http://doi.org/10.7207/twr23-01

Preserving Geospatial Data

Meagan A. Snow

DPC Technology Watch Report

July 2023

DigitalPreservationCoalition

ISSN: 2048-7916

First published in Great Britain in 2023 by the Digital Preservation Coalition.

This is the 2nd edition of Preserving Geospatial Data. The original report was authored by Guy McGarva, Steve Morris and Greg Janée and published in 2009 (available at: http://doi.org/10.7207/twr09-01).

V2.1 with minor changes to formatting and presentation.

This report is published by the Digital Preservation Coalition (DPC). The DPC is an international charitable foundation which supports digital preservation and helps its members around the world to deliver resilient long-term access to digital content and services. In addition to the publication of reports on a range of themes which cover the state of the art in digital preservation, the DPC also supports its members through community engagement, targeted advocacy work, training and workforce development, identification of good practice and standards, and through good management and governance. Its vision is a secure digital legacy.

Discover the DPC's publications, including the latest updates and revisions, at: https://www.dpconline.org/digipres/discover-good-practice/tech-watch-reports

Find out more about the DPC, the support it offers and how to become a member at: https://www.dpconline.org/about/join-us

Foreword

The *DPC Technology Watch Reports* identify, delineate, monitor, and address topics that have a major bearing on ensuring our collected digital memory will be available tomorrow. They provide an advanced introduction in order to support those charged with ensuring a robust digital memory, and they are of general interest to a wide and international audience with interests in computing, information management, collections management, and technology. The reports are commissioned after consultation among DPC members about shared priorities and challenges; they are commissioned from experts; and they are thoroughly scrutinized by peers before being released. The authors are asked to provide reports that are informed, current, concise and balanced; that lower the barriers to participation in digital preservation; and that they are of wide utility. The reports are a distinctive and lasting contribution to the dissemination of good practice in digital preservation.

This report was authored by Meagan Snow, Geospatial Data Visualization Librarian at the Geography & Map Division of the Library of Congress, Washington, D.C. It is a revision of an existing *DPC Technology Watch Report* of the same title which was written by Guy McGarva, Steve Morris and Greg Janée and published in 2009. The report is published by the DPC in association with the UK Nuclear Decommissioning Authority who raised geospatial data preservation as a priority topic as part of a collaborative digital preservation project with the DPC in 2020.

Acknowledgements

Bringing together a wide domain of knowledge into one report relies on the generous support of many members of the digital preservation and geospatial data communities, particularly those who have been thinking and working towards geospatial preservation issues since the very beginning. I am indebted to my colleagues, especially Tim St. Onge, Rachel Trent, and Butch Lazorchak, for their conversation and partnership in working through geospatial preservation issues at the Library of Congress. Thanks also to Dr Paulette Hasier and Sundeep Mahendra for supporting my work on this report.

I am thankful for the generous support of Guy McGarva, Steve Morris, and Greg Janée, the authors of the original *Technology Watch Report* of this title. Though this field has developed in new directions since 2009, some of their original text was still relevant today and needed little updating.

Additionally, an earlier draft of this report benefited from an insightful and detailed peer review from Guy McGarva of the University of Edinburgh and Kieron Niven of the Archaeology Data Service.

Finally, thank you to Jen Mitcham for her indispensable shepherding of this project.

Meagan Snow, Library of Congress, July 2023

Contents

1 Executive Summary

The world of geospatial data is diverse, spanning both traditional vector data (point, line, and polygon) and raster data (continuous cell-grid data, such as satellite imagery or elevation models), mappable coordinate data (often in spreadsheet form), as well as emerging forms of location-based data and dynamic web services.

There are key preservation issues that relate specifically to geospatial data, including:

- the diversity and complexity of data formats, data structure, and the GIS software which reads them, all of which continue to evolve over time.
- widespread use of proprietary geospatial formats.
- the difficulty of preserving the technical and social contexts in which the data exists.
- the increasing importance of dynamic (and ephemeral) web-based cartography and web services which are difficult to preserve.

Though standards for geospatial metadata have been defined at both national and international levels, metadata often becomes disconnected from datasets, or is incorrect, non-standard in nature, or not created in the first place. Additional considerations in preserving geospatial data include spatial reference systems, cartographic representations, topology, project files, and data packaging. A number of technologies and tools that are, or may be, of relevance to geospatial data preservation efforts have emerged, though there is no single tool or technology that will be relevant in all cases.

While preservation recommendations provide a basic checklist of issues to be considered when preserving geospatial data, the wide variety of data formats and use-cases will mean that preservation decisions will vary across different contexts. Preservation is a still-developing area in the geospatial world and good practices are not universal.

2 Abstract

This report is designed as a resource for use by librarians, archivists, and digital preservation specialists who may be new to the realm of geospatial data but want a practical understanding of the geospatial data files they encounter in their collections. It may also be useful to geographers, cartographers, academics, and researchers who are increasingly involved in the preservation decisions around their own research data or mapping products. The report focuses on describing challenges specific to the preservation and management of geospatial data.

3 Why Preserve Geospatial Data?

As we navigate the world around us, we interact with geospatial and location-based data daily in a multitude of ways, whether we realize it or not. Mapping applications can be used in everything from planning a route across town, searching for a nearby restaurant, getting traffic or travel-time estimates, and even the targeted ads that display online as we complete these tasks. When natural disasters strike, we consult web maps to check wildfire locations, predicted hurricane tracks, or potential storm surge risks by neighbourhood. Buildings we live and work in sit in tax parcels neatly mapped out by our county assessor and those boundary files are likely available online in shapefile form, alongside files delineating trash pick-up zones or point files storing the location of every tree. Nearly everything we measure in a specific location on earth has the potential to be stored as geospatial data, and this spatial information is a key component in decision-making processes large and small, and in planning efforts across a broad range of industries. The quantity and variety of available geospatial data is rapidly increasing and much of this data is at risk of being lost or becoming unusable. So much of the information once preserved in static paper maps is today offered only in born-digital formats: any given map layer likely exists today as a dataset which can be symbolized or represented in anything from a digital map image to a dynamic web service (or both). Librarians, researchers, and archivists need to make critical scoping decisions around what to keep and collect from the huge volume of born-digital maps and geospatial data available.

When discussing geospatial data, it is useful to understand Geographic Information Systems (GIS), the software used to render, symbolize, edit, transform, and analyse geospatial data. There are a wide variety of GIS software products available from commercial suppliers as well as open-source projects. They range from simple, general purpose, web-based clients to large, highly complex, integrated systems aimed at specific application domains. The sheer variety of types of geospatial data, taken together with the variety of software and platforms used to manipulate that data, is one of the main factors making the preservation of geospatial data a highly complex issue.

4 Geospatial Data Preservation Issues

There are two principal data types associated with geospatial data: vector data (comprised of points, lines, and polygons); and raster data (comprised of a continuous cell grid). Some preservation challenges apply across both formats of data (such as spatial reference systems), while other issues are specific to just one form of data.

Many geospatial projects will contain both raster and vector data files, and some data formats (such as geodatabases or GeoPackage files) can contain both raster and vector data within a single file. Common challenges across geospatial data formats include spatial reference systems (which define how locations on the earth are modelled), cartographic representations of data (how data is symbolized as a map), the spatial relationships within and between datasets, the formats used to define project files used by different systems, and how geospatial data is packaged to contain all required elements.

The net result of these characteristics is that there is no single, easy, or universal solution to the preservation of geospatial data. There are many formats and applications, all of which have overlapping but different capabilities. Because conversion of geospatial data across formats, data structures, and applications often results in loss of data or data alteration, the migration of geospatial formats over time is not easily automated, but instead must be evaluated on a case-by-case basis.

Furthermore, the general preservation problem for geospatial data is only getting more complicated. Satellites, drones, sensor networks, and mobile applications are producing more geographic information than humans can reasonably sort through or save. How machine learning or artificial intelligence (AI) technology may be applied to filtering, analysing, or describing this data is still being understood.

4.1 Scale and Generalization

In contrast to textual information, which can be successfully modelled using multi-page textual documents as the sole granule size, geospatial data is regularly created at varying levels of granularity. Geospatial data is produced for visualization and analysis at multiple scales or resolutions: some datasets work best to capture the globe overall, while others are scaled for local and regional use. Local datasets typically contain data at more detailed geographic units and may seek to capture the world in as much detail as possible and may serve as source datasets from which small-scale datasets are created. By contrast, small-scale maps or datasets will purposely generalize data to cover more geographic territory. This generalization could take the form of boundary lines stripped of detail, or raster datasets produced at lower resolutions.

Geospatial data can be aggregated, disaggregated, and operated on with fluidity. Every scale of data has its uses, affords different functionalities, and poses different preservation challenges. Therefore, there is no *single* preservation problem for geospatial data; instead, choosing which level or levels of granularity to address, and therefore identifying the preservation problem(s), is a first step in the process. When choosing what data to preserve, it may be helpful to consider that thematic datasets of the same topic but produced at different scales or resolutions often have distinct practical use-cases.

4.2 Spatial Reference Systems

A spatial reference system is a model of the earth which allows for a common system of identifying the location of features on the earth. All geospatial data utilizes *some* type of spatial reference system to locate the data in the correct location on the surface of the earth. Spatial reference systems are typically comprised of multiple elements including a spheroid, datum, and a geographic coordinate system, and may also include a projected coordinate system. Spheroids model the shape of the earth, while datums define how that model relates to the very centre of the earth. Datums are part of a dataset's geographic coordinate system, which also includes a prime meridian and an angular unit of measurement.

Geographic coordinate systems model the earth as a three-dimensional globe and use a system of coordinates (commonly known as longitude and latitude) to precisely locate features. They provide a common geographic language of location to use when recording and measuring the world. Projected coordinate systems move spatial data from a three-dimensional model of the earth into a two-dimensional depiction of the earth, capable of being displayed on a flat surface and with linear map units.

Casual users or creators of spatial data, especially data that is formatted in terms of latitude and longitude, may not have a deep understanding of the spatial reference system that underlies the data or coordinates they are producing. It is critical that the spatial reference system used to generate coordinates or create spatial data is always included with spatial data, whether embedded in the file itself or stored as metadata. Without spatial reference systems, it would be impossible to accurately measure distances, analyze relationships, and locate features on the Earth's surface.

Some geospatial data formats (both raster and vector) directly contain information about the spatial reference system the dataset is based on, either embedded as part of the file itself (for example, a GeoTIFF or GeoPDF file) or as an additional, tightly bundled file (for example, a '.prj' projection metadata file stored as part of a shapefile (Esri, n.d.-a)). However, some other formats that are spatially referenced do not directly contain coordinate reference system information, for example, plain TIFF image files. A plain TIFF file may have an associated TIFF World File (TFW), but the association is loose, and it is easy for the two files to become separated. If a dataset includes a stated bounding box (in coordinates), coordinate system information should be provided for both the dataset and the bounding box. Geospatial files that get disconnected from their spatial reference information will fail to locate themselves in GIS software. Thus, it is critical to keep all elements of these multi-part files together.

There are several ways of specifying a coordinate reference system. The EPSG Geodetic Parameter Dataset (EPSG, 2022) maintains a list of spatial reference systems, datums, ellipsoids, and units of measurement with corresponding EPSG codes (for example, 'ESPG:4326' for WGS 1984). These EPSG codes can function as identification codes in GIS software. Unique codes for spatial reference systems are often referred to as Spatial Reference System Identifiers (SRIDs). Some commercial GIS software vendors maintain and utilize their own in-house SRIDs for spatial reference systems.

4.3 Georeferencing and Rectification

Raster data may undergo a process of georeferencing and rectification to bring the data into a known coordinate system. This is especially common for digitized historic (analogue) maps that are georeferenced for use in GIS software or web mapping applications.

Lineage information for georeferenced raster data or map images should describe the process taken, and the software used, to georeference the image. Furthermore, any data digitized from georeferenced images should cite the underlying georeferenced source, particularly in cases where vector feature digitization is based on an underlying georeferenced analogue map.

For imagery data, such as satellite or aerial photography, a further process of ortho-rectification corrects for scale differences due to surface topography and requires a digital elevation or terrain model. In the latter case, since the accuracy of the data is dependent on the accuracy of the elevation model, the source elevation model and ortho-rectification software should be recorded as part of the data's lineage.

4.4 Mosaicked Raster Data

Raster data is often used to represent continuous phenomena (for example, surface elevation), but for convenience of data management and delivery it is frequently packaged into fixed-size tiles divided along arbitrary tile boundaries. This means that it is often desirable to mosaic the tiles back together into a seamless whole, and to thereby allow users to browse and crop out just the portion of the dataset that is of interest to them. A similar process can be used to turn digitized and georeferenced analogue map sheets into one continuous digital map image.

The ramifications of mosaicking for preservation purposes depend greatly on the implementation specifics. If the raster tiles are stored as files in a filesystem, for example as GeoTIFFs, each independently carrying metadata and georeferencing information, and if the mosaicking system is entirely automated, then the preservation problem may be no more difficult than the problem of preserving a collection of files. In this case, the mosaic can be viewed purely as an access mechanism. Preservation of the raster tile files and coordinates is sufficient to recreate the mosaic in the future.

However, sophisticated mosaicking systems often perform edge alignment and colour balancing across tile boundaries, and even allow for fully manual and/or manually directed adjustments. In the case of digitized and georeferenced historic maps (originally produced in analogue), mosaicked datasets could contain map sheets created across different time periods. In this case, the mosaicked image effectively becomes a new data product derived from diverse source raster tiles, and merits preservation and documentation in its own right. Such data products would benefit from detailed descriptions in lineage metadata describing the manual and/or automated choices made in the creation of the data product.

4.5 Topology

Topology is the spatial relationship between features, such as their connectivity (for example, a road network) or adjacency (for example, countries sharing a boundary). Vector datasets can model these relationships in different ways depending on the software being used and the data model being implemented. Topological data models can range from what are termed ‘spaghetti’ or ‘unstructured’ datasets, where there are no explicit relationships between vector features, to ‘fully’ or ‘structured' topological datasets in which every feature has a relationship. The ability for a geospatial file format to store topology is often (but not always) tied to proprietary vector formats and data structures. Topological files allow for automated methods to check for and clean topological relationships. If converting between proprietary and open file formats, the conversion process or data transfer may result in a loss of information. Thus, it is critical to understand which formats can store topology and to be aware of this when transferring or converting data between formats.

4.6 Project Files

GIS software ‘project files’ are complex digital documents produced by GIS software programs in the process of creating digital cartographic products or spatial analysis. Project files typically document the source and format of geospatial data opened in GIS software, how that data is symbolized, can contain overall project metadata or documentation, and may record a history of data transformations or operations, whether through data models or scripts. A key feature of a project file is a data view in which a combination of data is presented in a tailored manner that involves classification, symbolization, and annotation based upon the data content, and potentially a layout view, where the data is assembled into a final map product. These data views typically appear as maps, charts, or tables, or some combination thereof. For an end user to render this content, they must have the original project file, the software that supports the project file, related components (possibly including software add-ons or extensions), as well as the underlying data. The required use of specific software, the complexity of the project file formats, and the tenuous links to the actual data, which is often simply pointed to, put these project files at very high risk of failure.

Examples of project files include the Esri ArcView .apr file, the Esri ArcMap .mxd file, the Esri ArcGIS Pro .aprx file, and the QGIS .qgs file. It should be noted that simply preserving a project file does not preserve the underlying data or auxiliary files that are needed to display and use that data. The same is true for Esri layer files (.lyr), which are intended to preserve layer symbology, but which need the underlying data file to function.

There is recognition in the GIS community of the need to be able to archive not just data but also projects and their various components in order to preserve the ability to revisit how different data processes and analyses were carried out. Yet project file incompatibilities with software upgrades point to large preservation challenges in maintaining this content, and vendor commitment to forward compatibility of current project files with future software releases remains inconsistent.

4.7 Cartographic Representation

Cartographic representations are a common output produced as a result of working with geospatial data in GIS software, often created through GIS project files. Cartographic representations are, at their most simple, map images. These map images can take different forms, from digital image formats to digital images that are spatially referenced, allowing for the use of the map as an overlay in GIS software or web-based geospatial applications. Cartographic outputs may also take the form of more complex documents such as PDF or GeoPDF files, which allow for more advanced user interaction with the resulting digital map. These finished information products typically do not include underlying source data used to produce the document, though certain formats, such as GeoPDF, support the retention of some amount of data intelligence derived from the original data.

Preservation of data and preservation of documents created from the data comprise two separate and non-exclusive objectives. The data itself is required to recreate a work product or research project, or to build new work from original source data. The finished product, whether a map, chart, or other output, is an information product in its own right. Decision-makers for preservation may be tasked with preserving an end-product cartographic representation, the underlying data contained in the representation, or potentially both. Individual preservation decisions and metadata considerations are needed for both items.

4.8 Data Packaging

Geospatial data files are often complex, multi-file, multi-format objects. A ubiquitous example of this is the shapefile, which is always comprised of at least three required files (the main .shp file, a .shx index file, and a .dbf database table), but which can also have up to a dozen different component files (such as a .prj projection file, .xml metadata file, .fbn/.fbx spatial index files, and .sbn/.sbx read-write spatial index files). Complex geospatial objects may include elements such as georeferencing files, metadata files, licensing information, and other ancillary documentation or supporting files. The absence of a standard scheme for content packaging and accounting for all possible elements makes transfer and management of these complex data objects difficult both for archives and for users of the data.

The Federal Geographic Data Committee (FGDC) developed a Geographic Information Framework Data Standard for seven themes of geospatial data (transportation, hydrography, boundaries, geodetic control, elevation, orthoimagery, and cadastral). This data standard establishes common requirements for describing data and aims to do so as efficiently as possible (Federal Geographic Data Committee, 2008). The framework is required to be used by US federal data products that fall within National Spatial Data Infrastructure theme categories.

In widespread practice, the geospatial community often utilizes compressed file formats (such as zip files) to serve as content packages for multi-file datasets or groups of related datasets. Such archive files typically lack data intelligence about file relationships and functions within the data bundle. In some cases, formalized approaches to the use of zip files do exist, for example KMZ files that are used to package KML files and their ancillary components.

4.9 Compression

An issue that arises in preserving geospatial raster data specifically is sensitivity to lossy compression. Lossy image compression techniques subtly change data values. For example, JPEG, by design, changes data values in ways that are not readily noticeable to human vision, but which may be very significant to the analytic functions the data is intended to support. Generally, if the data is intended to support analysis, only lossless compression should be used.

5 Geospatial Metadata

Metadata plays a central role in the current and future use of geospatial data by making data discoverable through data portals and search systems, by providing the means for prospective users to evaluate the data for use, and by allowing data producers to better manage their data holdings and encourage use of the data in the way it was intended. Metadata also provides end users with key information about geographic positioning information including coordinate reference information (such as projection and datum), entity and attribute information, data quality, provenance and rights information that are essential for proper use of the data.

5.1 Metadata Challenges for Archives

Geospatial metadata, either by its presence or its absence, can create numerous archival challenges. Common pitfalls include metadata not being created by the original data producer, metadata being created but not distributed or preserved alongside the data itself, metadata that is not concurrent with the data, metadata that does not adhere to standards, or metadata that exists in different versions or from multiple sources.

If metadata is not available or has not been created, a recipient archive can attempt to assemble a metadata record. Many elements of metadata records may be auto extracted by software, and metadata templates for different producer agencies or data collections can further help aid the metadata production process. Unfortunately, many portions of a metadata record, including data quality information, lineage information, and detailed explanations of the meaning of attribute information, can only be provided by the data producer and cannot be recreated if they are not originally provided.

When metadata does exist, archives may find it necessary to normalize the structure of the metadata to adhere to standards or schemas, update out-of-date metadata, or fix errors. Archives may alternately choose to preserve metadata as it was originally produced, even if the original metadata production has errors or does not adhere to standards. After all, metadata quality could reflect the process which resulted in the creation of the dataset, and a lack of metadata or error-prone metadata may provide interesting information in and of itself for a future user or researcher examining or evaluating that dataset for use.

Incomplete or missing metadata may primarily pose a challenge for discoverability of geospatial datasets, depending on the technical infrastructure available for accessing and discovering geospatial data. Missing metadata may lead to incomplete or bare bones cataloguing of datasets, and a lack of embedded metadata in geospatial files can lead to substantial work on the part of libraries or archives to make the data discoverable within systems that rely on, or assume, the existence of embedded metadata.

A particular challenge with some pre-ISO geospatial metadata standards created before the arrival of XML has been the absence of standard methods of encoding metadata. The lack of consistent structure to metadata records makes receipt and management of metadata from other sources difficult.

5.2 Metadata Standards

In 2003 the ISO standard *19115 Geographic Information – Metadata* (International Organization for Standardization, 2003)was finalized, providing a new international standard for geospatial metadata. Prior to that, several national metadata standards had emerged around the world, providing several years of initial experience as a starting point to inform the development of the international standard. For example, in the United States the Federal Geographic Data Committee (FGDC) *Content Standard for Digital Geospatial Metadata* (Federal Geographic Data Committee, 1994) was released in 1994, with version 2.0 released in 1998. United States federal agencies were mandated to begin using the standard in 1995, and it came into wide use by state agencies and commercial data producers shortly thereafter. The FGDC currently endorses ISO geospatial metadata standards in support of global interoperability, having formally endorsed ISO 19115 in 2010 (Federal Geographic Data Committee, n.d.).

In the United Kingdom EDINA developed and implemented a metadata profile based on the ISO 19115 standard but with extensions to support the needs of the UK academic community, called AGMAP (Academic Geospatial Metadata Application Profile). This was succeeded by AGMAP2, which contained updated metadata requirements for compliance with the Go-Geo Portal (Mathys and Kamel Boulos, 2011). However, over time the AGMAP standard has fallen into disuse. A more popular alternative is the UK GEMINI (Association for Geographic Information, n.d.) metadata standard, which is also based on ISO 19115. It is published by the Association for Geographic Information (AGI) and is the required format for UK INSPIRE geospatial datasets (www.gov.uk, 2023).

In 2015, over 20 academic libraries in the United States, including Princeton University, Stanford University, and Massachusetts Institute of Technology, founded an initiative called OpenGeoMetadata (OGM), an online repository for sharing geospatial metadata for discovery platforms. Participating institutions manage their own GitHub-based metadata repositories, which are discoverable by other institutions (OpenGeoMetadata, n.d.-a). Multiple metadata schemas are accepted within the OGM project, though OGM does have its own metadata schema, termed OGM Aardvark. Aardvark is based on Dublin Core with additional spatial values added, is formatted as JSON, and designed for discovery, but does not provide complete technical documentation of items (such as the lineage of a dataset or its processing history) (OpenGeoMetadata, n.d.-b).

Desktop and online tools are available for creating metadata in appropriate standards including: Esri's ArcGIS Pro's catalog functionality and GeoNetwork (GeoNetwork, 2023), an open-source solution from the Open Source Geospatial Foundation.

5.3 Rights Documentation Metadata

Common geospatial metadata standards lack some features which would be useful in the archival management of data. A major element that libraries or archives should associate with geospatial data wherever possible is archival rights information, either in free text form or in a rights expression language.

Metadata for data rights should specify access and copyright status for a dataset. If access or use of a dataset is restricted, rights metadata should document how, why, and for what period of time restrictions are in place. If a dataset has no use or access restrictions, that is equally important to document. Where possible, data providers or depositors should be encouraged to document dataset rights using an open data framework, such as an Open Database License (Open Knowledge Foundation, n.d.-a) or a Creative Commons (Creative Commons, 2023) licence. Both provide standardized legal language which allows non-legal experts to easily specify terms for how a user can use and credit their data source under a range of possible open licence terms.

5.4 Preservation Metadata

Notably, geospatial metadata standards do not provide a wrapper function that would allow additional technical or administrative metadata elements to be associated with (rather than replace) the data producer-originated metadata. Important examples of potential preservation metadata include:

- Administrative metadata related to the manner of the data acquisition.
- Technical metadata related to the actual transfer of the data, including provision of assurances about data integrity.
- Metadata related to any transformations carried out by the archive post-acquisition.
- The outcomes of any assessments of data validity or any assessments of risk associated with the data.

In the digital library community, efforts have been made to use a combination of METS (Metadata Encoding and Transfer Standard; Library of Congress, 2023a) and PREMIS (Preservation Metadata: Implementation Strategies; Library of Congress, 2023b) to address the metadata wrapper need.

6 File Formats

6.1 Vector Data

One of the most common forms of geospatial data is vector data, which models features on the earth's surface as points, lines, and polygons. Vector data is typically structured such that each geometric feature is tied directly to a row in an attribute table, which uses columns to record information relevant to each feature. For example, a line section representing a portion of a street might have attribute information for characteristics such as 'street name', 'number of lanes', 'speed limit', and so on. Attribute data may either be stored directly within the vector dataset or stored externally in a spreadsheet or database.

Real-world features that are represented by vector data are typically subject to change and the corresponding data may be updated accordingly. The updated dataset typically replaces the previous version and, unless a snapshot of that earlier dataset is set aside and archived, it becomes impossible to look at historical changes in the data.

In some cases, the dataset itself may be designed to store the historical changes within the active dataset. In some spatial databases, previous versions of a vector dataset may be stored along with the current dataset but with a different date attribute. There is also the possibility that only changes to features in a dataset are provided, and the receiving system is required to apply updates to that dataset.

6.1.1 Shapefile

Arguably the most ubiquitous geospatial file format, the shapefile has come into wide-spread use as a distribution format that is supported by a range of both commercial and open-source tools. Since shapefiles do not support advanced features such as topology (the spatial relationships between features), they have a simple data structure and lend themselves to rapid drawing and analysis.

While the shapefile format was developed by the commercial geospatial company Esri, it is openly documented and supported in a variety of software tools.

A shapefile consists of at a minimum three files, a .shp file (feature geometry), an .shx file (index of the feature geography), and a .dbf file (a dBase database file that stores the attribute information of the features). Additional files can also be included, including projection files (.prj), metadata files (.xml) and spatial index files (.sbx and .sbn). Although the shapefile format is arguably an 'old' file format, it is also an enduring format that is still widely used and supported.

There are practical limitations to shapefiles: the overall file size is limited to 2 gigabytes, attribute column names are limited to ten characters, their date fields do not store hourly times, they do not allow for null values, there is no specified x,y tolerance (the amount of distance before two points are considered the same point), and they do not support domains, topology, or subtypes (Esri, 2021).

6.1.2 Geodatabase

A geodatabase is a set of data stored in a format based on Esri file format specifications, which include the use of multiple files such as feature classes, attribute tables, raster datasets, or topologies, and which can handle data validation and versioning. Additionally, the geodatabase can also include a range of non-spatial data formats such as charts and tables. The geodatabase currently serves as the default data storage option in Esri's ArcGIS Pro desktop software.

Geodatabases came into use in the late 1990s with the advent of the ArcGIS software environment (for more information on early geodatabase formats, see 6.1.9.2). Beginning with ArcMap version 9.2, the file geodatabase was introduced as a standalone format not requiring a commercial back-end database. All information is stored in a directory of files. There is no limit to the overall size of a geodatabase, though individual feature classes or tables hit capacity at one terabyte in size.

They can store more data than shapefiles (which have a 2 GB file size limit), are more efficient at data storage than shapefiles, and typically have faster performance times. Some transformation toolboxes inside ArcGIS software *require* the use of a geodatabase to generate an output.

Geodatabases are not human-readable when viewed in a file directory: you cannot tell what a geodatabase contains (i.e., the quantity of feature classes or the names of vector or raster data contained within) unless you are viewing the geodatabase in compatible GIS software capable of reading its internal contents. The format specifications of the file geodatabase have not been made publicly available by Esri. However, reverse engineering-specifications authored by Even Rouault are available online (Rouault, 2023) and have served as the basis for GDAL tools (see 8.7.2) related to geodatabases (Library of Congress, 2021).

ArcGIS enterprise geodatabases used in ArcGIS software later than ArcGIS 9.2 are capable of a feature known as 'archiving' (Esri, n.d.-b) Previously to this, data changes could only be tracked by managing transactional versions of the data, and the history of the data could easily be lost if the versions were deleted or if versioning was disabled. Geodatabase archiving supports the creation of an historical version that represents the data at a specific moment in time and provides a read-only representation of the geodatabase. The ArcGIS 'History Viewer' (or in ArcGIS Pro, the 'History Pane') allows user examination of data at specific points of time. Critically, for archiving of geodatabases to work, archiving must be actively *enabled* on the dataset, and not all users may know to do this.

6.1.3 OGC GeoPackage

The OGC GeoPackage is a file-based, open-standard format for storing and exchanging geographic information developed by the Open Geospatial Consortium (OGC). The structure of a GeoPackage is based on the SQLite relational database, and it includes one or more tables that store feature data, tile data, and/or raster data. Each table contains a set of columns and the values for each row. These columns and values are used to store the geographic information, such as the coordinates of a point, the boundary of a polygon, or a raster image.

GeoPackage has two required elements: a spatial reference systems table with coordinate system documentation, and a contents table which describes the content of the package. Vector data will contain a geometry columns table. A key selling point of the GeoPackage is its interoperability, allowing for portability of geospatial data between open-source and commercial GIS applications. GeoPackages do not have overall file size constraints, allowing them to store larger datasets than (for example) shapefiles.

There are constraints to GeoPackage files, as they do not allow for domain or attribute rules. Attribute rules are defined by a data user or creator to meaningfully constrain attribute fields to maintain the integrity of the data. For example, domain rules might require a field to only contain numbers within a specific range. GeoPackages also do not compress file geometry, which can lead to larger overall file sizes than geodatabases (Ho, 2020). OGC maintains an Encoding Standard for GeoPackage (Open Geospatial Consortium, n.d.-a).

6.1.4 KML/KMZ

KML, formerly known as Keyhole Markup Language, is an XML language focused on geographic visualization, including annotation of maps or images in digital globe or mapping environments. KML provides support for both feature data in the form of points, lines, and polygons, and image data in the form of ground and photo overlays. KML files may be associated with images, models, or textures that exist in separate files. KML was initially developed by Google for use solely within Google Earth but is now an open specification with an OGC standard (Open Geospatial Consortium, n.d.-b) that is supported in a range of software environments.

KML files may refer to external resources and other KML files via ‘network links’ (a link to a local or remote resource), which are used to link related data files and to facilitate data updates. Large data resources such as imagery datasets may be divided into many smaller image files which are then made available via network links on an as-needed basis. KML presentations using network links pose a preservation challenge in that any data available via the links may no longer be available in the future.

When working with KML files in ArcGIS Pro, users may notice that KML files operate differently than other vector formats. All content in a KML file is drawn as a single layer (with nodes showing up as sublayers), features cannot be selected, and files cannot be edited. Additionally, some Google Earth-specific features, such as Tracks, Tours, and Sky Data, do not carry over into traditional GIS software (Esri, n.d.-c).

KMZ files allow one or more KML files to be bundled together along with other ancillary files required for the presentation, allowing for ease of transfer of the entire collection. KMZ files are compressed in the ZIP archive format, resulting in reduced file size.

6.1.5 GeoJSON

GeoJSON is a geographical data format commonly used for exchanging geographic data between web services or for storing and exchanging geographic information in web and mobile applications. Based on JavaScript Object Notation (JSON), GeoJSON files are capable of encoding points, lines, or polygons, but do not store topology. GeoJSON files always use World Geodetic Survey 1984 as the spatial reference system, with decimal degrees as a map unit. While they cannot be directly opened in ArcGIS Pro (they must be converted into a geodatabase feature class using the 'JSON to Features' toolbox), QGIS and ArcGIS Online will both open GeoJSON files directly.

The structure of GeoJSON is very similar to other JSON-based formats, and it is composed of two main components: the 'type' and the 'geometry'. The type of element defines what kind of geographical feature is being represented, such as a point, line, or polygon. The geometry element consists of the spatial information associated with the feature, such as the coordinates of the points, the lengths of the lines, and the coordinates of the polygon vertices. The Internet Engineering Task Force (IETF) put together a working group in 2015 to update previous 2008 GeoJSON format specifications. Updated standards were published in 2016 as RFC 7946 (Butler *et al.*, 2016).

6.1.6 GML

Geography Markup Language (GML) (Open Geospatial Consortium, n.d.-c) is a standard first introduced in 2000 by the Open Geospatial Consortium (OGC) and subsequently published as ISO standard 19136. The GML specification outlines many elements and attributes intended to support a wide range of capabilities. Since the scope of GML is so wide, profiles that deal with a restricted subset of GML capabilities have been created to encourage interoperability within specific domains that share those profiles. GML is not so much a single format as it is an XML language for which there are a wide range of different community implementations. These are embodied by specific GML profiles associated with specific GML versions, and for which different application schemas might be available. The GML specification is highly complex, and that complexity, combined with the diversity of profiles and application schemas, can present a barrier to vendor and tool support.

The US Census Bureau conducted a 2006 pilot project to test implementation of a potential TIGER/GML release format (Open Geospatial Consortium, 2006). The pilot project found significant issues with the GML format, including large file size (which led to slow processing) and significant complications when determining the naming structure of interrelated GML documents (Guo, 2013). As of 2023, the Census Bureau has yet to move forward with a public release of TIGER/GML data, and the project is presumed to be discontinued. A prominent ongoing example of GML implementation is the Ordnance Survey's OS MasterMap Topography Layer, which is supplied in GML 2.1.2 (Ordnance Survey, 2017).

6.1.7 OpenStreetMap (OSM)

OpenStreetMap (OSM) is an open-source XML-based file format used to store and exchange geospatial data as contained in the Open Street Map project. Open Street Map is a global collaborative web-based mapping project maintained by volunteers and supported by the OpenStreetMap Foundation. OSM is notable in that it is offered under an open database licence, making the data freely available for download and reuse.

OSM is a vector format which uses tags to store information about geographical features such as roads, buildings, or water bodies. The data is organized into three elements: nodes, ways, and relations, each of which can be described using text tags comprised of key-value pairs. Nodes are used to represent point features, ways connect nodes, and relations help organize ways and/or nodes (Wiki OpenStreetMap, 2023). Elements are assigned tags – some tags may be used to explain the real-world function of a node or way (for example, 'school') or may describe the process of creating the element (such as user, uid, timestamp, version, changeset).

OSM data can be exported directly in small geographic areas from the OpenStreetMap website and opens directly (without transformation needed) in QGIS.

6.1.8 Spatial Databases

Spatial databases reach a higher level of complexity than individual data files, as they can store multiple datasets along with dataset relationships, behaviours, annotations, and data models, all of which are hosted in a relational database system. Spatial databases have played an increasingly prominent role in data production and management, while dataset-oriented formats are often still used for data distribution.

A variety of commercial database management systems, some using spatial extensions, can store geospatial data. These include Oracle Spatial and Graph (Oracle, 2023), Amazon Redshift (Amazon Web Services, 2021), IBM DB2 Spatial Extender (IBM, 2023), and Microsoft SQLServer (Microsoft, 2022). A prominent open-source option is the PostgreSQL-based PostGIS spatial database (PostGIS, 2023). These spatial extensions generally allow the user to store raster and vector data by adding spatial data types to the database that supports storing and querying of spatial data. Access to the spatial data in these databases can be directly through the database or, more commonly, through a connection to a desktop or web-based client.

Spatial databases have several features in common, including support for:

1) Continuous (large geographic extent) datasets.
2) Large volumes of data (raster and vector).
3) Complex data models (spatial data and business models).
4) Long transactions, multi-user editing, and versioning.

Increasingly, spatial databases also seek to store spatio-temporal data, a complex challenge that not all database products do well with.

These features make the long-term preservation of data in spatial databases much more complex, as it is often not possible to extract and transfer individual components of this data into other systems without losing some information. Preserving geospatial databases in general is likely to be particularly challenging, as all the problems of preserving relational databases are inherited: the need to take snapshots of running databases; storage of snapshots in proprietary database dump formats; complex dump formats; and large, monolithic sizes of snapshots.

6.1.9 Legacy Formats

6.1.9.1 ArcInfo Coverage/Export Files

The Esri ArcInfo Coverage file (Library of Congress, 2008) is a legacy proprietary vector data format which preceded the shapefile and was replaced by the geodatabase. ArcInfo Coverage files can include information such as topology and annotation, which cannot be directly transferred to shapefiles in data conversions. Coverage files are more difficult to manage than shapefiles because coverages have a multi-directory structure that makes the data susceptible to corruption in data transfers.

The .e00 format (sometimes known as an ArcInfo Export file) is a coverage export format that can store data related to a specific coverage (for example an Esri Grid or TIN) in a single data file. In instances where multiple interchanges files are necessary to contain a coverage product, the file extension may increase numerically (that is, .e01, .e02) (Library of Congress, 2020a).

ArcInfo Coverage and Export files are common to early GIS data projects and are likely to be encountered on data saved to floppy disks or CD-ROMs. ArcGIS Pro does not support the import or transformation of .e00 files, though early versions of ArcMap did. Esri has never officially released file format specifications for the ArcInfo coverage format, though outside analysis of the file has been performed and that documentation is available online (Morissette, 2000).

6.1.9.2 ArcSDE/Personal Geodatabases

The geodatabase as we know it today is technically termed a ‘file geodatabase’. Before the file geodatabase, there were two other forms: ArcSDE geodatabases and personal geodatabases. ArcSDE geodatabases store the data in a relational database management system (RDBMS) and support multiple users. By contrast, personal geodatabases are stored in Microsoft Access (.mdb) and cannot be larger than two gigabytes in size.

6.1.9.3 SDTS

An outcome of an early effort to define a means for open exchange of data was the Spatial Data Transfer Standard (SDTS), which was created in the 1990s by the United States Geological Survey (USGS) to support exchange of geospatial data. While the SDTS was used extensively by USGS and other US federal government agencies to distribute vector data and digital elevation model data in the mid-1990s, the format never gained traction in the wider geospatial data community and is no longer in active use.

The SDTS allowed the transfer of both data and metadata through modules, often encoded using ISO/IEC 8211. Typical file extensions include .ddf, and potentially further sequential extensions such as .ddg and .ddh (Library of Congress, 2020b).

6.1.9.4 National Transfer Format (NTF)

The National Transfer Format is officially British standard BS 7567 *Electronic transfer of geographic information* (NTF) (British Standards Institution, n.d.) and was primarily used by Ordnance Survey. NTF defines several levels of differing complexity that support different types of features and data, from Level 1 for simple vector features, to Level 5 which allows users to define their own data model and is used to transfer data such as Digital Terrain Models (DTM). The Ordnance Survey has discontinued its use of NTF data.

6.2 Raster

Raster geospatial data is comprised of continuous cell grid data. Such data can generally be characterized by the number of dimensions (most often two, but occasionally three); the number of bands (that is, the number of coincident layers); and the data type of the cell values in each layer (whether drawn from a continuous or discrete domain, or categorical in nature). For continuous and discrete data types, the data can be further characterized by the range (or ‘depth’) of the data values. For categorical data types, the category semantics (category labels such as ‘desert’ or ‘ocean’) are critical metadata without which the data loses meaning. Depending on the file format, such metadata may be stored with the raster data, in a separate data dictionary, or along with external metadata. Topology is not a concern with raster data since the relationships between cells are inherent in the raster itself.

Raster data is closely related to image data and many of the issues associated with the use and preservation of images pertain also to raster geospatial data. In some cases, raster data *is* imagery; that is, the raster bands contain c information such as visible wavelength radiances. In other cases, such as with elevation or bathymetric data, image formats provide a natural and convenient way to store such data. Therefore, many of the issues that arise with the preservation of image data apply to geospatial raster data as well. The evaluation factors identified (Library of Congress, 2022) for still images – clarity (resolution and bit depth) and colour maintenance – apply to raster data as well, though colour maintenance generalizes here to maintenance of the semantics of the data.

Because raster data is continuous over an area it can require several orders of magnitude more storage than the equivalent area represented through a vector data source. There is little that can be done about this. While it is possible to convert data from raster to vector representation (and vice versa), doing so is a highly analytic, lossy process that changes the essential character and functionality of the data. Thus, raster data generally must be preserved as raster data.

Compounding the problem of size is that automatic capture methods such as digital aerial photography and satellite remote sensing make it possible to quickly amass volumes of raster data that are large by any measure. For example, Sentinel-2, (Thales, 2016) an ongoing satellite observation and imagery program run by the European Space Agency, acquires two and a half terabytes of imagery per day alone. Raster data access mechanisms may impose additional storage requirements. For example, raster pyramids (.ovr or .rrd files created to support efficient panning and zooming of large images) add at least 30 per cent to the data size.

Consequently, in comparing raster data to vector, preservation of raster data is a quantitatively larger problem to such a degree that it is a *qualitatively* different problem. Large raster datasets will generally require custom-engineered storage and processing systems. If raster data is stored in a spatial database, the preservation problems due to size may compound the inherent migration and snapshot problems of preserving spatial databases. As with vector data, there are several formats in common use for raster data.

6.2.1 GeoTIFF

TIFF has emerged as a common format for storing and delivering raster data owing to its open standard (the standard is controlled by Adobe, but openly published and not subject to licence), its flexibility in describing multiple bands and data types, its extensible framework for embedded metadata ('tags'), and its popularity in the desktop publishing world. TIFF itself defines the semantics of a few tags; GeoTIFF (GeoTIFF, 2018) is an open standard that defines additional tags applicable to geospatial raster data, including complete coordinate reference information. Today GeoTIFF is arguably the most robust format for geospatial raster data, due to its widespread use and support.

GeoTIFF files can contain almost anything, from scanned and georeferenced analogue maps or map sets to born-digital cartographic products to raster datasets built for spatial analysis (such as population grid rasters or bathymetry files). Casual users unfamiliar with the potential of geo-enabled data may assume a GeoTIFF is simply a traditional TIFF file: indeed, GeoTIFFs will open in basic image software, though GeoTIFFs built for analysis will often have a file size of a gigabyte or more.

6.2.2 JPEG 2000

JPEG 2000 (JPEG. n.d.) (.jp2) is a standard that supports progressive, wavelet-based compression. It offers a wealth of other features, including lossy and lossless compression techniques, selective and adaptive compression. JPEG 2000 also allows arbitrary XML metadata to be embedded in image files, and the OGC has defined a standard for embedding GML documents in JPEG 2000 (Open Geospatial Consortium, n.d.-d). Exploiting the full capabilities of GML opens the possibility of embedding in image fields not just coordinate reference system information, but also coverage metadata, annotations, and even vector features. However, JPEG 2000 files are not as well-supported across software programs as GeoTIFFs are: for example, web browsers such as Google Chrome and Mozilla Firefox do not provide application support, though QGIS and ArcGIS Pro can read and write them.

6.2.3 GeoPDF

PDF is commonly used to provide end-user representations of data in which multiple datasets may be combined and other value-added elements may be included such as annotations, symbolization, and classification of the data according to data attributes (Adobe, 2021). While these finished data views, typically maps, can be captured in a simple image format, PDF provides some opportunity to add additional features such as attribute value lookup and toggling of individual data layers.

GeoPDF, which specifies a method for geopositioning of map frames within a PDF document, originated as a proprietary format developed by TerraGo Technologies, a strategic partner of Adobe. GeoPDF has proven to be a powerful format for presentation of complex geospatial content to diverse audiences who are not familiar with geospatial technologies, or audiences looking to use detailed geospatial navigation products in areas without traditional internet service, such as off-grid hiking or military use. The OGC publishes a 'Best Practices' document for GeoPDF standards (Graves and Reed, 2009). In parallel, Adobe introduced its own method for geo-registration into the ISO standards process for PDF.

The preservation challenges that apply to complex PDF documents will affect GeoPDF as well. While the PDF/A specification has been developed to define an archive-friendly version of PDF, some of the more advanced functionality that is used in geospatial implementations is not supported by the current PDF/A specification. The history of complex geospatial PDF documents is rather short, and risks associated with external dependencies (for example, fonts) and reliance on specialized software require close attention by the preservation community. Additionally, some GeoPDF files may not immediately be recognizable as having enhanced geospatial properties until they are examined in specialized software.

6.2.4 Stereo, Oblique, and Ground-Level Imagery

Discussion of raster data so far has focused on data that can serve as a representation of the earth's surface, and hence is suitable for projection and layering. Stereo and oblique imagery are types of imagery that are captured at varying angles to the vertical and are used to create stereo pair images and 3D models. Such imagery requires additional metadata to describe its 3D spatial orientation.

Ground-level photographs are not suitable for projection, but they can be point georeferenced. The EXIF (Japan Electronics and Information Technology Industries Association, 2002) metadata standard defines a means of capturing coordinate reference system information in JPEG files and is compatible with GPS systems that are often the source of such metadata.

6.2.5 NetCDF

NetCDF (Network Common Data Form) is a set of software libraries and self-describing formats used for storing and exchanging geographic data. It is an open-source format developed by the Unidata Program Center at the University Corporation for Atmospheric Research (UCAR). The netCDF file format structure is based on a hierarchical data model, where data is stored in a tree-like structure.

Each node in the tree contains a data set, and each data set is composed of a set of variables and attributes (Esri, n.d.-d).

NetCDF files are most often used in conjunction with GIS software when a dataset is comprised of multiple dimensions; for example, data measured at varying elevations at multiple locations, and taken at routine time intervals. Climate and weather data are the most common use-cases in a geographic context due to the huge amount of data that can be generated very quickly when a large area is measured at monthly, weekly, or even daily intervals. NetCDF provides an efficient storage mechanism for 'stacking' dimensions that describe the same geographic area as a time-series. Esri's 'Space-Time Cubes' are stored as NetCDF files (Buie, 2020). To be opened in GIS software, a software user typically needs to define the specific time (within a time-series) of the data they would like to view.

6.2.6 Cloud Raster Format

The Cloud Raster Format (.crf) is an Esri-developed raster format for large raster files which is comprised of smaller tile bundles packaged together for distributed processing. CRF files can be created from NetCDF files and support multi-dimensional data.

6.2.7 Legacy Raster Formats

Early GIS software utilized a number of simple raster formats, some dating back to the days when data was read directly from tape drives. Examples include BIL (band interleaved by line), BIP (band interleaved by pixel), and BSQ (band sequential), all formats for multi-band raster data, though they could also be described as generic data organization techniques employed by formats. ArcInfo ASCII GRID and USGS DEM (US Geological Survey, n.d.) are other examples of simple, open, ASCII formats for single-band raster data. Each format lists raster cell values in left-to-right and top-to-bottom order, augmented with georeferencing information in the header and/or trailer records. These formats still find use in converting and processing raster data. From a preservation perspective, the main challenge of simple raster formats will increasingly be verifying support for older formats in current-day software packages. Esri maintains a useful list of raster file formats with notations as to the level of support provided by ArcGIS Pro and includes a short description for each in a long list of file types, many of them legacy (Esri, n.d.-e).

6.3 Dynamic Web Services

Geospatial web services allow end-user applications and server applications to make requests for sets of data over the web. Requests might also be made for specific data processes, such as finding a route or locating a street address.

In web service client applications, data is drawn from one or multiple sources and presented in map form to the user. These mapping environments take the burden of data acquisition and processing away from the user. While it is typically possible for the user to save service state (such as map area or zoom level), it is usually not possible to save the state of the data within the service, creating a substantial preservation challenge for capturing such interactions. An example of this is the Cooper Hewitt, Smithsonian Design Museum's additions of Stamen Design's Watercolor Maptiles to its permanent collection in 2021. Cooper Hewitt collected 56 million map tiles (in PNG format) in addition to maintaining a live, duplicate website as a 'Smithsonian version' of the data layer (Smithsonian, 2021).

Due to the timely nature of the data in web services and APIs, useful and frequently updated products are increasingly accessible to make, and serve an important function. At the same time, new challenges in maintaining data persistence are also created. In some cases, the availability of web services-based access to data has decreased the incentive to replicate data resources to additional locations that might otherwise retain copies of the data.

6.3.1 Web Map Service (WMS)

The Web Map Service (WMS) is a lightweight web service at the core of which is the 'Get Map' request, which allows a client application to request an image representation of a specific data layer (often, but not exclusively, PNG, GIF, or JPEG). The OGC WMS 1.0 specification was released in 2000 and by virtue of its simplicity, gained wide adoption and vendor support. Requests can be made from individual clients such as desktop GIS software, web browsers, as well as other map servers which might blend data sources from several different servers. The Web Map Context specification was developed by the OGC to formalize how a specific grouping of one or more maps from one or more map servers can be described in a portable, platform-independent format. The Styled Layer Descriptor (SLD) profile of the Web Map Service provides a means of specifying the styling of features delivered by a WMS using the Symbology Encoding (SE) language. If preservation of the cartographic representation of a map delivered by a WMS is important, then it may be necessary to preserve the associated SLD (if there is one). WMS tiling efforts have come as a response to the experience of Google Maps and other commercial map services, which demonstrated the speed with which static tiled imagery could be presented in user applications. Efforts to develop a standard approach to provide access to static map tiles resulted in the creation of the OGC-produced Web Map Tile Service (WMTS) Implementation Standard (Open Geospatial Consortium, n.d.-e). The last WMS update, 1.3.0, was released by OGC in 2006 (Beaujardiere, 2006).

6.3.2 Web Feature Service (WFS)

The Web Feature Service allows for a client application to request vector geographical features at a detailed level. Where WMS returns images, WFS streams vector data with geometry and attributes (typically, but not exclusively, in GML format). It allows for a range of operations, including querying, editing, and updating of vector features. First released as a standard by the OGC in 2002, it currently defines 11 operations: GetCapabilities, DescribeFeatureType, GetPropertyValue, GetFeature, GetFeatureWithLock, LockFeature, Transaction, CreateStoredQuery, DropStoredQuery, ListStoredQuery, and DescribeStoredQueries (Open Geospatial Consortium, n.d.-f). It is compatible with some desktop GIS software programs, such as ArcGIS Pro and QGIS.

6.3.3 Web Coverage Service (WCS)

A Web Coverage Service allows a client application to request a spatio-temporal geographic coverage. It is similar in function to a WMS and a WFS but returns data in a coverage format which is useful for content such as satellite imagery, digital aerial photos, digital elevation data, or multidimensional raster data. WCS allows for coverage in a range of formats, including (but not limited to) GeoTIFF or GML.

6.3.4 Web Map Tile Service (WMTS)

A Web Map Tile Service (WMTS) utilizes pre-defined map tiles to a client application. As mentioned earlier (see 6.3.1), the OGC published the first set of WMTS specifications in 2010. WMTS tiles have a set resolution and extent. Where WMS utilizes a large single map image for each request, WMTS serves tiled map images which can be pre-rendered by a server and stored as a cached image by the client application.

6.3.5 Vector Tile Services and APIs

Vector Tile APIs have gained substantial popularity as a method for serving geospatial data on the web as web services have become increasingly API driven. Popular web platforms such as Mapbox (Mapbox, n.d.), ArcGIS Online (Esri, n.d.-f) and Ordnance Survey basemaps make use of Vector Tile Services, and the reliability and currency of real-time data often make these service layers invaluable. Vector Tile Services are an output format comprised of vector features layers which can be symbolized and adapt to display device resolution. The OCG has produced a Tile standard which documents specifications for today's 'resource-oriented Web API' style web service (Masó and Jacovella-St-Louis, 2022).

7 Geospatial Format Registries and Resources

Format registries support preservation by maintaining knowledge of file formats. The following subsections detail active major format registries that include documentation related to geospatial formats.

7.1 Sustainability of Digital Formats

The Sustainability of Digital Formats website is an online resource documenting digital content formats that is maintained by the Library of Congress. The website has been in operation since 2004 and is regularly updated. Its goal is to support strategic planning of digital formats for long-term preservation, identify emerging file formats, and describe the long-term sustainability of individual file formats. Though the resource is outward-facing and available to the public, the regular updating of the resource is intended to benefit and inform internal decision-making at the Library of Congress.

Each file format contained in the Sustainability of Digital Formats website seeks to document file formats across eight factors relevant to the long-term sustainability of the format: disclosure, adoption, transparency, self-documentation, external dependencies, impact of patents, and technical protection mechanisms. File formats can be browsed by content categories, including a geospatial content category (Library of Congress, 2023c) which lists all available geospatial file types documented by the website.

The geospatial section of the website grew out of the National Geospatial Digital Archive project (see Section 11.1.1), though geospatial format file registry details were enhanced by contractors of the Library of Congress (Fleischhauer, 2011) and continue to grow over time.

7.2 Recommended Format Statement

Library of Congress maintains a Recommended Format Statement (RFS) which organizes digital file formats by content type, then further groups subject-specific file formats into 'preferred' and 'acceptable' collecting categories. These categories are based on the assessed long-term sustainability of the format as determined by teams of technical subject-matter experts within the Library. The RFS was developed as an internal tool but is made available on the Library of Congress website for use by other institutions as an external resource. The RFS is updated on an annual cycle by the Library and is sent out for public comment and feedback each spring, allowing for revision. The goal of the RFS is to make sure Library collection development policies use digital formats best poised for long-term survival and accessibility. The RFS can be understood in part as an applied use of the Sustainability of Digital Formats listing, which documents sustainability factors and technical specifications of individual file formats.

Geospatial file formats are addressed in the 'GIS, Geospatial, and Non-GIS Cartographic' (Library of Congress, n.d.) section of the RFS, with subcategories detailing GIS Vector Data, GIS Vector and Raster Combined formats, GIS Raster and Georeferenced Images, and Non-GIS Cartographic specifications. Each subcategory describes preferred and acceptable collecting formats, preferred delivery method, metadata standards, and any preferred technological measures.

7.3 PRONOM

Developed and maintained by the Digital Preservation Department of The National Archives (UK), PRONOM is an online resource detailing the technical details of both data formats and software programs. It is intended to support long-term preservation of electronic files and records and serves as an authoritative resource on the technical details of file formats and software products. Though PRONOM does not provide a direct list of geospatial file formats and software, the web resource is searchable by file type (The National Archives, n.d.) and does contain information on some geospatial file formats, though there are also major omissions.

PRONOM contains a written description of each file format with additional fields, including date released, date supported until, the format risk, who the file was developed and supported by, a source date and description, and date last updated. These additional fields are largely not filled out on geospatial data entries. PRONOM also allows for searching of software by name or by vendor name.

7.4 Geospatial Software and File Format Documentation Web Archive

In 2023, the Geography & Map Division of the Library of Congress launched their Geospatial Software and File Format Documentation Web Archive, which seeks to capture software documentation and file format specifications relevant to the long-term preservation of geospatial data and which document the development and evolution of geographic information systems (GIS) technology. This includes, but is not limited to, versioned documentation relating to GIS and remote sensing software and online applications used for the creation, retrieval, analysis, transformation, and display of geospatial data. It also covers white papers, technical documentation, and geospatial community commentary as related to the geospatial file formats included in, but not limited to, the Library's Geospatial Recommended Format Statement (Library of Congress, n.d.) and the Sustainability of Digital Formats geospatial data formats list (Library of Congress, 2023d). This web archive will seek a one-time capture of static sources of documentation, as well as the ongoing capture of any live documentation which continually evolves, thus aiding the documentation of the history of cartographic production and GIS technology. The web archive is currently being built and available only on-site at the Library of Congress, though some of its archive is expected to be available off-site starting in the summer of 2024.

8 Technology and Tools

The diverse universe of geospatial file formats is accompanied by a wide range of specialized geospatial software often required to read and write geospatial file formats.

8.1 ArcGIS

ArcGIS is a widely used commercial geographic information system (GIS) software suite developed and maintained by Esri, a privately owned corporation based in Redlands, California. First launched in 1999, it is a suite of software applications which allows users to create, manage, analyse, and visualize geographic information and spatial data. Over time, the primary desktop GIS software program at the core of the Esri suite has evolved from ArcInfo to ArcMap, and more recently to ArcGIS Pro. ArcMap 10.8.1 is the final release of the ArcMap software program, which will be fully discontinued from March 2026 (Angel and Hynes, 2020). ArcGIS products work only on Windows-based operating systems and require a paid licence for operation.

The cloud-hosted extension of ArcGIS desktop software is ArcGIS Online, a web platform for hosting geospatial data layers, creating web maps, building web mapping applications, and publishing multi-media projects which may utilize web maps (such as the ArcGIS Online product StoryMaps). ArcGIS Online also offers the ability to create open data sites (such as through their ArcGIS Online Hub platform) that are commonly used by state and local government agencies, as well as by some federal agencies, as geospatial data-sharing platforms. ArcGIS Online offers a limited free public account for non-commercial use which allows a user to create and store a limited amount of content. For advanced use and analysis, a paid account is required.

8.2 QGIS

QGIS is the leading open-source GIS desktop application for the creation, visualization, and analysis of geospatial data with an OGC Web Server application, a web browser client, and developer libraries. QGIS works on both Windows, Linux, and OS operating systems, and is licensed under the GNU General Public License as an official project of the Open Source Geospatial Foundation. The QGIS website provides direct software downloads, as well as software documentation and tutorials for use.

8.3 GRASS GIS

Geographic Resource Analysis Support System (GRASS) GIS is an open-source software program that operates under the GNU Public License and works on Windows, OS, and Linux operating systems. First started in 1982, GRASS was developed by the US Army Corps of Engineers' Construction Engineering Research Laboratory, though development work was eventually transferred to other groups. GRASS GIS is a founding member of the Open Source Geospatial Foundation. The most recent version of GRASS GIS is 8.2.1, which was released in January 2023 (Neteler, 2023).

8.4 TerrSet

TerrSet (formerly known as IDRISI), is a software program developed by Clark Labs designed for spatial analysis, earth systems modelling, and image processing, and requires a paid licence to operate. The first version of the software was released in 1987 by Dr J. Ronald Eastman of the Clark University Department of Geography. IDRISI developed the IDRISI Raster Format (.rst) file format, which is native to the software.

TerrSet's Land Change Modeler (LCM) has tools built specifically for modelling land change, such as the Reducing Emissions from Deforestation and Forest Degradation (REDD), which can model the effect of land cover change on carbon emissions. Clark Labs developed two version of their Land Change Modeler (LCM) extension for ArcGIS (Clark Labs, 2014).

The most recent version of TerrSet is TerrSet 2020.

8.5 ERDAS IMAGINE

ERDAS IMAGINE is a paid software program currently owned by Hexagon AB (Hexagon, n.d.). Primarily a raster processing program, it supports remote sensing and photogrammetry, LiDAR analysis, radar processing, orthorectification, image interpretation, and basic vector data processing. ERDAS was first released in 1978 (as ERDAS 4) and found traction with government agencies. In 1988

ERDAS released an interoperable link to Esri's ArcInfo (Beaty, 2009). A few file formats were developed, such as ERDAS 7.5 GIS (.gis), ERDAS 7.5 LAN (.lan), and ERGAS IMAGINE (.img), which have mixed levels of support in other GIS software programs.

8.6 Web-Based Geospatial Platforms

A number of web-based platforms host geospatial data for the purpose of making web maps. Increasingly, these platforms also offer data analysis tools (competing with traditional desktop GIS) and may serve as a data discovery tool within a group or organization. Examples include ArcGIS Online, Mapbox, and CARTO. Data uploaded into web-based platforms are typically converted into new file formats by the platform itself. ArcGIS Online converts vector data into hosted feature layers and Mapbox renders data in Mapbox Vector Tiles. For more information, see 6.3).

8.7 File Format Transformation Tools

Given the large range of raster and vector file formats, it is not unusual to convert spatial data between formats for practical or preservation purposes. Sometimes this transformation occurs during routine spatial analysis operations: for example, ArcGIS Pro defaults all output files to become a feature class in a geodatabase, even when the input data is a shapefile. On the other hand, if an openly documented format is desired for preservation purposes, one may want to convert from geodatabase feature class to shapefile.

Options for conversion between common commercial formats exist as built-in features within desktop GIS software, sometimes as a function provided by open-source conversion tools. Due to the complexity of the data, migration from a proprietary or poorly supported data format into another, more preservation-friendly, format can lead to distortion or loss of data. It is up to the best judgment of the person preserving a dataset to weigh the pros and cons of a desired file format against potential data loss. The key to making this judgment call is often having good information about what loss can occur. Tool sets are available to help in this process; one such example is the Geospatial Data Curation Toolkit (Kernik, 2023), which provides an automated method to check if a geodatabase has functionality that would be lost in conversation to shapefile.

8.7.1 FME Software

Feature Manipulation Engine (FME) is an advanced data transformation and manipulation platform developed by Safe Software Inc., first released in 1996. A licensed software program, its primary function is the transformation and conversion of geospatial data from one file format to another, with data readers and writers available across both raster and vector formats. It also allows users to transform and integrate data from multiple sources into a single output format and allows for the creation of automated workflows for data transformation and processing. FME also provides data validation and verification workflows specific to spatial data, such as topology and geometric inconsistency validation.

8.7.2 GDAL

GDAL is a translator library for raster and vector geospatial data formats released by the Open Source Geospatial Foundation under an X/MIT open-source licence. It is widely used for reading, writing, and manipulating geospatial data, including satellite imagery, aerial photography, digital elevation models, and vector data. Capable of both reading and writing over 200 raster and vector formats, GDAL can be used for geospatial file processing. Its website provides a comprehensive list of supported raster and vector drivers. GDAL can be integrated with FME software, QGIS, and other GIS desktop applications.

9 Standards Bodies and Working Groups

The following is a selection of international standards bodies and working groups that are addressing the issues of geospatial data preservation.

9.1 Open Geospatial Consortium (OGC)

The Open Geospatial Consortium (OGC) (formerly the Open GIS Consortium) is an international industry consortium of companies, government agencies and universities that work together to develop publicly available interface specifications. It operates with a goal of making spatial data findable, accessible, interoperable, and reusable. Some notable members include the European Space Agency, Ordnance Survey, NASA, USGS, the US National Geospatial-Intelligence Agency, Google, Microsoft, Esri, and Amazon Web Services.

OGC specifications support interoperable solutions that 'geo-enable' the Web, wireless and location-based services, and mainstream information technology. It has published more than 30 standards (Open GeoSpatial Consortium, n.d.-g), including specifications for GML, KML, GeoPackage, WCS, WMS, WMTS, and 3D Tiles. The OGC has a close relationship with ISO/TC 211, which addresses standardization in the field of digital geographic information, and a subset of OGC standards are now ISO standards. It also works with other international standards bodies such as W3C, OASIS, WfMC, and the IETF.

9.2 US Federal Geographic Data Committee (FGDC)

The FGDC is a United States federal interagency committee empowered by the Geospatial Data Act of 2018 (Federal Geographic Data Committee, 2018) (see 10.3.1) to oversee the coordinated development, use, sharing, and dissemination of geospatial data within the United States federal government. A wide range of stakeholder organizations participate in FGDC activities representing the interests of state and local government, industry, and professional organizations. The FGDC also directs the National Spatial Data Infrastructure (NSDI), which maintains technology, standards, and policies to promote geospatial data among all US government agencies (at all levels) and institutions of higher education.

9.3 International Organization for Standardization (ISO)

The International Organization for Standardization (ISO) is an international non-governmental organization which develops and maintains standards for technical and industrial products and services. ISO has developed several standards related to geospatial data, including the ISO 19115 standard for geographic information metadata, the ISO 19109 standard for geographic information services (application schema), and the ISO 19136 standard for geographic information data processing. These standards provide the framework for the exchange and use of geospatial data. ISO 19115 was first finalized in 2003 and reviewed and confirmed as recently as 2019 (International Organization for Standardization, 2014).

10 Legal Landscape

The legal framework within which geospatial data is made available is a key component of our ability to preserve and make use of geospatial data in the future. Frequently, intellectual property rights in geospatial data are often protected or not documented by the rights-holder. Many geospatial datasets originate with an underlying dataset licensed from a third party, such as a mapping agency or satellite imagery supplier. This means that many geospatial datasets have an implied dependence on a third-party supplier who may take a view on preservation and access. Consequently, archivists and repository managers would be well advised to examine the licences under which data is presented to them to understand how the data can be used and redistributed in the future.

10.1 Open Geospatial Data

As a response to the complex licensing issues arising from geospatial data produced by national and local governments, private companies, and others, there is a strong and growing culture for availability, openness, and transparency in licensing geospatial data, including making data more accessible and with less restrictive licensing terms. Examples include the US federal government's GeoPlatform, which 'embodies the principles and spirit of Open Government, emphasizing government-to-citizen communication, accountability, and transparency'.

Several licensing initiatives have been created, including Creative Commons and the Open Data Commons Licences (Open Knowledge Foundation, n.d.-b), that aim to achieve these goals. These licences let data creators specify less restrictive licensing conditions up to, and including, putting the work into the public domain. Initiatives such as OpenStreetMap have adopted this approach, with user-contributed geospatial data currently being licensed under an Open Database Licence (ODbL). It is increasingly common to see research data products licensed under an open content framework. When data creators assign a specific open licence to their products, it enables data users and data archivists to understand in clear terms exactly how that data can be used, stored, cited, and redistributed.

10.2 United Kingdom Legal Landscape

The United Kingdom has historically had significant legal issues around the access and re-use of geospatial data, particularly data that is produced by, or on behalf of, government agencies and protected by Crown Copyright. Ordnance Survey data has at times been very restricted, requiring a paid licence which entitled a user access to Ordnance data for a period. More recently, the Ordnance Survey has launched OS Open Data, which provides a set of geospatial datasets under an Open Government Licence, allowing for full use of the data with attribution to Crown Copyright.

Preservation of Ordnance Survey data for the long term is carried out by The National Archives (TNA) (Kowal and Holmes, 2014). However, the UK Legal Deposit Libraries have an agreement with Ordnance Survey whereby they receive an updated snapshot copy every year of detailed mapping, including OS MasterMap. The Legal Deposit Libraries provide a facility whereby users in the libraries can view contemporary and historic versions of OS MasterMap (National Library of Scotland, n.d.) (British Library, n.d. -a) and Land-Line (the precursor dataset of OS MasterMap) going back to 1998 as online mapping and to print out small extracts.

10.2.1 INSPIRE

INSPIRE (Infrastructure for Spatial Information in Europe) is an initiative of the EU 'to trigger the creation of a European spatial information infrastructure that delivers to the users integrated spatial information services'. INSPIRE was established formally through Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007. One of the first deliverables of the INSPIRE initiative was the development of regulations and rules regarding the implementation of geospatial metadata to describe relevant datasets.

After the UK formally left the European Union, a Statutory Instrument was introduced, *The INSPIRE (Amendment) (EU Exit) Regulations 2018* (INSPIRE, 2018), to allow for continued operability with EU spatial data infrastructure standards. Currently, INSPIRE requires that relevant geospatial data is published using a GEMINI metadata standard (Data.gov.uk, n.d.).

10.3 United States Legal Landscape

In the United States, geospatial data produced by federal agencies is openly available, as works by the US government are not eligible for copyright protection. Federal agencies produce large volumes of high-quality geospatial data, much of which is global in scope, such as NASA's public domain satellite imagery and climate data.

While public agency data is typically in the public domain, there are a few rights-related issues that can complicate preservation. Public Records Law varies from state to state, and even within a single state interpretation may vary widely. Restrictions on commercial use or resale of data can result in restrictions on open secondary redistribution of that data. In general, there has been a trend towards more open access to data in recognition of the positive societal benefit that derives from free data access, and the negative burden on local agencies related to mediated or fee-based data request handling. In 2005, the National Spatial Data Infrastructure (NSDI) created guidelines to help identify dissemination of geospatial data products that could pose a risk to national security (Federal Geographic Data Committee, 2005).

10.3.1 Geospatial Data Act of 2018

The Geospatial Data Act of 2018 (GDA) is a foundational piece of US national legislation that establishes a comprehensive framework for the collection and use of geospatial data by the US federal government. The GDA establishes clear responsibilities for federal agencies in managing geospatial data, including required annual reporting and evaluation from all covered agencies. It also mandates reporting from the FGDC to Congress based on individual agency reports. The GDA is designed to modernize the federal government's approach to geospatial data, ensuring that the most up-to-date and accurate data is used to inform decision-making (Folger, 2018).

The GDA affirms the FGDC as 'the lead entity in the executive branch for the development, implementation, and review of policies, practices, and standards relating to geospatial data' and outlines 13 key duties for the FGDC to undertake. Among these duties is the role of establishing geospatial data standards for the federal government, reviewing agency compliance with said standards, managing data themes, ensuring the operation of GeoPlatform, and '… promot[ing] cost-effective data collection, documentation, maintenance, distribution, and preservation strategies'. Implementation of the GDA to date has focused largely on the creation of an agency reporting structure without much detail yet available as to long-term preservation plans or strategies of individual agencies.

10.4 Australia and New Zealand Legal Landscape

In Australia and New Zealand, the ANZLIC Committee on Surveying & Mapping (ICSM) provides leadership in the management and use of spatial data, including the development of national standards.

As in the United Kingdom, Australia has historically seen Crown Copyright asserted over government-produced spatial data (Alexander and Jankowska, 2017). For example, in 2009, Google Australia spoke out about Crown Copyright interfering with its ability to overlay real-time data from the Commonwealth Fire Authority into Google Maps (Braue, 2009). In 2010, the Australian federal government released a *Government 2.0 Task Force Report* which recommended that all Australian public sector information (PSI) be released under a Creative Commons framework by default (Park, 2010). Currently, datasets produced by the Australian Bureau of Statistics, data.gov.au, and Geoscience Australia are all offered under a Creative Commons framework (for information on governments and Creative Commons, see https://creativecommons.org/Government). In New Zealand, the government also provides its geospatial datasets under Creative Commons attribution (Toitū Te Whenua Land Information New Zealand, n.d.).

11 Discontinued Preservation Activities

The early 2000s and 2010s saw several short-term projects related to geospatial data preservation, some of which were funded by the National Digital Information Infrastructure and Preservation Program (NDIIPP), a Library of Congress project. This section details a short history of previous geospatial preservation groups that have since been disbanded.

11.1.1 Managing and Preserving Geospatial Electronic Records (GER):

The Managing and Preserving Geospatial Electronic Records project (Center for International Earth Science Information Network, n.d.) was funded through the National Historical Publications and Records Commission of the US National Archives and Records Administration (NARA) in 2003 and managed by the Center for International Earth Science Information Network (CIESIN) of Columbia University. The goal of the project was to examine requirements for electronic records generated by GIS software, as well as to identify recommended policies, standards, and practices.

The project resulted in publication of the *Data Model for Managing and Preserving Geospatial Electronic Records* (Center for International Earth Science Information Network, 2005), which provides recommendations regarding retention of metadata and related information to support the management and preservation of geospatial data records.

11.1.2 National Geospatial Digital Archive (NGDA)

The National Geospatial Digital Archive (NGDA) was a project between University of California-Santa Barbara's Map and Imagery Laboratory and Stanford University's Branner Earth Sciences Library that began in 2004 with National Digital Information Infrastructure and Preservation Program (NDIIPP) funding. The goal of the project was to create an archive of at-risk geospatial data and to develop and document best practices for geospatial data collection.

Both libraries collected geospatial data in line with their library's existing collection development policies and collecting strengths and suggested that the size and scope of geospatial data collection would require 'cooperative strategies' among library and other archives in order to succeed (Erwin *et al.*, 2009). The project also noted the huge importance of file format to the long-term success of preservations efforts and created a file format registry to document format specifications. This registry formed the geospatial section of the Library of Congress's 'Sustainability of Digital Formats' resource (see also 7.1) (Directions Magazine, 2012).

11.1.3 North Carolina Geospatial Data Archiving Project (NCGDAP)

The North Carolina Geospatial Data Archiving Project (NCGDAP) ran from 2004–2009 as a partnership between North Carolina State University Library and the North Carolina Center for Geographic Information and Analysis, with a focus on preservation of state and local agency digital geospatial data. Funding for the project came from NDIIPP. NCGDAP was carried out as a component of the NC OneMap initiative, which focused on cultivating seamless access to state, federal, and local data covering the state. Objectives for the program included acquiring at-risk geospatial data, developing a digital repository architecture for spatial data, and the creation of preservation metadata.

11.1.4 Geospatial Multistate Archive and Preservation Partnership (GeoMAPP)

The Geospatial Multistate Archive and Preservation Project (GeoMAPP) ran from 2007 to 2012 as a partnership between the state geospatial agencies and state archives of North Carolina, Kentucky, and Utah, and was funded through NDIIPP to test the findings of NCGAP (Lazorchak, 2012). In addition to directly addressing the selection, appraisal, and preservation of at-risk geospatial data, GeoMAPP focused on engagement of state archives within the spatial data infrastructure of each respective state. One of GeoMAPP's main arguments was that preservation of data within government could only be possible if the value of preserving that data was immediately apparent (the data must have a use-value). The project drafted a business use document for preserving geospatial data (Bethune, Lazorchak and Nagy, 2009).

11.1.5 Geospatial Data Preservation Resource Center (GDPRC)

The Center for International Earth Science Information Network at Columbia University carried out an NDIIPP-funded project, the Geospatial Data Preservation Resource Center (GDPRC). The goal of the Center was to serve as a central resource for those interested in the preservation of geospatial data. GDPRC launched a website in 2011, which provided resources for education and training, tools and software, and policies and benefits of collecting geospatial data. The GDPRC website is still online but has not been updated since 2012.

11.1.6 FGDC Users/Historical Data Working Group (UHDWG)

The FGDC Users/Historical Data Working Group was established to build awareness in geospatial communities about the ephemeral nature of geospatial data, especially among US federal agencies. The role of the Working Group was to 'promote an awareness among federal agencies of the historical dimension to geospatial data; to facilitate the long-term retention, storage, and accessibility of selected historically valuable geospatial data; and to establish a mechanism for the coordinated development, use, sharing, and dissemination of historically valuable geospatial data which have been financed in whole or part by Federal funds'. The group also encouraged federal agencies to deposit geospatial datasets with NARA and the Library of Congress.

The Working Group played a coordinating role in the development of a Historical Collections community within the national Geospatial One Stop Portal (Federal Geographic Data Committee, 2015).

11.1.7 OGC Data Preservation Working Group

In December 2006, the Open Geospatial Consortium's Data Preservation Working Group was formed 'to address technical and institutional challenges posed by data preservation, to interface with other OGC working groups that address technical areas that are affected by the data preservation problem, and to engage in outreach and communication with the preservation and archival information community'. A goal of the group was to 'create a dialogue with the broad spectrum of geospatial community and archival community constituents that have a stake in addressing data preservation issues'. The work of the group focused on identifying points of intersection between data preservation issues and OGC standards efforts, and to introduce temporal data management use-cases into OGC discussions.

12 Conclusions and Recommendations

Preserving geospatial data is an ongoing challenge that is only increasing in complexity over time. For micro-decision making, there is no one single best approach to preserving geospatial data. Each decision will be contextual, and depend on the type of geospatial data produced, the format it is stored in, the collection development policy of the institution, and the technical solutions available for opening, storing, and/or converting datasets. That said, geospatial data is valuable and faces similar risks and vulnerabilities as other types of data do. Some risks can be offset by the adoption and adaptation of good practice for preservation, and geospatial data strategies need to be incorporated into the mainstream of digital preservation planning. Specific actions to be considered are:

12.1 Formats

- Vector data
    - Retain data in its original format. If the data is produced in more than one format, consider collecting both formats.
    - If proprietary, legacy, or not widely supported, consider migration into widely supported (and openly documented) format.
- Raster data
    - Retain data in its original format. If the data is produced in more than one format, consider collecting both formats.
    - If proprietary, legacy, or not widely supported, consider migration into widely supported (and openly documented) format and compression scheme.
    - If possible, retain pre-processed and processed data.
- Spatial databases
    - Where databases are continually updated, save routine snapshots.
    - Employ ability to record and access all edits to a database when possible.
    - Extract individual datasets (for example, feature classes) into stable format.
- Dynamic data and web services
    - Take snapshot copies of data and service state and save locally.
    - Consider snapshot copies of any associated cartographic web mapping applications or cartographic products (that is, how the rendered data is symbolized and interacted with).

12.2 Metadata

- Maintain technical and administrative metadata in addition to geospatial metadata.
- Implement ISO descriptive keywords and ISO 19115X metadata.
- Retain original metadata and remediate or normalize metadata where possible.
- Where appropriate, describe geospatial datasets in clear language suitable for non-specialists.

12.3 Systems

- Keep archival data in live access systems.
- Provide access to superseded datasets.
- Avoid ‘atomization’ of data in digital repository systems.
- Capture data as well as cartographic representations deemed of value.
- Maintain independence of data from specific storage/repository environment.

12.4 Legal

- Secure archival rights and rights for access to older data and make sure incoming collections have well-documented rights.

- Develop appropriate rights mechanisms so that future users of the data can be presented with suitable background information.

12.5 Community Actions

- Promote the importance of preserving geospatial data to wider audiences.
- As geospatial preservation actions take form in individual libraries and archives, communicate collection development strategies around geospatial data to the wider community.
- Consult format registries and pay attention to file format when collecting data.
- Work with the data producer community to cultivate best practices for creation of metadata and frequency of capture of key data layers.
- Obtain and document data rights from donors to the maximum extent possible.

13 Glossary / Terminology

Acronym	**Meaning**
API	Application Programming Interface
DEM	Digital Elevation Model
DPC	Digital Preservation Coalition
FGDC	Federal Geographic Data Committee
FME	Feature Manipulation Engine
GDA	Geospatial Data Act of 2018
GDAL	Geospatial Data Abstraction Library
GIS	Geographic Information System
GML	Geography Markup Language
INSPIRE	Infrastructure for Spatial Information in Europe
ISO	International Organization for Standardization
KML	Keyhole Markup Language
METS	Metadata Encoding and Transmission Standard
NARA	National Archives and Records Administration
NCGDAP	North Carolina Geospatial Data Archiving Project
NDIIPP	National Digital Information Infrastructure and Preservation Program
NGDA	National Geospatial Digital Archive
NTF	National Transfer Format
OGC	Open Geospatial Consortium

OS	Ordnance Survey (GB)
OSM	OpenStreetMap
PREMIS	Preservation Metadata Implementation Strategies
NSDI	National Spatial Data Infrastructure
SDTS	Spatial Data Transfer Standard
USGS	United States Geological Survey
WCS	Web Coverage Service
WFS	Web Feature Service
WMS	Web Map Service
WMTS	Web Map Tile Service

14 Case Studies

As archives and libraries begin to implement geospatial data preservation policies, a few are documenting their process for the benefit of the wider community. Recommendations for two of those case studies follow:

A Proposed Geospatial Data Preservation Strategy for DOE's Office of Legacy Management, December 2021. Denise R. Bleakly, Joshua Linard, and Matthew J. Cuneo. https://www.osti.gov/servlets/purl/1855335

A report detailing the development of a geospatial data preservation strategy by the Office of Legacy Management within the US Department of Energy. Bleakly *et a*l. address both key management responsibilities and summarize their recommended best practices.

Placing Data in the Land of 10,000 Lakes: Navigating the History and Future of Geospatial Data Production, Stewardship, and Archiving in Minnesota, January 2016. Kevin Dyke, Ryan Mattke, Len Kne, and Shawn Rounds. Journal of Map & Geography Libraries, 12 (1), pp. 52–72. https://doi.org/10.1080/15420353.2015.1073655.

A short history of geospatial data preservation efforts in the state of Minnesota, followed by an outline of a potential data preservation framework for collecting based on the state's existing strengths and resources.

15 References

Adobe (2021) *Acrobat Developer Resources*. Available at: https://web.archive.org/web/20230707072144/https://opensource.adobe.com/dc-acrobat-sdk-docs/

Alexander, I., and Jankowska, M. (2017) *Rights in Geospatial Information: A Shifting Legal Terrain*. Melbourne University Law Review 41 (3) pp. 957–98. Available at:

https://web.archive.org/web/20230105161234/https://law.unimelb.edu.au/__data/assets/pdf_file/0010/2771497/01-Alexander-and-Jankowska.pdf

Amazon Web Services, Inc. (2021) *Amazon Redshift Announces Native Support for Spatial GEOGRAPHY Datatype*. Available at: https://web.archive.org/web/20230204013746/https://aws.amazon.com/about-aws/whats-new/2021/11/amazon-redshift-spatial-geography-datatype/

Angel, J. and Hynes, W. (2020) ArcGIS Blog, Esri. *ArcMap Continued Support*. Available at: https://web.archive.org/web/20230531184740/https://www.esri.com/arcgis-blog/products/arcgis-desktop/announcements/arcmap-continued-support/

Association for Geographic Information (n.d.) *UK GEMINI*. Available at: https://web.archive.org/web/20230309012707/https://www.agi.org.uk/uk-gemini/

Beaty, P. (2009) *The Field Guide: A Brief History of ERDAS IMAGINE*. The Field Guide (blog), April 15, 2009. Available at: https://web.archive.org/web/20230731100918/http://field-guide.blogspot.com/2009/04/brief-history-of-erdas-imagine.html

Beaujardiere, J. de la (2006) Open Geospatial Consortium, Inc. *OpenGIS® Web Map Server Implementation Specification*. Available at: https://web.archive.org/web/20230720071154/https://portal.ogc.org/files/?artifact_id=14416 (opens as a pdf)

Bethune, A., Lazorchak B., and Nagy, Z. (2009) *GeoMAPP: A Geospatial Multistate Archive and Preservation Partnership*. Journal of Map & Geography Libraries 6 (1), pp. 45–56. Available at: https://doi.org/10.1080/15420350903432630 [accessed 20 July 2023]

Bleakly, D., Linard, J. and Cuneo, M. (2021) *A Proposed Geospatial Data Preservation Strategy for DOE's Office of Legacy Management*. Sandia National Lab. (SNL-NM), Albuquerque, NM. Available at: https://web.archive.org/web/20230731100732/https://www.osti.gov/biblio/1855335/

Braue, D. (2009) 'Australian government limited Google's bushfire map'. CNET. Available at: https://web.archive.org/web/20230528191425/https://www.cnet.com/tech/services-and-software/australia-government-limited-googles-bushfire-map/

British Library (n.d. -a) *Collections Guides: Digital Mapping*. Available at: https://web.archive.org/web/20230328180414/https://www.bl.uk/collection-guides/digital-mapping

British Library (n.d. -b) *Ordnance Survey Mapping*. Available at: https://web.archive.org/web/20230708130433/https://www.bl.uk/collection-guides/ordnance-survey-mapping

British Standards Institution (n.d.) *S 7567 Electronic Transfer of Geographic Information (NFT) Package*. Available at: https://web.archive.org/web/20230401040423/https://webstore.ansi.org/standards/bsi/bs7567electronictransfer

Buie, L. (2020) ArcGIS Blog. Esri. *Explore Your Raster Data with Space Time Pattern Mining*. Available at: https://web.archive.org/web/20230731100258/https://www.esri.com/arcgis-blog/products/arcgis-pro/analytics/explore-your-raster-data-with-space-time-pattern-mining/

Butler, H., Daly, M., Doyle, A., Gillies S., Schaub, T. and Hagen, S. (2016) *The GeoJSON Format*. Request for Comments (RFC7946). Internet Engineering Task Force. Available at: https://web.archive.org/web/20230323171359/https://www.rfc-editor.org/info/rfc7946

Center for International Earth Science Information Network (CIESIN) (n.d.) *Geospatial Electronic Records - About this Project*. Available at http://www.ciesin.columbia.edu/ger/about.html [accessed 4 June 2023]

Center for International Earth Science Information Network (CIESIN) (2005) *Data Model for Managing and Preserving Geospatial Electronic Records Version 1.00*. New York: Columbia University. Available at: https://web.archive.org/web/20230731095848/http://www.ciesin.columbia.edu/ger/DataModelV1_20050620.pdf

Clark Labs (2014) *Clark Labs Receives Grant from Esri*. Available at: https://web.archive.org/web/20230323175854/https://clarklabs.org/clark-labs-of-clark-university-receives-grant-from-environmental-systems-research-institute-inc-esri-to-implement-its-land-change-modeler-extension-to-the-esri-cloud-platform/

Creative Commons (n.d. -a) *When We Share, Everyone Wins*. Available at: https://web.archive.org/web/20230730220856/https://creativecommons.org/

Creative Commons (n.d. -b) *Government*. Available at: https://web.archive.org/web/20230731095512/https://creativecommons.org/Government

Data.gov.uk (n.d.) *Inspire*. Available at: https://web.archive.org/web/20221007072012/https://guidance.data.gov.uk/publish_and_manage_data/harvest_or_add_data/inspire/

Directions Magazine (2012) *Setting Storage Standards for Digital Media: An Interview with Julie Sweetkind-Singer, Assistant Director of Geospatial, Cartographic and Scientific Data and Services, Stanford University*. Interview with *Campus Technology Magazine*. Available at: https://web.archive.org/web/20230731095409/https://www.directionsmag.com/article/1683

Dyke, K., Mattke, R., Kne, L. and Rounds. S. (2016) *Placing Data in the Land of 10,000 Lakes: Navigating the History and Future of Geospatial Data Production, Stewardship, and Archiving in Minnesota*. Journal of Map & Geography Libraries 12 (1), pp. 52–72. Available at: https://doi.org/10.1080/15420353.2015.1073655 [accessed 20 July 2023]

EDINA (n.d.) *Digimap*. Available at: https://web.archive.org/web/20230610221501/https://digimap.edina.ac.uk/

EDINA (2023) *EDINA*. Available at: https://web.archive.org/web/20230701195609/https://edina.ac.uk/

EPSG (2022) *EPSG.Io: Coordinate Systems Worldwide*. Available at: https://web.archive.org/web/20230728123113/https://epsg.io/

Erwin, T., Sweetkind-Singer, J. and Larsgaard, M. (2009) *The National Geospatial Digital Archives—Collection Development: Lessons Learned*. Library Trends 57 (3), pp. 490–515. Available at: https://doi.org/10.1353/lib.0.0049 [accessed 20 July 2023]

Esri (n.d.-a) *Projection File properties*. Available at: https://web.archive.org/web/20230328135949/https://pro.arcgis.com/en/pro-app/latest/arcpy/functions/prj-file-properties.htm

Esri (n.d.-b) ArcGIS Pro. *What Is Archiving?* Available at: https://web.archive.org/web/20230529125342/https://pro.arcgis.com/en/pro-app/latest/help/data/geodatabases/overview/what-is-archiving-.htm

Esri (n.d.-c) ArcGIS Pro. *What Is KML?* Available at: https://web.archive.org/web/20230731094809/https://pro.arcgis.com/en/pro-app/3.0/help/data/kml/what-is-kml-.htm

Esri (n.d.-d) ArcGIS Pro. *Fundamentals of netCDF data storage ArcGIS Pro 3.1* Available at: https://web.archive.org/web/20230714141143/https://pro.arcgis.com/en/pro-app/latest/help/data/multidimensional/fundamentals-of-netcdf-data-storage.htm

Esri (n.d.-e) ArcGIS Pro. *Raster File Formats*. Available at: https://web.archive.org/web/20230530021306/https://pro.arcgis.com/en/pro-app/latest/help/data/imagery/supported-raster-dataset-file-formats.htm

Esri (n.d.-f) ArcGIS Online. *Tile Layers*. Available at: https://web.archive.org/web/20230606103856/https://doc.arcgis.com/en/arcgis-online/reference/tile-layers.htm

Esri (2008) *What Is a coverage*. ArcGIS Desktop Help 9.3. Available at: https://web.archive.org/web/20230731094552/https://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=What_is_a_coverage

Esri (2021) ArcGIS Desktop. ArcMap. *Geoprocessing considerations for shapefile output*. Available at: https://web.archive.org/web/20230609203849/https://desktop.arcgis.com/en/arcmap/latest/manage-data/shapefiles/geoprocessing-considerations-for-shapefile-output.htm

Federal Geographic Data Committee (n.d.) *ISO Geospatial Metadata Standards*. Available at: https://web.archive.org/web/20230321173856/https://www.fgdc.gov/metadata/iso-standards

Federal Geographic Data Committee (1994) *Content Standard for Digital Geospatial Metadata*. Available at: https://web.archive.org/web/20230529125808/https://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/index_html

Federal Geographic Data Committee (2005) Homeland Security Working Group. *Guidelines for Providing Appropriate Access to Geospatial Data in Response to Security Concerns*. Available at: https://web.archive.org/web/20230202030551/https://www.fgdc.gov/policyandplanning/Access%20Guidelines.pdf

Federal Geographic Data Committee (2008) *Geographic Information Framework Data Content Standard. Part 0 – Base document*. Available at: https://web.archive.org/web/20230201231455/https://www.fgdc.gov/standards/projects/framework-data-standard/GI_FrameworkDataStandard_Part0_Base.pdf

Federal Geographic Data Committee (2015) *Users/Historical Data Working Group*. Available at: https://web.archive.org/web/20230202014852/https://www.fgdc.gov/organization/working-groups-subcommittees/hdwg/index_html

Federal Geographic Data Committee (2018) *Geospatial Data Act of 2018*. Available at: https://web.archive.org/web/20230314203645/https://www.fgdc.gov/gda

Fleischhauer, C. (2011) *Digital Formats, Part 2: Geospatial Formats Added to LC Web Site.* Library of Congress Blogs. The Signal: Digital Happenings at the Library of Congress. Available at: https://web.archive.org/web/20230323211627/https://blogs.loc.gov/thesignal/2011/12/digital-formats-part-2-geospatial-formats-added-to-lc-web-site/

Folger, P. (2018) *The Geospatial Data Act of 2018*. Congressional Research Service. Washington, D.C. Available at: https://web.archive.org/web/20230731094006/https://sgp.fas.org/crs/misc/R45348.pdf

GeoNetwork (2023) *Find and get information*. Available at: https://web.archive.org/web/20230722093549/https://geonetwork-opensource.org/

GeoTIFF (2018) *Information and Support*. Available at: https://web.archive.org/web/20230706034139/https://trac.osgeo.org/geotiff/

Gil, E. and Holmes, J. (2004) *Ever wondered what Ordnance Survey (OS) stores on behalf of the National Archives?* The Cartographic Journal, 41 (1), pp. 55–57. Available at: https://www.tandfonline.com/doi/abs/10.1179/000870404225020015 [accessed 20 July 2023]

Graves, P. and Reed, C. (2009) Open GeoSpatial Consortium, Inc. *GeoPDF Encoding Best Practice Version 2.2*. Available at: https://portal.ogc.org/modules/admin/license_agreement.php?suppressHeaders=0&access_license_id=3&target=https://portal.ogc.org/files/?artifact_id=33332 (downloads as a pdf) [accessed 20 July 2023]

Guo, L. (2013) *Pioneering GML Deployment for NSDI — Case Study of USTIGER/GML*. ISPRS International Journal of Geo-Information, 2 (1), pp. 82–93. Available at: https://web.archive.org/web/20230731092929/https://www.mdpi.com/2220-9964/2/1/82

Hexagon (n.d.) *ERDAS IMAGINE*. Available at: https://web.archive.org/web/20230601082844/https://hexagon.com/products/erdas-imagine

Ho, D. (2020) ArcGIS Blog. *How to Use OGC GeoPackages in ArcGIS Pro*. Available at: https://web.archive.org/web/20230602114422/https://www.esri.com/arcgis-blog/products/product/data-management/how-to-use-ogc-geopackages-in-arcgis-pro/

IBM (2023) *Documentation, Db2 Spatial Extender*. Available at: https://web.archive.org/web/20230603124449/https://www.ibm.com/docs/en/db2/11.1?topic=data-db2-spatial-extender

INSPIRE (Amendment) (EU Exit) Regulations 2018. No. 1338 [online] London: His Majesty's Stationery Office. Available at: https://web.archive.org/web/20230731092529/https://www.legislation.gov.uk/uksi/2018/1338/contents/made

International Organization for Standardization (ISO) (2003) *ISO 19115:2003: Geographic information –Metadata*. Available at: https://www.iso.org/standard/26020.html [accessed 11 January 2023]

International Organization for Standardization (ISO) (2014) *ISO 19115: 1:2014 Geographic information — Metadata — Part 1: Fundamentals.* Available at: https://www.iso.org/standard/53798.html [accessed 31 July 2023]

International Organization for Standardization (ISO) (2015) *ISO 19109:2015: Geographic information –Rules for application schema*. Available at: https://www.iso.org/standard/59193.html [accessed 6 February 2023]

Japan Electronics and Information Technology Industries Association (2002) *Exchangeable image file format for digital still cameras Exif Version 2.2*. Available at: https://www.digicamsoft.com/exif22/exif22/html/exif22_1.htm [accessed 6 March 2023]

JPEG (n.d.) *Overview of JPEG 2000*. Available at: https://web.archive.org/web/20230406111907/https://jpeg.org/jpeg2000/

Kernik, M. (2023) *Geospatial Data Curation Toolkit*. Available at: https://web.archive.org/web/20230731092209/https://github.com/mkernik/geodct

Kowal, K. and Holmes, J. (2014) *Library vs publisher vs archive: managing the UK topographic record*, Open Repositories Conference 2014. Available at: https://web.archive.org/web/20230731092020/https://www.doria.fi/bitstream/handle/10024/97574/Library_vs_publisher_vs_archive.pdf

Lazorchak, B. (2012) *GeoMAPP and the Future of Digital Geospatial Preservation*. Library of Congress Blogs: Signal – Digital Happenings at the Library of Congress. Available at: https://web.archive.org/web/20230731091850/https://blogs.loc.gov/thesignal/2012/05/geomapp-and-the-future-of-digital-geospatial-preservation/

Library of Congress (n.d.) *Preservation: Recommended Formats Statement VII GIS, Geospatial and Non-GIS Cartographic*. Available at: https://web.archive.org/web/20230511111707/https://www.loc.gov/preservation/resources/rfs/geo-carto.html

Library of Congress (2020a) *ESRI ArcInfo Interchange File (E00)*. Available at: https://web.archive.org/web/20230731091657/https://www.loc.gov/preservation/digital/formats/fdd/fdd000291.shtml

Library of Congress (2020b) *Spatial Data Transfer Standard*. Available at: https://web.archive.org/web/20230511113339/https://www.loc.gov/preservation/digital/formats/fdd/fdd000286.shtml

Library of Congress (2021) *ESRI Arc Geodatabase (file-based)*. Available at: https://web.archive.org/web/20230511112125/https://www.loc.gov/preservation/digital/formats/fdd/fdd000294.shtml

Library of Congress (2022) *Still Images – Quality and Functionality Factors.* Available at: https://web.archive.org/web/20230511111519/https://www.loc.gov/preservation/digital/formats/content/still_quality.shtml

Library of Congress (2023a) *METS Metadata Encoding and Transmission Standard (METS) Official Web Site*. Available at: https://web.archive.org/web/20230710185419/https://www.loc.gov/standards/mets/

Library of Congress (2023b) *PREMIS: Preservation Metadata Maintenance Activity Official Web Site*. Available at: https://web.archive.org/web/20230727162452/https://www.loc.gov/standards/premis/

Library of Congress (2023c) *Format Descriptions for Geospatial Data Formats*. Available at: https://web.archive.org/web/20230731091741/https://www.loc.gov/preservation/digital/formats/fdd/gis_fdd.shtml

Library of Congress (2023d) *Sustainability of Digital Formats: Planning for Library of Congress Collections*. Available at: https://web.archive.org/web/20230720183001/https://www.loc.gov/preservation/digital/formats/

Mapbox (n.d.) *Mapbox Tiling Service*. Available at: https://web.archive.org/web/20230606223150/https://docs.mapbox.com/mapbox-tiling-service/guides/

Masó, J. and Jacovella-St-Louis. J. (Eds) (2022) *OGC API-Tiles – Part 1: Core.* Open Geospatial Consortium. Available at: https://web.archive.org/web/20230715233557/https://docs.ogc.org/is/20-057/20-057.html

Mathys, T. and Kamel Boulos, M. (2011) *Geospatial Resources for Supporting Data Standards, Guidance and Best Practice in Health Informatics*. BMC Research Notes 4 (19). Available at: https://web.archive.org/web/20230731090715/https://bmcresnotes.biomedcentral.com/articles/10.1186/1756-0500-4-19

Microsoft (2022) *SQL Server 2022*. Available at: https://web.archive.org/web/20230627152622/https://www.microsoft.com/en-us/sql-server/sql-server-2022

Morissette, D. (2000) *Arc/Info Export (E00) Format Analysis*. Available at: https://web.archive.org/web/20230630183911/http://avce00.maptools.org/docs/v7_e00_cover.html

National Archives, The (n.d.) *Welcome to PRONOM*. Available at: https://web.archive.org/web/20230712065253/http://www.nationalarchives.gov.uk/pronom/

National Library of Scotland (n.d.) *OS MasterMap*. Available at: https://web.archive.org/web/20230327033620/https://www.nls.uk/collections/maps/subject-info/os-mastermap/

Natural Resources Canada (2013) GeoGratis/GéoGratis *Topographic Information*. Available at: https://web.archive.org/web/20230605073117/https://natural-resources.canada.ca/science-and-data/science-and-research/earth-sciences/geography/topographic-information/10785

Neteler, M., and GRASS Development Team (2023) *GRASS GIS 8.2.1 released*. GRASS GIS. Available at: https://web.archive.org/web/20230401211314/https://grass.osgeo.org/news/2023_01_21_grass_gis_8_2_1_released/

OpenGeoMetadata (n.d.-a) *About OpenGeoMetadata Repositories*. Available at: https://web.archive.org/web/20230731085024/https://opengeometadata.org/about-ogm-repositories/

OpenGeoMetadata (n.d.-b) *About OGM Aardvark.* Available at: https://web.archive.org/web/20230731084911/https://opengeometadata.org/about-ogm-aardvark/

Open Geospatial Consortium (n.d.-a) *GeoPackage Encoding Standard*. Available at: https://web.archive.org/web/20230724161358/https://www.ogc.org/standard/geopackage/

Open Geospatial Consortium (n.d.-b) *Standards KML Overview*. Available at: https://web.archive.org/web/20230630183405/https://www.ogc.org/standard/kml/

Open Geospatial Consortium (n.d.-c) *Geography Markup Language*. Available at: https://web.archive.org/web/20230610220748/https://www.ogc.org/standard/gml/

Open Geospatial Consortium (n.d.-d) *GML in JPEG 2000 for Geographic Imagery Encoding*. Available at: https://web.archive.org/web/20230610214941/https://www.ogc.org/standard/gmljp2/

Open Geospatial Consortium (n.d.-e) *OpenGIS Web Map Tile Service Implementation Standard*. Available at: https://web.archive.org/web/20230709152021/https://www.ogc.org/standard/wmts/

Open Geospatial Consortium (n.d.-f) *Web Feature Service*. Available at: https://web.archive.org/web/20230623041540/https://www.ogc.org/standard/wfs/

Open Geospatial Consortium (n.d.-g) *Standards*. Available at: https://web.archive.org/web/20230219035037/https://www.ogc.org/standards/

Open Geospatial Consortium (2006) *OGC Prepares the United States Census Bureau for a New Era*. Available at: https://web.archive.org/web/20230606192559/https://www.ogc.org/press-release/ogc-prepares-the-united-states-census-bureau-for-a-new-era/

Open Knowledge Foundation (n.d.-a) Open Data Commons. *Legal Tools for Open Data*. *Licenses*. Available at: https://web.archive.org/web/20230228064713/https://opendatacommons.org/licenses/

Open Knowledge Foundation (n.d.-b) Open Data Commons. *Legal Tools for Open Data*. *Open Data Commons Open Database License (ODbL)*. Available at: https://web.archive.org/web/20230220222711/https://opendatacommons.org/licenses/odbl/

OpenStreetMap. Available at: https://web.archive.org/web/20230221000635/https://www.openstreetmap.org/#map=6/39.085/-83.733

Oracle (2023) *Oracle® Spatial: Spatial Developer's Guide.* Available at: https://web.archive.org/web/20230731083944/https://docs.oracle.com/en/database/oracle/oracle-database/21/spatl/index.html

Ordnance Survey (n.d.) *OpenData Support.* Available at: https://web.archive.org/web/20230731083811/https://www.ordnancesurvey.co.uk/business-government/tools-support/open-data-support

Ordnance Survey (2017) *OS MasterMap Topography Layer – Product Overview*. Available at: https://web.archive.org/web/20230731083703/https://www.ordnancesurvey.co.uk/products/os-mastermap-topography-layer

Park, J. (2010) *Australian Government Commits to Open Access*. Creative Commons Weblog. Available at: https://web.archive.org/web/20230731083553/https://creativecommons.org/2010/05/17/australian-government-commits-to-open-access/

PostGIS (2023) *Documentation*. Available at: https://web.archive.org/web/20230115123446/https://postgis.net/documentation/

QGIS – The Leading Open Source GIS. Available at: https://web.archive.org/web/20230108042450/https://www.qgis.org/en/site/about/index.html

Rouault, E. (2013) GitHub. *FGDB Spec*. Available at: https://web.archive.org/web/20230109060408/https://github.com/rouault/dump_gdbtable/wiki/FGDB-Spec

Smithsonian (2021) *Watercolor Maptiles Website Enters Permanent Collection of Cooper Hewitt*. Available at: https://web.archive.org/web/20230322163504/https://www.si.edu/newsdesk/releases/watercolor-maptiles-website-enters-permanent-collection-cooper-hewitt

Thales (2016) *Sentinel-2: So Much Data, So Little Time*. Available at: https://web.archive.org/web/20230603142410/https://www.thalesgroup.com/en/critical-information-systems-and-cybersecurity/news/sentinel-2-so-much-data-so-little-time

Toitū Te Whenua – Land Information New Zealand (n.d.) *Copyright*. Available at: https://web.archive.org/web/20230606140750/https://www.linz.govt.nz/copyright

UK AGMAP (Academic Geospatial Metadata Application Profile) (n.d.) *Metadata Standards Catalog*. Available at: https://web.archive.org/web/20230731083051/https://rdamsc.bath.ac.uk/msc/m79

US Geological Survey (n.d.) *What Is a Digital Elevation Model (DEM)?* Available at: https://web.archive.org/web/20230516123109/https://www.usgs.gov/faqs/what-digital-elevation-model-dem#multimedia

Wiki OpenStreetMap (2023) *Elements.* Available at: https://web.archive.org/web/20230502144410/https://wiki.openstreetmap.org/wiki/Elements

www.gov.uk (2023) Cabinet Office and Geospatial Commission. *UK Geospatial Data Standards Register*. Available at: https://web.archive.org/web/20230629131845/https://www.gov.uk/government/publications/uk-geospatial-data-standards-register/national-geospatial-data-standards-register