Scottish Science Advisory Council

Future Landscapes: Report on Geospatial Knowledge

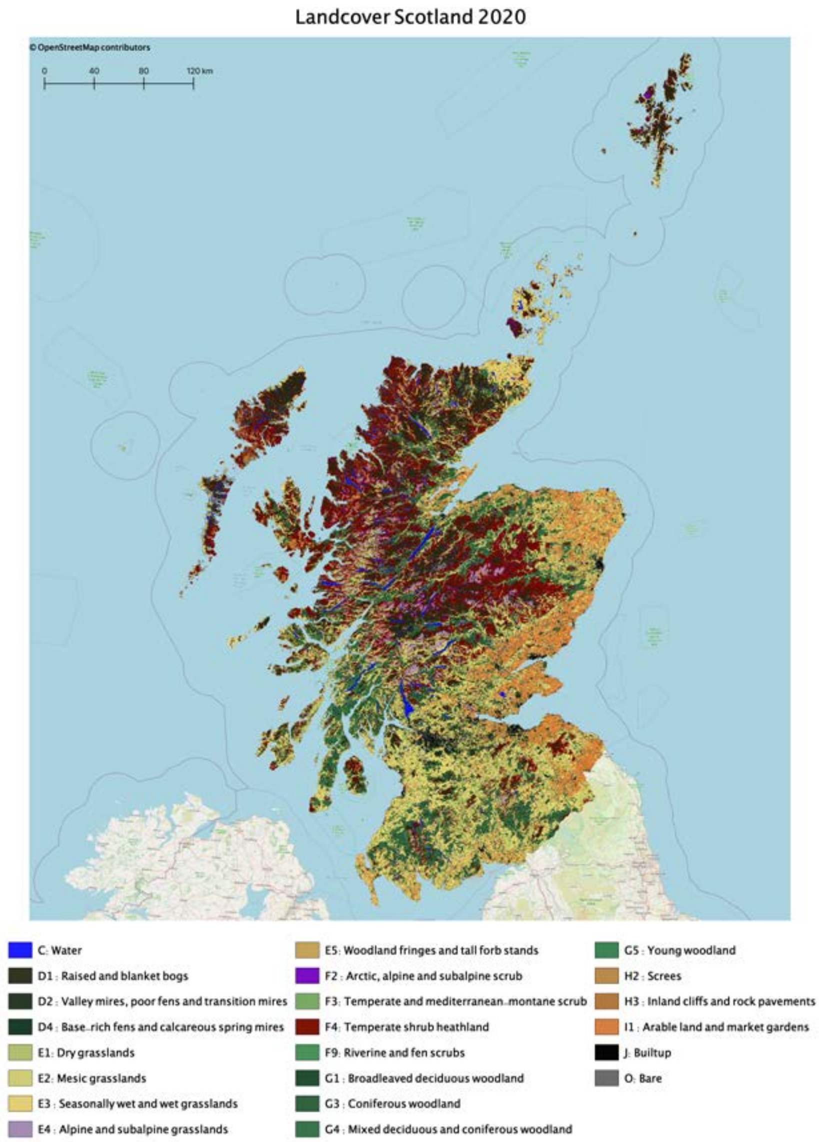
Geospatial data - information referenced by space and time - guides and informs human activity, from meeting friends to planning a spaceport. Though vast stores of geospatial data are added to daily, key to their optimal usage is that they are: Findable, Accessible, Interoperable, and Reusable - the so-called FAIR principles. This report: i) examines to what extent geospatial data available to the Scottish Government comply with these principles and ii) makes recommendations on how identified deficiencies can be remedied.

Contents

Foreword

The climate change emergency and the COVID-19 pandemic have brought centre-stage the need for spatially-referenced information (geospatial data), and the value and insight it can bring to managing global and local challenges. Satellite Earth observations give an impression of an increasingly all-seeing eye on natural ecosystems, while networks of remote observing platforms in the oceans, on land and in the atmosphere validate and challenge our numerical simulations of the geosphere, providing us with vast stores of information. However, seeing is not knowing: there exists an essential job to be done in the digital structuring of our geospatial knowledge base, ensuring wide dissemination and correct usage, and up-skilling of people to do this. This is not an inevitability, it requires sustained and supported effort as outlined in the evidence-based recommendations of this report.

In the Climate Emergency challenge of our time there is now consensus that greenhouse gas emission reductions must go hand-in-hand with carbon capture, and that the vast scale of the latter demands nature-based as well as tech-based solutions. As a precursor to nature-based solutions our government, regulators and agencies need an accurate, accessible geospatial knowledge inventory, with an embedded functional understanding to permit error-bound assessments of future actions and impacts.



Scotland's 2020 land cover as mapped by Space Intelligence.
Credit: Space Intelligence Ltd.

The Scottish Government is currently evaluating spatially-referenced data use for land-based policies, related activities and decision-making. A new Rural and Environment Science and Analytical Services Division (RESAS) Research position has been created to "identify sources of spatially referenced data relevant to land-based policy evolution in Scotland and identify gaps in data provision and spatial data which might be useful in the future". By identifying the more generic geospatial data problems and providing specific recommendations, this report provides a framework for setting the evidence base underpinning three key areas of Scottish Government policy[1].

This report tacitly accepts the aspiration that artificial intelligence (AI) and digital twins will accelerate our use and understanding of the vast and ever-growing archives of geospatially referenced environmental data (e.g. "Digital technology and the planet", Royal Society, 2020)[2]. We begin this report with one such example of good practice - Space Intelligence, but our focus remains on how we make the transition from the present fragmented landscape of geospatial knowledge to a structured basis for unlocking the potential of AI and digital twins.

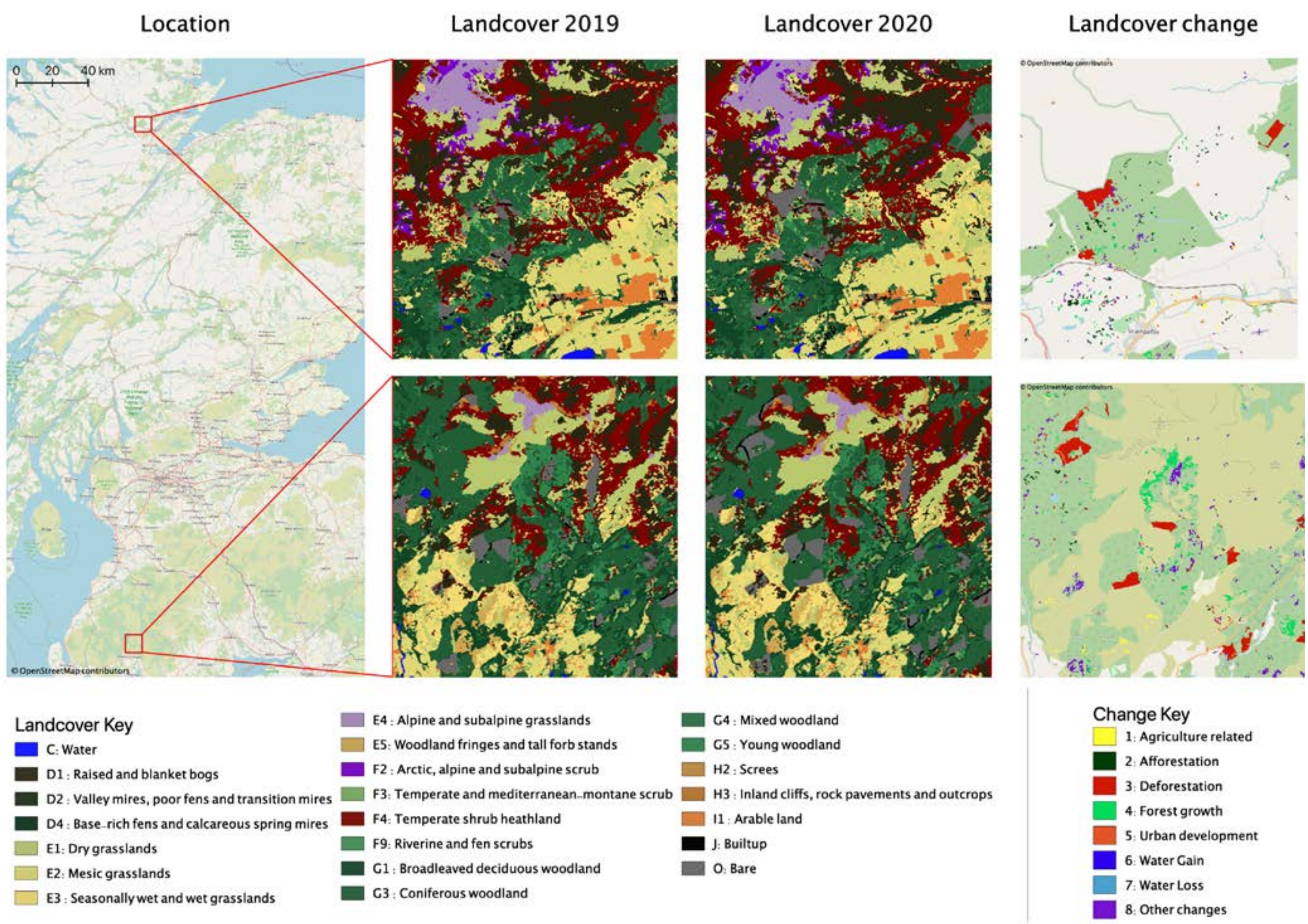
1 Land use - getting the best from our land: strategy 2021 to 2026; A changing nation: how Scotland will thrive in a digital world; and Transforming Places Together: digital strategy for planning

2 Digital Technology and the Planet

Good practice - Space Intelligence

In 2020 Scottish Enterprise, through a “Can Do Innovation Challenge Fund” initiative managed a programme to develop AI-enabled technology to address the climate change emergency. Space Intelligence, a satellite data analytics company based in Edinburgh, thus embarked on a journey working with NatureScot with the aim of developing machine learning techniques to interpret satellite images and identify different land cover types, providing land cover maps at very fine scales and measuring changes over time.

After a successful Phase I, which mapped land cover in the Cairngorms National Park, they have now moved on to Phase II, which entails classifying the whole of mainland Scotland into European Nature Information System (EUNIS) classes. Going forward, they aim to detect areas for restoration of peat, forest or wetlands and to improve the wealth of natural capital across Scotland. This is just one example of how geospatial information and new approaches can help inform our understanding of the environment, assess baseline changes, and inform future approaches. Through continued skills training, greater access to technology and more data availability, the project provides an illustrative example of excellent practice that takes geospatial work in Scotland to the next level.



Change in land cover from 2019 to 2020 at two sites in Scotland. Credit: Space Intelligence Ltd.

Executive Summary

The purpose of this report is to provide an assessment of the bottlenecks to realising the potential for geospatial data to aid the Scottish Government's aspirations for sustainable economic growth in the context of the Global Climate Emergency and the COVID-19 pandemic recovery. With some limited exceptions, we found that geospatial data, though generally good quality, resides in a fragmented landscape, limiting its potential. Action is needed to build trustworthy, cross-sectoral, publicly available data infrastructures. Detailed recommendations are made within the report for stakeholders across all sectors, covering improving data quality, increasing data accessibility, and implementing data standards. A summary of these recommendations, together with the organisations which could be responsible for implementation, is found in Appendix 1. High-level recommendations include, but are not limited to the following:

1. Scotland needs a central data directory for all spatially-referenced data with links to existing cross-sector data portals: this is seen as government responsibility;
2. All government-funded projects which generate new geospatial data should be required to provide free public access to the new data;
3. Health professions (NHS) should improve the geospatial context and Geographic Information Systems GIS) compatibility of their data, and a skills-gaps analysis undertaken to identify the cause(s) of gaps in their data;
4. Government-held or funded data catalogues should provide clearer user guidance, and follow GEMINI/INSPIRE standards;
5. A Scotland-wide (public, private, academic) mapping exercise of skilled data analysts, data gaps, skill gaps, and data science training opportunities would help to identify gaps in priority training which should be commissioned via a Scottish Government agency. There is a need both to up-skill non-specialist data generators to improve their data offering, and for specialist data scientists to make data repositories more user-friendly;
6. A Scotland-wide geospatial framework should be established to combine expertise and datasets across policy areas, starting with the Environment, involving a consortium of cross-sectoral partners, allowing for expert input to policy decisions of a geospatial nature. ERAMMP, co-funded by the Welsh Government, is a prime illustrative example which could provide the basis for a similar group in Scotland [https://erammp.wales/en];
7. Long-term data collection projects should be given adequate priority by public funders, with in-built flexibility in the collection process, to sustain continued dynamic data collection.

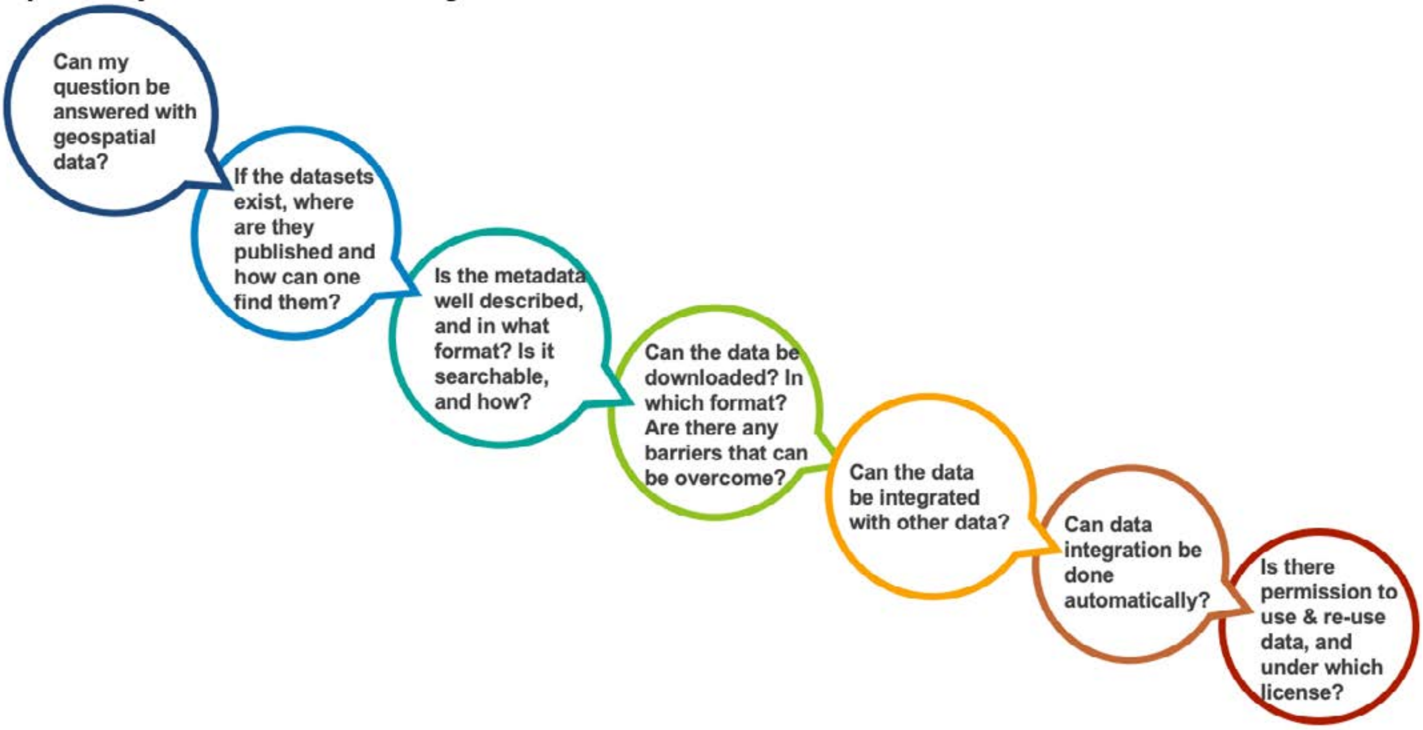
1. Background

Scottish Government (SG) is developing a Geospatial Strategy to identify Scotland's geospatial data priorities and the continued development of a spatial data infrastructure, which has fed into the UK Geospatial Commission strategy (Geospatial Commission 2020). At present, geospatial data around Scotland is mostly seen as good quality, and users generally use and trust the data. Current initiatives around geospatial data such as the Research Data Scotland's Metadata Catalogue, Geovation Scotland, SG's Network Accelerator, and the data maturity model from DataLab are also making further improvements to geospatial data quality. However, key areas of Scottish geospatial data need have been identified as priorities of the Scottish Geospatial strategy, such as the collaboration between public, private and academic sectors, but are not being presently addressed.

The aim of this report is to provide advice and recommendations to the geospatial community on how Scottish geospatial data can be used more effectively to contribute to SG's aspirations for sustainable economic growth and help dealing with the climate emergency. The specific objectives are: 1) to map out what data availability and infrastructure are already in place for selected examples; and 2) to give recommendations on how to improve data sharing/flow between public, private and academic bodies and more generally.

Through focused interviews it was investigated which geospatial datasets are available across Scotland, along with an assessment of whether they meet guidelines for good practice on data management, such as the FAIR principles (Wilkinson et al. 2016). It was identified where bottlenecks or limitations exist in data access, integration, analysis (**Figure 1a**), and data quality (**Figure 1b**) and recommendations to move past current limitations were provided.

a) Data process & usability



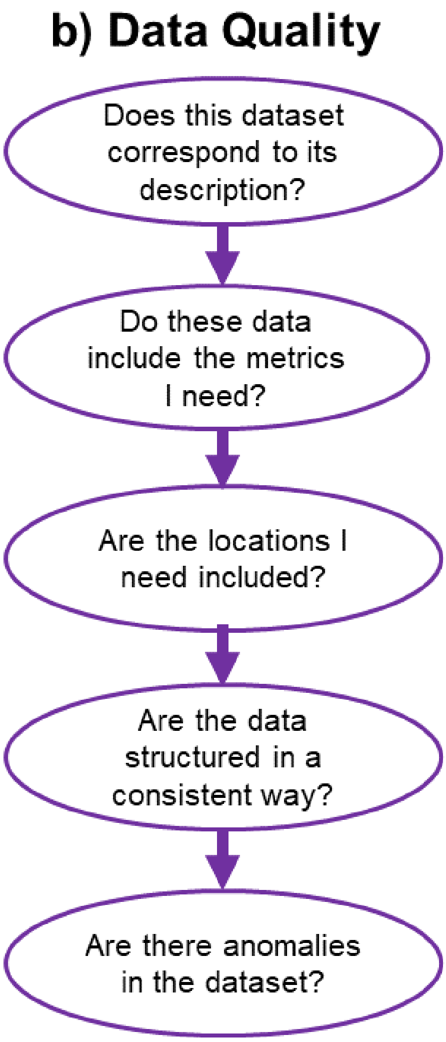


Figure 1. General guiding questions about a) the data access and usability process for geospatial data users to consider when trying to obtain and analyse geospatial data, and b) the data quality and suitability for the user's purposes.

2. Methods

To understand the data landscape across topical areas for the Scottish context, four case studies were defined, as described in Section 3. We identified nine experts or stakeholders working on each of these four case studies across the private, public and academic sectors (total = 36 topical interviews, **Figure 2b, c, e, f)**, and carried out a standardised structured interview protocol with 20 questions related to their particular experience working with geospatial data and their impressions of data sharing and accessibility across Scotland (see Appendix 2 for the questionnaire).

In addition, we carried out 8 general non-structured interviews with relevant stakeholders to help us gain better understanding of the data sharing and accessibility landscape across Scotland (**Figure 2d**). In total, 44 interviews were carried out (**Figure 2a**). We tried to achieve parity in the representation of academic, public and private sectors per case study. While this was not always possible and greatly reflected sectoral involvement per thematic area, we sought further contacts in other sectors to ensure that nine interviewees took part per case study.

The selection of interviewees was not an open consultation, and thus the interviewees were not self-selected. We identified interviewees following referrals from experts in the area and from previous interviewees. The majority of the identified recommendations in Section 5 were indicated by most respondents and supported by the evidence cited in this report and expert opinion. A further peer-review with one geospatial data expert took place to ensure the suitability of the suggested recommendations and the adequate representation of current limitations.

3. Case Studies

Given that geospatial data and analytics is a vast area, this report is focused on four specific case studies. By area, Scotland is approximately 70.7% natural land, 26.4% farmland, 2.1% built-on and 0.9% green-urban (Corine Land Cover, updated from European Environment Agency, 1995); and thus there is a natural environment weighting in the selected case studies.

Outlined below are four case studies with particular topical questions that guided the process of understanding data sharing and analysis practices. The objective was not to answer these hypothetical questions directly, but rather to understand data availability and the processes of data access, management and analytics relevant to these particular types of data. Insights from these exercises can be applicable to other questions in similar fields or data environments.

The four case studies and their hypothetical research questions were:

1. Wildfires (**Figure 2b**). How can we predict fire frequency and severity (including impacts on biodiversity) across Scotland particularly under a changing climate?
2. Public health and green spaces (**Figure 2c**). What is the effect of access to urban green spaces on public health across Scotland?
3. Coastal erosion (**Figure 2e**). Considerable knowledge outputs have been recently produced through the work of Dynamic Coast (Hanson et al. 2017). Therefore, this is a benchmark case study where responses to topical questions are not hypothetical. Consequently, our approach to this topic was to ask different questions to interviewees depending on their particular expertise, including 'What are the expected effects of changing climate on coastal erosion intensities and magnitudes?' or 'Which archaeological sites are most at risk of collapsing due to coastal erosion?'.
4. Ash dieback (**Figure 2f**). Which areas in Scotland have been most impacted by ash dieback and why?

Figure 2. Number of interviews carried out **a)** in total, **b)** on wildfires, **c)** on green spaces and public health, **d)** for general interviews, **e)** on coastal erosion, and **f)** on ash dieback

4. Limitations

There were multiple limitations to working effectively with geospatial data that applied to all four case studies in terms of data sharing, data quality and working with geospatial data. Additionally, case-specific limitations were identified which could potentially occur on similarly related topics and data environments.

General limitations

4.1. Geospatial data sharing

4.1.1. Data users are very often not aware of which data are available for a given topical area, and who holds particular data, especially in the public and private sectors. Data access often comes down to personal connections and experience, with proximity to data sources increasing data accessibility. Data inaccessibility without personal connections hinders data access to early career researchers, those new to a particular thematic area or those less connected to a topical network.

4.1.2. Generally, a difficult relationship of data sharing exists between different sectors (i.e., academic, public and private), with no straight-forward connections and no basic data sharing agreements implemented between sectors.

4.1.3. The process of arranging data sharing through contracts and licences is often extremely time-consuming, often delaying the start of projects and/or the publication of final outputs. Similarly, accessing data through request submissions can be a long and bureaucratic process.

4.1.4. Any form of personal data needs to comply with the UK General Data Protection Regulation (GDPR) legislation. Thus, obtaining geospatial data with a personal element can be a long and sometimes challenging process, occasionally resulting in users abandoning their research ideas.

4.1.5. Data collectors seem to be generally protective of their data. This might be due to commercial reasons and fear of being judged on data quality, among other reasons.

4.1.6. Some data are only available by paying expensive fees, and often users need to pay for whole databases or spatial tiles when they only need part of them.

4.1.7. Public institutions find data sharing challenging because of high IT security systems. The use of conventional data sharing platforms is usually not allowed, but there is no appropriate platform in place allowing for data transfer.

4.2. Geospatial data quality

4.2.1. Geospatial data quality is highly variable between sources (e.g., private versus academic sector collected data), but also within similar types of sources (e.g., different local authorities) when no standardised data collection protocol exists.

4.2.2. Available data often can only be accessed at a resolution that does not suit the analysis task at hand. Different datasets in a common analysis should share similar resolution to ensure compatibility.

4.2.3. Generally, data are collected assuming that end-users will work with a particular software and in specific data formats. Thus, those working with different software will have to spend considerable time and effort adapting data into a format that is compatible with their particular software.

4.2.4. Writing metadata is one of the main barriers to making data publicly available. Particularly, following metadata standards with long and complex documentation poses challenges to understanding and to translation into specific actions. Certain standards seem overly complicated to the average user so they end up not being implemented, and there are generally no implementation procedures ensuring that data collectors follow particular guidelines when writing metadata. When data analysts are not familiar with data principles, additional staff members specialized in data and metadata standards would be needed, but this is not always possible.

4.2.5. Particular research institutions or funders specify that data produced as part of e.g., UKRI-funded projects follow specific formatting guidelines that require great time investment even for those well versed on geospatial data. Some standards end up being a barrier to data sharing because of their complexity, given that implementing data standards requires funded time, which is not always included in a project, given the ever-changing nature of funding in academia.

4.2.6. Sometimes data include inappropriate metadata or no metadata at all. Thus, data users often need to contact the original data collectors seeking clarification on the available data, even after having purchased these data. This causes further delays and time losses for both the data user and collector which wouldn't occur if appropriate metadata were originally included.

4.2.7. There are limitations to the available time periods of geospatial data. Often, particular types of data are only available for recent years, while baseline or historical data are inaccessible or non-existent, thus requiring extrapolation. Similarly, time series data are often lacking even for dynamic processes.

4.3. Working with geospatial data

4.3.1. Computing power, hardware, access to software, staff numbers, and availability of skills are consistent issues across the board. Likewise, time and mental space is needed for particular analyses, which is not always accessible to everyone due to the fast-paced nature of many jobs.

4.3.2. Working with geospatial data can be greatly improved if data producers follow data guidelines and write descriptive metadata. However, there is an irreplaceable element of system knowledge that is needed to fully understand and correctly interpret a given dataset.

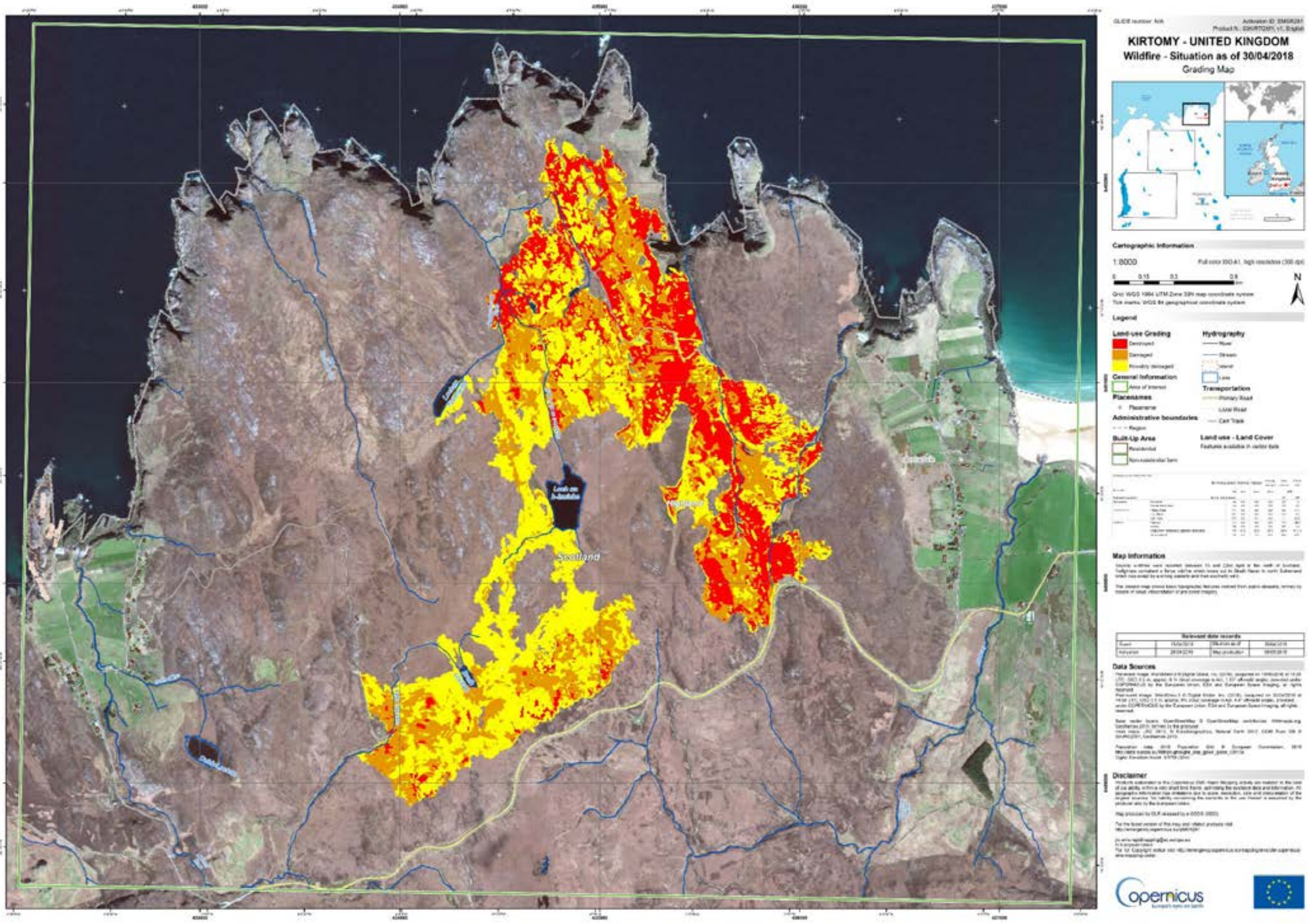
4.3.3. It is sometimes easier to collect data than to extract data from certain databases for certain data users. This is particularly true when available data have been collected following different methodologies, thus not being fit for the data user's purposes.

4.3.4. Projects that work with multidisciplinary teams tend to speak different languages when it comes to data structure and sharing. Different areas of expertise employ distinct approaches and standards when working with geospatial data.

4.3.5. There is a time lag (up to 6 months) when working with large-scale data (e.g., Earth Observation and remote sensing data) between the time when data are collected and the time when end-users can access and analyse the data.

4.3.6. Integrating data from different sources, formats and methods remains one of the biggest hurdles when working with geospatial data.

4.3.7. The increasing amount of citizen science data is very welcome in different topical areas and contributes to the expansion of the available geospatial data pool. However, volunteer-provided data is of variable quality and requires checks prior to making data available.



Map showcasing the location and severity of 2018 wildfires in Sutherland using Copernicus data.
Credit: NatureScot.

Case study-specific limitations

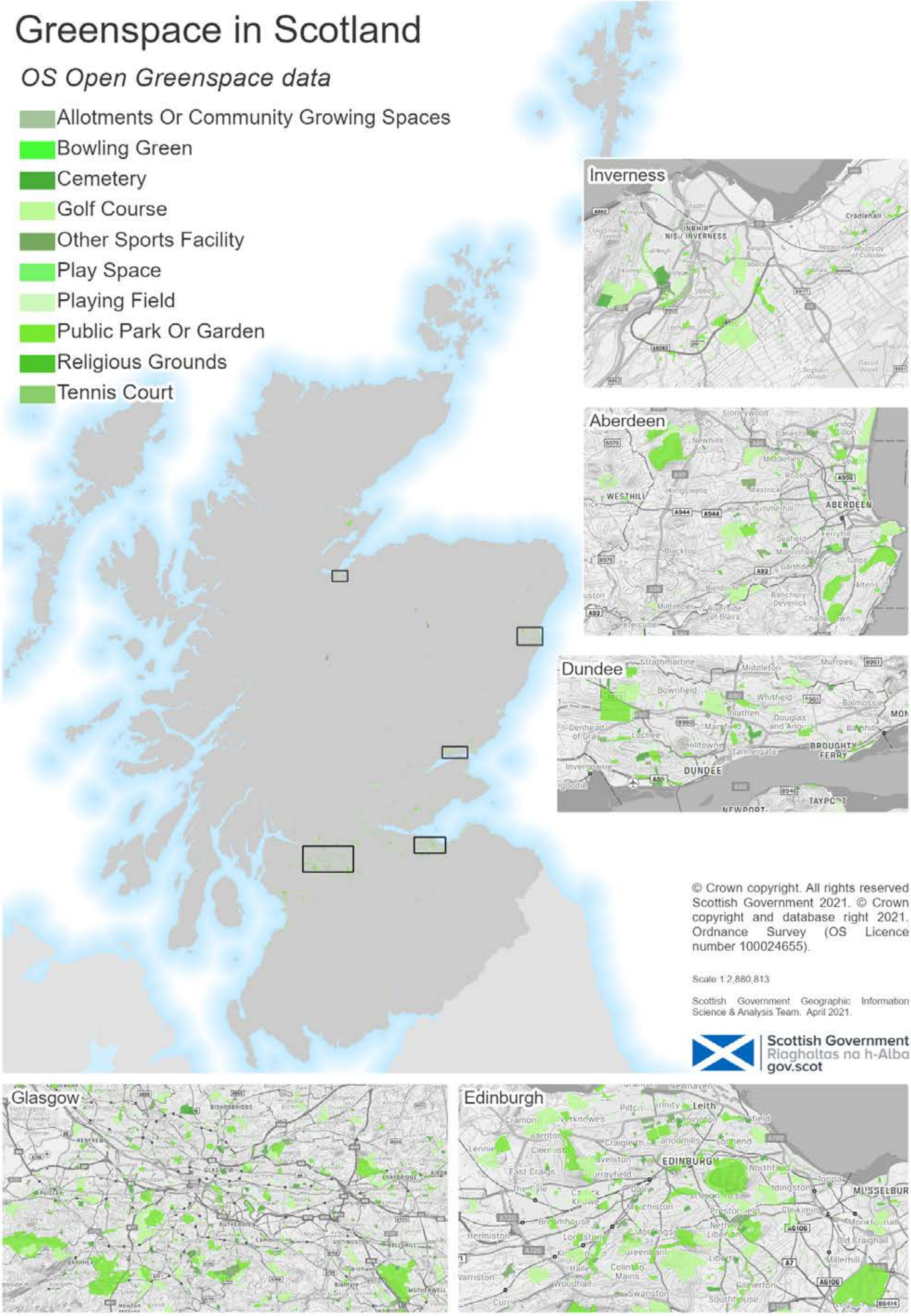
Wildfires

- Many forest wildfires happen at the same time of the year, creating an issue of internal capacity and skilled staff to respond to the workload.
- It remains challenging to make predictions of wildfires under changing climatic conditions since wildfires are not only dependent on climate but also on land use and human behaviour change, which are difficult to predict.
- The location of wildfires is sometimes inaccurate in collected data, given that the reported coordinates usually belong to where fire brigades trucks were parked and not the actual location. Equally, perimeter and area data are usually not available, and hectares reported sometimes don't match the full wildfire extent.
- Wildfire reports frequently contain lengthy word descriptions of the fires but no geospatial data or maps.
- Contextual data about the surrounding landscape and environmental conditions is important to understand wildfires, but these are often lacking.
- When working with optical remote sensing data, fires smaller than 30 hectares are usually not captured, which leads to data users missing a large proportion of fires.
- Some raw data on fires (not aggregated) are not available because of confidentiality issues. Data users need access to relevant contacts to obtain these kinds of data.

- For certain types of geospatial data, most wildfire data is only available from 2009. Thus, extrapolation is needed to understand past trends, with the associated levels of uncertainty that this process entails.
- The time series element to wildfires is lacking in many cases, with most data being static and the absence of change-over-time data.
- Muirburn is usually reported under ‘controlled wildfires’, so there are no specific data sources for this type of burning.
- There are differences in compatibility between new and older datasets and between different methods of land classification.
- Storage capacity is an important issue because of the large size of wildfire data.
- The nature of the topic requires collaboration with different stakeholders, which can mean sometimes data are not presented in the most accessible way to a variety of audiences.

Public health and green spaces

- Public health data are generally not available at a fine-scale resolution due to GDPR issues. Health data is often aggregated and not always appropriate for analyses with green space data, which is available at finer resolutions.
- Some green space maps are not updated due to the evolving nature of urban spaces.
- A lot of personal data (e.g., alcohol consumption) needs to be bought from the private sector and is not easily found or accessible, even when funds are available to acquire the data.
- The health data application process is lengthy and requires express authorization. This bottleneck was particularly evident during the COVID-19 pandemic.
- Some basic-level health data such as GP catchment areas are not available to users yet.
- Large amounts of health data are only available in spreadsheet format and/or lacking any geospatial element. A lot of users’ time is usually invested in converting these to geospatial data.
- Most NHS data are very fragmented as there is no overall geospatial department. Users usually need contacts to identify who to reach out to when trying to obtain public health data.



Different types of green spaces mapped across cities in Scotland. Credit: Scottish Government.

Coastal erosion

- Acquiring historical data on coastlines remains challenging. These data are either not available, or have an expensive paywall attached and/or a licence for one use only.
- Volunteer data from citizen science is growing and public involvement is increasing, but the quality of the data is widely variable.
- Due to the extensive number of data sources needed for certain research questions, thorough checks on the quality of the data are needed prior to performing analyses.
- Working at large scales, such as the entire coast of Scotland, is a complex process due to issues derived from data gaps and modelling power.
- Finding useful data at the right resolution and that is up to date enough, particularly in the fast-moving process of coastal erosion, remains complicated.
- Many available datasets are over 10 years old, and thus do not represent the current status of Scottish coastlines.
- Large data gaps exist on parts of Scotland's coast. Additionally, different sources provide data at different resolutions and formats, possibly creating further gaps.
- Finding bathymetry data to incorporate on coastal erosion models is problematic.

Ash dieback

- The location of ash trees is not regularly surveyed, so it remains difficult to establish a baseline of where the ash trees are. Thus, identifying where the disease might appear is even more challenging.
- GDPR is often an issue since some trees are in private gardens or land. This means that locations can't always be fully displayed in datasets.
- Further data collection is required to assess the full extent of ash dieback, since already collected data is not comprehensive. However, there is a substantial lack of resources and personnel for on-the-ground surveys, and repeat surveys with different spatial and temporal coverage are needed.
- While the plant disease community is relatively small and well-networked, data are fragmented and there is a lack of knowledge on the work of different stakeholders. Further collaborative work is needed.
- Record collection is heavily reliant on citizen science, but more consistent surveying from tree health officers is needed.

5. Recommendations

Based on feedback collected from interviewees, we outline recommendations aimed at different sectors and Scottish Government in order to address the aforementioned challenges posed by geospatial data. Many of these recommendations are simple steps that institutions can take to improve data quality and the processes of data sharing and working with geospatial data. Other recommendations involve an element of cultural change and will require long-term progress to achieve behavioural shifts. Together, these recommendations aim at improving data accessibility and to streamline the process of working with geospatial data, ultimately facilitating higher-level work and analyses with geospatial data across the geospatial sector. A summary of all the proposed recommendations, together with their indicative timeframes and responsibilities for implementation, can be found in table format in Appendix 1.

5.1. Geospatial data sharing

5.1.1. **A central directory listing all existing data portals** would vastly improve data accessibility across Scotland. Since a variety of data are hosted in websites belonging to different institutions, and physically hosting and updating all data to a single portal would take vast amounts of time and resources, we suggest a central directory which stores links to other host websites where the data are held. It is recommended that the many governmental data catalogues (i.e., Remote Sensing Data Portal, Scottish Data Portal, Scottish Spatial Data Infrastructure, Metadata Catalogue, Spatial Hub by Improvement Service, etc.) are all listed in a clear and visible manner, ensuring that portals can harvest metadata between them and are optimized for search engines. This directory will provide a centralized entry-point to different hubs, greater (inter-) connectivity of the data portals, and the reuse of public sector data. All hubs included in the directory should meet a certain set of criteria, thus increasing the overall quality of accessible data.

5.1.2. The **Public Sector Data Catalogue** is a commitment set out in the Scottish Digital Strategy (https://www.gov.scot/policies/digital/) and the remit could be widened to incorporate private and academic data or at the very least provide guidelines. This would increase the visibility of data generated from all three sectors and help enable greater potential for the integration of public, private and academic data being used together.

5.1.3. We recommend that all data portals (and the central directory mentioned in 5.1.1) **continuously assess user requirements** with different levels of expertise to identify ways to improve their portals. This could be done through feedback forms and questionnaires that users can fill in after downloading data, or through wider stakeholder consultations if needed. Best practice guidelines on engaging with data users are outlined by the UK Geospatial Commission (UK Government 2020a). Currently, some SG portals are not catalogued in a way that can be easily accessed. Suggestions for improvement from interviewees include following the format of the European Space Agency, which is map-driven and allows for focus on topical areas, while also searching for keywords on an issue basis (not departmental). For example, users should be able to access peatland data by using “peat” rather than “restoration” or other departmental keywords.

5.1.4. All data catalogues should have more **guidance and instructions** so that users with different expertise can use the portals more effectively to access their required data. This also applies to SG data catalogues such as, but not limited to, Scottish Spatial Data Infrastructure (SSDI).

5.1.5. **Data sharing agreements** can be made more straight-forward by creating a template contract that can be adapted depending on the needs of the two parties. Standardisation of various types of data agreements is currently underway through the Open Government Licence (The National Archives, 2021). Similarly, standardised licence templates for geospatial datasets would be helpful in streamlining data access, and the type of licence assigned to each dataset should be displayed prominently so the user is certain about the type of permissions they have depending on their sector. Equally, a clear conversation in the early stages of data sharing where all parties involved express their needs can speed the process of data sharing and improve the workflow. This is particularly important for any projects that involve personal data, which should involve data protection teams from the start. It should also be established that projects won’t begin until the data agreement has been finalized.

5.1.6. SG has a role to play in incentivising **making data publicly available**, particularly from public institutions and publicly owned datasets. There is a growing acceptance from the geospatial community towards making data publicly available, so there should be capitalization on this momentum at the current time (Coetzee et al. 2020). Making data available will benefit data collectors by reducing the amount of time spent on responding to requests, obtaining feedback on data issues, improving visibility and transparency, and feeding back into the physical and technological world, ultimately contributing to better decision making (Arzberger et al. 2004). This focus on reuse of public sector data has been noted in the Scottish Digital Strategy Consultation, which sets out the development of a Digital Transformation Framework. The Framework will set out practical, achievable and actionable steps for organisations, helping them improve data maturity and increase data sharing and accessibility.

5.1.7. While private companies are increasingly releasing publicly available data or arranging Memoranda of Understanding with different institutions, it is usually not possible to make whole commercial catalogues available if data offer is one of the main financial sources for the private sector. However, commercial companies could still **provide metadata and data extent** to improve visibility and for potential data users to be able to purchase their data.

5.1.8. Data producers that charge for data could implement in their business plans different **options of data acquisition**. For example, smaller tiles or datasets at different resolutions could increase sales to fit the needs of particular clients.

5.1.9. Institutions should research which data they need at the beginning of a project and clearly **budget in funds to acquire data** if these are behind a paywall.

5.1.10. A push towards **health professionals** making their data available in a geospatial format is needed and very much encouraged.

Coastal erosion at Skara Brae: 0.4 m per year retreat. Credit: Alistair Rennie.

5.2. Geospatial data quality

5.2.1. Geospatial data should be made available at the **best possible resolutions**. Data collection is usually carried out with one particular purpose in mind, which sometimes does not allow for data re-use. This can be overcome by having organisations better understand user needs and ensuring that data is fit for purpose. This means researching and understanding users' needs, and making data available at the best possible resolutions so users can process and adapt the data according to their needs. Another issue is that good quality data might not be what the end-user was looking for, and this is usually noticed after the data has been downloaded and/or paid for. Essential to overcome this is to write appropriate and useful metadata, which should be available even if the data are not, so the user can understand the actual data content, resolution and format and decide if they wish to download it.

5.2.2. **End-user requirements** around quality, formatting, type of data and resolution should be taken into consideration as part of any future data collection exercise. Engaging end-users from the beginning will help understand who will use data and for which purpose, and ensure co-production. This includes making data available in common open formats that most users will be able to work with, following FAIR principles and INSPIRE regulations. Guidelines on considering end-user requirements have been set in The Government Data Quality Framework (UK Government 2020b). This recommendation, among others, is about cultural change, involving the data collector to think about the whole lifecycle of the data, and the possible uses that data can have after collection. This will affect the way data are collected in the first place and extend the longevity and variety of uses of data.

5.2.3. While particular **data standards for data and metadata** are only implemented at certain institutions, all interviewees agree that data standards improve their experience of data use, workflow and productivity, and provide confidence on the reliability of the dataset, particularly when working with data they are not familiar with. While writing metadata can take a substantial amount of time and data guidelines can be cumbersome, metadata remains essential for discoverability and access, and will provide benefits for both the data collector and data users in the long term (Foulonneau and Riley 2008). Institutions should follow data standards which appropriately describe the data and which allow for data to be machine-harvested. Likewise, since data are dynamic, metadata should be too, and metadata should be continuously updated to avoid project setbacks and erroneous analyses. Good guidance on writing metadata has already been produced by the UK Geospatial Commission (UK Government 2020a).

5.2.4. There is a missing step between wanting to make data available and actually making data available because of the challenges of implementing data standards, including standard complexity, time and cost. Thus, providing specific suggestions and **best practices on data and metadata implementation** in an accessible format for the end user will help data producers make their data publicly available (IGGI 2002, IGGI 2005). We encourage SG and the SG GIS team to clearly communicate that FAIR-inspired standards should be adopted (see Appendix 2) and that the national standard that should be followed is GEMINI 3.0 (as transposed into national legislation to be compatible with the European INSPIRE legislation) (AGI 2018). Likewise, SG should provide guidance on following standards, in the format of templates or examples of best practices on filling metadata that institutions can use. Further guidance on what exactly should be included in different metadata fields will streamline the process across different sectors.

5.2.5. As for the **academic sector**, there are discrepancies between the different data standard requirements placed on researchers by funders, institutions and scientific journals. We recommend that research institutions and funders standardise their requirements to follow GEMINI/INSPIRE in order to ensure that the academic sector is less fragmented and more consistent and compatible with other sectors.

5.2.6. Data standards (e.g., GEMINI/INSPIRE) need to be clearly **specified into contracts** between commissioning bodies and sub-contractors.

5.2.7. Greater **training on data standards** and metadata would improve adhesion to data standards. Collaboration between data managers with specialist expertise on data standards and staff familiar with the data can facilitate improved data standard use and process automatization. There is a need to both upskill non-specialist data generators to improve their data offering, and for specialist data scientists to make data repositories more user friendly. Organisational efforts are needed to map out skilled data people, skill gaps, data standards, data gaps, national standards and making the data machine-readable. This is best undertaken through a Scotland-wide piece of work commissioned by Centres of Expertise or Scottish Government Geographic Information - Science and Analysis Team (GI-SAT).

5.2.8. Users would like to see **standardisation of Council level data**, and work is underway through the Improvement Service, the Local Government Digital Office and the new Planning Strategy (The Scottish Government, 2020). Continued progress on this issue will ensure better use and implementation of Council level geospatial data.

5.2.9. New mapping and surveying approaches with Earth Observation should be **validated with field observations** (Loew et al. 2017), with an understanding of the key questions being asked by policy frameworks.

5.3. Working with geospatial data

5.3.1. **Data expertise has to input into policy decisions**. Policies that will be adopted need to consider underlying evidence, and the way data are collected for particular purposes needs to be discussed with those that will be collecting and analysing data to support evidence-based policy-making (The Scottish Government 2018). This should be happening particularly at the start of policy-writing, rather than later in the process, and should also form a continuous dialogue. Specific policy areas which are data-heavy and require lots of data like climate or land use are particularly encouraged to define their data needs early. Equally, 'data people' are encouraged to speak to policy-makers about their data to make sure their insights are considered in policy-making. Engaging with centres of expertise (e.g., Scotland's Centre of Expertise for Waters; CREW) is a good example of ensuring both that the scientific process is more policy-driven and that policy-making continues being evidence-based.

5.3.2. Consideration needs to be given to the dynamic nature of data. Many aspects of these case studies are time-sensitive (e.g. coastal erosion, ash dieback) so a data snapshot in time (e.g. a one-off data collection) will not be helpful to understand long-term processes. **Sustainability in long-term data collection** needs to be factored in when funding the collection of dynamic data (Hughes et al. 2017). Long-term projects with associated funding are needed to obtain appropriate long-term data showing clear patterns, rather than short-term projects with little funding.

5.3.3. **Forming stakeholder groups** with different topical foci which meet frequently is an excellent way of mapping out available data and fostering collaborations. A recent example is the Ash Dieback Risk Group formed across Scotland, comprising representatives of different sectors. Some interviewees suggest SG-led networks, but any coordinator that is well placed to establish a good spread among users (e.g. technical users, software providers, private companies) would be suited for the task. We suggest that the SG GIS team showcases examples of well-established groups to support those that want to set up their own groups and bring together different groups of stakeholders.

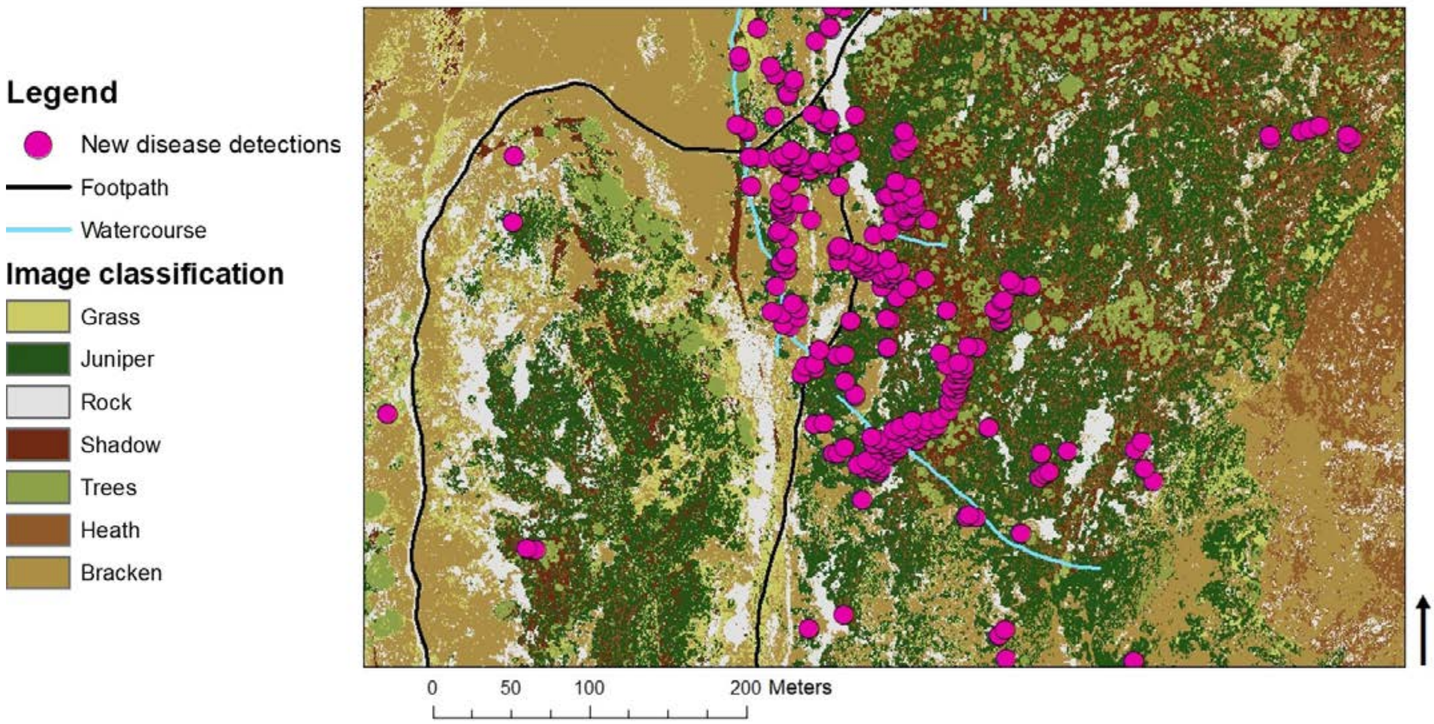
5.3.4. A more advanced version of the stakeholder groups would be the **establishment of a framework** with a Scotland-wide scope to combine expertise and datasets across thematic areas. A good example of such a framework is the Environment & Rural Affairs Monitoring and Modelling Programme (ERAMPP) from the Welsh Government, which involves a large consortium of partners that makes use of activities across the monitoring and modelling community (ERAMPP 2020). Setting the foundations of a similar framework would make the processes of obtaining data and expertise, and liaising with the researcher and practitioner communities, much more streamlined and sustainable in the long-term than at present.

5.3.5. The efficient **management of data queries** is strongly recommended. Institutions should appoint a data manager or at least specify a contact that can answer questions about the data. This could be a general institutional address that is monitored in case individuals move on, and the contact should be properly advertised on institutional websites and clearly specified in the metadata. Clear contacts would make for a smoother process when data users have questions about the data.

5.3.6. The use of tools such as dynamic maps or web mapping services are encouraged in order to increase **public engagement** and communicate geospatial data. The vast quantity of geospatial work that is undertaken should be communicated to make geospatial information accessible to the public.

5.3.7. More **interdisciplinary research** is encouraged, particularly crossing bridges between the social sciences and the geosciences. Partnering with others will provide numerous benefits for sharing data and expertise.

5.3.8. To address the limitations of data sharing in the public sector and improve the possibility of cross-sectoral work, a **joint work space for the public sector** can be put together by SG through the Collaborative Workspace and Analytical Tool Bench.



Location of pathogen detections in trees across Birk Fell Site of Special Scientific Interest identified from aerial photographs. Credit: Flora Donald.

6. Conclusions

Available geospatial data in Scotland is generally of good quality and users from all sectors use and trust the data, but more remains to be done to improve data accessibility and discoverability by considering user needs. While some users and organisations are working at very high levels with geospatial data (particularly in the private sector), the vast majority of geospatial users have an average level of expertise. Improving data literacy skills, plus the implementation of the recommendations within this report will lead to an increase in geospatial data expertise, together with more advanced and refined analyses and technologies across Scotland. Consequently, this could bring greater use and uptake of geospatial data, much clearer business and organisational intelligence, and better expertise around metadata and standards. Ultimately, improving the quality of publicly held geospatial data and easing their access and use is estimated to unlock a further £6-11 billion per year for the UK economy (Cabinet Office 2018).

Data accessibility in Scotland has greatly improved in the past few years, and both data generators and users agree that sharing data is essential for evidence-based decision-making. In the framework of the current challenges of the COVID-19 pandemic and the Brexit process, there is uncertainty on the future of data sharing, access and standards, with consequent delays to projects. However, more advances and improvements are being made through the RDS Metadata Catalogue, SG's Network Accelerator, or the data maturity model from DataLab, among others. The future of geospatial data is bright, and improving the current landscape of data accessibility and quality could release the vast potential of geospatial data in Scotland to new heights.

7. References

Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G. Casey, K., Laaksonen, L., Moorman, D., Uhlir, P., Wouters, P. 2004. An International Framework to Promote Access to Data. Science 303, 1777-1778.

Association for Geographic Information (2018). UK GEMINI standard and INSPIRE implementing rules. Accessible at: https://www.agi.org.uk/gemini/40-gemini/1037-uk-gemini-standard-and-inspire-implementing-rules

Cabinet Office (2018). An Initial Analysis of the Potential Geospatial Economic Opportunity. Accessible at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/733864/Initial_Analysis_of_the_Potential_Geospatial_Economic_Opportunity.pdf

Coetzee, S., Ivánová, I., Mitasova, H. and Brovelli, M.A. (2020). Open Geospatial Software and Data: A Review of the Current State and A Perspective into the Future. International Journal of Geo-Information, 9:90.

ERAMMP (2020). Environment and Rural Affairs Monitoring & Modelling Programme. Accessible at: https://erammp.wales/en

European Environment Agency. (1995). Corine land cover. EEA, Copenhagen

Foulonneau, M. and Riley, J. (2008). Metadata for *Digital Resources. Implementation, Systems Design and Interoperability*. Chandos Publishing: Oxford, England.

Geospatial Commission. (2020). *Unlocking the power of location: The UK's geospatial strategy*. UK Cabinet Office.

Hansom, J.D., Fitton, J.M., and Rennie, A.F. (2017) Dynamic Coast - National Coastal Change Assessment: National Overview, CRW2014/2.

Hughes, B. B., Beas-Luna, R., Barner, A. K., Brewitt, K., Brumbaugh, D. R., Cerny-Chipman, E. B., Close, S. L., Coblentz, K. E., De Nesnera, K. L., Drobnitch, S. T., Figurski, J. D., Focht, B., Friedman, M., Freiwald, J., Heady, K. K., Heady, W. N., Hettinger, A., Johnson, A., Karr, K. A., Mahoney, B., Moritsch, M.M., Osterback, A-M.K., Reimer, J., Robinson, J. Rohrer, T., Rose, J.M., Sabal, M., Segui, L.M., Shen, C., Sullivan, J., Zuercher, R., Raimondi, P.T., Menge, B.A., Grorud-Colvert, K., Novak, M. and Carr, M.H. (2017). Long-Term Studies Contribute Disproportionately to Ecology and *Policy. BioScience*, 67(3), 271–281.

Intra-governmental Group on Geographic Information (2002). The Principles of Good Metadata Management. IGGI: UK.

Intra-governmental Group on Geographic Information (2005). The Principles of Good Data Management. IGGI: UK.

Loew, A., Bell, W., Brocca, L. Bulgin, C.E., Burdanowitz, J., Calbet, X., Donner, R.V., Ghent, D., Gruber, A., Kaminski, T., Kinzel, J., Klepp, C., Lambert, C., Schaepman-Strub, G., Schröder, M., Verhoelst, T. (2017). Validation practices for satellite based Earth observation data across communities. *Reviews of Geophysics* 55(3): 779-817.The National Archives (2021). UK Government Licensing Framework. Accessible at: https://www.nationalarchives.gov.uk/information-management/re-using-public-sector-information/uk-government-licensing-framework/

The Scottish Government (2018). Evaluation for policy makers - A straightforward guide. Scottish Government. Evaluation for policy makers - A straightforward guide - gov.scot (www.gov.scot)

The Scottish Government (2020a). Update to the Climate Change Plan 2018 – 2032. Securing a Green Recovery on a Path to Net Zero. Scottish Government. Securing a green recovery on a path to net zero: climate change plan 2018–2032 - update - gov.scot (www.gov.scot)

The Scottish Government (2020b). Transforming Places Together - Scotland's digital strategy for planning. Scottish Government. Transforming Places Together: digital strategy for planning - gov.scot (www.gov.scot)

The Scottish Government (2021a). Scotland's Third Land Use Strategy 2021-2026. Getting the best from our land. Scottish Government. Land use - getting the best from our land: strategy 2021 to 2026 - gov.scot (www.gov.scot);

The Scottish Government (2021b). A changing nation: how Scotland will thrive in a digital world. Scottish Government. 602f83f8804c3_CES-climatechange-final (2).pdf; A changing nation: how Scotland will thrive in a digital world - gov.scot (www.gov.scot);

UK Government (2020a). Best practice guidance and tools for geospatial data managers. Accessible at: https://www.gov.uk/government/collections/best-practice-guidance-and-tools-for-geospatial-data-managers

UK Government (2020b). The Government Data Quality Framework. Accessible at: https://www.gov.uk/government/publications/the-government-data-quality-framework/the-government-data-quality-framework

Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3: 1–9.

8. Acknowledgements

Mariana García Criado, Project Intern, PhD in Plant Macroecology at the University of Edinburgh (Interviewer, primary author)

Professor Mark Inall, Member of SSAC (Project Coordinator, co-author, main Editor)

Shona Nicol, Scottish Government (Intern co-supervisor, Editor)

SAGES (Scottish Alliance for Geoscience, Environment and Society), Project Support

We are very grateful to all interviewees who provided their time and expertise to this project. With thanks also to an expert peer-reviewer for revising this report.

9. Appendix 1

The proposed recommendations in Section 5 are summarised below in table format, together with their indicative timeframes and responsibilities for implementation.

Indicative timeframes

🍒 Low-hanging fruit, short time frame.

🕓 Worthy but more complex/time-consuming/costly, medium term.

🖼 Needs further discussion, longer term.

Responsibilities for implementation

L Lead Role

S Supporting role

Report reference	Recommendation	Time frame	Responsible Actor				
			Scottish Government	Public Sector	Private and Third Sector	UKRI / SFC	Academia
1. Geospatial data sharing							
5.1.1	Central data directory	🍒	L	S	S	S	S
5.1.2	Public Sector Data Catalogue	🕓	L	S	S	S	S
5.1.3	Assess user requirements	🍒	L	L	L	L	L
5.1.4	Guidance on using data catalogues	🍒	L	L	L	L	L
5.1.5	Data sharing agreements	🍒	L	L	L	L	L
5.1.6	Making data publicly available	🕓	L	S	S	S	S
5.1.7	Provision of metadata and data extent	🕓	S		L		
5.1.8	Options of data acquisition	🕓			L		
5.1.9	Budget data acquisition funds	🕓	L	L	L		L
5.1.10	Health professional push for geospatial data	🖼	S	L			

Report reference	Recommendation	Time frame	Responsible Actor				
			Scottish Government	Public Sector	Private and Third Sector	UKRI / SFC	Academia
	2. Geospatial data quality						
5.2.1	Best possible data resolutions		L	L	L		L
5.2.2	End-user requirements		L	L	L		L
5.2.3	Data and metadata standards		L	L	L	L	L
5.2.4	Best practice on data standards		L	S	S	S	S
5.2.5	Data standards in academia					L	L
5.2.6	Specify standards into contracts		L	L	L	L	L
5.2.7	Training on data standards		L	L	L	S	L
5.2.8	Standardise Council level data		S	L			
5.2.9	EO validation with field observation				L		L
	3. Working with geospatial data						
5.3.1	Data expertise input to policy		L	L	L		L
5.3.2	Sustainability of long term data acquisition		L			L	
5.3.3	Form stakeholder groups		L	L	L		L
5.3.4	Create stakeholder framework		L	S	S	S	S
5.3.5	Management of data queries		L	L	L	S	L
5.3.6	Public engagement		S	L	S	L	S
5.3.7	Interdisciplinary research		S	L	L	L	L
5.3.8	Joint workspaces for the public sector		L	S			

10. Appendix 2

Questionnaire for interviewees

Your experience of using geospatial data

Please give answers firstly from the point of view of a geospatial problem you have already tackled, and then from the perspective of a question you may need to tackle in future (i.e., the proposed question). If only one of these two perspectives is relevant to you, please just answer from the relevant perspective.

1. What is your experience broadly with geographically spatial data for the topic X? Please provide a specific example of a geospatially-based objective/question that you have tried to, or need to answer.
2. If you were to try to investigate Question X, where would you start?
3. Where would you try to obtain data to answer this Question and how would you find the data?
4. Would the metadata contain an appropriate and useful description of the data?
5. Would the data be accessible to you?
6. Could the data be downloaded directly or is there an authentication or paywall to access them?
7. Do you have the appropriate software and hardware to download, store, process and analyse this data?
8. Do you have the skills to process and analyse the data once it has been downloaded?
9. Can the data be integrated with your other data to carry out analyses?
10. Can you re-use the data after using it in your analyses for this project?
11. What bottlenecks did you find during the process?
12. What recommendations do you have to improve these limitations?

Data accessibility in Scotland

These questions aim to assess how easy or otherwise you find data access, the quality control of those data, the quality of metadata, and intercompatibility with other data types you wish to merge/compare with.

1. Which differences did you find between different data sources in terms of data download, quality and integration?
2. Did you find all the data you needed for your project/question or are there remaining gaps of data that you identified?
3. What do you find to be the relationship between the public, private and academic sectors in terms of data sharing?
4. Do you think it's easier for a user to access data from the public, private or academic sector?
5. In which sector do you think most of the work is happening with geospatial data?
6. Are data producers making their data available? Do you know of any particular datasets that would be useful but are not available to the public?
7. Are there any accountability processes checking that data producers are making their data available?
8. Are there any processes in place to make sure that the produced data follows some sort of data guidelines (e.g. Inspire or FAIR)?
9. Are you aware of institutions that follow some data principles (e.g. Inspire or FAIR)?
10. Does compliance to these guidelines/standards make a difference in your workflow as a researcher or on how useful the data are?

11. Appendix 3

FAIR Data Principles

The FAIR Data Principles are a set of guidelines for good practice when managing data that was designed by stakeholders in academia and the private sector (GO FAIR 2021). The emphasis is on enhancing the possibility of machines finding and using the data automatically, in addition to the re-use of data by individuals. FAIR principles are based on transparency, reproducibility and reusability and apply to data, metadata and supporting infrastructure (algorithms, tools and workflows). The FAIR acronym stands for Findable, Accessible, Interoperable and Reusable.

Coined in the scientific article from Wilkinson et al. (2016), the FAIR principles do not offer details on readily implementable procedures. Instead, they provide a common framework that different users can take for a unified approach to data management – therefore, different interpretations and implementations of the FAIR principles exist. As a complementary guide to FAIR, the CARE Principles for Indigenous Data Governance have also been released (Carroll et al. 2020). These are based on Collective benefit, Authority to control, Responsibility, and Ethics.

A large number of private and public institutions have adopted the FAIR principles. National organisations around the world are also abiding by FAIR principles – for example, the Swiss National Science Foundation's policy on open research data is underpinned by FAIR principles, and so is the UK Wellcome Trust's policy. In the United States, the National Academies of Sciences, Engineering, and Medicine abides by FAIR principles, while the National Institute of Health has endorsed the FAIR principles and led the Big Data to Knowledge (BD2K) programme, which supports efforts towards making datasets FAIR. The National Science Foundation (NSF) encourages more effective practices for data management and a move towards open science, for example through the Project Open Data. While the data principles that NSF follows are similar to FAIR's in general, their policy does not specify that researchers should abide by FAIR principles necessarily.

International examples of organisations abiding by FAIR principles include the International Virtual Observatory Alliance and the Association of European Research Libraries. In the European Union, the European Commission mandates that Horizon 2020-funded project beneficiaries manage their data according to FAIR principles, and has published a report giving recommendations on implementing the FAIR principles in practice (European Union, 2018). We provide a series of questions to understand whether one's data complies with FAIR principles (**Figure A1**).

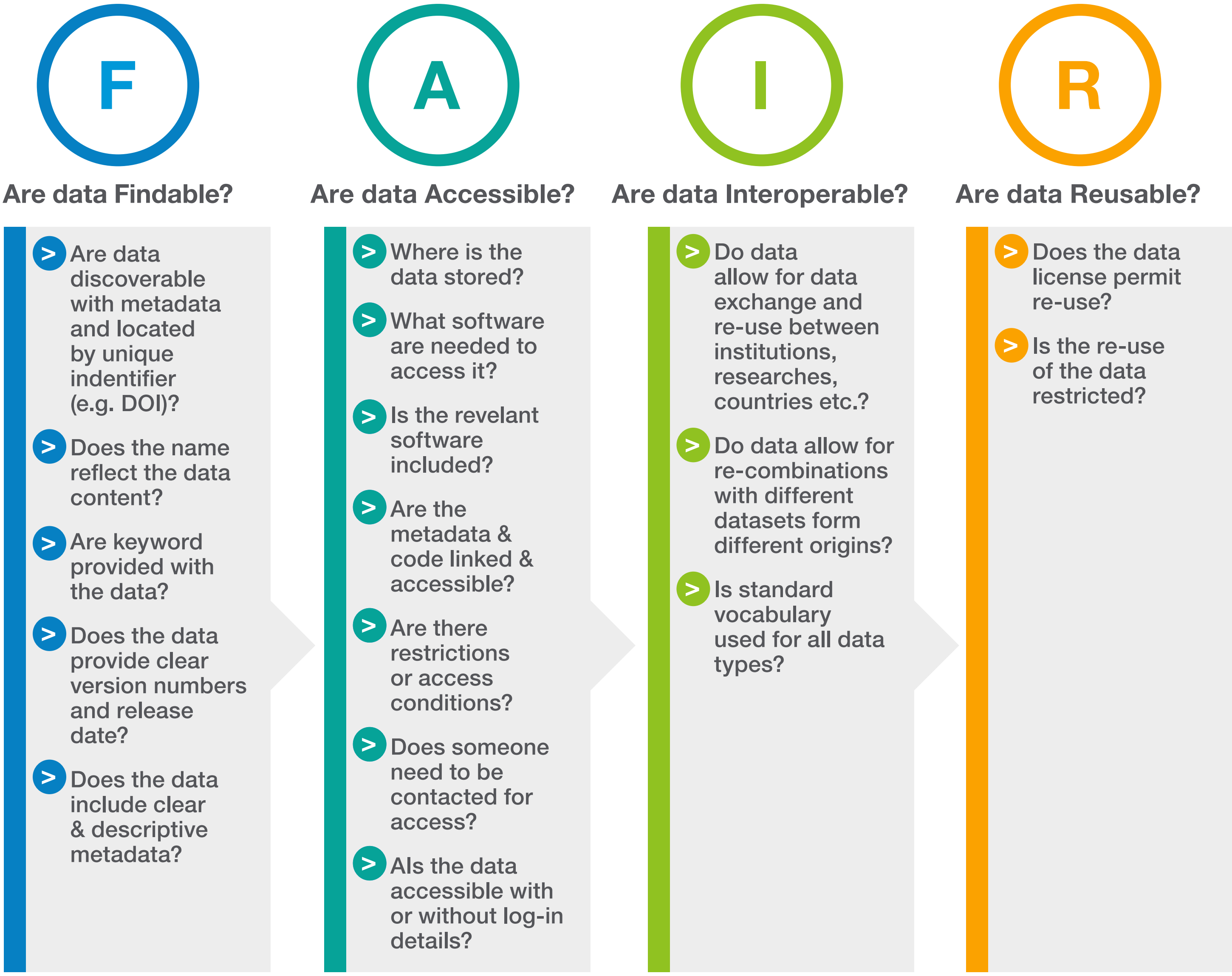


Figure A1. Diagram defining the questions that the researcher should consider to understand whether their data follows FAIR principles. Adapted from the Horizon 2020 data management plan.

References

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J., & Hudson, M. (2020). The CARE Principles for Indigenous Data Governance. *Data Science Journal*, 19(1), 43. https://doi.org/10.5334/dsj-2020-043

European Union (2018). *Turning FAIR into reality*. Luxembourg: Publications Office of the European Union.

GO FAIR (2021). GO FAIR. Accessed at https://www.go-fair.org/ January 2021.

Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. - Scientific Data 3: 1–9.

Produced for Scottish Science Advisory Council by APS Group Scotland, 21 Tennant Street, Edinburgh EH6 5NA
(August 2021)

scottishscience@gov.scot