Building an Enterprise version of GeMS (formerly NCGMP09 map schema)

By Jennifer Athey, Michael Hendricks, and Patricia Gallagher
Alaska Division of Geological & Geophysical Surveys
3354 College Road
Fairbanks, Alaska 99709

Jennifer Athey
Telephone: (907) 451-5028
email: jen.athey@alaska.gov

Michael Hendricks
Telephone: (907) 451-5029
email: michael.hendricks@alaska.gov

Patricia Gallagher
Telephone: (907) 451-5039
email: patricia.gallagher@alaska.gov

Abstract

The Alaska Division of Geological & Geophysical Surveys (DGGS) was awarded a U.S. Environmental Protection Agency (EPA) – Environment Information Exchange Network grant in 2016. One of the grant deliverables includes the development of a multi-map, multi-user "enterprise" database model based on the single-map Geologic Map Schema (GeMS) developed by the USGS and state geological surveys (NCGMP, 2010). The enterprise version of GeMS in development is ultimately intended for national use, as is a pilot data-sharing protocol to be developed with the model. To date, DGGS and other stakeholders in the geologic community have determined specifications for the enterprise database, begun testing the PostgreSQL-ArcGIS for Server environment, and initiated re-scoping of the GeMS data model to meet the specifications. The development process is designed to be collaborative; interested persons are encouraged to contact DGGS for information.

Slide 1. DGGS constantly reviews its business practices to identify ways to meet customer expectations and future-proof products and processes while more efficiently spending resources such as time, staff, and money. In other words, we strive to "do more with less." DGGS has identified enterprise-style data management as a way to increase both output and efficiency. Changes in data management will likely take many forms affecting internal and external data users, such as moves toward increased standardization, data centralization, interoperability, and accessibility. However, fundamental changes in DGGS business processes will happen gradually.

Slide 2. One big driver of change is a 3-year EPA Exchange Network grant awarded to DGGS in 2016. The grant has three goals, each with a related deliverable. Goal 2 focuses on the development of a multi-map, multi-user geologic database schema and pilot data-transfer protocol, while goals 1 and 3 concentrate on radon data management and visualization. The creation of the enterprise geologic database will help position DGGS to make progress on and ultimately achieve some of our long-term data-management objectives and data products, such as geologic map compilations and applications to serve out geologic spatial data.

Slide 3. The main purpose of the Exchange Network is to more easily share data among the EPA and its partner agencies, and data sharing is encouraged on a national scale. The program considers spatial data themes in the Office of Management and Budget (OMB) Circular A-16 such as geology as important to conducting analyses that support environmental and health issues. In keeping with the national-scale goals of the Exchange Network, DGGS intends to develop an enterprise database model and data-sharing protocol that might fulfill the data management needs of other geologic organizations as well as our own, to be developed in collaboration with the geologic community. Community input is sought during monthly tele-meetings, usually the second Monday of the month. DGGS thanks the many participants who have thus far contributed to informative discussions and the National Geologic Map Database for providing tele-meeting support.

Slide 4. The geologic database project is based on implicit specifications inherited from the IT infrastructure of DGGS and general goals of the project proposed to the EPA. The schema in development should be translatable to other database platforms and IT environments. We intend for this work to be well documented and available to others in the geologic community.

Slide 5. Project Goal 2 is described as developing an "enterprise" database. Community participants defined enterprise database largely based on the Esri definition.

Slide 6. The in-development enterprise database is also described as a "multi-map" and "multi-user" database, for which community participants created definitions.

Slide 7. DGGS believes an enterprise-style database will increase efficiency, after a period of more intensive training and creation of new workflows. New business processes will be developed over time to reap the benefits of centralized, standardized data. Employees will be able to find, utilize, and create data more quickly. Public users of geologic data will see increased accessibility, a greater number of near real-time data services, more robust applications, and faster publication of new data.

Slide 8. Community participants listed and voted on explicit specifications for the enterprise database, where a score of 1 is very important, score of 2 is somewhat important, and score of 3 is not important or not important at this time. Scores were averaged for each specifications. Finally, participants discussed each specification and agreed to accept, reject, or optionally accept the specification with the knowledge that not every survey will implement it. The specifications serve as guidelines and goals for the project, and go beyond the basic description of the enterprise database in the proposal to the EPA. Not all of the specifications may be addressed during the project timeline. Accepted specifications will be implemented before optional specifications.

Some specifications may appear to conflict with each other or with the project, for example, "ease of use for staff." Creating a more complex data management system will complicate work for data managers and administrators; however, the overall affect will be to make data management easier for all staff.

Slide 9. Most accepted specifications are related to the nature of the database model (flexible, scalable, interoperable with other databases), types of data that it will hold (bibliographic, original, analytical, unpublished, field, and ephemeral data and FGDC/ISO metadata), and its ability to interface with the data (topology, queryable). Although some changes and additions to the GeMS model will obviously be necessary to accommodate the specifications, the model will be built with

GeMS as its base and include GeMS tables and relationships such as common vocabularies, Glossary, DataSources, and DescriptionOfMapUnits. An additional optional specification is to accommodate 3D data, as well as multi-scale and multi-temporal data, although this may be difficult to accomplish.

Slide 10. Other technical, workflow, and tool specifications were accepted, except for those referring to cartography and map editing. Community participants concluded that the implementation of cartographic tools and editing workflows, although important, were outside the scope of the enterprise database. By using the national GeMS standard to create a common enterprise database standard, we hope that any tools and workflows developed by community members will be applicable to and shared with other organizations.

Slide 11. During the past 10 months or so, DGGS has created the IT environment necessary to complete the project—ArcGIS for Server/ArcMap with PostgreSQL as the backend database. We are currently testing Arc data management functions such as versioning on several DGGS datasets, importing the complex USGS Alaska map into the GeMS schema and evaluating the ability of the GeMS model to contain the data, and using DbSchema to map to diagram the database schema and view the schema through pgAdmin (PostgreSQL tools) and ArcGIS.

Slide 12. DGGS has identified several unresolved philosophical or technical issues through work to date on the schema and technical environment.

1. Although making minimal modifications to the single-map version of GeMS as described in the documentation will help ensure that map databases are more interoperable, a key design strategy of GeMS is flexibility of structure to accommodate differing map data. Further, an enterprise version will require additional fields and tables to be most effective, and each organization adopting an enterprise version of GeMS will have at least a slightly different implementation. If an ultimate goal of designing an enterprise database for the community is to share data amongst ourselves and with sister agencies, is there a point at which too much schema flexibility will hinder our ability to share data?
2. In a single-map or multi-map GeMS database, multiple spatial features (points, lines, polygons) may reference multiple data sources (many-to-many relationship), although the current schema allows only one DataSources table reference per feature (1-to-many relationship) for the sake of simplicity. Should the enterprise data model attempt to capture complex feature data sources, such as with many-to-many relationships?

   Further, the GeMS schema includes a DataSourcePolys feature class related to the DataSources table that stores the map footprints of data sources. An option would be to make the DataSources table into a feature class to store spatial data; however, not all data sources in the DataSources table may have polygonal footprints, such as point-based analytical data or a dictionary.
3. When working with large numbers (spatial locations) in PostgreSQL stored as Esri floating point numbers, DGGS found inconsistencies in area calculations and rounded numbers. To remove the inconsistencies, the spatial locations were recreated in additional fields as strings and calculations were performed on the strings instead.

Slide 13. Another technical issue was identified regarding spatial calculations around the International Date Line (IDL), 180 east/180 west longitude, for which we do not yet have a workaround. Alaska's Aleutian Islands are bisected by the IDL, which often makes display of the entire state problematic in GIS. Spatial calculations are performed differently in various GIS software. Esri software uses flat-surface-model geometric calculations to display data. Within Esri products, locations from the PostgreSQL database display correctly around the IDL. However, when viewed through PostGIS, the same geometric-based spatial data in the PostgreSQL database displays incorrectly. PostGIS, but not Esri, supports spherical-model geographic calculations, which do display the data correctly via PostGIS.

Why would you want to use spatial data directly from PostgreSQL/PostGIS instead of through an Esri product? Non-spatial data are transferred faster via PostgreSQL than through Esri software. Users will see an increase in application speed when accessing large datasets with simple spatial data, for example, a point-based geochemical dataset. Alternatively, a dataset with complex polygons and limited attributes would more quickly be transferred and displayed using Esri software.

Slide 14. A geologic map stored in the single-map GeMS format is typically a cartographically-correct publication with well-defined feature classes. The polygons represented in the MapUnitPolys feature class are the polygons displayed on the map, and ancillary data such as overlay polygons that coincide with the map unit polygons are stored in a separate feature class. In a multi-map enterprise-version of GeMS, the database becomes a storage container for a multitude of map data rather than a snapshot of one particular published map. As such, definitions become blurry among what are the primary map unit polygons, coincident overlay polygons, coincident subclasses of polygons, and derivative products based on these features, since any one of these layers could be symbolized and published as a cartographically-correct map. Consequently, we will need to decide how topologically-related features are best stored in the schema, and apply some measure of consistency to each geologic dataset that is ingested into the enterprise database.

Slide 15. Providing an empty GeMS enterprise database and applicable software to DGGS employees is not enough to ensure increases in agency efficiency and the best use of resources. Instead, employees need tools and knowledge to take advantage of the enterprise database and personal investment in the system and goals of the program to ensure that they stay engaged. Even several years before the enterprise geologic database is truly available, DGGS is making an effort to create a knowledge base and new tools and workflows that will benefit the agency. New tools/workflows being investigated include an agency-wide system for collecting digital field data without internet connectivity, semi-automatic templates for map surrounds, and dropdown lists via domains, subtypes, and feature templates for more efficient data creation. Geologists are also voluntarily attending weekly staff-led training sessions to learn about and discuss the GeMS data model, ArcGIS tricks and tips, and DGGS data management practices.

Slide 16. As the geologic database project moves forward, next steps are to continue work on the database model and begin discussions on the specifications of the data-sharing protocol. An update on the project will be provided at DMT in 2018.

Slide 17. If you are interested in participating in the project, please contact me at jennifer.athey@alaska.gov or 907.451.5028. The next tele-meeting is tentatively scheduled for Monday, July 10, 2017 at 10AM Eastern Time, but it will likely be rescheduled due to a conflict with the Esri User Conference. We also have a public wiki that chronicles the project at http://137.229.113.30:8080/jamwiki/.

REFERENCE

National Cooperative Geologic Mapping Program, USGS (NCGMP), 2010, NCGMP09—Draft Standard Format for Digital Publication of Geologic Maps, Version 1.1, *in* Soller, D.R., Digital Mapping Techniques '09—Workshop Proceedings: USGS Open-File Report 2010-1335, accessed at http://pubs.usgs.gov/of/2010/1335/pdf/usgs_of2010-1335_NCGMP09.pdf.

Building an Enterprise version of GeMS (formerly NCGMP09 map schema)

JENNIFER E. ATHEY, MICHAEL D. HENDRICKS, AND PATRICIA E. GALLAGHER

ALASKA DNR/DIVISION OF GEOLOGICAL & GEOPHYSICAL SURVEYS (DGGS)

DIGITAL MAPPING TECHNIQUES 2017, MINNEAPOLIS, MINNESOTA, MAY 21-24
PRESENTED BY MICHAEL HENDRICKS, MAY 22, 2017

Overview of EPA Exchange Network project

Goal 1

Develop radon database for Alaska and data-sharing schema

Goal 2

Develop "enterprise" version of GeMS and data-sharing protocol

Goal 3

Create predictive geology-radon web map with radon "heat" map overlay

+

=

3-year project, Oct 2016 – Sep 2019

Enabling Geospatial Data Exchange

EPA and its partners **use geospatial data** in tandem with programmatic data, through geospatial information systems and browsers, **to conduct analyses** in a geographic or place-based context.

Office of Management and Budget (OMB) Circular A-16

The **geologic spatial data theme** includes all geologic mapping information and related geoscience spatial data that can contribute to the **National Geologic Map Database** as pursuant to Public Law 106-148.

Community participants:

Charlie Cannon	Mike Hendricks
Dave Soller	Ralph Haugerud
Evan Thoms	Ric Wilson
Greg Barker	Sean Eagles
Jen Athey	Suzanne Luhr
Jeremy Crowley	Tracey Felger
Lina Ma	Trevor Ellis
Mark Yacucci	Trish Gallagher

Implicit specifications

IT infrastructure and software

- Esri ArcGIS for Server/SDE version 10.4
- PostgreSQL database version 9.5
- Unix server environment
- Use the multi-user functionality in SDE

Database model

- Hold multiple and overlapping geologic maps
- Common data structure across multiple maps
- Test database model with two geologic maps
- Use of General Lithology to describe geologic units to laypeople

Documentation

- Schema, scripts, and other reusable components through EPA's Reusable Component Services

Others

- *Project wiki http://137.229.113.30/jamwiki/*
- *Future… NGMDB website and GeMS documentation?*

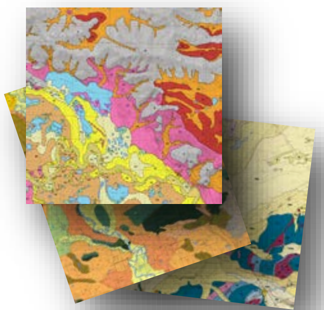
Definition of enterprise database

A spatial database with versioning, defined user roles, and stored procedures built on a relational database structure.

For the purposes of this project, which will use Esri products, Esri defines an enterprise geodatabase as being separated into two tiers:

- The application sphere is where you have all of your ArcObject and ArcSDE software to manage stored procedures, versioning, distributed data, and attribute and spatial validation.
- The data storage tier would be an RDBMS server, holding a database which allows storage, security and backup and recovery. This repository is a set of tables and stored procedures from the RDBMS which supports the geodatabase.

More definitions...

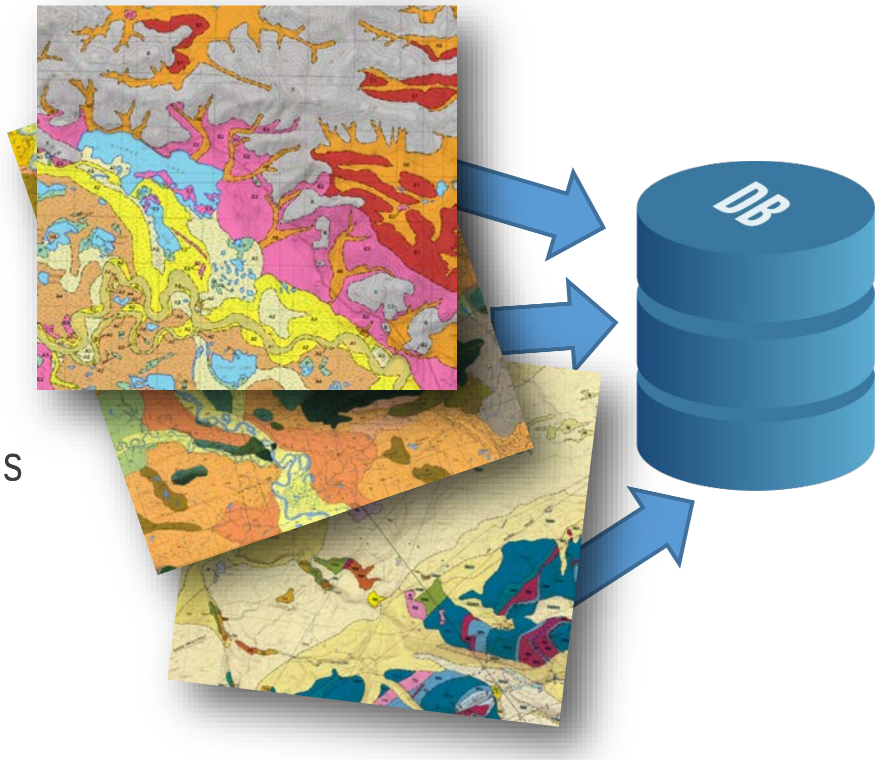


- **Multi-map database**: in the enterprise database, multi-map will refer to maps of different subjects, different geographical areas, different scales, different times and different lineages.

- **Multi-user database**: for the enterprise database the users can be separated into viewers, editors, creators, and administrators. These roles would have attending limitations of their ability to insert, modify or delete records on a table-by-table basis, or change the database structure itself.

Why an enterprise database?

- **A controlled container** for agency-wide spatial data
- **A vehicle to standardize** geologic data, increasing accessibility and enabling digital products
- **A way to increase efficiency** through standard procedures for data collection, map production, analysis, compilation, and archiving



Explicit specifications

Oct-Nov 2016

Voted

1. very important
2. somewhat important
3. not important or not important at this time

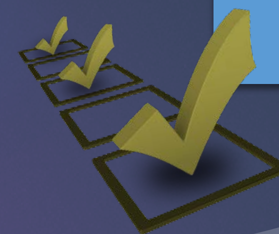
Reviewed

- Accepted (Y)
- Rejected (N)
- Optional (O): accepted specification that organizations may decide not to utilize

General

- Y 1.31: **Ease of use** for staff
- Y 1.62: Create **compilation maps more efficiently**
- Y 1.69: Provide **standardization** across geologic data sets in multiple organizations
- Y 1.69: Allow for **tools and scripts** to be built to increase efficiency

Explicit specifications: Database model



- Y 1.08 Topologic consistency
- Y 1.15 Data are queryable across multiple maps
- Y 1.23 Flexible model
- Y 1.31 Manage multi-scale, multi-temporal data sets
- Y 1.38 Have the database structure and/or scripts enforce QA/QC
- Y 1.38 Scalable
- Y 1.42 Allow single and multi-map unit descriptions
- Y 1.46 Ability to integrate with data in other databases
- Y 1.46 Common vocabularies stored as tables in the database
- Y 1.46 Manage bibliographic information and metadata
- Y 1.62 Reuse GeMS 1:many tables for multiple maps
- Y 1.69 Common unit descriptions
- O 1.69 Manage multi-scale, multi-temporal, and multi-dimensional (3D mapping) data sets
- O 1.77 Manage original data
- O 1.85 Manage analytical data
- Y 1.92 Manage unpublished data
- O 2.08 Manage field data
- N 2.23 Non-proprietary format for data archiving
- N 2.23 Online and offline connection to field devices for data collection
- O 2.38 Manage ephemeral interim products and processes

Other explicit specifications

Technical considerations

- Y 1.38 Reasonable speed of access to data (draw time)
- Y 1.62 Low administrative and technical overhead
- Y 1.62 Facilitate data services (WMS, WFS)
- Y 1.69 Enable metadata to be harvested by other data portals

Documentation and workflows

- Y 1.31 Schema, scripts, and other reusable components will made available
- N 1.54 Protocols for map editing and cartography

Tools and scripts

- Y 1.31 Tool to check data sets/structure for errors
- O 1.54 Tool to create FGDC or ISO metadata
- N 1.85 Tool to speed up cartography

Technical efforts from July 2016 – May 2017

- Set up ArcGIS for Server-PostgreSQL environment
- Created GeMS schema in PostgreSQL with editor tracking and global id field
- Using DbSchema to map out GeMS schema and testing connection to Arc
- Testing versioning, multi-user editing, and other features with different PostgreSQL map databases
- Started importing data from USGS AK map

Thoughts on schema and setup

- OK to add new attribute fields to GeMS tables
- DataSources table
  - What is the best way to tie a compilation map back to its sources?
  - Should it be a many-to-many relationship?
  - Should the table be a feature class instead and store map footprint?
- Arc spatial locations are floating numbers. In PostgreSQL, the numbers are always changing a little bit, especially for very large numbers. This may cause values like area calculations and rounded numbers to be inconsistent.

Spatial calculations around dateline

Accessed through PostGIS

geometry (Cartesian measurements, flat surface model) supported, displays incorrectly

geography (geodetic measurements, spherical model) supported, displays correctly

Accessed through ArcGIS

geometry (Cartesian measurements, flat surface model) supported, displays correctly

geography (geodetic measurements, spherical model) not supported

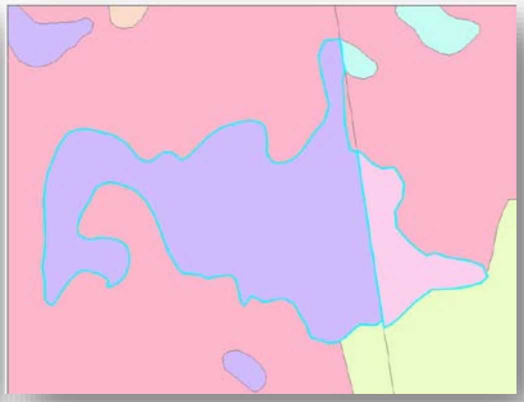
Topological consistency among map features

Overlay polygons

	CLASS	QCLASS *	NSACLASS *
	719	4551	455
	110	4550	455

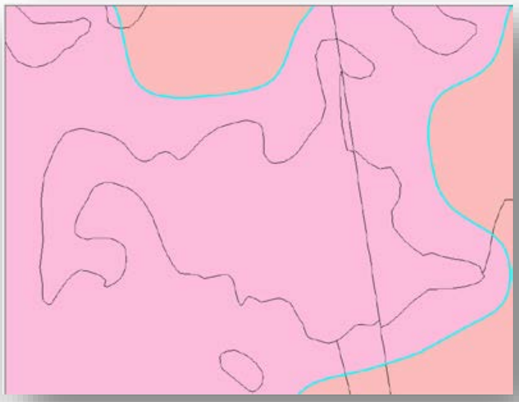
Primary map layer

	CLASS	QCLASS *	NSACLASS *
	719	4551	455
	110	4550	455



Derivative product

	GROUP_ID	GROUP_LABEL
	1	QTs



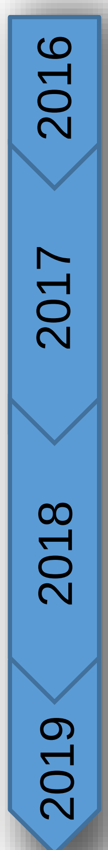
DGGS mapping environment

Making things easier for geologists will help DGGS adopt an enterprise database

- Investigating field data collection scenarios for no internet connectivity
- Creating map surround templates
- Investigating domains, subtypes, and feature templates
- Tweaking GeMS schema to make sure it works for geologists
- Regular Arc training and data management classes

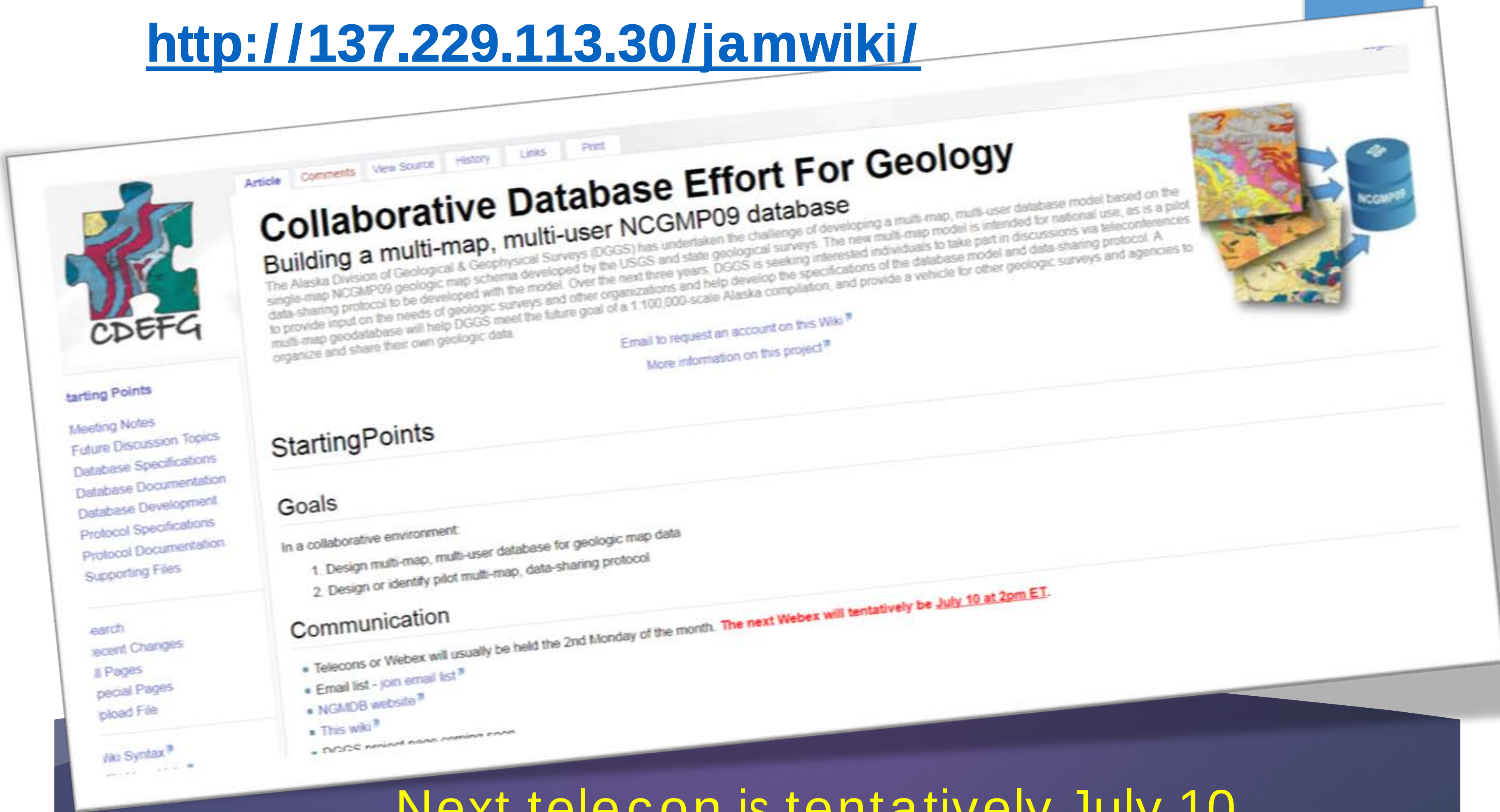
A way forward

Many heads are better than one



- 2016 DMT: Initial input and discussion
- 2016-17 NCGMP09 workgroup meetings
  - **Database model** development and testing with 2 data sets
- **2017 DMT: Update on database model**
- 2017-18 NCGMP09 workgroup meetings
  - **Data-sharing protocol** development and testing
- 2018 DMT: Update on pilot data-sharing protocol
- 2018-19 Documentation
- 2019 DMT: Update on project
- 2019 Code and models will be posted to EPA's repositories

http://137.229.113.30/jamwiki/



Next telecon is tentatively July 10

For questions, contact jennifer.athey@alaska.gov, 907.451.5028