

CDI SSF Category:
Management, Policy, and Standards

Metadata Wizard: An Easy-to-Use Tool for FGDC Metadata Creation in ESRI's ArcGIS Software

Applicants/Principal Investigator(s):

Drew Ignizio, Contractor with Cherokee Services Group (CSG)
USGS Fort Collins Science Center, 2150 Centre Ave., Fort Collins, CO 80526
Phone (970) 226-9213 Email dignizio@usgs.gov

Colin Talbert, USGS Fort Collins Science Center, 2150 Centre Ave., Fort Collins, CO 80526
Phone (970) 226-9425 Email talbertc@usgs.gov

Michael O'Donnell, USGS Fort Collins Science Center, 2150 Centre Ave., Fort Collins, CO 80526
Phone (970) 226-9407 Email odonnellm@usgs.gov

Abstract:

Creating FGDC-compliant metadata is mandated for all federal GIS professionals and a best practice for the geospatial data community. However, the complexity of the FGDC standard and the limited availability of easy-to-use tools continue to make metadata creation a daunting challenge. Staff at the U.S. Geological Survey Fort Collins Science Center are working to develop a Python toolbox for ESRI Desktop to facilitate a mostly automated application to create and update metadata records, all from one easy-to-use Graphical User Interface (GUI). The USGS metadata tool automatically populates many metadata elements: the spatial reference, spatial extent, native system/processing environment, and certain GIS attributes. Once the software auto-populates these elements, the user can easily add attribute definitions and other relevant information in a GUI to create an FGDC compliant metadata record. The tool offers a simple interface, free of esoteric metadata language that has the potential to save many government and non-government organizations a significant amount of time and costs, as well as facilitating the development of FGDC compliant metadata for all GIS users.

Total funding amount requested: \$18,917.73

Total in-kind funding: \$14,372.70

Specific Datasets Exposed/Augmented/Distributed:

No specific datasets will be distributed with this project.

Geographic/geologic/ecosystem/habitat/taxonomic/other context:

FGDC Metadata in ESRI Desktop
Metadata Participation among Geospatial DataUsers,
Automated FGDC Metadata Creation

Type of Product(s) Generated:

Python-based GUI for ESRI ArcGIS
USGS Publication and Tool Release

Summary

Introduction and Background:

Creating metadata records compliant with the Federal Geographic Data Committee (FGDC) standard poses an enormous challenge for many geospatial data users. This has resulted in many scientific data sets being poorly or incompletely documented; in many cases, in spite of best practices or formal requirements, data producers simply forego metadata creation. The work outlined in this proposal immediately addresses this problem by way of a tool whose design and implementation has been dramatically shaped by the shortcoming of other metadata tools heretofore. Based on the awareness that a large majority of geospatial users within the scientific community are working within the ESRI ArcGIS environment, the tool described is a Python-based toolbox designed to be added to ArcCatalog. By automating much of the process of creating metadata, overcoming the technical difficulties associated with working with metadata in ESRI ArcMap 10, and through a conscientious effort to capture critical user information while avoiding the technical language of the FGDC standard, the tool represents an all-in-one resource for metadata creation/review. The tool provides functionality that other have lacked (such as automatically retrieving information inherent to geospatial data sets, providing a semi-automated form to describe tabular field values, and a built-in viewer to preview a formatted XML metadata record). The application can be easily installed on a standard version of ESRI ArcGIS and dovetails with other metadata efforts and existing workflows by allowing easy export of metadata records.

CDI SSF Category: Management, Policy, and Standards

Title: Metadata Wizard: An Easy-to-Use Tool for FGDC Metadata Creation in ESRI's ArcGIS Software

Contacts:

Drew Ignizio

Contractor with Cherokee Services Group (CSG)

USGS Fort Collins Science Center, 2150 Centre Ave., Fort Collins, CO 80526

Phone (970) 226-9213

Email dignizio@usgs.gov

Colin Talbert

USGS Fort Collins Science Center, 2150 Centre Ave., Fort Collins, CO 80526

Phone (970) 226-9425

Email talbertc@usgs.gov

Michael O'Donnell

USGS Fort Collins Science Center, 2150 Centre Ave., Fort Collins, CO 80526

Phone (970) 226-9407

Email odonnellm@usgs.gov

This tool will constitute a much-needed resource for individuals within the USGS and in the greater scientific community tasked with creating FGDC-compliant metadata for geospatial data sets.

Scope

Project Description

This project consists of developing, testing, and making publicly available one simple tool that will enable geospatial data users to easily create and edit FGDC metadata records, all from within ESRI's ArcGIS software. Such a tool will address a grave need within the USGS and the greater scientific community at large while contributing significantly to the goal of the Community for Data Integration (CDI) to advance data and knowledge management, policies, standards, and best practices.

Quality metadata greatly increases the value and scientific reliability of geospatial data. An innovative tool to automatically generate metadata content and make the overall metadata creation process easier for scientists will increase the amount of data documentation taking place among the entire geospatial community and provide enormous benefit to future data users. This tool will contribute to virtually every aspect of the Science Support Framework (SSF), but specifically the processing, analysis, preservation, and publishing/sharing components with regard to scientific data. As such, this project constitutes a major step toward 'advancing scientific data and information management and integration capabilities' as outlined in Category 1 of the SSF.

Partners/participants

The proposal applicants are the primary developers for the project and work at the U.S. Geological Survey Fort Collins Science Center. Additional testing support will employ other contracted data specialists; all contractor staff are managed through task orders which are reviewed and processed by the Contracting Officer's Representative (COR) at the U.S. Geological Survey Fort Collins Science Center.

This project has been developed in support of, and is partially funded through a collaborative relationship with the U.S. Geological Survey Core Science Analytics and Synthesis (CSAS) metadata team. This tool will meet a different user need than tools currently being developed by CSAS (i.e., our targeted audience will specifically be ESRI users), but the workflow and tool outputs have been conceived such that this tool will complement and dovetail with the efforts of the CSAS team.

Other participants and beneficiaries will ultimately include principal investigators (PI's), scientists, and other geospatial data users within the USGS and wider scientific community.

Goals

The design and function of the tool proposed in this project is motivated by three primary goals:

1. Automate as much of FGDC metadata creation as possible to reduce user burden and eliminate input error.

Given the fact that multiple components of the FGDC metadata standard require information that is intrinsic to geospatial data sets, and that much of this information can be retrieved by accessing the properties of a data set with the ESRI ArcPy module for Python, this constitutes a very realistic goal, and one that to date has not been fully implemented in other tools.

The proposal applicants have been successful in their initial efforts to extract information on the spatial reference, spatial extent, geospatial data format, vector feature count or raster column/row data, and the tabular attributes associated with a particular data set. This information is subsequently 'mapped' to the appropriate XML element (XPath) corresponding to a particular FGDC metadata component within the

output XML document that is created. This not only makes metadata creation much simpler, but it eliminates the potential for user input error in some of the most critical components of a metadata record.

2. Provide a simple user interface to capture necessary metadata content as required by the FGDC standard, while avoiding the technical language of the standard.

After extracting all the properties that can be read directly from a data set and writing them to the output FGDC record, the Arc toolbox then launches a GUI in which a user can add additional metadata information that cannot be automatically populated. Rather than having the complex standard dictate the design of the interface, a concerted effort to create a simple, usable form is employed. Complex metadata elements whose presence or absence may determine a secondary set of required information are encapsulated and simplified in the design and dynamic behavior of the tool.

3. Handle format conversions automatically and eliminate the complexity caused by changes in the ArcGIS 10 environment.

The changes in the way that metadata is handled in ESRI's ArcGIS 10 software have complicated (and in some cases, made unworkable) the workflows of many users tasked with creating and updating metadata. The tools needed to work with metadata now require even more in-depth knowledge to properly handle metadata format transformations, import and export stand-alone XML records, and account for the fact that multiple users accessing files (such as on a network location) may result in undesired and unnoticed changes being made to metadata records. To the greatest extent possible, this tool addresses these issues. Metadata records in both the FGDC format and the ESRI Arc 10 format are handled. Updated metadata is associated with data sets without requiring a user transformation, an unmodified backup copy of a data set's original metadata is preserved in a separate location, and a stand-alone XML record in FGDC format can be generated with a simple toggle option. This workflow greatly reduces the burden on users in the ESRI environment and lends itself elegantly to other metadata needs, such as the submission of stand-alone metadata records to a metadata clearinghouse.

Scope of Work and Requested Funds

The funds requested in this proposal will cover the time needed to complete development of this tool, conduct a robust testing and bug resolution phase, and the preparation and submission of the final tool as a downloadable USGS product. Funding to cover internal USGS publication costs are also budgeted. For additional information please see the 'Technical Approach,' 'Budget,' and 'Timeline' sections.

Products and Outcomes

The outcome of this project will be an all-in-one resource for ESRI users to easily develop metadata for geospatial data in compliance with federally mandated standards (i.e., FGDC).

This effort will

- Deliver an immediate benefit to solve an existing data integration challenge;
- Demonstrate a methodology and/or solutions architecture that can be repeated/replicated for other data or research projects;
- Provide a benefit to scientists;
- Promote standards and best practices for data management; and
- Complement other SSF categories, elements, and CDI projects.

Ultimately, the software will be made available as a component of a USGS publication and/or as a download from the ESRI ArcGIS Resources website. In either case, the tool will be accessible as a relatively small download from a persistent URL on the web and available to the CDI, USGS, and others. A final round of testing by staff at the U.S. Geological Survey Fort Collins Science Center will involve downloading and using the deployment package to ensure proper installation and functionality. For additional details, please see the ‘Technical Approach’ and ‘Timeline’ sections of the proposal.

Technical Approach

The process or software development life-cycle for the ArcGIS metadata toolbox includes assessing the requirements, design, implementation, testing/validation, deployment, and maintenance components. We have been working with researchers, data developers and FGDC coordinators to assess the requirements of the mostly automated FGDC compliant metadata creator tool described above. Based on these criteria, we designed a workflow to develop an easy to use GUI with minimal metadata jargon and an intuitive interface. The implementation/development of the tool is largely completed, but development for handling certain GIS data remains to be incorporated into the software.

While the initial development of such a workflow has been largely successful, robust tools require comprehensive testing to identify and address bugs and challenges, both known and unforeseen. To this end, a full development suite of test environments and data sets will be systematically employed to ensure a stable product. Eight virtual machines will be configured as test environments to run the tool in all combinations of the Windows XP and Windows 7 operating systems with either ESRI Desktop 10.0 and 10.1, in both 32 bit and 64 bit. Against these machines, tests will be run with the tool using a wide array of spatial data including large and small vector data sets, discrete and continuous raster data in various formats (e.g., ESRI grid, TIFF, and IMG files), feature data sets from personal and file geodatabases, and stand-alone tabular data.

With the aforementioned data types, three different metadata scenarios will be created for each: one in which the file has an associated FGDC metadata record, one in which the file has a metadata record in the Arc10 format, and one in which no metadata is associated. Testing will involve opening each file on the different machines with the tool and carefully tracking any issues with viewing, modifying, or creating metadata. The output XML records generated by the tool will be visually inspected to check for data omission or improper Xpath assignment for any metadata values. The metadata parser utility (MP), developed by USGS, will also be run on outputs to identify problems that arise from a conflict with the FGDC metadata standard. Results from this testing period will guide the second 6-week period of bug resolution, as outlined in the proposal (please see the ‘Timeline’ section).

The tool relies on standard Python modules, all of which are present on a user’s machine as part of a standard ESRI installation, and included executable files that are run directly from a user’s machine which require no external dependencies. This is by design as such an application can be more easily run across different platforms and greatly simplifies the installation and use process for tool users. The software deployment of this product will likely be a zipped folder, which will include the necessary components to run on a Microsoft Windows operating system with an installation of ArcGIS. To run the tool, users simply acquire the ArcToolbox package (zip file) and add the toolbox to the ESRI desktop application.

Project Experience

The applicants are uniquely suited to the objectives outlined in this proposal and possess expert knowledge in the various areas needed to complete this task. The team members all have an established track record of Python programming, GUI development with VB.NET, and integrating tools and workflows to process and document geospatial data in the ESRI environment. In addition to these skills all three applicants regularly work with metadata and are intimately familiar with the FGDC metadata standard; several team members on this proposal serve as the metadata development and review group for the U.S. Geological Survey Fort Collins Science Center and also as metadata reviewers for the U.S. Geological Survey Core Science Analytics and Synthesis (CSAS) Metadata Clearinghouse.

The applicants have worked extensively with PI's and other individuals tasked with creating metadata but unfamiliar with the FGDC standard. These experiences and the feedback garnered in this process over the past years have lent invaluable insight toward the design of a tool that minimizes the complexity of the process of creating metadata, while capturing the necessary information to produce a compliant (and complete) FGDC metadata record.

Commitment to Effort

The tool produced in the scope of this project will be an incredibly valuable resource and a key utility to be integrated into the current (and ongoing) metadata creation and review work at the U.S. Geological Survey Fort Collins Science Center. While the metadata tool will be made available to other USGS personnel and the general public, the need of staff and scientists for such a tool in Fort Collins has been long standing and acute. As such, there is a vested interest on behalf of the proposal applicants to not only develop this resource, but to maintain it into the future as new hardware and software requires. While this proposal accounts only for the finalization, publication, and release of the tool as outlined here, there are no insurmountable obstacles to releasing updated version of the tool in the future as needed.

In more technical terms, the proposal presented outlines the specific system/software requirements for which the metadata tool will be tested and made available (please see the 'Technical Approach' section above). However there are several possible changes that may present themselves in the future, requiring updates to the tool. These include the eventual migration to the Windows 8 operating system, new versions of ESRI's ArcMap software (subsequent to 10.1), and the possible transition to the ISO metadata standard as the required format for federal geospatial data users. The applicants are well aware of these changes and the influence they may have on the tool as currently proposed.

Insofar as any of these developments would affect the performance of the proposed tool, staff at the U.S. Geological Survey Fort Collins Science Center will continue to research and implement changes to this tool and overall metadata workflows as necessitated by the ongoing metadata needs that exist at the center and agency level. Provided future funding exists for additional testing and formal publication costs for any new versions of this tool, any updated versions of this tool could be made available to the USGS and the public.

Budget

The proposed budget is outlined below in Table 1.

Budget Category	Federal Funding "Requested"	Matching Funds "Proposed"
-----------------	-----------------------------	---------------------------

1. SALARIES (inc. number of hours and hourly rate):

Personnel		
<i>Michael O'Donnell</i>	\$1,680.00 (40hrs @42.00)	\$
<i>Colin Talbert</i>	\$	\$2,000 (40hrs @50.00)
Contract Personnel		
<i>Drew Ignizio</i>	\$8,528.50 (185hrs @46.10)	\$6,915.00 (150hrs @46.10)
<i>Aaron Freeman</i>	\$1,500.00 (40hrs @37.50)	\$
Total Salaries:	\$11,708.50	\$8,915.00

2. FRINGE BENEFITS:

Total Fringe Benefits:	\$0	\$0
------------------------	-----	-----

3. TRAVEL EXPENSES*:

Total Travel Expenses:	\$ 0	\$0
------------------------	------	-----

4. OTHER DIRECT COSTS: (itemize)

Support from Computer Operations for Tool Testing	\$	\$1,500.00
Supplies	\$	\$
Training	\$	\$
Publications	\$2,000.00	\$
Office supplies	\$	\$
Communications Costs	\$	\$
Total Other Direct Costs:	\$2,000.00	\$1,500.00

Total Direct Costs:	\$13,708.50	\$10,415.00
Indirect Cost (%)	\$5,209.23 (38%)	\$3,957.57 (38%)
GRAND TOTAL:	\$18,917.73	\$14,372.70

Table 1. Proposed Budget

Timeline

All project work will be completed by the end of July 2013. Submission of the final USGS Open File Report (OFR) for publication in the USGS Publications Warehouse will take place no later than August 1, 2013. For a detailed breakdown, please see Table 2.

Task	Week(s) After Award Date							
	1-2	3-4	5-6	7-8	9-10	11-12	13-14	15-16
Update and complete Metadata Wizard GUI, including dynamic 'Help' tips	X	X						
Address known issues/bugs, integrate contact lookup, etc.		X	X	X				
Organize and Assemble Test Data Suite for Test Case Matrix				X				
Open, view, and edit Metadata for each test data set and in each test environment using the Metadata Wizard				X	X			
Address the bugs identified with the test data sets in each test environment					X	X	X	
Finalize a version of the ArcToolbox package for the Metadata Wizard						X	X	
Preparation and submission of publication to USGS and/or ESRI ArcUser. Code will be made available on GitHub, ArcGIS Resources, or similar.							X	X

Table 2. Proposed Timeline with Scheduled Task Completion

Appendices

Appendix 1.

Letter of Support from Peter Schweitzer for 'An Easy-to-Use Tool for FGDC Metadata Creation in ESRI's ArcGIS Software'

Date: 8 November 2012

From:

Peter N. Schweitzer, Ph.D.
Eastern Mineral and Environmental Resources Science Center
USGS National Center
Reston Virginia

To: CDI proposal reviewers

Subject: Metadata tool of Talbert and Ignizio

This note encourages the CDI to support a proposal by Colin Talbert and Drew Ignizio to fund some of their time to enable them to bring to USGS users a metadata editing tool they have been working to develop.

In my ad-hoc work in support of effective documentation for scientific data, one of the most common complaints I've heard is that creating good quality metadata for data stored within ArcGIS has become significantly more complicated and difficult with the advent of ArcGIS 10. Tempting as it is to assume that ESRI will solve all such problems, I think that does not address the real need of USGS to engage this issue thoughtfully and thoroughly.

Colin and Drew have been working within the ArcGIS software environment to develop an attractive tool for creating and maintaining metadata that draws from the data and the GIS such information as can be taken with confidence, and provides the GIS user with sensible interfaces for entering or modifying the other, more important information which only a knowledgeable person can provide.

I was sufficiently impressed by their work that I took time to make substantial modifications to some of my metadata software so that my tools could be used more effectively by theirs. Specifically I re-vamped the error reporting system used by mp so that error messages can now be output in XML or plain text, with the location of errors specified using an XPath expression. I modified two of my web services to generate XML on request so these might be available to users of XML tools such as Colin and Drew's editor.

Both of these authors bring to this effort a wealth of personal experience with actual use of software to create scientific data sets. I have seen that they understand first-hand how small, seemingly insignificant aspects of the user interface can color a user's experience and strongly affect the attitude with which the work is done.

Most of the goals of the CDI depend heavily on good quality documentation of data, often at many different levels of detail. The work of these two promises to advance those goals in very practical ways. I urge CDI leadership to provide funding to support the completion of this much needed work.

Appendix 2.

Note of Support from Vivian Hutchison for 'An Easy-to-Use Tool for FGDC Metadata Creation in ESRI's ArcGIS Software'

Date: 11/8/2012

From:
Viv Hutchison
Data Management Program Coordinator
US Geological Survey
Core Science Analytics and Synthesis (CSAS)
Denver Federal Center
Building 810, Mail Stop 302
Denver, CO 80225
303.202.4227 (office)
571.481.7799 (BB cell)
vhutchison@usgs.gov

Subject: Arc Metadata Tool CDI Proposal

To encourage scientists to create formal metadata records about their data, it is critical to have appropriate and easy to use tools available. Many of our scientists in the Bureau are using ArcCatalog, but the "out-of-the-box" metadata creation process has some challenges. Completing a metadata tool add-in for ArcCatalog that can support our scientists by making documentation of their GIS data more stream-lined and easier to do is an important cause, and one that the Core Science Analytics and Synthesis Program supports. Our work with scientists on data management issues is growing, and this tool will help with our ability to offer documentation options for our scientists in an application they are already familiar with in their work.