

The Alaska Division of Geological & Geophysical Survey's Metadata Policy Development and Implementation

By Carrie L. Browne¹, Larry K. Freeman¹, and Gina R.C. Graham²

¹Alaska Division of Geological & Geophysical Surveys

3354 College Road

Fairbanks, AK 99709-3707

Telephone: (907) 451-5031, (907) 451-5027

Fax: (907) 451-5050

e-mail: carrie_browne@dnr.state.ak.us; larry_freeman@dnr.state.ak.us

²Gina R.C. Graham

2575 Cresbard Ct.

Fairbanks, AK 99709

Telephone: (907) 474-3647

e-mail: graben@gci.net

ABSTRACT

In December 2001, the Alaska Division of Geological & Geophysical Surveys (DGGS) identified an emerging problem concerning the documentation of our geospatial datasets. For example, DGGS released 20 publications in 2001, of which 17 required metadata; however, only 1 actually included metadata. DGGS has published an enormous number of geospatial datasets in recent years, and the number of publications with geospatial datasets increases every year. Until recently there was no incentive, penalty, or process for documenting that data. To remedy this, DGGS launched a division-wide metadata policy. This policy stipulates that all future DGGS publications with geospatial data must have Federal Geographic Data Committee (FGDC) compliant metadata before they are published.

In January 2002, the DGGS director assembled a metadata committee to devise an efficient approach to generating metadata for our geospatial datasets. The committee proposed the establishment of an interim "transition period" of 6 months, January through June 2002, during which a minimum metadata standard would be met prior to publication. After June 2002, complete FGDC-compliant metadata would be required for all geospatial datasets prior to publication. Throughout the transition period, the metadata committee met semimonthly to familiarize themselves with the details of the FGDC Content Standard for Digital Geospatial Metadata Workbook version 2.0, and to discuss how the standard should be applied to DGGS. The committee's first priority was to determine which of the FGDC metadata elements specifically apply to the various types of geospatial datasets

published by DGGS. Another task during the transition period was to research existing metadata generation tools and select the best tool for DGGS. The staff's diverse preferences in GIS software and geospatial data formats drove the metadata committee to recommend that DGGS create its own metadata tool specifically designed for the unique needs of its staff.

Metadata is now an integral part of the publication process at DGGS, as evidenced by the fact that, in 2002, 64 DGGS publications that required metadata did include it. We are continuing to improve the methods and have established an initial DGGS-specific metadata text template. We expect to have a user-friendly Web-based application available to DGGS personnel by October 2004.

INTRODUCTION

Since April 11, 1994, when President Clinton signed Executive Order 12906, the issues of data archiving and documentation have been brought to the attention of most local, state, or federal government agencies. One consequence of this Executive Order is that agencies must develop plans and allocate resources to generate metadata for geospatial datasets. If an organization introduces a metadata policy into their geospatial dataset creation process, managing the collection of their geologists' datasets becomes less cumbersome in terms of knowing what data they have, how it was generated, when, and by whom. One obvious scenario that illustrates this would be if a geologist were to leave a state survey after many field seasons. If the state survey required metadata documentation for all its geospatial datasets, it would be able to

protect and manage the geologist's data left behind. The background information on how the dataset was generated would be preserved and the dataset would not be discredited by possible data misuse. Of course, the initial reaction of an organization's members learning they are now responsible for the generation of metadata might not be optimistic. Fortunately, this reaction is generally short lived because no one wants to have spent valuable time compiling a geospatial dataset only to have the data users discredit the dataset for lack of metadata documentation.

ESTABLISHING A METADATA POLICY

The Alaska Division of Geological & Geophysical Surveys (DGGS) has published geospatial datasets for more than 10 years and has, like most agencies, experienced challenges with its staff regarding the need to create metadata for all publishable geospatial datasets. Although DGGS staff knows why metadata is necessary and how to generate it for their geospatial datasets, very few older published geospatial datasets included metadata documentation. There were no real consequences to the staff for this lack of metadata, and so geospatial datasets commonly were published without adequate documentation. When a digital dataset was requested, metadata would have to be generated, sometimes years after the dataset was created and by individuals who didn't actually produce the data.

The DGGS metadata policy emerged in December 2001 after numerous public requests for digital geospatial datasets from DGGS publications that lacked metadata documentation. An investigation into our geospatial dataset documentation process was initiated to determine why none of the requested geospatial datasets included metadata. Our findings were staggering. During the 2001 publication year, only 1 of 17 publications containing geospatial datasets included proper metadata. Moreover, less than 5 percent of our published geospatial datasets over the previous 10 years included FGDC-compliant metadata.

We reviewed how DGGS had been addressing metadata creation and found that the agency was creating metadata for requested digital geospatial datasets on an *ad hoc* basis. Unfortunately, generating metadata this way was rapidly becoming overwhelming as the volume of geospatial data without metadata increased. We also realized that the originators of the geospatial data either had left the agency or were committed to other projects and were not available to assist in resurrecting information needed to write the metadata.

The lack of proper geospatial dataset documentation and the increasing pressure of public requests for our metadata were brought to the attention of the director of DGGS. In response, the director established a formal metadata policy stating that as of January 2002, all new geospatial datasets must include FGDC-compliant

metadata before they will be published by DGGS. This formal policy was met with opposition and frustration at the survey. In addition, the initial metadata policy was flawed because it did not include an implementation plan. In response to the resistance to this policy, the director assembled a metadata committee from his staff to establish an effective work plan for generating FGDC-compliant metadata at DGGS. As a result of the staff discussions on metadata, it was clear that one goal of the committee would be to produce instructions and methodology for the generation of metadata that could be clearly communicated to project teams at DGGS.

METADATA TIMELINE FOR DGGS

The metadata committee consisted of DGGS managers and various staff members. The committee agreed that a strategy was needed to establish a transition into systematically generating metadata for our geospatial datasets. The recommended timeline was distributed at a survey-wide staff meeting in January 2002, and is presented in figure 1.

METADATA COMMITTEE MEETINGS

Once the survey staff agreed on the timeline, a second metadata committee was assembled. This committee was also staff-based and consisted of at least two members of each DGGS section (Minerals, Energy, Engineering Geology, and Geologic Communications). The committee met semimonthly for 6 months, during the timeline-defined "transition period," to become familiar with the details of the FGDC Content Standard for Digital Geospatial Metadata Workbook version 2.0 (Federal Geographic Data Committee, 2000) and to determine how to apply the FGDC standard to DGGS. Many of the committee members were not familiar with the FGDC workbook or its contents and had never completed a full FGDC-compliant metadata file for a geospatial dataset. To help get past our lack of metadata knowledge we used examples of metadata from USGS geologic mapping products, as well as some from other state surveys, and communicated by e-mail with Peter Schweitzer of USGS regarding metadata formatting and development questions that arose.

The first lesson for the committee members was to learn how to read the FGDC workbook and interpret its contents. Once everyone on the committee was familiar with the FGDC workbook we determined that our main concern was to select which FGDC metadata elements applied to specific types of geospatial datasets published by DGGS.

As the committee evaluated various metadata elements within the FGDC workbook, we began interpreting and clarifying relevant metadata element definitions to relate specifically to DGGS geospatial datasets. Our idea was that this strategy would assist data originators (DGGS

Timeline for DGGs Metadata Requirements

The following DGGs metadata requirements and timeline were discussed and agreed upon in a meeting on December 20. Present at the meeting were ADGGs director, section chiefs, and other interested members of the survey.

January 1, 2002 to June 30, 2002: Transition Period

- Minimum required fields for FGDC compliant metadata will be completed for geospatial data layers that were created for an ADGGs publication. This metadata will be considered preliminary and will include the fields listed on the "Interim ADGGs Minimum Required Metadata Fields" document (see attached template and example **).
- Preliminary metadata will be submitted to metadata coordinator for review to ensure that all interim minimums required metadata fields are properly documented.
- Publications submitted during this transition period will require a work-plan to be submitted by the author to Geologic Communications section chief documenting the planned future date of completion of the final FGDC compliant metadata. The date of completion must be on or before October 31, 2002.
- Once the metadata has been reviewed by metadata coordinator, edited by publications editor, and the work plan documenting the future date of completion of the final FGDC compliant metadata has been approved by Geologic Communications section chief, the publication containing the data can be released, provided all other publication requirements are met.
- During this transition period, the Geologic Communications members with input from ADGGs staff will develop ADGGs-specific guidelines and templates to assist ADGGs staff in the creation of complete FGDC compliant metadata.

July 1, 2002: FGDC Compliant Metadata Required

- Beginning July 1, 2002, FGDC compliant metadata will be required for all geospatial data prior to its release in an ADGGs publication.

** The Interim template is very similar to Figure 2, see below

Written 1/3/02

Edited 1/4/02

Figure 1. DGGs timeline for implementing metadata policy.

geologists, staff, and other data contributors) in generating FGDC-compliant metadata for their unique geospatial datasets by helping determine when and why a particular element would apply to their dataset. The committee also supplemented metadata element domains with DGGs "boilerplate" text where appropriate or clarified element definitions so that data originators could clearly understand what was an appropriate entry for that element.

The committee thoroughly researched available metadata generation tools to determine the best tool for

DGGs. We concluded that DGGs should create its own metadata tool designed specifically for its needs because the currently available metadata generation tools are software- and geospatial dataset-format specific, which does not allow for much flexibility. The metadata tool DGGs created is specialized to our diverse preferences in GIS software (ESRI ArcGIS, MapInfo, ERMapper, AutoCAD) and formats of geospatial datasets (geologic maps, data tables, databases).

The DGGs-specific metadata generation tool was

produced in two phases. The first phase, completed in July 2002, consisted of constructing a text-formatted file or text template that includes all FGDC workbook elements and their DGGSS-specific definitions applicable to DGGSS datasets (fig. 2). In addition, DGGSS provides training and support for those using this metadata template. The

second, ongoing phase entails the production of a user-friendly interface application that is (1) not software specific, and (2), will output an FGDC-compliant metadata file when the data originator has finished entering information into the working metadata file. This tool is tentatively scheduled for completion by October 2004.

Full FGDC Metadata Template:

```

Identification_Information: COMPOUND-
  Citation: COMPOUND-
    Citation_Information: COMPOUND-
      Originator: ALWAYS-REPEAT AS NEEDED- The name of an organization
      or individuals (authors) that developed the data set. If the names of
      editors or compilers are provided, the name must be followed by "(ed.)"
      or "(comp.)" respectively.
      Publication_Date: ALWAYS- The date when the data set was
      published or otherwise made available for release. It MUST be in the
      YYYY format, or OPTIONAL- in the YYYYMM or YYYYMMDD format if the month
      and/or day of publications are known.
      Publication_Time: IF APPLICABLE TO DATA SET- This is the time of
      day when the data set is released (used if essential to the dataset,
      any time sensitive datasets)
      Title: ALWAYS- The name by which the data set is known, the title
      of the map or publication that the dataset is the source for.
      Edition: IF APPLICABLE TO DATA SET- The version or edition of the
      publication (In the case of more than one version, modification or
      edition, this field needs to be filled in.)
      Geospatial_Data_Presentation_Form: ALWAYS- The mode in which the
      geospatial data are represented, "atlas" "audio" "diagram" "document"
      "globe" "map" "model" "multimedia" "presentation" "profile" "raster
      digital data" "remote-sensing image" "section" "spreadsheet" "tabular
      digital data" "vector digital data" "video" "view" (-You can string
      more than one together with commas.)
      Series_Information: COMPOUND-
        Series_Name: ALWAYS- Long version of the publications series
        name the data set is a part of, "Raw Data File", "Geophysical Report",
        "Preliminary Interpretive Report", "Report of Investigations",
        "Professional Report", "Special Report", "Miscellaneous Publication",
        "Guidebook", "Information Circular"
        Issue_Identification: ALWAYS- The publication number by which
        the data set will be referred to in a bibliography. "RDF 2002-01", "GPR
        2002-01", "PIR 2002-01", "RI 2002-01", "RI 2002-01", "PR 100", "SR 01",
        "MP 01", "GB 01", "IC 01"
        Publication_Information: COMPOUND-
          Publication_Place: Fairbanks, AK
          Publisher: State of Alaska, Department of Natural Resources,
          Division of Geological & Geophysical Surveys
          Other_Citation_Details: ALWAYS- This includes other information
          required to complete the citation such as number of sheets, pages,
          disks, or other information.
          Online_Linkage: ALWAYS- REPEAT AS NEEDED- The name of an online
          computer resource that contains the data set. Entries should follow the
          Uniform Resource Locator convention of the Internet. Example:
          \\craton\archive1\fbxpdfdata\covers\fbxmatld, the purpose of this is so
          that we know where the dataset resides should the originators forget.
          "NONE" is an option but it should be a last resort.
  
```

Figure 2. Metadata text-formatted template, including all FGDC sections/elements and their definitions, to be used when generating metadata for a geospatial dataset.

DEFINING A DATA UNIT

As the committee gained more knowledge about how to produce FGDC-compliant metadata, the question arose of how to divide a geospatial dataset to produce more helpful FGDC-compliant metadata. Establishing how many metadata files were needed for a large and diverse geospatial dataset was difficult. No one wanted to generate too many or too few metadata files for a geospatial dataset. A DGGs geologic publication consists of multiple layers or themes, each made up of a coverage or shapefile. If each coverage or layer of a published dataset were documented with an individual metadata file there would be numerous metadata files containing redundant data for each unique publication. On the other hand, if a single metadata file documents the dataset as a whole there is the legitimate concern that important pieces of metadata, essential to a specific layer or theme, may be lost. The metadata committee discussed three main options to address this problem.

Option 1: Publication Set = Data Unit

We proposed generating one metadata file for each published set of geospatial data and came to the following conclusions:

- 1) Only one metadata file would need to be generated for the dataset for this option.
- 2) The metadata file has the potential for being enormous and including a lot of information not applicable to any one thematic layer.
- 3) The types of links or relationships between elements of different sections of the FGDC standard necessary for this type of metadata file do not exist. This means a loss of data integrity between metadata sections/elements and specific thematic layers.
- 4) If a thematic layer was copied and reused from this publication set there is no way to parse the applicable metadata elements for that layer.

Option 2: Thematic Layer = Data Unit

Next we suggested generating one metadata file for each thematic layer composing a geospatial dataset. We concluded:

- 1) Each thematic layer would be fully documented by metadata with no loss of data integrity.
- 2) Many metadata files would have to be generated for one publication set.
- 3) There would be some redundant data contained in each metadata file within a subset of the same publication set.

Option 3: Distribution Determination = Data Unit

The last option that we considered was generating our metadata files as a function of how a digital geospatial dataset will be distributed. The following are our conclusions:

- 1) It would be easy for the originator of the dataset to determine how to break out the geospatial dataset into data unit(s).
- 2) Each geospatial dataset is fully documented by metadata with no loss of data integrity.
- 3) We may need to generate many metadata files that will include redundant data when a publication is distributed by thematic layer.

The metadata committee concluded that Option 3 was the best fit for DGGs. This means that the number of data units that make up the dataset (and thus, the number of metadata files required for a specific geospatial dataset) will be determined by how a digital geospatial dataset will be distributed by DGGs. For example, if a published geologic map will be distributed as a stand-alone map only, then the data unit would be the entire map and only one metadata file is needed for the publication. On the other hand, if a published geologic map will be distributed on a thematic layer basis when requested, then one data unit is assigned to each thematic layer.

METADATA GENERATION PROCESS

Once the text-formatted file including all FGDC workbook elements and definitions (fig. 2) became available for DGGs in July 2002, the survey began generating full FGDC-compliant metadata for our geospatial datasets. The metadata generation process is illustrated below in figures 3 and 4.

DGGs has an assigned Metadata Coordinator to check all metadata files for errors using the mp (metadata parser) tool (created by Peter Schweitzer of USGS). This saves time and effort by not trying to train everyone in how to interpret the error file mp produces. This mp program is a compiler to parse formal metadata, checking the syntax against the FGDC Content Standard for Digital Geospatial Metadata and generating output suitable for viewing with a web browser or text editor. It runs on UNIX systems and on PC's running Windows 95, 98, or NT. The mp tool generates a textual report indicating errors in the metadata, primarily in the structure but also in the values of some of the scalar elements (i.e. those whose values are restricted by the standard). We also have an on-site publications editor who thoroughly reads the metadata files for content and grammar to help eliminate errors.

Once the metadata file(s) have been generated for

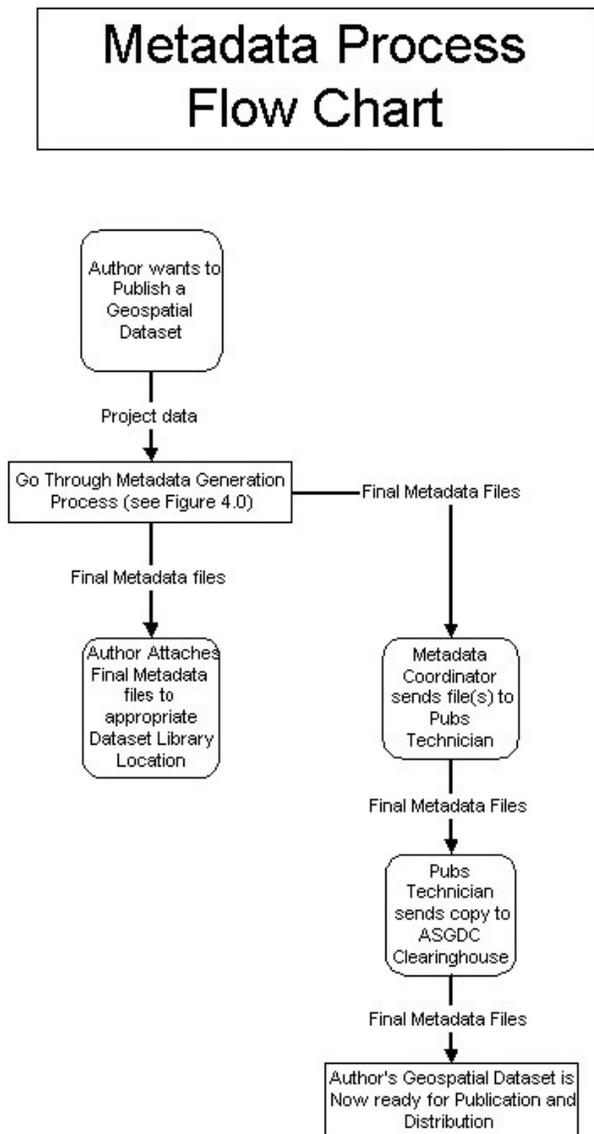


Figure 3. Metadata process flow chart providing an overview of DGGs metadata generation process.

a geospatial dataset they are sent to the Alaska State Geospatial Data Clearinghouse (ASGDC) to be posted on their website <<http://www.asgdc.state.ak.us/>>.

WHAT WE HAVE LEARNED

DGGs was not in a position in January 2002 to immediately begin generating proper FGDC-compliant metadata for its geospatial datasets. Even though the director made it a priority and policy that FGDC-compliant metadata must be included with all geospatial datasets published by DGGs beginning January 2002, the survey needed a strategy specifying how this new priority and policy would be implemented. This need is evidenced by the fact that metadata generation has occurred with greater

frequency and precision since the metadata committee's strategy was introduced to the survey.

It is clear that a need exists for a user-friendly, non-software-specific application to help generate FGDC-compliant metadata. In all the research done by the second metadata committee, participants were astonished that with a federal metadata mandate in place there are few applications if any available that are non-software-specific and assist a user in generating metadata files.

The "FGDC Content Standard for Digital Geospatial Metadata Workbook version 2.0" is an invaluable resource for generating metadata for geospatial datasets. It is also a great resource for training personnel on what is expected of them when producing a metadata file. In DGGs, we decided to bring the workbook into a personalized language that the staff could understand and relate to in the form of a text-formatted metadata template.

The more documentation completed throughout the geospatial dataset production phase, the easier it is to generate metadata prior to the publication process. Once people understand what specific information metadata files require, they become more organized throughout the data gathering and documentation process of their project, resulting in more efficient metadata generation.

CONCLUSIONS

The DGGs metadata policy established in January 2002 sparked rapid development of FGDC-compliant documentation of all geospatial datasets ready for publication by DGGs. In the publication year 2002 DGGs generated 64 publications that contain or were produced from geospatial datasets; each is documented with metadata. The director-appointed division-wide metadata committee was successful in learning what is required to complete FGDC-compliant metadata. DGGs created a specialized metadata tool to aid in the completion of FGDC-compliant metadata and is currently producing a user-friendly, non-software-specific application that will be ready for use in October 2004. We have proposed a project and are applying for funding to generate metadata files for the legacy datasets from DGGs publications released prior to January 2002. If the project is funded, our entire geospatial dataset collection will include associated metadata files and be properly documented, archived, and protected for future use.

REFERENCES

- Federal Geographic Data Committee, 2000, Content Standard for Digital Geospatial Metadata Workbook Version 2.0, Washington D.C., <http://www.fgdc.gov/metadata/meta_workbook.html>.
- Schweitzer, Peter N., 1995, MP: A compiler for formal metadata (2.7.6 ed.): Reston, Virginia, U.S. Geological Survey, <<http://geology.usgs.gov/tools/metadata/>>.

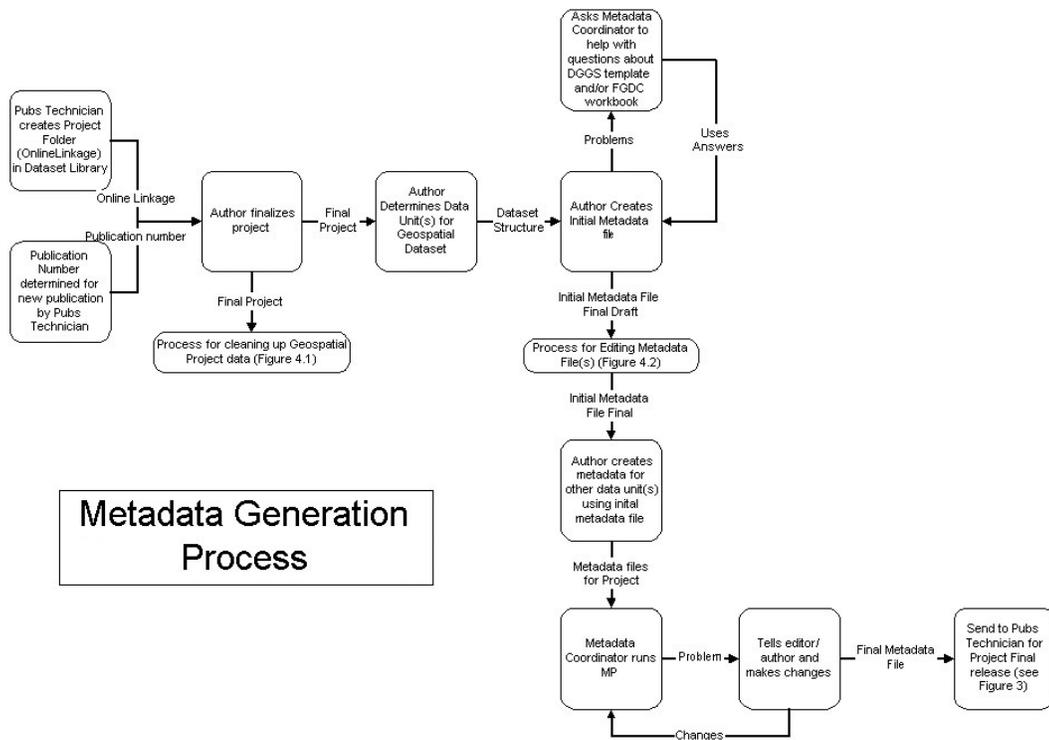


Figure 4.0. Metadata generation flow chart detailing all steps to be used when generating FGDC-compliant metadata for a geospatial dataset.

Geospatial Data Cleanup Process

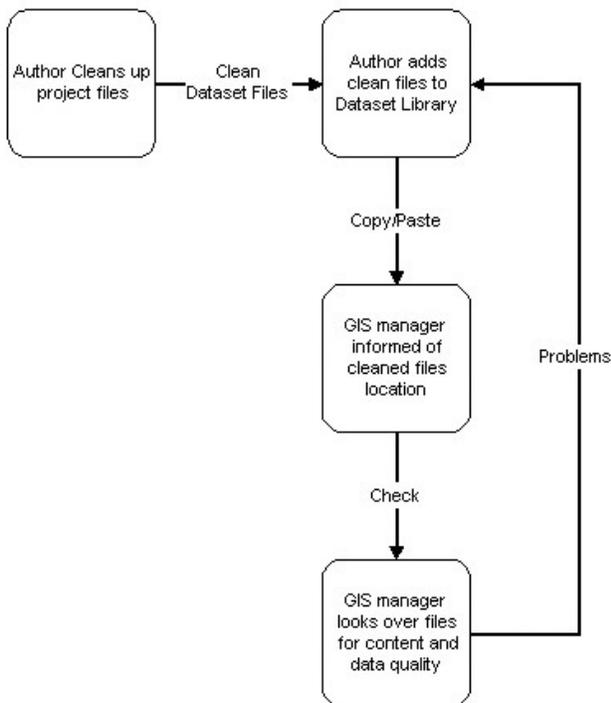


Figure 4.1. Geospatial data cleanup process detailing the checks and balances of our geospatial data files prior to storage in the dataset library.

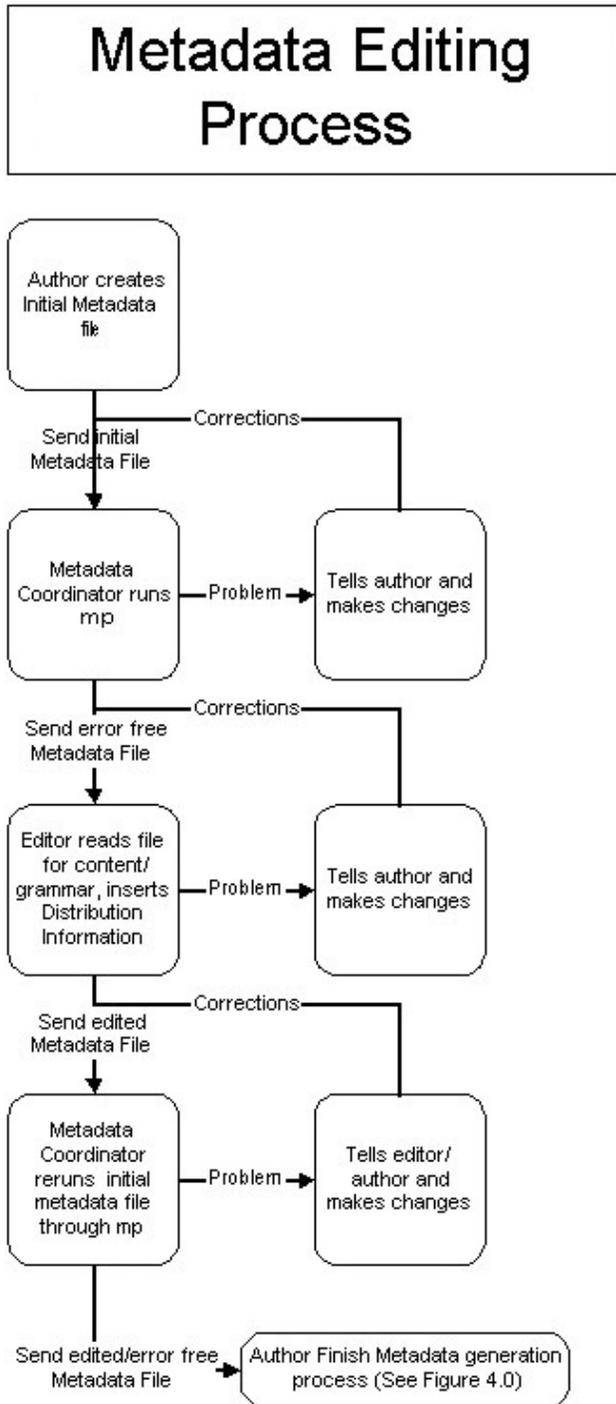


Figure 4.2. Metadata editing process detailing the steps a metadata file goes through at DGGs before it is considered FGDC-compliant metadata.