

An Introduction to the GIS/Geospatial Recommended Formats Statement

Meagan Snow
Tim St. Onge
Rachel Trent

Geography & Map Division



Introduction

— — —

- Library of Congress has over **5.5 million maps**
- We help build and manage geospatial collections in the Geography and Map Division
- We run the *Congressional Cartography Program* and support the *Congressional Research Service*, both part of the Library's GIS analysis services to Congress



What is the Recommended Formats Statement (RFS)?

— — —

“Recommended Formats Statement identifies hierarchies of the physical and technical characteristics of creative formats, both analog and digital, which will best meet the needs of all concerned, maximizing the chances for survival and continued accessibility of creative content well into the future.”

<https://www.loc.gov/preservation/resources/rfs/>

How to read the RFS

VII. GIS, Geospatial, and Non-GIS Cartographic

<https://www.loc.gov/preservation/resources/rfs/geo-carto.html>

VII. GIS, Geospatial and Non-GIS Cartographic

i. Geographic Information System (GIS) - Vector Data

i. Geographic Information System (GIS) - Vector Data		
	Preferred	Acceptable
A. Formats	<p>Most complete data (all layers, appendices), even if proprietary, with a preference for preserving the native format and projection of the data</p> <p>Vector formats compatible with widely adopted GIS including,</p> <ul style="list-style-type: none">➤ Shapefile, which is comprised of at least a SHP, SHX, and DBF file and optionally a PRJ (highly recommended), XML (highly recommended), SBN, and/or SBX.➤ Esri File Geodatabase➤ OGC GeoPackage	<ul style="list-style-type: none">➤ GeoJSON (may have scalability issues)➤ KML➤ GML
B. Delivery method, in order of preference	<ol style="list-style-type: none">1. Public download URLs2. Automated private download URLs with any necessary API keys or credentials3. Hard drives	
C. Metadata	<ul style="list-style-type: none">➤ For metadata information see 191xx ISO standards and Federal Geographic Data Committee (FGDC)➤ To the extent allowed by the underlying format, include available information about how the data was created, collected and any post-processing which has been applied	<p>Project and layer files (.mxd, .qgs, .lry) may be acquired with deposits of content data to assist in reviewing materials during deposit process</p>
D. Technological measures	<p>Files must contain no measures (such as digital rights management or encryption) that control access to or prevent use of the digital work.</p>	

RFS vs Sustainability of Digital Formats

Introduction

Sustainability Factors

Content Categories

Format Descriptions

Contact

Format Description Categories >> Browse Alphabetical List

ESRI Shapefile

Table of Contents

- Identification and description
- Local use
- Sustainability factors
- Quality and functional factors
- File type signifiers
- Notes
- Format specifications
- Useful references

>> Back

Format Description Properties

- ID: fdd000280
- Short name: ESRI_shape
- Content categories: dataset, gis
- Format Category: file-group
- Other facets: structured, symbolic
- Last significant FDD update: 2020-05-29
- Draft status: Full

Identification and description

Full name	ESRI Shapefile
Description	<p>The ESRI Shapefile format was developed by Esri, formerly the Environmental Systems Research Institute, Inc., and abbreviated as "ESRI," and published in 1998. Although proprietary, the intention behind publishing the format was to encourage its use for interoperability among geographic information system (GIS) applications. The Shapefile format stores nontopological geometry and attribute information for spatial features in a data set. A Shapefile consists minimally of a main file, an index file, and a dBASE table.</p> <p>In the main file, the geometry for a feature is stored as a shape comprising a set of vector coordinates. This main file is a direct access, variable-record-length file in which each record describes a shape with a list of its vertices. In the index file, each record contains the offset of the corresponding main file record from the beginning of the main file. Attributes are held in a dBASE format file. The dBASE table contains feature attributes with one record per feature. Attribute records in the dBASE file must be in the same order as records in the main file. Each attribute record has a one-to-one relationship with the associated shape record.</p> <p>The shapefile format can support point, line, and area features. Area features are represented as closed loop, double-digitized polygons.</p> <p>Instances of the Shapefile format have often been used as a data exchange format from Esri formats to non-Esri applications. The format is most useful for writing simple features and attributes quickly as there are limitations inherent in the Shapefile format related to both geometry and attributes. As outlined elsewhere in this description, these limitations may cause loss of data when using shapefiles to contain or exchange complex geometry or attributes. The Shapefile format may be used as an intermediary between data creation applications and more functionally capable GIS formats and applications, albeit with the limitations noted in the Dataset/Normal Dataset section.</p> <p>The cluster of files is typically stored in the same file directory or project workspace, with all component files having the same filename (prefix) and identified by individual file extension (suffixes). Three components are mandatory: a main file that contains the feature geometry (.shp), an index file that stores the index of the feature geometry (.shx), and a dBASE table (.dbf) that stores the attribute information of features. A comprehensive list of component files follows:</p> <ul style="list-style-type: none">• <i>shp</i> -- Main file (mandatory): a direct access, variable-record-length file in which each record describes a shape with a list of its vertices.• <i>shx</i> -- Index file (mandatory). In the index file, each record contains the offset of the corresponding main file record from the beginning of the main file. The index file (.shx) contains a 100-byte header followed by 8-byte, fixed-length records.

<https://www.loc.gov/preservation/digital/formats>

LIBRARY
LIBRARY
OF CONGRESS

GEOGRAPHY AND MAP DIVISION

5

What criteria does the format consider?

— — —

The RFS considers two main sets of criteria:

- Global community sustainability factors (external)
- Local sustainability factors (internal)

The Library maintains a matrix of both sets of criteria against which each format is evaluated and a final decision rendered

Global/Community (External) Format Sustainability Factors

— — —

- External dependencies
- Impact of patents
- Technical protection mechanisms
- Disclosure
- Adoption
- Transparency
- Self-documentation



Local (Internal) Format Sustainability Factors

— — —

- Internal staff experience and expertise
- Software/Hardware/OS availability
- Representation in LC collections
- Established workflow/functionality



What geospatial formats are currently approved? - Vector

- Preferred:

- Shapefile
- Esri File Geodatabase
- OGC GeoPackage



- Acceptable:

- GeoJSON
- KML
- GML

GEOJSON



What geospatial formats are currently approved? - Raster

— — —

- Preferred:

- GeoTIFF
- OGC GeoPackage



- Acceptable:

- TIFF with TIFF world file (.tfw)
- GML in JPEG 2000
- GML



What geospatial formats are currently approved? Vector and Raster Combined

— — —

- Preferred:
 - Esri File Geodatabase
 - OGC Geopackage
- Acceptable:
 - TerraGo GeoPDF
 - Geospatial PDF



Geospatial PDF
quick start guide

File type example: Open Geospatial Consortium Geopackage

— — —

GeoPackage is an open, standards-based, platform-independent, portable, self-describing, compact format for transferring geospatial information.

“A GeoPackage is a platform-independent SQLite database file that contains data and metadata tables with names and structures having definitions, integrity assertions, format limitations and content constraints as described in the OGC GeoPackage Encoding Standard from the Open Geospatial Consortium.”

Source: LOC Sustainability of Digital Formats



File type example: OGC Geopackage - Global Factors

— — —

- External dependencies: No dependencies beyond software to unpack the data structure
- Impact of patents: No concerns.
- Technical protection mechanisms: Encryption not provided for within the GeoPackage specification
- Disclosure: Open standard, documented in freely available specifications.
- Adoption: Not locked into any vendor-specific implementations, robust user groups.
- Transparency: Can be analyzed with basic tools
- Self-documentation: Format defines an optional table to hold descriptions of data fields

File type example: OGC Geopackage - Local Factors

— — —

- Staff Expertise: LC staff are familiar with the format (but are not experts!)
- Software/Hardware: Format is supported in existing software
- Representation in Collections: None
- Established workflow: None

File type example: Open Geospatial Consortium Geopackage

— — —

The Library of Congress Recommended Formats Statement (RFS) designates GeoPackage as a "preferred" format for:

- GIS Vector Data
- GIS Vector and Raster Combined data
- GIS Raster and Georeferenced Images



Thank you!

— — —

The Recommended Format Statement is currently **open for public comment** until April 15, 2022.

Please direct any feedback (including questions or comments) about the RFS for GIS, Geospatial, and Non-GIS Cartographic materials to:

Meagan Snow (msnow@loc.gov)