Alaska-GeMS Multi-map Database Schema Changes from the Federal GeMS Standard

JENNIFER ATHEY, MIKE HENDRICKS,
PATRICIA EKBERG, AND CHRIS WYATT
ALASKA DNR/DIVISION OF GEOLOGICAL & GEOPHYSICAL SURVEYS (DGGS)

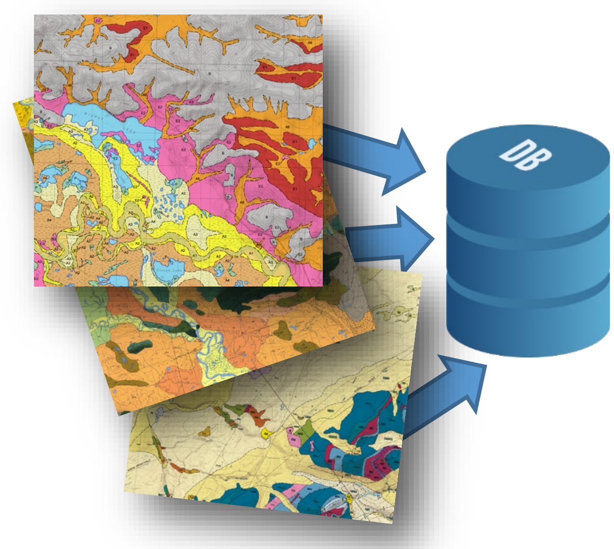
DIGITAL MAPPING TECHNIQUES 2020, VIRTUAL, JUNE 8-10
PRESENTED BY JENNIFER ATHEY, JUNE 10, 2020

Project goals

EPA-funded research project, Oct 2016-Sept 2019

Various USGS grants and agreements, Oct 2019-August 2021

- Develop a schema for a multi-map database
- Easy to use by geologists
- Design with care
- Be true to the implicit relationships of the data
- Emphasizes geologic data over cartography

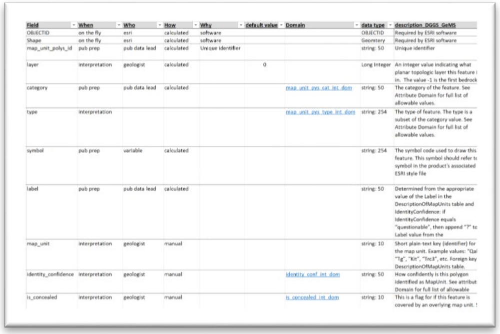


Collaboration

Goals and products benefit from the **collective wisdom** and points of view of a variety of people.

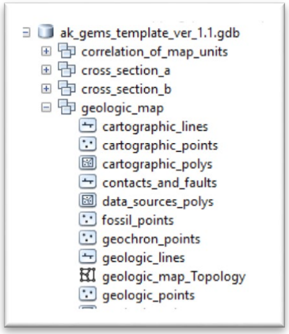
We are smarter together.

- Ideas could benefit other agencies
- Can be used for single-map geodatabase
- Files on https://dggs.alaska.gov/gemswiki/



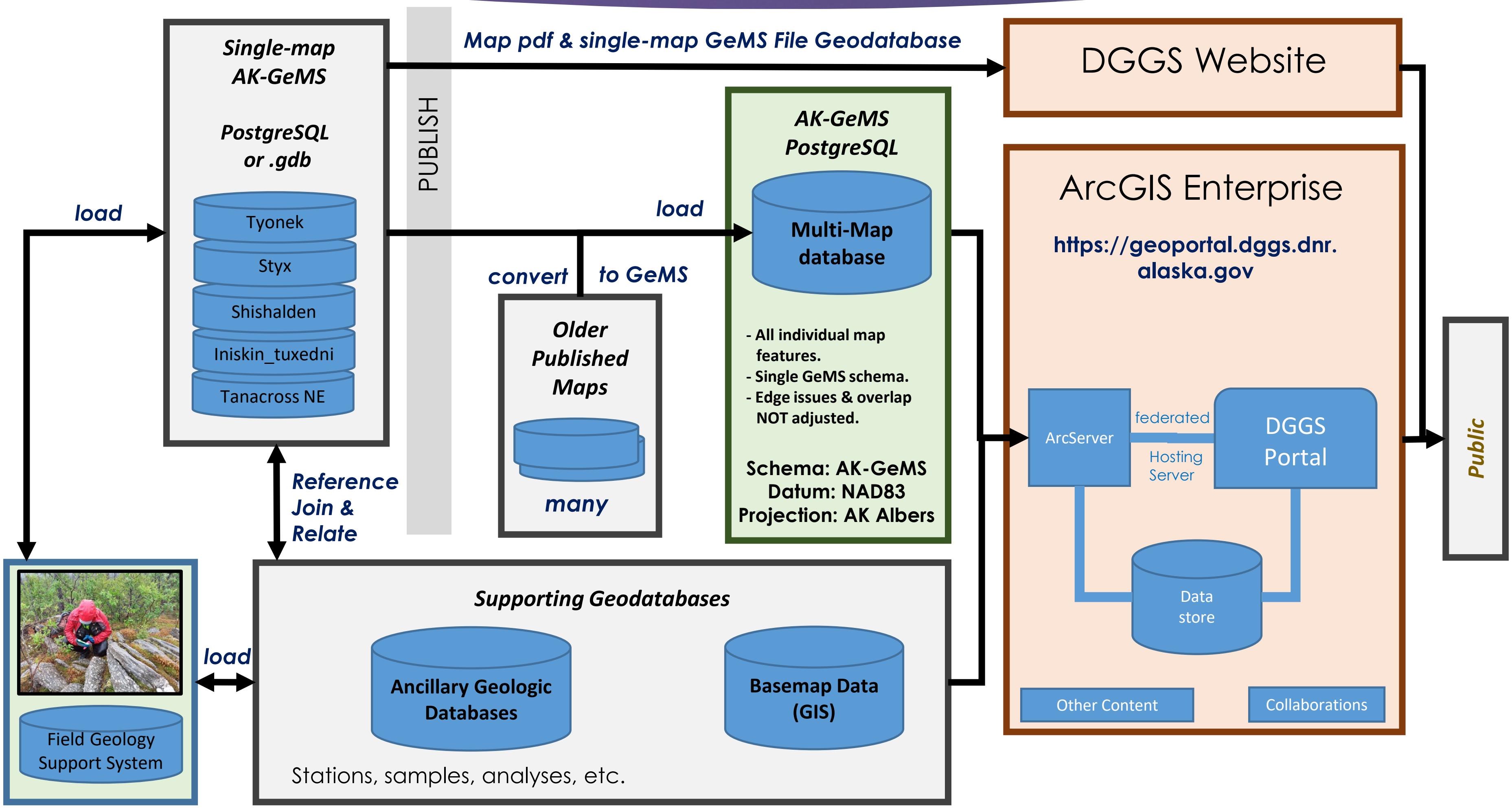
Field	When	Who	How	Why	default value	Domain	data type	description_DGGS_GeMS
OBJECTID	on the fly	esri	calculated	software			OBJECTID	Required by ESRI software
Shape	on the fly	esri	calculated	software			Geomtery	Required by ESRI software
map_unit_polys_id	pub prep	pub data lead	calculated	Unique identifier			string: 50	Unique identifier
layer	interpretation	geologist	calculated		0		Long Integer	An integer value indicating what planar topologic layer this feature i in. The value -1 is the first bedrock
category	pub prep	pub data lead	calculated			map_unit_pys_cat_int_dom	string: 50	The category of the feature. See Attribute Domain for full list of allowable values.
type	interpretation					map_unit_pys_type_int_dom	string: 254	The type of feature. The type is a subset of the category value. See Attribute Domain for full list of allowable values.
symbol	pub prep	variable	calculated				string: 254	The symbol code used to draw this feature. This symbol should refer to symbol in the product's associated ESRI style file
label	pub prep	pub data lead	calculated				string: 50	Determined from the appropriate value of the Label in the DescriptionOfMapUnits table and IdentityConfidence: if IdentityConfidence equals "questionable", then append "?" to Label value from the
map_unit	interpretation	geologist	manual				string: 10	Short plain-text key (identifier) for the map unit. Example values: "Qal "Tg", "Kit", "Trc3", etc. Foreign key DescriptionOfMapUnits table.
identity_confidence	interpretation	geologist	manual			identity_conf_int_dom	string: 50	How confidently is this polygon identified as MapUnit. See attribut Domain for full list of allowable
is_concealed	interpretation	geologist	manual			is_concealed_int_dom	string: 10	This is a flag for if this feature is covered by an overlying map unit.

Data Dictionary



Schema v1.1

AK DGGS Geologic Map Production & Management System



AK-GeMS Universal Changes

- Change to “snake_case” from “PascalCase”
- Fields are nullable while editing
- Additional choices for domains

cf_cat_int_dom	**Contacts and Faults Category Integer Coded Domain**	
code	**description**	**glossary_definition**
100	contact	A linear feature indicating where two map units meet
200	fault	A linear feature indicating where two map units have moved in relation to each other
300	boundary	A linear feature indicating where one map unit ends without an adjacent map unit
997	unprovided	no specific knowledge available to provide a valid entry. Normally used only in data conversion projects
998	unknown	not known to the mapping geologist
999	other	value not provided in the field’s attribute domain list. See notes field for details

Schema Comparison, #1

Federal GeMS	Single-Map AK-GeMS	Examples
Items with no real-world existance; only for cartographic display; non-geologic ancillary information		
CartographicLines	cartographic_lines	cross-section traces, pipeline route
	cartographic_points	mine adit, drill hole
	cartographic_polys	polygon for overlay pattern
various annotation	annotation	
Geologic features associated with a map unit; polygons share topology with contacts_and_faults		
	map_unit_lines	long map-unit polygons too thin to show at scale; dike, cliff outcrops
MapUnitPoints	map_unit_points	map-unit polygons too small to show at scale
MapUnitPolys	map_unit_polys	map-unit polygons big enough to show at scale
MapUnitOverlayPolys		overlay polys related to a map unit
Geologic features associated with rock deformation on both a large and a small scale		
ContactsAndFaults	contacts_and_faults	
OrientationPoints	orientation_points	bedding attitudes, foliation attitudes
	structure_lines	fold hinge-surface traces, boundary of basin
	structure_polys	fault-breccia zone, shear zone

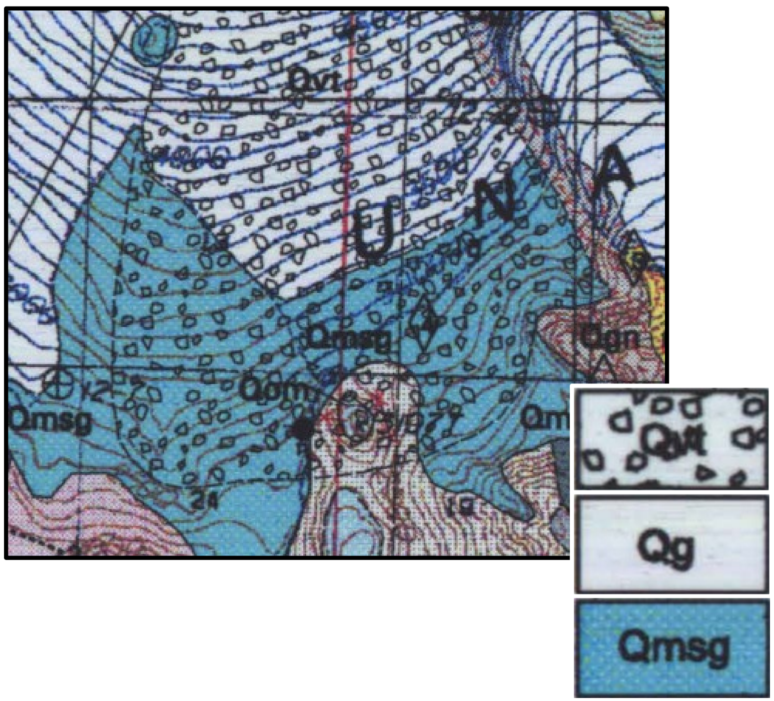
CDEFG

Schema Comparison, #2

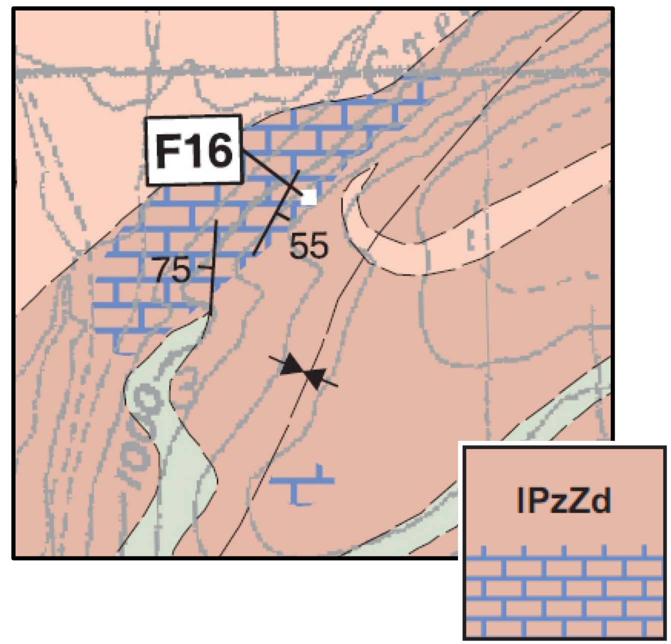
Federal GeMS	Single-Map AK-GeMS	Examples
Geologic features not covered by normalized feature classses (not fossil locations, not related to structure, not a map unit, etc.)		
GeologicLines	geologic_lines	scarp, key beds that are not map units (fold hinges to structure lines)
GenericPoints	geologic_points	pingo, glacial erratic
	geologic_polys	outcrop area of key bed, hummocky topography
Field obervation, sampling, and analytical points		
Stations	stations	
GeochronPoints	geochron_points	
FossilPoints	fossil_points	
SamplePoints/GenericSamples		denormalized analysis data
Miscellaneous feature classes		
OverlayPolys	overlay_polys	hornfels, alteration zones (not map unit but modifying a map unit)
IsoValueLines	iso_value_lines	geobarometry contours, isopach contours of coal seam thickness
DataSourcePolys	data_sources_polys	

Overlay Polygons and Layers

Overlay polygon related to a Map Unit



Overlay polygon NOT related to a Map Unit



Schema Comparison, #3:
Non-spatial tables

Federal GeMS	Single-Map AK-GeMS
non-spatial tables	
DataSources	data_sources
DescriptionOfMapUnits	description_of_map_units
GeoMaterialDict	geo_material_dict
Glossary	glossary
	location_confidence_lookup
	orientation_confidence_lookup
MiscellaneousMapInformation	product_info
RepurposedSymbols	repurposed_symbols

Contacts_and_faults symbols and confidence

REF NO	DESCRIPTION	SYMBOL
		1.1—Contacts
1.1.1	Contact—Identity and existence certain, location accurate	————————
1.1.3	Contact—Identity and existence certain, location approximate	— — — — — — —
1.1.5	Contact—Identity and existence certain, location inferred	- - - - - - - - - -

Location Confidence Method Integer Coded Domain		
code	**description**	**glossary_definition**
1	generalized	The feature's location confidence is primarily identified with the ordinal classification scheme used in the location_confidence field.
2	specified	The feature's location confidence is primarily identified with a numeric value provided by the geologist in the location_confidence_meters field.
3	measured	The feature's location confidence is primarily identified with a numeric value provided by GPS in the location_confidence_meters field.
997	unprovided	no specific knowledge available to provide a valid entry. Normally used only in data conversion projects
998	unknown	not known to the mapping geologist

location_confidence_lookup_TABLE

scale_denom	accurate_min	accurate_default	accurate_max	approximate_min	approximate_default	approximate_max	inferred_min	inferred_default	inferred_max
100000	0	40	40	40	100	100	100	200	Null
63360	0	25	25	25	65	65	65	130	Null
25000	0	10	10	10	25	25	25	50	Null

Capturing Map Metadata

MiscellaneousMapInfo

Table

MiscellaneousMapInformation

	MapProperty	MapPropertyValue	Miscella
	Title	Lidar-revised geologic map of the D	MMI01
	Authors	Rowland W. Tabor and Derek B. Bo	MMI02
	Year of publication	2017	MMI03
	Publisher	U.S. Geological Survey	MMI05
	Series	Scientific Investigations Map	MMI06
	Number	3384	MMI07
	Geology mapped by	D.B. Booth, 1986-1998, 2004; H.H.	MMI09
	Revision by	Rowland W. Tabor (2008-2015)	MMI10
	Base-map culture int	Rowland W. Tabor (2008-2013)	MMI11

0 (0 out of 17 Selected)

MiscellaneousMapInformation

product_info

	A	G
1	**Field**	**description_DGGS_GeMS**
2	product_id	Unique ID derived from AK DGGS GERILA database
3	product_guid	
4	name	
5	project_id	
6	project_guid	
7	type	Type of product
8	pub_date	
9	citation_id	Unique ID derived from AK DGGS GERILA database used to identify
10	citation_link	
11	product_map_link	
12	product_db_link	
13	project_link	
14	notes	
15	status	
16	distribution_policy	
17	map_scale_denominator	
18	product_theme	

Extending GeMS Feature Classes

- 'category' and 'type'
- symbol_alt
- 'label' is calculated
- layer
- modifier
- group_id
- draw_policy
- distribution_policy
- data_sources_method

Structure Lines Type Integer Coded Domain

code	description	glossary_definition	USGS Carto Std
101	fold, anticline	As defined in AGI Glossary	5.1.1-5.1.16
102	fold, anticline, asymmetric	As defined in AGI Glossary	5.3.1-5.3.16
103	fold, anticline, inverted	As defined in AGI Glossary	5.3.33-5.3.48
104	fold, anticline, overturned	As defined in AGI Glossary	5.3.17-5.3.32
105	Etc.		

Data Sources Method Integer Coded Domain

code	description	glossary_definition
1	Feature unmodified from source material	Feature unmodified from source material
2	Feature modified from source(s), primarily field mapping	Feature modified from source(s), primarily field mapping
3	Feature modified from source(s), primarily basemap data	Feature modified from source(s), primarily basemap data
4	Etc.	

Specific changes

- MapUnit => map_unit_associated and map_unit_observed
- Data_sources_id: Athey, 1999
- Glossary: "As defined in AGI Glossary of Geology"
- Linking orientation points: associated_feature_id
- Contacts_and_faults: age

age_label
age_type
age_oldest
age_youngest

Single-map to multi-map conversion

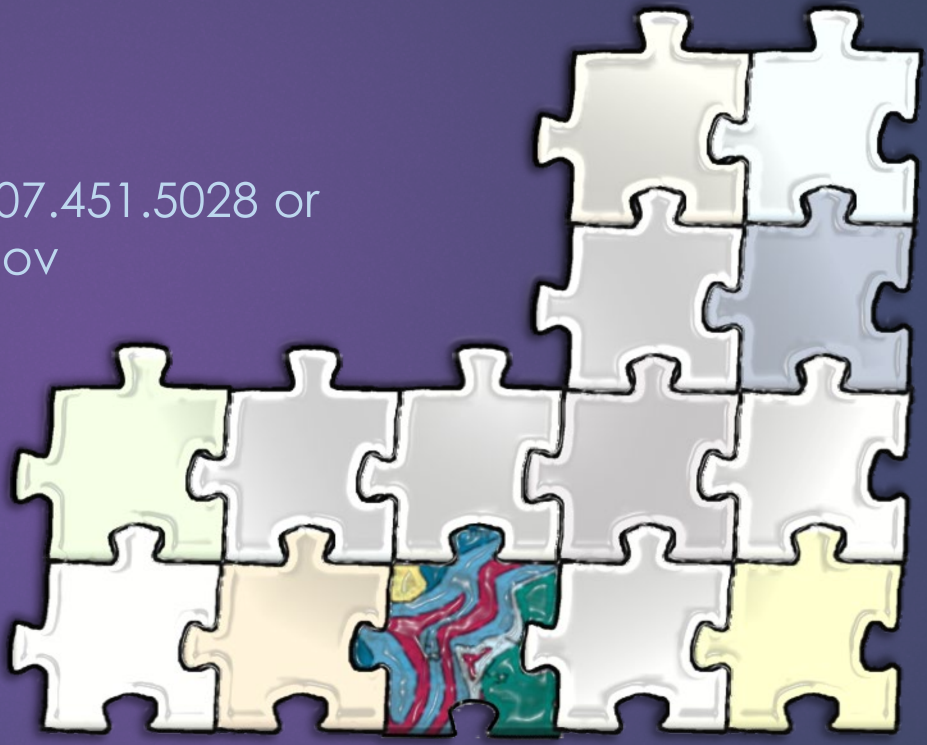
- Population of GUIDs
- Convert enumerated domains to coded domains
- Burn overlay polygons into map_unit_polys ‘modifier’ field
- Remove annotation and maybe cartographic feature classes
- Create many-to-many data sources cross-ref table
- Append map and data elements into single schema (e.g., one contacts_and_faults table, one DMU table)

CDEFG

Join the CDEFG discussion

- Monthly telecons
- Project wiki
- Questions?

Contact Jen Athey at 907.451.5028 or jennifer.athey@alaska.gov



https://dggs.alaska.gov/gemswiki/

Alaska-GeMS Multi-map Database Schema Changes from the Federal GeMS Standard

By Jennifer Athey, Mike Hendricks, Patricia Ekberg, and Chris Wyatt
Alaska Division of Geological & Geophysical Surveys
3354 College Road
Fairbanks, Alaska 99709

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

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

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

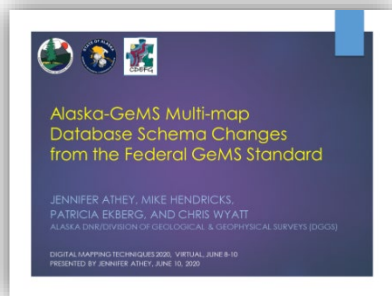
W. Christopher (Chris) Wyatt
Telephone: (907) 451-5034
email: mailto:chris.wyatt@alaska.gov

Abstract

The Alaska Division of Geological & Geophysical Surveys (DGGS) is in the final stages of developing a multi-map, multi-user "enterprise" database model based on the single-map *Ge*ologic *M*ap *S*chema (GeMS) developed by the USGS and state geological surveys (https://pubs.er.usgs.gov/publication/tm11B10). In additional to in-house use, Alaska's enterprise database model is intended for adoption by state and federal geological survey programs needing multi-map, multi-user functionality in their agencies' geologic map databases. To support this extended functionality, the model necessarily departs from the federal version of GeMS in a number of ways. This presentation is an overview of the "hows" and "whys" of changes to the federal data model. Additional information and details are available in documentation at the project wiki (https://dggs.alaska.gov/gemswiki/), which are evolving into formal, published documentation as the project progresses. Interested persons are encouraged to contact the DGGS authors for more information about this collaborative project.

This information was presented at the Digital Mapping Techniques 2020 meeting, which was held virtually from June 8-10, 2020. The presentation was delivered by Jennifer Athey, June 10, 2020. This explanatory text is meant to be viewed alongside the presentation slides.

Slide 1.



Work on this collaborative multi-map database model has been informed by many people inside and outside of Alaska's geological survey. In Alaska, Jennifer Athey is the Principal Investigator, and Mike Hendricks is the Technical Lead. Hendricks runs weekly committee meetings where GIS team members Patricia Ekberg, Chris Wyatt, and geologists working on GeMS-based geologic maps discuss and problem-solve issues with the schema and develop workflows to increase efficiency. This committee is currently working on finalizing domain lists, compiling and documenting standard FGDC and custom symbology into a comprehensive style file, and documentation for the multi-map database model. As of June 2020, we have significant work to show. Documentation published after June 2020 are also referenced in this explanatory text for completeness.

Slide 2.

Alaska's geological survey has been interested to build a geologic map-GIS database since the late 1990's. Recent database development began with the promise of a national geologic map database standard (GeMS; formerly NCGMP09) and EPA-funded research into extending GeMS to contain multiple maps from 2016-2019. Work continues through a cooperative agreement funded by the USGS National Geologic Map Database (https://ngmdb.usgs.gov) and is informed by geologic map GeMS-format conversions funded by the USGS National Geological and Geophysical Data Preservation Program and new mapping. All of the geologic mapping at DGGS feeds into the iterative development, evaluation, and documentation of the multi-map schema and process of creating and publishing geologic maps. For DGGS, the ultimate goal is an efficient geologic mapping and GeMS-map-conversion process that leads to a standardized, statewide geologic map database from which to build a 1:100,000-scale compilation of Alaska and other derivative maps.

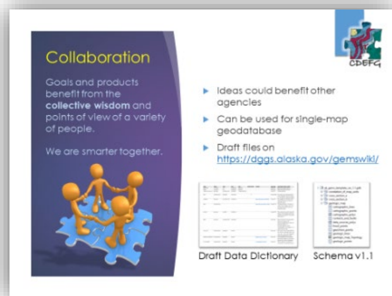
DGGS adheres to several basic tenets that we believe will benefit the long-term viability of the database schema and help realize the potential of the project.

- The mapping and publishing process, including the schema, must be as simple as possible while still fulling the needs of the geologists and organization. Further, the schema needs to provide efficiencies where possible, such as logical and vetted domain values.
- The multi-map schema must be designed with care. There are many details that matter, and missteps will rapidly become apparent as the schema is populated and the database grows. Fixing problems down the road will likely be harder than getting them correct at the start.
- Be true to the implicit relationships of the data. For example, there are many-to-many relationships in geologic map data (e.g., many maps to many data sources) that are currently challenging to implement successfully with Esri software. However, it is better to capture the complete data now, even if it is challenging, rather than let the current version of the software

determine what data are captured. The software will likely handle complicated relationships better in the future as it continues to mature.

- The multi-map schema emphasizes geologic data over cartography. DGGS anticipates that most symbolization of the multi-map data will be performed on the fly, likely in online applications. Given this, it's critical that the information behind traditional cartographic decisions is recorded in the database for both impromptu symbolization and robust querying. The AK_GeMS_symbology style file provides consistency for the data management process (Ekberg and others, 2021).

Slide 3.



It is important to note that from the beginning of this project in 2016, DGGS has collaborated with other state geologic surveys, the USGS, and Canadian geologic agencies to define terminology and develop specifications to help steer and drive the model's development. From the original group of about nine agencies, DGGS-led communication around GeMS implementation (through open Collaborative Database Effort for Geology [CDEFG] monthly meetings and conference presentations) has grown to about 36 entities! Then and now, we hope that by encouraging the participation of other state and federal agencies, the multi-map schema and mapping system will be more suitable for adoption into their GIS and Information Technology systems. Therefore, we suggest partner agencies should think about Alaska's schema (AK GeMS) as an "Extended GeMS" (developed for all agencies), and not as "AK" GeMS (solely implementable in Alaska's northern climes, jigsaw-puzzle geology). Further, as geologic mapping is quite challenging in Alaska for many reasons (e.g., project audiences from volcanologists to coastal managers to resource explorationists; various, unstudied, and even unique geologic materials; 180th meridian GIS software conundrums), a data model that works for Alaska has a good chance of being successful in other geologic terrains.

In 2016, project participants defined several important terms that continue to resonate.

➢ 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 uses Esri products, Esri defines an enterprise geodatabase as being separated into two tiers: an application tier and data storage tier. The application tier is where you have all of your Esri software to manage stored procedures, versioning, distributed data, and attribute and spatial validation. The data storage tier is a Relational Database Management System (RDBMS) server, in this case PostgreSQL, holding a database which allows data storage, security, and backup and recovery.

➢ Multi-user: for the enterprise database, users can be separated into viewers, editors, creators, and administrators. These roles have limitations in their ability to insert, modify, or delete records on a table-by-table basis, or change the database structure itself. The end-user may gain access to the data through a portal available on the web, interact with the data in a web mapping application, or simply download the data from the portal. For example, roles may be developed by task, such as database design (administrator, creator), compilation (creator, editor), quality assurance/quality control (editor, viewer), or scientific review (viewer), and users would have access to certain management tools, tables, and SQL queries to facilitate those tasks. Policies govern how data from multiple users and edits are managed in workflows and the database.

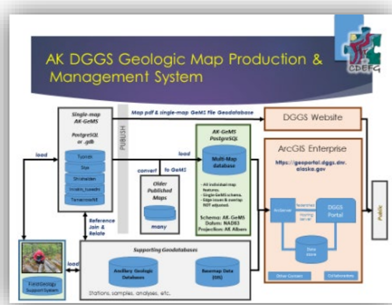
➢ Multi-map: in the enterprise database, multi-map refers to maps of different subjects, different geographical areas, different scales, different times, and different lineages that are stored in a single schema. Gapless coverage is the goal for this phase of the data management process. No efforts are undertaken toward removing boundary faults or merging unit descriptions beyond those of the original authors.

The term "seamless" has also been introduced to describe geologic databases. Thorleifson (2020) defined seamless geology as a "synthesis of multiple published maps, in which steps have been taken toward harmonization, and facilitation of query." Therefore, a multi-map enterprise database as described here is one of the first efforts needed to create seamless geology, as it facilitates compilation of geologic unit descriptions, resolution of geology mismatches across boundaries, and robust querying through map standardization such as constrained domain values. In fact, because the original map data are preserved in a multi-map database, multiple compilation and derivative maps can be conflated from the multi-map database.

Currently, the U.S. geologic community's need for geologic data standards, including a common vocabulary and framework with which to discuss issues, is escalating due to (1) recent interest in national and state geologic map databases fueled by increased USGS National Geologic Map Database ─ Phase/Task Three funding for 2D and 3D national coverage (now known as the U.S. GeoFramework Initiative) and (2) USGS grant deliverables required in GeMS format that will ultimately feed into the same national dataset. The utility of both the GeMS single-map schema and Alaska's multi-map database schema is highlighted. It should be noted that quite a few geologic organizations already maintain a statewide geologic map database and in-house schema (whether for multi-map or single-map applications) and intend only to translate their schemas to the federal GeMS format as needed.

DGGS published the schema and data dictionary for Alaska's GeMS extension (AK GeMS). The data dictionary and an XML Workspace Document that can be used to generate an empty geodatabase are available at https://dggs.alaska.gov/pubs/id/30669. These documents are much more comprehensive than the information highlighted in this presentation.

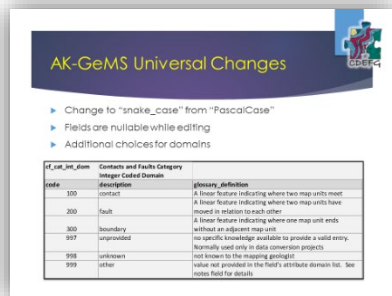
Slide 4.



DGGS has developed two different versions of the extended database schema. First, the single-map version, which can be used for a stand-alone map, contains the tables and fields necessary for the map to be merged into the multi-map database. Second, the multi-map version of the schema is very similar to the single-map version but is optimized for large amounts of data, e.g., constrained text domains are converted to integer-based domains to allow for faster data processing. Use of both of the schemas creates an efficient workflow, where multiple geologic teams can work on many disconnected file geodatabases or PostgreSQL geodatabases at the same time, including new mapping and legacy map conversions. Once any of the new geologic map databases are peer reviewed, quality controlled, approved, and published, they can be loaded into the multi-map database. Since published maps being converted to GeMS have already been peer reviewed and approved, these GeMS products are quality controlled, loaded into the multi-map database, and more quickly made available to the public. We are also working on scripts and other data management techniques to speed up repetitive processes such as quality control, GeMS conversion tasks, and database loading.

In this explanation, we mostly discuss the single-map AK GeMS version. Slide 4 shows how the single-map schema fits into DGGS' entire, but simplified, map production and management process. Alaska's extended schema is very similar to the USGS' single-map GeMS but departs in some important ways that will be discussed in the next slides. For clarity in this explanatory text, Alaska's extended schema is referred to as "AK GeMS" and the USGS' single-map GeMS is referred to as "federal GeMS".

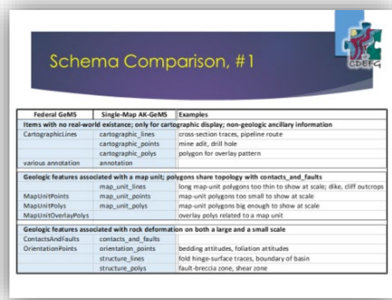
Slide 5.



DGGS has made a few universal and other changes to the federal GeMS standard in its implementation of the schema. Scripts and other automations will be used to convert databases constructed using AK GeMS to federal GeMS-compliant databases for delivery to the USGS. Map databases created using AK GeMS and converted to the federal schema pass Level 3 GeMS compliance checks.

- Table names and fields are changed to 'snake_case' (all lower case with underscores) instead of 'PascalCase' (first letter of words capitalized) to make the database compatible with the multi-map PostqreSQL database. Snake_case is the most common naming convention used in computing as identifiers for variable, function, and file names. It is also convention to use snake_case with PostgreSQL, as the database platform is case insensitive. Although Esri software could handle the translation of capital letters between the spatially-enabled PostgreSQL database and its geodatabase mirror inside the software, having manipulated table and field names would complicate any custom programming that accessed the PostgreSQL database directly instead of through Esri software. (In this explanatory text, italicized table or field names in PascalCase refer to federal GeMS; alternatively, italicized table or field names in snake_case refer to AK GeMS.)
- Most fields are nullable while geologists are editing. Once the data are finalized, fields are returned to the not-nullable state, such as in the multi-map database and federal GeMS deliverables.
- Many fields now have domains with value choices, along with integer-based domain values for use in the multi-map database. Instead of only an "other" category to be used when a not-nullable field has no value, we wanted to be able to provide more detail about the information not being added to the record. "Unprovided" is commonly used in data conversion projects to show that the editor did not have the information available to populate the record, though the map author may have originally had the information. "Unknown" may be used to show that the value is not known, e.g., the age of a poorly preserved fossil. DGGS plans to harvest data coded as "other" and the provided note field for subsequent updates to the schema's domains, such as the need to add a new geochronologic dating method.

Slide 6.



GeMS was designed to be flexible. Some feature classes are mandatory while other feature classes can be added as needed for different types of geologic data. Alaska's extension to GeMS contains the same mandatory feature classes, with prescribed additional feature classes to contain a broad range of geologic map information. The multi-map database requires a stable schema to be most effective for data management. Through comprehensive testing and documentation, we strive to remove the ambiguity of where different kinds of data should be stored in the schema.

Slides 6 and 7 compare the feature classes between federal GeMS and AK GeMS. Federal GeMS feature classes are displayed in the first column, and corresponding Alaska single-map GeMS feature classes are shown in the second column. The third column contains examples of different kinds of geologic data that would be contained in the feature classes. The federal GeMS example in this explanation is based on the draft version 1.1 GeMS documentation prior to Techniques and Methods 11-B10 and the Des Moines quadrangle, WA database at https://pubs.er.usgs.gov/publication/sim3384.

<u>Cartographic points, lines, and polygons</u>

Federal GeMS prescribes a cartographic line feature class whereas AK GeMS has feature classes for cartographic points, lines, and polygons. We added these two feature classes to contain features on the map that are extrinsic to the geologic data, e.g., features needed for locational purposes, features from other databases, incidental information, and symbolization needed for cartographic purposes where the geologic feature's critical metadata is stored elsewhere in the database. For example, an anticline fold axis symbolized with a line must have anticline arrows showing the dip of the fold limbs. Sometimes it is necessary for the cartographer to exactly place the anticline arrows along the fold hinge instead of using a multi-part line-and-arrow symbol. In this case, the placed anticline arrow points would be stored in the *cartographic_points* feature class. Alternatively, a symbology decoration is stored in the *orientation_points* feature class if it is based on an observation or measurement collected at the location. The *cartographic_polygons* feature class will be discussed further in slide 8.

<u>Map unit points, lines, and polygons</u>

Map_unit_polys feature class is the primary map unit feature class described in the federal schema documentation. A *map_unit_points* feature class is described in the GeMS documentation as well. *Map_unit_points* are observed geologic units at a scale too small to be mappable. AK GeMS includes a *map_unit_lines* feature class in the extended schema. M*ap_unit_lines* describes geologic units with mappable length but not mappable width at the given scale, e.g., dikes and vertical outcrop faces along rivers. As in the federal schema, objects in map-unit feature classes must reference a map unit in the Description of Map Units (DMU) table.

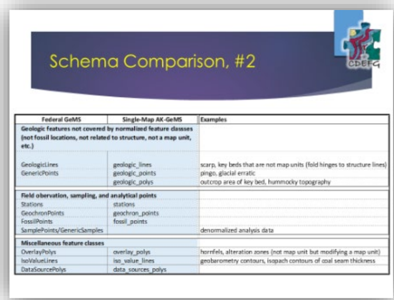
<u>Map unit overlay polygons</u>

Federal GeMS feature class *MapUnitOverlayPolys* is not used in AK GeMS. See slide 8 for more information.

Structure-related feature classes

DGGS normalized the database structure by adding two additional specific feature classes for structure data: *structure_lines* and *structure_polys*. The federal schema already contains an *orientation_points* feature class, which is useful for describing structural phenomena as points (i.e., bedding, joints, foliation). Having all of the structure data in expected tables will help facilitate queries, and structure symbols in the FGDC Digital Cartographic Standard (2006) are automatically assigned to these feature classes (Ekberg and others, 2021).

Slide 7.



Continuing the comparison of AK GeMS to federal GeMS:

Geologic points, lines, and polygons

The "geologic" feature classes are catch-alls for real-world geologic features that don't fall nicely into other feature classes. For this grouping, DGGS redefined the federal GeMS feature class *GenericPoints* as *geologic_points* and is more specific about its contents. Most surficial geologic features and their FGDC Digital Cartographic Standard (2006) symbols are assigned to these feature classes.

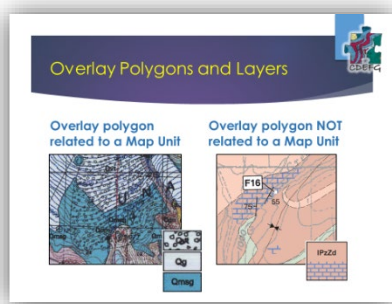
Field observation, sampling, and analytical points

Different kinds of stations, sample points, and analytical data translate well between AK GeMS and federal GeMS. Because the federal GeMS standard is flexible, additional feature classes can always be added for new types of information. DGGS has specified several standing feature classes that work well for our current IT-database situation and the kinds of data we most often display on maps: *stations, geochron_points,* and *fossil_points*. Other points such as drill hole locations and miscellaneous analyses that might have been assigned to *GenericPoints, SamplePoints,* or *GenericSamples* are instead assigned to the *cartographic_points* feature class in AK GeMS. For example, DGGS maintains a separate comprehensive geochemical database for the state at https://maps.dggs.alaska.gov/geochem/. Any geochemical data on a geologic map would refer back to DGGS' primary geochemical database for feature-level metadata.

Miscellaneous feature classes

There are no substantial differences between the federal and Alaska schemas for the *OverlayPolys, IsoValueLines*, or *DataSourcePolys* feature classes. *OverlayPolys* is discussed further in slide 8 in comparison with the *MapUnitOverlayPolys* feature class.

Slide 8.



Federal GeMS documentation describes two kinds of overlay polygons: a feature directly referencing a map unit (*MapUnitOverlayPolys*), and a feature that describes a modification of a defined map unit (*OverlayPolys*).

MapUnitOverlayPolys

On the slide's left is a scenario where federal GeMS' *MapUnitOverlayPolys* would be used, as there is an area of volcanic ash (map unit Qvt) overlying a glacier (Qg) and subglacial volcanic flows (Qmsg). In fact there are three stacked units, as the glacier is also overlying the subglacial volcanic flows. According to the federal schema, the area of volcanic ash would be a polygon in the *MapUnitOverlayPolys* feature class.

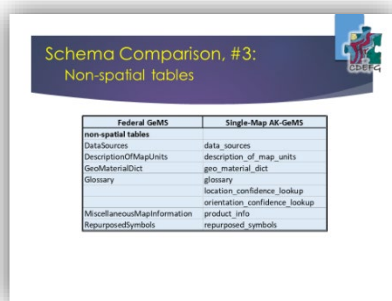
AK GeMS takes a different approach with stacked geologic map units and does not use the *MapUnitOverlayPolys* feature class. All polygons referring to a unit in the DMU are recorded in the *map_unit_polys* feature class. We believe this approach will simplify querying for map units in the multi-map database. Where the stacked units' relationship is known, the polygons are assigned layer values (via a new *layer* field), where a higher number indicates the polygon is stratigraphically higher in the stack. A couple presentations are available describing DGGS' approach to layers in the extended AK GeMS database: https://ngmdb.usgs.gov/Info/dmt/docs/DMT19_Hendricks1.pdf (Hendricks, 2019) and https://dggs.alaska.gov/gemswiki/lib/exe/fetch.php?media=start:CDEFG_3D%20presentations_by_Illinois_and_Alaska-20200413_1807-1.mp4 (starts at 24:08; Hendricks and others, 2020).

<u>OverlayPolys</u>

On the slide's right, a pattern depicts a limy area in a dolomite map unit, a modification of map unit lPzZd's general unit description. In the AK GeMS schema, this limy-patterned polygon is stored in the *overlay_polys* feature class—same usage as federal GeMS *OverlayPolys* feature class.

When the single-map database is eventually loaded into the multi-map database, the overlay polygon area is "burned into" or merged with the *map_unit_polys* feature class and becomes an area of the map unit with a modifier (via a new *modifier* field) of "limy". In another example, a map unit modifier can also be used to store critical information for mineral industry exploration, such as specific types of alteration related to ore deposits. An exploration geologist could easily query for potassic alteration in granite, anticipating that the map unit polygon fields are populated systematically with constrained values. The process simplifies multi-map queries among map units and modifiers, and produces results that are expected.

Slide 9.

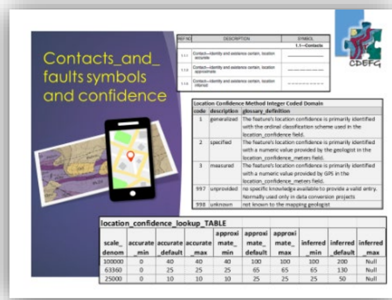


Most of the non-spatial tables (*DescriptionOfMapUnits*, *GeoMaterialsDict, Glossary,* and *RepurposedSymbols)* are largely the same for both the federal and Alaska-extended schemas, and are used in similar ways in the single-map schemas. In the multi-map database, these will be universal tables that link back to individual maps with a Globally Unique Identifier (GUID) for each record, i.e., glossary terms for all of the maps will be found in the one Glossary table.

- Having all of the map units in one *DescriptionOfMapUnits* table will help facilitate the combination of geologic units from different maps into compiled map unit descriptions. Units may be involved in more than one compilation since the original map unit definitions are preserved. This could be arranged by a many-to-many relationship between *DescriptionOfMapUnits* and a future map unit compilation table.
- DGGS does not usually take advantage of the *RepurposedSymbols* table. We prefer to create new symbols that can be permanently assigned to a feature (Ekberg and others, 2021).

The following slides explain additional differences between the schemas in more detail, and how many of the non-spatial tables are involved in data management.

Slide 10.



AK GeMS improves the mapper's ability to flexibly assign location confidence values and symbology to contacts and faults. Location confidence is also recorded for other line and point feature classes.

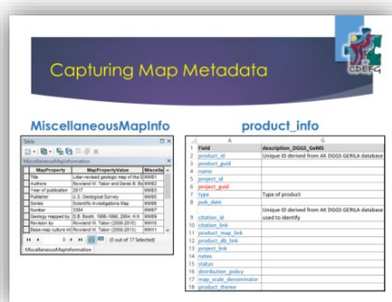
- DGGS has mapped various combinations of feature type, location confidence, identity confidence, existence confidence, "is concealed", and map scale to specific FGDC Digital Cartographic Standard (2006) line symbols. A symbology publication for AK GeMS by Ekberg and others (2021) provides this information in detail. Using Esri's Data Reviewer extension, we can check that the appropriate symbol was selected for the entered feature-level metadata. In the future, we may use ArcGIS Pro's attribute-driven symbology to automatically display data using these fields.
- Location confidence may be assigned in a number of ways using a new field *location_confidence_method* in many of the feature classes.

  If "generalized" is selected from the domain, the feature's location confidence value will be automatically populated from the *location_conf_lookup_table* using map scale and the default value for the symbol type. This method is preferred by geologists who would rather assign an accurate, approximate, or inferred symbol to depict the confidence.

  A value of "specified" allows the mapper to enter a specific value for the location confidence. Mappers who don't want to use the default confidence value or have values other than the default can enter their own values.

  A value of "measured" indicates a device-generated location confidence value, such as from a GPS for point locations. This option is important because the location confidence value of data used for digital mapping can be much smaller than the intended scale of the digital map, unlike in the paper map scenario where the minimum location confidence value of all features defaults to the line width of the printed symbology.

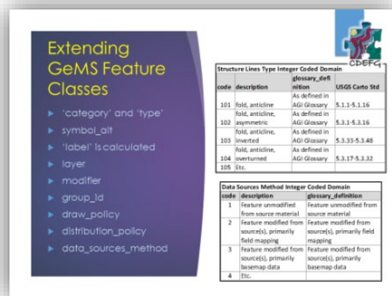
Slide 11.



Ancillary map metadata are handled differently by the federal GeMS and Alaska extended schemas. The federal GeMS' *MiscellaneousMapInfo* table is structurally designed to handle information from only one map, while the AK GeMS *product_info* table contains fields for similar information, to be compatible with a multi-map database implementation. The *product_info* fields are essentially transposed from *MiscellaneousMapInfo* to hold primary information for many maps.

The *product_info* table also introduces the concept of a geologic "product". In DGGS' publication PostgreSQL database that is related to but separate from the multi-map PostgreSQL database, a project (parent) can have multiple child publications and multiple grandchild products. Products can be PDF maps, digital geodatabases, text documents, other database types, spreadsheets, etc. PDF maps and digital map geodatabases are both relevant to the processing of GeMS geologic data and population of the multi-map database.

Slide 12.

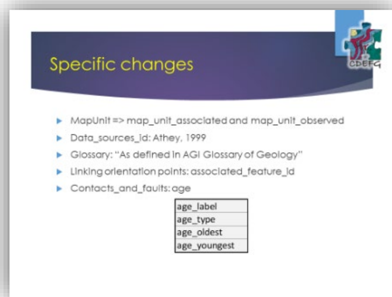


Most of the AK GeMS feature classes contain additional fields needed for various purposes.

- The *category* field is added as a parent-level hierarchal organizer to the *type* field. DGGS assigned most FGDC Digital Cartographic Standard (2006) symbols in our style file to specific feature classes (Ekberg and others, 2021). We needed a way to organize the large number of geologic features and their symbols logically in the constrained domains. The *category* field adds needed structure to the organization, e.g., in the *contacts_and_faults* feature class category domain *cf_cat_dom*, there are six categories: contact, fault, boundary, unprovided, unknown, and other. In turn, categories "contact" comprises six types, "fault" comprises 21 types, and "boundary" comprises six types. Additional related symbols are calculated from values recorded in attributes *type, is_concealed, location_confidence_meters, existence_confidence, identity_confidence*, and the expected visualization scale. All values in domains are defined in the universal AK GeMS Glossary and refer directly to FGDC Digital Cartographic Standard (2006) symbols where relevant.
- The AK GeMS field *symbol_alt* provides a necessary alternative for specific or custom symbolization found on a single map, as opposed to feature symbolization for a web map or more generic presentation of a large number of features across multiple maps. DGGS uses the field *symbol* for more generic or web symbolization purposes. Field *symbol_alt* is used to capture the symbolization of the original map, if it is different from standard usage. This field perhaps better records the original intent of the authors.
- In Alaska's extended schema, the *label* field is calculated from the *DescriptionOfMapUnits* table and *IdentityConfidence* field, e.g., the label of a Tertiary granite with a low identity confidence is calculated as "Tg?".
- See slide 8's discussion of the *MapUnitOverlayPolys* feature class for more information about the *layer* field.
- See slide 8's discussion of the *OverlayPolys* feature class for more information about the *modifier* field.
- The AK GeMS *group_id* field is useful for grouping together discontiguous or planarized features, e.g., segments of a through-going, major faults can be grouped together, potentially even across multiple maps. The value used in this field is a GUID string rather than a feature's common name.
- Feature classes in AK GeMS contain a binary *draw_policy* field that controls whether the feature should be displayed on the map. For example, if many foliation measurements were collected in a small area, the author can choose a subset of the foliations to be represented on the map and still keep all of the measurements available in the *orientations_points* feature class for potential future use.

- The AK GeMS *distribution_policy* field controls whether features in the database are bundled with the data distribution package, e.g., proprietary age data from another agency could be used to help develop geologic interpretations, but it should not be made available to the public.
- The *data_sources_method* field captures how values in the *data_sources* field were used to develop the feature. The field's domain values include options for whether the feature was modified from the data source, based primarily on the data source, or not modified from the data source. Data sources may also be classified in *data_sources_method* as from field mapping, basemap data, or geophysical data. For example, a contact along a retreating coastline, revised by lidar data, could be attributed as "Feature modified from source(s), primarily basemap data". DGGS acknowledges the USGS Intermountain West project for the idea for this field.

Slide 13.



DGGS has also made a few other tweaks to the federal GeMS schema. Some of the more important changes are described below.

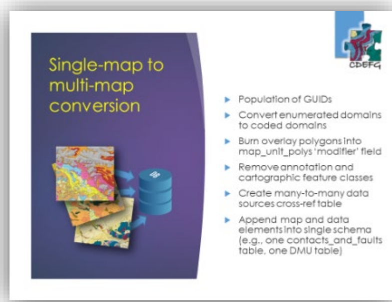
- The *MapUnit* field within point feature classes such as *OrientationPoints* can be problematic. Federal GeMS guidance states that *MapUnit* field values should be extracted from the *MapUnitPolys* feature class. However, there are geologic scenarios where the point's *MapUnit* value might be something other than the value of the surrounding polygon, e.g., a Tertiary basalt dike not mappable as a polygon at the map's scale (and possibly stored in *MapUnitPoints*) should have a MapUnit of "Tb", instead of the larger Tertiary gravel (Tg) polygon surrounding it. To handle this situation and alleviate possible confusion, AK GeMS does not use the *MapUnit* field. Instead, it contains *map_unit_observed* (the actual map unit observed [Tb] at that point location), and *map_unit_assoc* (the map unit of the surrounding polygon [Tg]).
- Most GeMS feature classes contain the foreign key of *DataSourceID* that maps to the *DataSources* table. Federal GeMS suggests using increasing values of DAS1, DAS2, etc. for the foreign key. For the mappers' ease of use and to more easily accommodate the multi-map database scenario, the AK GeMS version of the foreign key *data_sources* is formatted as an inline text reference (e.g., Athey, 1999). When maps are loaded into the multi-map database, the text IDs are converted to GUIDs.

  AK GeMS also allows features to reference multiple data sources, which are separated by a delimiter. For example, a contact's metadata may reference the original map source that it came from as well as the lidar dataset that was used to adjust the line. This usage creates a many-to-many relationship between the feature classes and the data sources table that is resolved by the multi-map schema when the data are loaded.

- DGGS' creation of a master *Glossary* table for all Alaska maps will ultimately save us time, but building the table's first draft could have been a major effort. Our *Glossary* currently contains more than 300 entries for geologic features and other terms, most of which are matched with symbols from the FGDC Digital Cartographic Standard (2006). We did not want to create custom definitions for all of these terms, and copying that many definitions directly from a non-open-source dictionary would be a copyright infringement. Instead, we have identified AGI's Glossary of Geology, 5th ed. online subscription as our primary reference, and we refer directly to the dictionary for term definitions where possible, e.g., DGGS' definition for contact is: As defined in the AGI Glossary of Geology, see term "contact". Terms that do not exist in the Glossary of Geology are modified from AGI definitions or defined by the author, e.g., the DGGS definition for gradational contact is: As defined in the AGI Glossary of Geology, see term "contact", but contact displays a gradual and progressive change in particle size between strata, usually from coarse at the base to fine at the top.

- In an effort to more robustly capture the relationships between geologic features in the GeMS schema, AK GeMS contains a new field *assoc_feature_id* in the *orientations_point* feature class. This relationship is particularly useful when the relationship between orientation data and the mapped feature is ambiguous. For example, if the geologist has collected a foliation on a limb of a mapped fold, the *structure_lines_id* (primary key) of the fold hinge line in the *structure_lines* feature class can be inserted as the *assoc_feature_id* value for the foliation point. The database then records that the foliation measurement was not randomly collected, but is measuring a particular feature.
- Additional fields were added to the *contacts_and_faults* feature class to capture ages of latest fault movement. The fields are primarily used to identify Quaternary and historical (less than 150 years) faulting in Alaska.

Slide 14.



Alaska's single-map schema is designed so that maps in the format can be merged more easily into the similar multi-map schema. DGGS has identified several processes that need to happen or differences in the schemas that need to be resolved when map data are uploaded into the multi-map database. We will automate as many of these repetitive processes as possible to make the upload process more efficient.

- Since primary keys are not often needed during map database creation, these fields are often calculated at the end of the process. The single-map schema uses text-based unique identifiers in GUID format for most primary keys, instead of federal-GeMS-suggested formats such as "MUP1" for a *MapUnitPolys_ID* value. As discussed in slide 13, the *data_sources_id* primary key is an exception to this rule. All primary keys in the single-map schema must be converted from text format to GUID, e.g., a *data_source_id* of "Athey, 1999" will be converted to GUID format {1051D9B2-9B8D-4997-9E3C-3EEE0836B781}, while preserving the relationships among the feature classes.
- Alaska's single-map schema uses text-based enumerated domains. The text values are converted to integer-based domains in the multi-map database to decrease the size of the database and increase data processing speed. For example, as described in slide 10, "contact" has a code of 100. All other types of contacts have codes from 101―199. All faults have domain codes in the 200s.

- As described in slide 8, *overlay_polys* feature class value-areas are "burned into" or merged with polygons and the modifier field in the *map_unit_polys* feature class. The *overlay_polys* feature class does not exist in the AK GeMS multi-map data schema.
- Annotation classes will not be loaded into the multi-map database. DGGS is still looking at whether the cartographic point, line, and polygon feature classes can be removed from the multi-map schema, as we need to ensure that all feature metadata described in cartographic feature classes as symbology are also contained as feature-level metadata in the primary features classes. Our goal is for purely cartographic information to be removed from the multi-map database, with the understanding that symbolization of multi-map data will be conducted on the fly and through convenient web services.
- As described in slide 13, AK GeMS allows a feature to have multiple data sources, which creates a many-to-many relationship between the feature classes and the *data_sources* table. During multi-map data loading, a cross-table is populated from the delimited list of map sources to accommodate the many-to-many relationship.
- Finally, the standardized schema facilitates appending new map databases to the multi-map tables. Having one *map_unit_polys* feature class will greatly enhance the users' ability to quickly and easily query information across multiple maps.

Slide 15

DGGS is continuing to work on documentation and process efficiency for the database schemas and our general workflow. We anticipate multiple future publications on these topics.

Monthly CDEFG meetings are open to the public and held online. Through these meetings and other venues, we hope to encourage conversation about digital geologic mapping and data management. Please contact Jen Athey at 907.451.5028 or jennifer.athey@alaska.gov if you would like to be added to the email list for this group. Meeting recordings, notes, and presentations are available at https://dggs.alaska.gov/gemswiki/.

ACKNOWLEDGEMENTS

DGGS greatly appreciates the many people who have helped us refine Alaska's extended database schemas through presentations and discussion. We also thank our funding sources for helping to make this work possible: Environmental Protection Agency Exchange Network grant OS 836470-01, USGS National Geological Map Database Cooperative Agreement award number G19AC00411, USGS National Geological and Geophysical Data Preservation Program grant number G19AP00077, and the State of Alaska General Fund.

REFERENCES

Ekberg, P.G., Hendricks, M.D., and Athey, J.E., 2021, AK GeMS symbology: A description of the AK GeMS style file: Alaska Division of Geological & Geophysical Surveys Miscellaneous Publication 169, 7 p., https://doi.org/10.14509/30584.

Federal Geographic Data Committee [prepared for the Federal Geographic Data Committee by the U.S. Geological Survey], 2006, FGDC Digital Cartographic Standard for Geologic Map

Symbolization: Reston, Va., Federal Geographic Data Committee Document Number FGDC-STD-013-2006, 290 p., 2 plates, https://ngmdb.usgs.gov/fgdc_gds/.

Hendricks, M.D., Ekberg, P.G., Athey, J.E., Wyatt, W.C., Willingham, A.L., and Naibert, T.J., 2021, AK GeMS data dictionary: A description of the Alaska geologic mapping schema: Alaska Division of Geological & Geophysical Surveys Miscellaneous Publication 170, 10 p. https://doi.org/10.14509/30669

Hendricks, Mike, 2019, Supporting multiple planar topologies in a GeMS geodatabase: A baby step toward a true 3D data model? Considerations in modeling a multi-layered geologic map: Digital Mapping Techniques Workshop (presentation), Butte, Montana, May 19─22, 2019, 12 p., https://ngmdb.usgs.gov/Info/dmt/docs/DMT19_Hendricks1.pdf.

Hendricks, Mike, Athey, Jen, Ekberg, Trish, Wyatt, Chris, and others, 2020, Layer support in AK GeMS: Moving toward 3D geologic data: Collaborative Database Effort for Geology (CDEFG; presentation), online, April 13, 2020, 19 p., https://dggs.alaska.gov/gemswiki/lib/exe/fetch.php?media=start:AK_GeMS_Layer_Support_to%20CDEFG_13_Apr_2020.pptx. Recording https://dggs.alaska.gov/gemswiki/lib/exe/fetch.php?media=start:CDEFG_3D%20presentations_by_Illinois_and_Alaska-20200413_1807-1.mp4, starts at 24:08.

Thorleifson, Harvey, 2020, Seamless geological mapping: proposed international collaboration: Digital Mapping Techniques Workshop (presentation), videoconferenced from Reston, Virginia, and Champaign, Illinois, June 8─10, 2020, 34 p., https://ngmdb.usgs.gov/Info/dmt/docs/DMT20_Thorleifson.pdf.

U.S. Geological Survey National Cooperative Geologic Mapping Program, 2020, GeMS (*Ge*ologic *M*ap *S*chema)—A standard format for the digital publication of geologic maps: U.S. Geological Survey Techniques and Methods, book 11, chap. B10, 74 p., https://doi.org/10.3133/tm11B10.