# LIBRARY OF CONGRESS COLLECTIONS POLICY STATEMENTS

## Digital Geospatial Materials

## I.   Scope

This Collections Policy Statement covers digital geospatial or Geographic Information Systems (GIS) data presented in a variety of digital file types and formats, including vector and raster representations; relational databases embedded with geographic information or topology; digitally scanned and georeferenced versions of analog maps; remotely sensed digital imagery; software used for the creation, retrieval, analysis and display of geospatial data; and web map applications.

See also the *Analog Geospatial Materials Collections Policy Statement*.

## II.   Diverse and Inclusive Collecting Statement

As the nation's de facto national library, the Library of Congress strives to build an expansive, yet selective, collection that records the creativity of the United States and is reflective of the nation's diversity and complexity. The Library's mandate is to have collections that are inclusive and representative of a diversity of creators and ideas. A priority includes acquiring material of underrepresented perspectives and voices in the Library's collections to ensure diverse authorship, points of view, cultural identities, and other historical or cultural factors. The Library also seeks to build a research collection that comprises a globally representative sample of international materials that are diverse in voice and perspective, relative to their places of origin, further supporting the Library's mission to sustain and preserve a universal collection of knowledge and creativity for Congress and future generations.

Diverse collecting is mentioned within many of the Library's Collections Policy Statements.  In addition, the Library has adopted several specific collection policies in an effort to ensure it is building an inclusive and representative collection.  For more information, see the Library's Collections Policy Statements on *Ethnic Materials*, *LGBTQIA+ Studies*, *Women's and Gender Studies*,  *Independently Published and Self-Published Textual Materials*, and *Countries and Regions with Acquisitions Challenges*.

## III.  Research Strengths

In order to maintain its unparalleled collection of cartographic materials, the Library of Congress develops and implements strategies led by the Geography and Map Division for the large-scale acquisition of digital geospatial data. A major portion of the sum of present-day human knowledge in the subjects of geography and cartography is expressed in digital geospatial data and related digital technologies. The Library collects geospatial data from a wide range of sources, in domestic and global geographic coverage, and in a variety of data file formats and themes. The strength of the digital data collection is gauged by how well the data meet the analytical needs of Congress and the Congressional Research Service, as well as the degree to which the data document historical events and the technological innovations of the time period that are incorporated into that documentation.

## IV.  Collecting Policy

The Library collects digital geospatial data that:

   a.  Support present-day and anticipated future research needs of Congress and the Congressional Research Service

   b.  Document and allow for analysis of historical and geographical events

   c.  Document and allow for analysis of issues of long-term scientific, historical and/or cultural importance to Congress and the American public

   d.  Document the development of geographic information systems technology

   e.  Broaden and enhance worldwide geographic and thematic coverage

   f.  Fill identified gaps in existing collections and map series

The Library acquires digital geospatial data on a world-wide basis and of varying digital file types, geographic coverages, scales, languages, or themes. Priority is given to acquiring official geospatial data publications from the U.S. Federal Government, U.S. state and local mapping bureaus, and national mapping agencies of all other countries. Collected digital data will be responsive to the research needs of Congress, the Congressional Research Service, and Library of Congress researchers. Additionally, the Library endeavors to capture the world's geographic, geospatial and cartographic heritage.

**Federal Government Data**

The acquisition of U.S. Federal Government digital geospatial data is well supported through the Library's participation in organizations concerned with the stewardship of federal geospatial resources, including the Federal Geographic Data Committee (FGDC). Geospatial data acquisition from Federal sources is also greatly facilitated through online geospatial data clearinghouses, most notably including GeoPlatform, which "provides shared and trusted geospatial data, services, and applications for use by

the public and by government agencies and partners to meet their mission needs."[1] In curating specific data acquisitions out of the overwhelming whole of federally-authored GIS data, the Library will use as a starting point the A-16 National Geospatial Data Asset Datasets list.[2] Each item included in this list is a data "asset that has been designated as such by the FGDC Steering Committee and meets at least one of the following criteria: (1) supports mission goals of multiple federal agencies; (2) is statutorily mandated; or (3) supports Presidential priorities as expressed by Executive Order or by OMB."[3] This list represents core geospatial datasets of the federal government that are worthy of preservation and facilitate access to researchers. Additionally, methods of incorporating active public use of federal datasets are being considered in the curation process, for example, acquiring and preserving the most popular datasets downloaded from GeoPlatform.gov (based on the web site's metrics) on a regular basis. The regular acquisition of federally authored datasets will be further supported by improved communications with GIS data stewards across federal agencies so that the Library is current in its geospatial data acquisitions.

**State and Local Government Data**

Every U.S. state has a GIS data clearinghouse or portal of some kind to disseminate state-level data, whether the clearinghouse is operated by a state government or a university. With the proper monitoring of state-level GIS data clearinghouse web sites, the Library can regularly acquire large collections of state-level GIS data through direct downloads. In addition to state-level GIS data portals, many local governments in the United States, at the county, city, or town levels, have GIS offices that author, compile, and disseminate GIS data. Through direct downloads, curation of important datasets, and knowledge of the state- and local-level geospatial data offerings unique to each state, the Library acquires important datasets of this variety for long-term preservation and use in geospatial analyses.

**Commercial Data**

With the proliferation of no-cost open data created by government entities, public interest groups, universities, and other organizations, there is a reduced dependence in the GIS field on data from commercial vendors that may entail high costs, as well as stringent licenses and terms of use. However, there are many important datasets that are still best acquired through commercial data suppliers who have expertise in geospatial data collection, aggregation, and management. Valuable commercial data may include original data products created by vendors that serve as "value-added intermediaries" which "take raw or native data generated by the suppliers and enhance it to provide time and space utility to it."[4] For example, by infusing street centerline data from a Federal source with more attributes and

---

[1] Federal Geographic Data Committee. 2016. "GeoPlatform." Accessed 24 June, 2016. https://www.geoplatform.gov.

[2] Federal Geographic Data Committee. 2016. "National Geospatial Data Asset (NGDA) Datasets." Accessed 24 June, 2016. http://www.fgdc.gov/ngda-reports/NGDA_Datasets.html.

[3] Federal Geographic Data Committee. 2015. "National Geospatial Data Asset Management Plan Lifecycle Maturity Assessment Tool" Accessed 24 June, 2016. https://cms.geoplatform.gov/sites/default/files/a16themeleads/1_NGDA_BaselineAssessment_01_IntroAndAssessment_FINAL2_Q12Q19_modified.pdf.

[4] Wilson, John P., and A. Stewart Fotheringham. 2008. *The Handbook of Geographic Information Science*. Oxford: John Wiley & Sons.

3

improved accuracy, the vendor could then sell this value-added product "in a more convenient or technically acceptable manner" to consumers.[5] Decisions to purchase and acquire commercial geospatial data are made with considerations given to the availability of no-cost alternatives, the quality and usefulness of raw or value-added commercial data, available fiscal resources, and the overall value of the data for near- and long-term spatial analysis and historical record.

**International Data Sources**

The Library collects geospatial data authored by international sources to maintain a digital data collection that reflects global geospatial technologies, perspectives, and heritage. Collecting data from international sources also helps ensure that the Library's geospatial data repository is responsive to spatial analysis needs in research focused on non-U.S. locales or requiring premier datasets of a certain theme that are authored by non-U.S. sources. Given the very large scale at which geospatial data is produced internationally, data curation based on data themes and the nature of sources is essential. Among proposed methodologies for the regular acquisition of international data is the collection of "core" geospatial data sets (defined by common mapping importance or download popularity) authored by national governments from a range of countries with robust national GIS data programs. Geospatial data created by non-U.S. based nonprofit organizations and universities are also considered for acquisition based on authoritativeness and value to modern spatial analysis and preservation of geospatial heritage.

**Thematic Data**

Thematic geospatial data refer to datasets intended for the examination of a special purpose or subject across geographic space.[6] It is important to distinguish this type of geospatial data and maps because they represent the harnessing of geographic theory and technologies to understand and study a specific phenomenon across space, in contrast to general or reference maps that typically convey tangible elements of the physical environment (roads, rivers, building footprints, topography, etc.) or human constructs to organize space (various kinds of political boundaries) "for their own sake."[7] The characteristics, themes, sources, and intended uses of thematic geospatial data are wide-ranging. Examples of thematic data would include census data of demographics and socioeconomic statistics (as compiled by the United States Census Bureau), data on common health indicators at a subnational level (as compiled by the United States Agency for International Development), or geo-informed data on locations of armed conflict (as compiled by the Armed Conflict Location & Event Data Project, a non-governmental organization).

Thematic data is critically important to modern geospatial analysis and often serves as the basis for focused research studying geography. Understanding the importance of thematic data as well as its diversity in format and source, the Library will broadly pursue the acquisition of thematic data and give strong consideration to source authoritativeness, compliance with geospatial metadata standards,

---

[5] Ibid.

[6] Thrower, Norman J. W. 2008. *Maps and Civilization: Cartography in Culture and Society*, 3rd ed. Chicago: University of Chicago Press.

[7] Ibid.

usefulness in the analysis of modern policy issues, and value for long-term preservation.

**Limiting Criteria**

**Storage Space:** Available server space for storing digital geospatial data may influence the scope of data collection decisions, especially considering the often exceedingly large file sizes of geospatial data as compared to many other digital file types, and the high rate of data production from geospatial data initiatives (e.g., the Landsat 8 Earth observation satellite captures 550 scenes of Earth, at about 2GB uncompressed per scene, every day).[8,9] Storage space limitations may favor a focus on notable subsets of large format data collections as opposed to comprehensive collections (e.g., Landsat 8 scenes that capture historical events). Increasing available server space in the Geography and Map Division and using cloud storage technologies for off-site data storage will allow for greater flexibility in acquiring digital geospatial data on increasingly larger scales.

**Licensing:** Licensing agreements for proprietary geospatial data directly impact how data can be seen and used by Members of Congress and their staff, the Congressional Research Service, Library of Congress staff, and the general public. Negotiating licensing agreements that open proprietary data to a wider audience through the Library may be prohibitively expensive.

**Formats and Resources**

**Vector Data:** Vector is a "coordinate-based data model that represents geographic features as points, lines, and polygons" and that often includes associated attributes that provide characteristics of each vector feature.[10] An example of a vector data product is TIGER (Topologically Integrated Geographically Encoded Reference) files produced by the U.S. Census Bureau to geographically represent roads, hydrology, U.S. Census Bureau statistical divisions, and other features. There are a variety of commonly used file types for this data format, including both proprietary (e.g., Esri's Shapefile [.shp] and Geodatabase [.gdb]) and open-source (e.g., Geography Markup Language standard for XML) formats. The Library collects vector geospatial data in whatever file types are popular, accessible and/or available, but will look towards open source formats and preservation workflows to help ensure the long-term usability of all vector data.

**Raster Data:** Raster data is a "spatial data model that defines space as an array of equally-sized cells arranged in rows and columns, and composed of single or multiple bands."[11] Common data products created in raster format include aerial photography, remote sensing imagery, digital elevation models (DEMs), and land-cover/land-use classification maps. Raster data comes in a variety of image-based file formats such as TIFF and JPEG. As with vector geospatial data, the Library collects in all available and

---

[8] National Aeronautics and Space Administration. 2013. "Landsat 8." Accessed 24 June, 2016. http://landsat.gsfc.nasa.gov/?p=3186.
[9] United States Geological Survey. 2015. "What Files Are Included When I Download a Landsat 8 Scene?" Accessed 24 June, 2016. http://landsat.usgs.gov/L8_files.php.
[10] Esri. "vector." *GIS Dictionary*. Accessed 24 June, 2016. http://support.esri.com/sitecore/content/support/Home/other-resources/gis-dictionary/term/vector.
[11] Esri. "raster." *GIS Dictionary*. Accessed 24 June, 2016. http://support.esri.com/sitecore/content/support/Home/other-resources/gis-dictionary/term/raster.

5

practical raster file formats and implements preservation workflows to ensure long-term usability.

**Web Applications:** The Library evaluates, selects, and catalogs web-based geospatial content that includes geospatial datasets deemed important and useful for analysis, as well as web map applications that represent notable expressions of cartography and data analysis worthy of long-term preservation. A common format for web-based geospatial data is a Web Map Service (WMS) which is an open-source standard that "provides a simple HTTP interface for requesting geo-registered map images from one or more distributed geospatial databases."[12] Files associated with Web Map Services, such as XML files, can be acquired and preserved.

## V.  Best Editions and Preferred Formats

For guidance regarding best editions for material acquired via the Copyright Office, see:  http://copyright.gov/circs/circ07b.pdf.

For guidance regarding recommended formats for material acquired via all other means; e.g., purchase, exchange, gift and transfer, see:  http://www.loc.gov/preservation/resources/rfs.

For information regarding electronic resources, open digital content, web archiving, and data sets, see the following Supplementary Guidelines:  http://www.loc.gov/acq/devpol/electronicresources.pdf, https://www.loc.gov/acq/devpol/opencontent.pdf, http://www.loc.gov/acq/devpol/webarchive.pdf, and  https://www.loc.gov/acq/devpol/datasets.pdf.

## VI.  Acquisition Sources

The range of digital geospatial data resources is highly diverse and requires that the Library of Congress employ a variety of methods of acquisition.

**Direct Download:** A plethora of online geospatial data portals allows for direct downloads of GIS data from a wide range of sources, including the Federal Government, with little or no direct coordination required with the source organization or agency (GeoPlatform is a primary example of such a data portal). A significant portion of geospatial data of interest to the Library can be acquired by downloading data hosted online.

**Government Deposit and Transfer:** The Library receives non-classified cartographic publications issued by Federal agencies. Given the expansive nature of GIS data being created across the Federal Government for a wide range of purposes, the Library should curate digital data for usefulness and relevancy rather than attempting to acquire all non-classified federally authored geospatial data.  It should also communicate to federal agencies the need for the deposit of selected geospatial data with the Library. Coordination with federal agencies is especially important when acquiring geospatial data that is not available on online data clearinghouses such as GeoPlatform.

---

[12] Open Geospatial Consortium. 2016. "Web Map Service." Accessed 24 June, 2016. http://www.opengeospatial.org/standards/wms.

**Purchase:** Appropriated funds are used to purchase retrospective and current digital geospatial resources.

**Copyright:** As currently devised, copyright deposit is a viable source for acquiring geospatial data and software issued on such optical media as CD-ROM and DVD's. The Copyright Law is not currently a viable resource for acquiring born-digital commercially produced or repackaged geospatial data.

**Donations:** The Library receives gifts of retrospective and current geospatial materials. The Geography and Map Division and its collections are the beneficiary of acquisition support provided by the Library's James Madison Council and the Geography and Map Division's Philip Lee Phillips Society.

## VII. Collecting Levels

Meeting the Library's Diverse and Inclusive Collecting Statement (see Section II) and the collecting levels outlined below requires continual evaluation of the publishing landscape, sources of expression, current events, and socio-cultural trends to thus maintain effective collecting policies and acquisitions methods. Changes in publishing or in the creation of materials covered by this policy statement may necessitate collecting efforts not explicitly referenced here. Such efforts will be handled on a case-by-case basis while the Library evaluates the need for policy statement updates.

For explanation of the Collecting Levels used by the Library, see
https://www.loc.gov/acq/devpol/cpc.html.

The goal is to acquire coverage at the research level, 4, for the United States (U.S. Federal Government, U.S. state and local mapping bureaus) and for non-U.S. countries (national mapping agencies).

Revised October 2016. Updated by the Collection Development Office April 2022.