**Term Project: Final Report**

*Word Count: 2,403 words (not including works cited)*

*Dataset Title: “National Forest System Trails (Feature Layer)”*

*Dataset URL:* <https://catalog.data.gov/dataset/national-forest-system-trails-feature-layer-f51e8>

ArcGIS Hub URL (Text): <https://data-usfs.hub.arcgis.com/datasets/usfs::national-forest-system-trails-feature-layer/about>

ArcGIS Hub URL (Map): <https://data-usfs.hub.arcgis.com/datasets/usfs::national-forest-system-trails-feature-layer/explore?location=47.586482%2C-123.247677%2C11.70>

*Data Repository:* Forest Service Research Data Archive

*Registry of Research Data Repositories URL:* <https://www.re3data.org/repository/r3d100011147>

*Repository URL:* <https://www.fs.usda.gov/rds/archive/>

**Data**

Description

This dataset is a “features layer” that describes trails located on U.S. National Forest System land, including their location and a rich set of descriptive characteristics. According to the Esri technical support page, a “features layer” is a data layer that provides information about features of a map, and which corresponds to a “map layer,” which provides images of maps directly to users for display. This features layer dataset allows users to explore characteristics of over 82,000 trails, including their location, length, type, accessibility and allowed use, and surface, among many other attributes. The dataset links to an [interactive map](https://data-usfs.hub.arcgis.com/datasets/usfs::national-forest-system-trails-feature-layer/explore), hosted on the U.S. Forest Service’s ArcGIS hub website, which allows users to explore the trails and their features in greater detail.

Origin and Provenance

The data were first published by the U.S. Forest Service in June of 2015, and were updated most recently in January of 2024. I initially located the dataset on Data.gov, and have also found it hosted on the Forest Service’s [Enterprise Data repository](https://data.fs.usda.gov/geodata/edw/datasets.php?xmlKeyword=trails#:~:text=The%20TrailNFS_Publish%20Layer%20is%20designed,are%20prohibited%2C%20allowed%20and%20encouraged.) and its [ArcGIS hub](https://data-usfs.hub.arcgis.com/datasets/usfs::national-forest-system-trails-feature-layer/about).

Key Stakeholders

Key stakeholders for this dataset include the U.S. Forest Service, which created the data; the U.S. Department of Agriculture, within which the Forest Service is housed; the U.S. government and tax payers broadly, whose funds supported the collection, curation, and hosting of the dataset; and the general public which may use the data to explore trails on Forest Service land. Theoretically, other more specialized users, such as outdoor tour companies, camp programs, and other companies and non-profit organizations looking to support the creation and maintenance of trails, or the recreational use of trails by various groups of people (e.g., students) may also find the data useful.

Data Files, Format, and Software Specifications

The dataset contains four data files, not including additional metadata files. These include:

* CSV file with 82,352 rows, each representing a unique trail, and extensive metadata describing each trail in 77 accompanying columns. I opened this file with Microsoft Excel.
* Keyhole Markup Language (KML) file that contains visual geographic information to render the maps and trails contained in the dataset. According to Google’s developers support page, KML is a “file format used to display geographic data in an Earth browser such as Google Earth.” I attempted to open this file directly in Google Earth, as advised, but the upload kept timing out and I was ultimately unsuccessful.
* A zipped “shapefile” that contains a six sub-files in, respectively, the .cpg, .shp, .prj, .dbf, .shx, and .xml filetpypes. According to Esri, a provider of Geographic Information System (GIS) software, a shapefile is “a simple, nontopological format for storing the geometric location and attribute information of geographic features. Geographic features in a shapefile can be represented by points, lines, or polygons (areas).” I was not able to access any of the above files without specialized software, such as Esri’s ArcGIS.
* A geoJSON file that contains JSON-formatted data about the trails and their characteristics included in the dataset. It appears to contain the same information as the CSV file. I was able to open this file as a .txt file in my MacBook’s native TextEdit program.

Usage and Access

The data set is available for public use, and does not include any specified usage restrictions or access limitations beyond the software specifications described above. The data are licensed under the Creative Commons Attribution.

**Metadata**

Metadata Overview

Metadata describing the dataset appears in multiple places. First, metadata about the overall dataset (including descriptive, administrative, and structural metadata) is located on the Data.gov repository page. Second, the “features layer” map data itself, hosted on ArcGIS and available in the downloadable data files, contains extensive metadata for each individual trail. Overall, the features metadata is extremely comprehensive, containing 77 fields that describe each of the trails included in the dataset, including lots of information of interest to users. However, the 26 fields that contain additional metadata describing the dataset overall are much more technical and difficult to understand for non-experts, including fields for Bureau Codes, Harvest Sources, Source Hashes, and more.

Metadata Location

The dataset’s metadata are available in text format on the same repository page within Data.gov as the data files themselves, while features metadata is available for download in JSON, geoJSON, and XML formats (as separate files), and linked in PHP format at data.fs.usda.gov/geodata/edw/dir\_trails.php.

Standard(s)

The trails-level metadata appears to be structured according to two standards: the “FGDC Content Standard for Digital Geospatial Metadata,” which is published by the Federal Geographic Data Committee and, according to the standard’s website, includes the following information:

* Identification
* Metadata reference
* Data quality
* Spatial data organization
* Spatial references
* Entity and attributes
* Distribution

Other metadata available to describe the dataset appears to follow the ISO-19139 standard instead. The standards are similar in some ways, different in others, and I am not exactly sure why both are present. It may be because the dataset’s creator, the U.S. Forest Service, used one standard, while the publisher and/or repository used another.

**Data and Metadata Enrichment**

The data and features-level metadata are already extremely rich, containing lots of information about the over 82,000 trails included in the dataset and the methods and means by which recreators might access and use the trails, down to details such as “surface firmness.” Features metadata might be enriched by including recommended travel times, or linking to information about campsites or other recreational sites connected to trails.

Improve Discoverability and Use

Metadata about the dataset itself (e.g., resource type, publisher, harvest object ID) is limited and highly technical, and could be improved to help users more easily discover and use the dataset. For example, metadata could include states or regions covered, types of recreation described (e.g., snowmobiling, hiking), or terrain types.

**Publications Associated with Dataset**

I searched Google Scholar, basic and advanced Google Search, and the University of Washington’s library discovery layer for the dataset by its name, and also for publications related to National Forest System trails more broadly. I could not find any publications that appeared to use this dataset, specifically. There were many publications related to National Forests and National Parks, opportunities and consequences of human interaction with them (e.g., economic development, crime, enjoyment), and ecological studies of them. I would expect various trail planning applications to use these data, but that was also difficult to determine.

**Recommended Data Citation**

U.S. Forest Service (2015): National Forest System Trails (Feature Layer). https://www.arcgis.com/home/item.html?id=0969eb1cbb2f4a1d861ee58fff587cc2&sublayer=0

**Repository Rationale**

I chose the Forest Service Research Data Archive (FSRDA) repository to host my selected dataset on National Forest trails and their features because it is specifically concerned with research data related to the U.S. Forest Service, data produced by Forest Service scientists, and “data sets that are broadly relevant to forest and grassland ecology, and the economic and social interactions of humans with these ecosystems” (re3data.org, FSRDA). My selected dataset is directly related to the “economic and social interactions” that people have with the trails on U.S. Forest Service land, providing map and features data that help people understand where trails are located, when they are accessible, what condition they are in, and how best to interact with them (e.g., by hiking, snowmobiling, etc.).

I searched extensively on the Registry of Research Data Repositories for repositories related to forests, trails, outdoor recreation, GIS, mapping, and other terms and concepts related to my dataset. The next closest repository I identified is the [National Park Service Geographic Information Systems](https://www.re3data.org/repository/r3d100012483). This repository does focus on geospatial data, but it is concerned with U.S. National Parks rather than the Forest Service. My choice to prioritize the repository’s terrain and domain focus, rather than its data format focus, is debatable.

**Data Submissions Policy**

The FSRDA offers 931 data publications, 831 hosted by the archive itself. The repository’s mission is to archive and share the research data generated by Forest Service Research & Development, to make those data available and useful today and to preserve them for long-term access and use. The repository is only open to data submissions from Forest Service researchers, Joint Fire Science Program-funded researchers, or “[a]nyone who has collected data using Forest Service R&D funding or on Forest Service R&D-managed lands” (USDA, *Submitting Data,* n.d.).

The FSRDA offers detailed guidelines for data submissions by eligible submitters, including a text overview, multiple slides and webinars, and downloadable forms and templates (USDA, *Submitting Data,* n.d.).

1. **Data Accepted:** Any data used for a publication, observational or historical data used for long- or short-term study, or secondary data that has been “substantially modified” (USDA, *Submitting Data,* n.d.). The repository notes that submitted data may include, but are not limited to: Excel files, CSV files, geodatabases or shapefiles, access databases, maps, photographs, word documents, other text files, or other media files. Supplemental “non-data” content may also be submitted, including but not limited to: maps, line drawings of plots or transects, photographs of sites, instrumentation, etc., computer code, analysis code and results, study plans and reports, and user guides.
2. **SIP:** Submitters must submit: data files, a prescribed metadata document in either Microsoft Word or XML format (more below), additional or supplemental files that should be archived with the data, associated journal articles or other publications, or links to them if available online, and a data submission form. The FSRDA offers two different data submission forms, one for Forest Service authors and another for all other authors. The two forms are mostly identical, requiring information about the author(s), submitter(s), affiliation, working title of data package, full citation and status for any manuscript(s) if applicable, approximate size of the data package, archiving preferences (e.g., open public vs. private repository), access preferences (e.g., download vs. link to alternative archive or website), and signed approval and authorization. The Forest Service-specific form asks the submitter to share FS-specific information about research station(s) and group/program/unit, and to indicate whether the data should be considered “Influential Scientific Information” (i.e., “scientific information FS R&D has determined will or does have a clear and substantial impact on important public policies or private sector decisions” or “Highly Influential Scientific Information” (i.e., ISI with “a potential impact >$500 million in any one year on either the public or private sector (Forest Service Research Data Archive Submission Form for Forest Service Authors, n.d.)).
3. **Human Assistance:** The FSRDA is an actively curated repository, and data submissions are reviewed “for content and conformance to the metadata standard” (USDA, *Submitting Data,* n.d.). The archive team may return review comments that submitters must resolve, with archive team assistance as needed, before publication. The archive team also assigns a DOI to final publications.
4. **Metadata Standard:** The FSRDA uses two metadata standards: the Federal Geographic Data Committee’s Content Standard for Digital Geospatial Metadata (CSDGM) for geospatial data, and the Biological Data Profile (BDP), which is a superset of the CSDGM and adds additional elements related to taxonomy, methodology, and analytical tools. The standard is available as a downloadable Word document, “Metavist” software output, or Esri ArcGIS output (for spatial data only). These standards require metadata on the following:
   1. Identification (i.e., what are the data, where/when/why were they collected, what tools were used to collect them)
   2. Data Quality (i.e., how data were collected, how reliable the data are, gaps in the data)
   3. Spatial Data Organization
   4. Spatial Reference
   5. Entity and Attribute (i.e., lists and descriptions of all files and all variables within each file)
   6. Distribution (i.e., how to access the data, data formats)
   7. Metadata Reference (i.e., how current are the metadata, who is responsible for the metadata)

**Data Access**

A login is not required for access to this repository, and all data publications are available to the public for free. Access and download options for data vary by publication, but typically include an option to download a zip file that includes metadata, a file index, data files, and supplemental files. Users may also view a publication’s file index as an HTML page. Metadata is displayed according to the CSDGM and BDP standards detailed above, and may be viewed as HTML files or downloaded along with file indexes. In some cases, users may directly query a data publication’s database, view an image library, or explore an external repository.

DIP formats vary, but all publications I reviewed allow users to download a zip file of all data files included in the publication, which follow one or more of the accepted data formats listed earlier.

**Data Preservation**

Overall, this dataset is vulnerable to degradation and obsolescence over time, and could be difficult to preserve given its large size, proprietary data formats, and the fact that the dataset itself is housed on the proprietary ArcGIS platform.

The data files themselves, whether accessed on Data.gov or the U.S. Forest Service’s ArcGIS hub, are very large and can be difficult to download at all without a sufficient computer and access to internet that can accommodate large file downloads. The large file sizes also complicate preservation, as they are too large to fit on certain freely available repositories (namely GitHub, in this project’s case), and so cannot be easily copied and transferred. Next, while some of the data files exist in non-proprietary formats (e.g., CSV), others are proprietary to the ArcGIS or Google Earth platforms (e.g., shapefiles, KML files) and require specialized software to open and access. In the case of the KML file, I was unable to access it even after I tried to open it directly on Google Earth, as the platform kept timing out (possibly due to the large file size). Lastly, users can most easily access and interact with the data on the USFS’s ArcGIS hub page. However, Esri, the company that publishes and sells ArcGIS software, is a private company and there is no long term guarantee that it will continue to exist, maintain the ArcGIS software, or continue to allow these specific file types to run on future versions of the ArcGIS platform.

**Copyright Considerations**

According to the dataset’s [ArcGIS hub page](https://data-usfs.hub.arcgis.com/datasets/usfs::national-forest-system-trails-feature-layer/about), the data are available under a [CC0 1.0 Universal license](https://creativecommons.org/publicdomain/zero/1.0/), meaning that users can “copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission.”

**Human Subject Considerations**

The dataset does not appear to contain any data or other information about human subjects, or any individually identifiable information. Rather, the data are about trails and their features. The dataset does not even contain usage statistics for any trails, for example.

**Citations**

1. Esri. (n.d.). *What is a shapefile?* ArcMap. <https://desktop.arcgis.com/en/arcmap/latest/manage-data/shapefiles/what-is-a-shapefile.htm>
2. Federal Geographic Data Committee. (n.d.). *Content Standard for Digital Geospatial Metadata (CSDGM)*. <https://www.fgdc.gov/metadata/csdgm-standard>
3. Forest Service Research Data Archive Submission Form for Forest Service Authors, n.d.
4. Forest Service Research Data Archive Submission Form for Non-Forest Service Authors, n.d.
5. FS R&D, Research Data Services. (n.d.). *How to prepare data and metadata for submission.* <https://www.fs.usda.gov/rds/archive/info/How_to_prepare_data_and_metadata_for_FSRDA.pdf>
6. re3data.org: Forest Service Research Data Archive; editing status 2024-01-31; re3data.org - Registry of Research Data Repositories. http://doi.org/10.17616/R3404J last accessed: 2024-02-14
7. re3data.org: National Park Service Geographic Information Systems; editing status 2023-01-16; re3data.org - Registry of Research Data Repositories. http://doi.org/10.17616/R3R79F last accessed: 2024-02-14
8. U.S. Department of Agriculture. (n.d.). *Submitting Data*. Research Data Archive. <https://www.fs.usda.gov/rds/archive/submittingData>
9. *What is the difference between a map service, feature service, and a hosted feature layer?* (2022, May 11). Esri Technical Support. Retrieved February 18, 2024, from <https://support.esri.com/en-us/knowledge-base/faq-what-is-the-difference-between-a-map-service-featur-000027304>
10. *What is a KML file?* (2023, November 3). Keyhole Markup Language. Retrieved February 18, 2024, from <https://developers.google.com/kml/documentation/kml_tut>