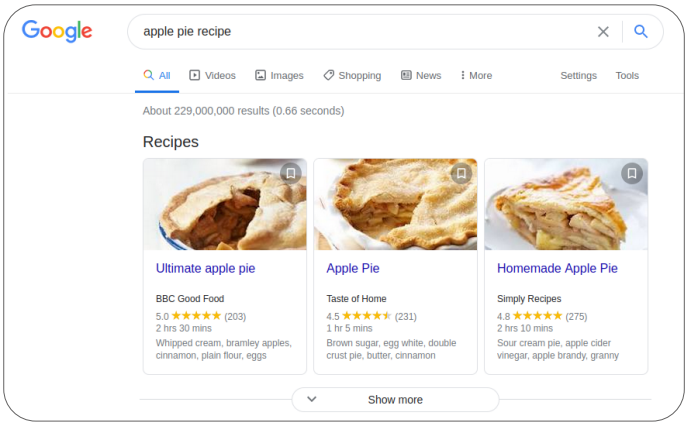
User stats on spatialdata.gov.scot

Structured Data

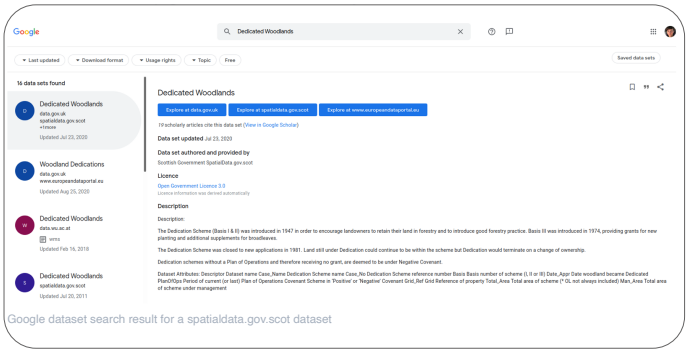
Furthermore, structured data is increasingly being used to inform the way that search results are displayed, and indeed how search engines identify what a page is about. That's how you get nicely formatted recipe cards if you search google for "apple pie recipe", for instance!

For any pages recognised as datasets, google will add them to the dataset search tool, which contains information about portals, licensing, citations and so on. This helps users differentiate canonical, or "source" datasets from other web pages where the dataset is merely mentioned.

We've incorporated this structured data into the Gemini 2.3 plugin for GeoNetwork, by mapping elements from the metadata standard to the schema.org dataset schema. This means that Gemini 2.3 records created with our plugin automatically include structured data when output from GeoNetwork.



Searching for “apple pie” in google



Google dataset search result for a spatialdata.gov.scot dataset

Data Quality

Finally, when search engines start to index your metadata portal, the quality of the metadata will directly impact on the ranking of your datasets in search results. In a throw-back to Goldilocks and the Three Bears, elements such as page titles or descriptions, both of which derive from the metadata, cannot be “too long” or “too short”. It’s also a bad idea to have metadata records with duplicated titles, as search engines tend to assume this is an attempt to game the system. However in a multi-agency portal such as the Scottish Spatial Data Infrastructure, several different organisations may legitimately have datasets of, for example, Conservation Areas, so we need naming conventions to ensure users know which dataset is which, and search engines don’t think we’re trying to cheat!

Here, we get into the realms of data custodians rather than portal maintainers, but we’ve cooked up some reports for the Scottish Government to show which records don’t meet the data quality standards for titles and descriptions, or perhaps don’t have all the right elements. We’ve also created reports showing records with duplicate titles. These are delivered via an interactive dashboard that directly links to the metadata catalog, and are kept automatically up to date.

Show all records where the title is greater than 50 characters in length, or the abstract is less than 150

settings

uuid	title	titlelength	abstractlength
4fa059bb-c89d-45e4-9e63-a48da593acc4	Dataset record for testing associated resources	47	13
af3ca0c3-e1c6-43a4-a709-d2eb724773e0	Demonstration and Research Marine Protected Areas (DR MPA)	58	589
92443592-7c35-4322-be5f-c4d011169bc1	CORPDATA.A_aa_Metatdata_testing_1	33	48
65f40d81-8179-47d8-b91f-13716c526575	Prestwick Airport Safeguarding Zones	36	60
8dc69503-23e1-42a2-8ce8-1f4ffef4f670	Polling Places	14	75
286fe523-7cf3-4ab1-8d4d-1d4d5d58d48f	Galloway Forest Park	20	92
53bc39c3-7c00-401b-9869-fc5f374c2802	South Ayrshire Health & Social Care Partnership Locality Boundaries	71	505
4e0573b2-3633-42e9-80ea-d734026c4f24	Vacant and Derelict Land	24	82

Interactive report showing records with long titles and short abstracts in spatialdata.gov.scot

We recently held a joint webinar with Scottish Government and the Improvement Service to introduce the new recommendations and changes in GeoNetwork to the users, enabling them to start creating high-quality search-engine-friendly metadata.

Coming Soon…

We’ll shortly have further GeoNetwork enhancements to talk about, so watch this space!

ASTUN TECHNOLOGY

Products

Astun offer a range of products based on open source technology. Our products can be grouped into three categories – iShare, Data Discovery and Astun Data Services.

Services

Astun Professional Services help you make the most of the iShare products, integrating with your existing systems, enabling data sharing and providing Open Source Support and Training.

Use Cases

Our Cloud and On-Premise solutions both let you put experts in place - providing services that deliver outstanding results in every sector.

astuntechnology.com