AK GEMS DATA DICTIONARY: A DESCRIPTION OF THE ALASKA GEOLOGIC MAPPING SCHEMA

Michael D. Hendricks, Patricia G. Ekberg, Jennifer E. Athey, W. Chris Wyatt, Amanda L. Willingham, and Travis J. Naibert

AK GEMS DATA DICTIONARY: A DESCRIPTION OF THE ALASKA GEOLOGIC MAPPING SCHEMA

Michael D. Hendricks, Patricia G. Ekberg, Jennifer E. Athey, W. Chris Wyatt, Amanda L. Willingham, and Travis J. Naibert

STATE OF ALASKA
Mike Dunleavy, Governor

DEPARTMENT OF NATURAL RESOURCES
Corri A. Feige, Commissioner

DIVISION OF GEOLOGICAL & GEOPHYSICAL SURVEYS
Steve Masterman, State Geologist and Director

Contents

Supplemental Files

The following supplemental file is available to download at: https://doi.org/10.14509/30669

    AK_GeMS_Data_Dictionary_ver_1pt1.xlsx

    AK_GeMS_XML_Workspace_Document_ver_1pt1.zip

AK GEMS DATA DICTIONARY: A DESCRIPTION OF THE ALASKA GEOLOGIC MAPPING SCHEMA

Michael D. Hendricks[1], Patricia G. Ekberg[1], Jennifer E. Athey[1], W. Chris Wyatt[1], Amanda L. Willingham[1], and Travis J. Naibert[1]

INTRODUCTION

To meet local geologic mapping requirements, the Alaska Division of Geological & Geophysical Surveys (DGGS) designed the Alaska Geologic Mapping Schema (AK GeMS), which is an extension to the Geologic Mapping Schema (GeMS) (U.S. Geological Survey National Cooperative Geologic Mapping Program, 2020). This document and accompanying Microsoft Excel workbook (hereafter referred to as the data dictionary) describes the organization and logic of AK GeMS. It is intended to serve as a multi-use, comprehensive reference document for an AK GeMS compliant geodatabase. A data dictionary catalogs and communicates the structure and content of data and provides meaningful descriptions for individually named data objects (U.S. Geological Survey, 2020).

Uses of this data dictionary may include:

- As a reference guide for geologic mappers populating relevant features and fields in an AK GeMS production geodatabase.
- As a reference guide for data leads while finalizing an AK GeMS production geodatabase.
- As a reference guide for AK GeMS GIS administrators while conducting quality control on an AK GeMS production geodatabase.
- As a form of metadata for users of any published AK GeMS geodatabases.

The accompanying data dictionary file is available to download from doi.org/10.14509/30669. Details on symbolizing data utilizing the AK GeMS extension can be found in a related report, AK GeMS symbology: A description of the AK GeMS style file (Ekberg and others, 2021, doi.org/10.14509/30584).

AK GEMS DESIGN CONSIDERATIONS

The following elements are key design considerations of the AK GeMS extension:

- Focus on modeling geologic features and information relationships (as opposed to graphic element display) to better support quality division-wide map production, database queries, and web delivery.
- Capability for exporting to the GeMS (hereafter referred to as Fed GeMS) without significant loss of essential information.
- Capability for supporting both single map production geodatabases and the DGGS multi-map PostgreSQL enterprise geodatabase.
- Support for multiple geologic layers (e.g., bedrock, surficial, others).
- Formalized pick-lists of attribute domains that are fully defined in a glossary table.
- Integration with DGGS's publication workflow and digital data management infrastructure, including the Alaska DGGS Geologic and Earth Resource Information Library of Alaska (GERILA) enterprise database (DGGS Staff and Papp, 2005).

To meet these design considerations AK GeMS supports various geodatabase implementations as part of the overall Alaska DGGS Geologic Mapping System (fig. 1). Each of these implementations has several unique characteristics:

- **Single Map AK GeMS Production Geodatabases**. This implementation of AK GeMS uses the full AK GeMS specifications, including string-based attribute domains, as described in the data dictionary.
- **Single Map AK GeMS Publication Geodatabases**. A public distribution version of an AK GeMS geodatabase. It excludes feature classes and tables that are empty or only required for production, such as the editor tracking fields.
- **Single Map Fed GeMS National Archive Geodatabases**. The Fed GeMS standard has several naming conventions that differ from AK GeMS, particularly the use of Pascal case instead of underscore word delimiters for table and field names.
- **A Multi-Map AK GeMS Repository Geodatabase.** A modified AK GeMS geodatabase hosted on PostgreSQL enterprise architecture. This geodatabase modifies the standard AK GeMS schema in several ways, such as the use of integer-based attribute domains and true Globally Unique Identifiers (GUID) type fields. Details of these modifications are not addressed directly in this publication.
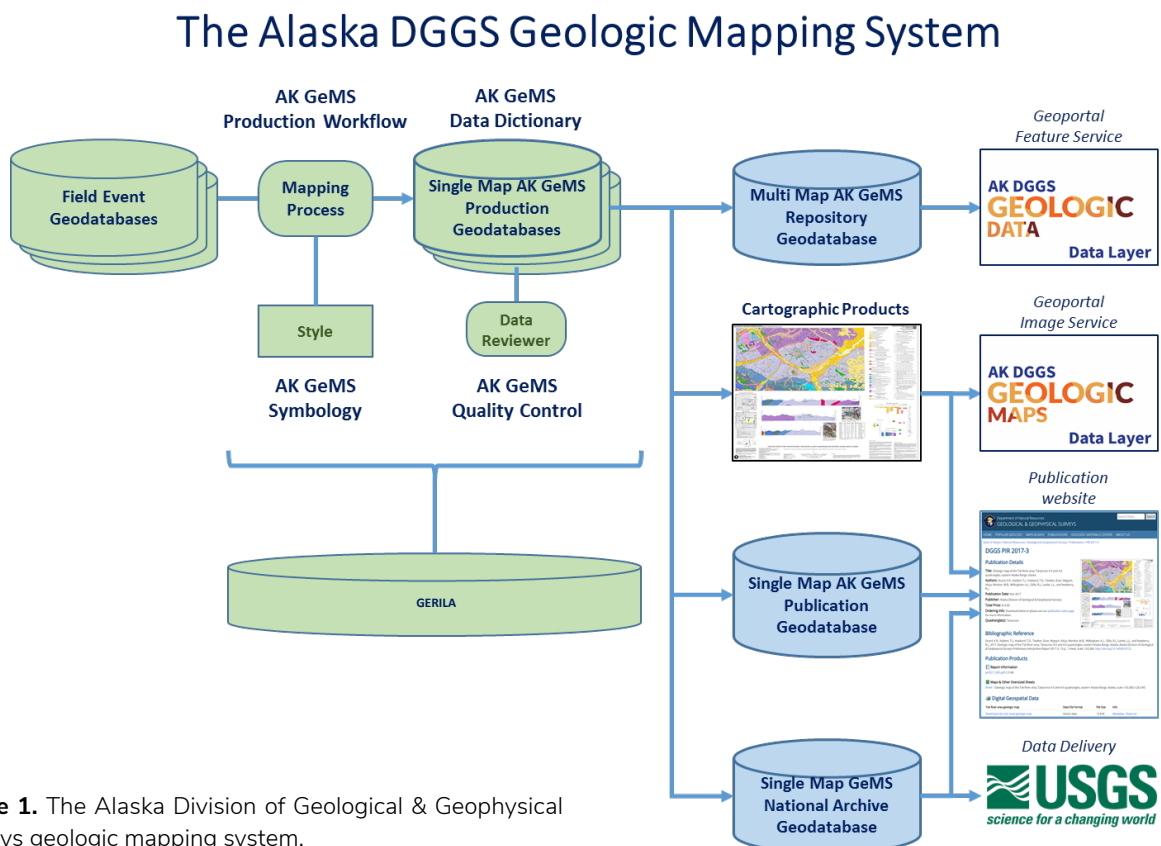
Figure 1. The Alaska Division of Geological & Geophysical Surveys geologic mapping system.

AK GEMS GEODATABASE DESIGN

AK GeMS organizes the feature classes and stand-alone tables into groups that include map units and boundaries, structure information, other geologic information, thematic points, and cartographic features (figs. 2 and 3). These feature classes and stand-alone tables include the required Fed GeMS classes as well as several custom feature classes and stand-alone tables to meet DGGS requirements. Details of each of these feature classes and stand-alone tables are provided in the accompanying data dictionary spreadsheet.

Naming Conventions

To enable multi-user editing on enterprise geodatabase architecture, specifically PostgreSQL, AK GeMS modified the Fed GeMS table and field names to use the snake case naming convention, which converts compound words to lower case words combined with an underscore. For example, "MapUnit-Polys" becomes "map_unit_polys."

Figure 2. Organization of the AK GeMS geodatabase features classes and tables.

```
ak_gems_template_ver_1.1.gdb
   correlation_of_map_units
      cmu_lines
      cmu_points
      cmu_polys
   cross_section_a
      csa_contacts_and_faults
      csa_map_unit_lines
      csa_map_unit_points
      csa_map_unit_polys
      csa_orientation_points
      csa_overlay_polys
   geologic_map
      cartographic_lines
      cartographic_points
      cartographic_polys
      contacts_and_faults
      data_sources_polys
      fossil_points
      geochron_points
      geologic_lines
      geologic_map_Topology
      geologic_points
      geologic_polys
      iso_value_lines
      map_unit_lines
      map_unit_points
      map_unit_polys
      orientation_points
      overlay_polys
      stations
      structure_lines
      structure_polys
   data_sources
   description_of_map_units
   geo_material_dict
   glossary
   location_confidence_lookup
   orientation_confidence_lookup
   product_info
```
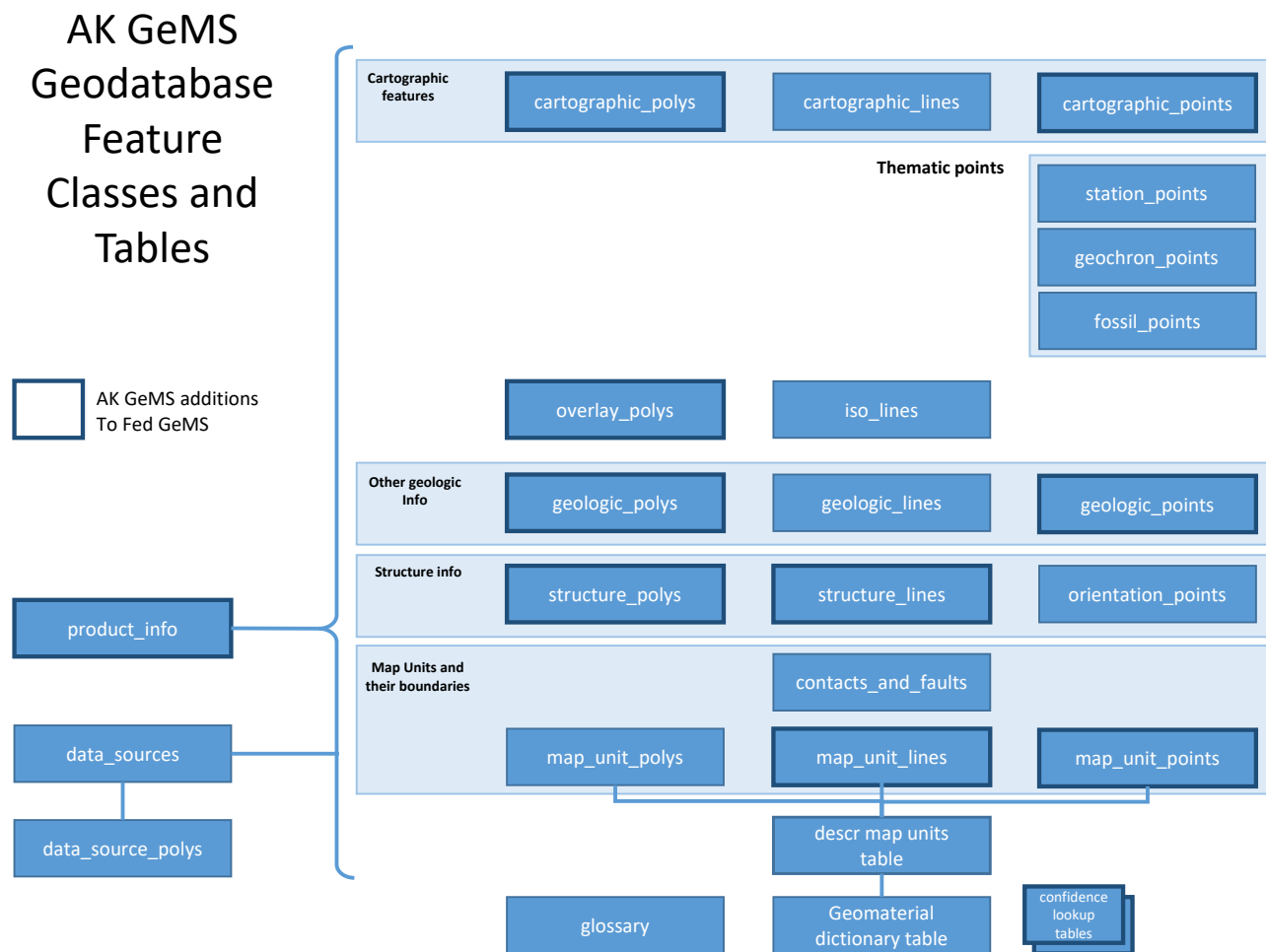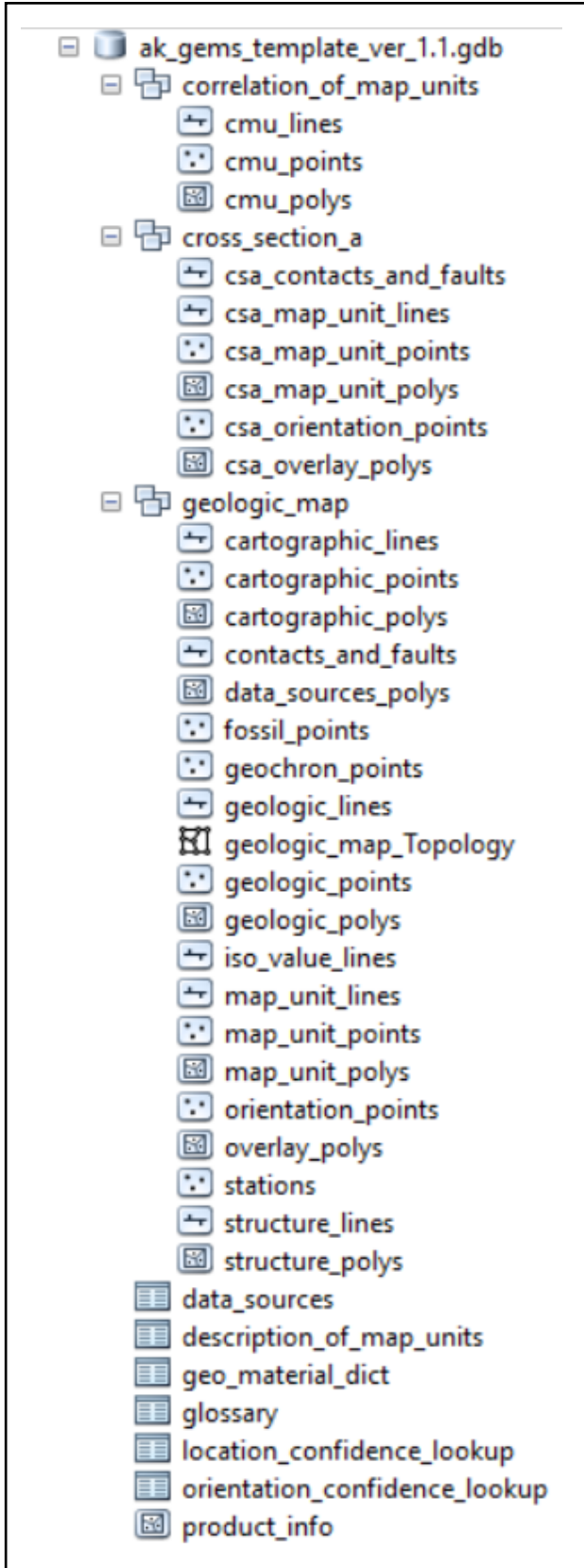
Figure 3. Feature classes and tables in the AK GeMS geodatabase.

Database Relationships

AK GeMS supports certain relationships among feature classes and tables. Appendix A provides an Entity_Relationship (ER) diagram of these relationships. These relationships are color-coded and categorized into the conceptual groups:

- Product Identification Relationships
- Data Source Relationships
- Map Unit Relationships
- Points Feature Class to Stations Relationships
- Associated Feature Relationships
- Relationships to the Alaska DGGS GERILA database

As part of the AK GeMS design, we implement GUIDs. GUIDs are 128-bit values that uniquely identify records in the database, for example {1051D9B2-9B8D-4997-9E3C-3EEE0836B781}. Since GUID field types can be difficult to understand by geologists and map makers, temporary human-readable ID values are used by DGGS during the production phase. As a result, in the Single Map Implementation of AK GeMS, ID fields throughout the geodatabase are set as String type with length 50. These values are overwritten by GUID values before publication. In a multi-map implementation, a true GUID field type is used.

Product Identification Relationships

To ensure that AK GeMS data can be easily integrated into a multi-map implementation, every feature class and table includes a field named product_id [string length 50 in a single map product geodatabase], which is a foreign key to the product_info table's product_id field. These relationship fields are color-coded green (without connecting lines) in the ER diagram provided in appendix A.

Data Source Relationships

Fed GeMS requires feature-level source metadata. Every feature class and table includes a data_sources field which is a foreign key to the source

field in the data_sources table. This relationship is inherently a many-to-many relationship. As a result, AK GeMS allows multiple, semicolon-delimited entries in the data_sources field of Single Map AK GeMS Production and Publication Geodatabases. The data_sources field is type String length 150. This field's expected value is an in-text citation format; for example, "Foster, 1992" or "Foster, 1992; Day and others, 2007" are valid data_ sources field values. We use this format instead of an integer or GUID value during production and publication of single map geodatabases to make citing feature-level metadata more intuitive for the geologist making the map and examining the data. Note that in the Multi-Map AK GeMS Repository Geodatabase implementation, a GUID-based many-to-many relationship class is built with this relationship. These relationship fields are color-coded salmon (without connecting lines) in the ER diagram provided in appendix A.

There are several deviations and additions to the basic data source relationship described above:

- The description of map units table has a description_source_id field in lieu of the data_sources field. In practice, this field supports only single values.
- The geochron_points feature class has a second source identification field, analysis_source_id, which is a foreign key to the sources field in the data_sources table. In practice, this field supports only single values.
- The fossil_points table has three additional fields: analysis_source_id, fossil_forms_source_id, fossil_age_source_id. In practice, these fields support only single values.

The data_sources_polys feature class can also be used to create an inset map indicating the geographic extent of data sources used to support geologic interpretations. There is a one-to-one relationship between the data_sources_polys feature class and the data_sources table via the data_sources_id field.
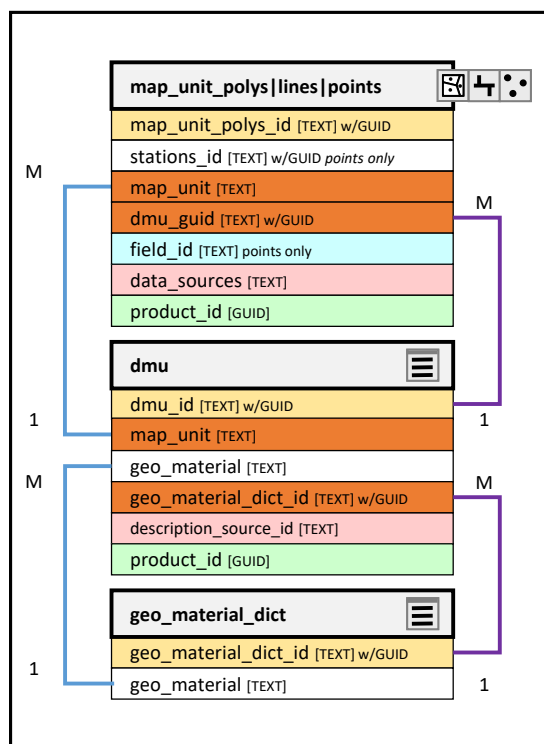

Figure 4. Relationships within the Single Map AK GeMS Production and Publication Geodatabase.

Map Unit Relationships

Within the Single Map AK GeMS Production and Publication Geodatabases, there is a many-to-one relationship between records in the map_unit feature classes (map_unit_polys, map_unit_lines, and map_unit_points) and the description_of_map_units table via the map_unit field. Also, there is a many-to-one relationship between the description_of_map_units table and the geo_material_dict table via the geo_material field, see light blue lines in figure 4 and appendix A.

These relationships require the use of unique identifiers in the Multi-Map AK GeMS Repository Geodatabase implementation. To ensure this, GUID-based relationships are established; see purple lines in figure 4 and appendix A.

Points Feature Class to Stations Relationships

Features within the stations point feature class typically represent the location where a geologist collected a measurement or observation. These stations are often referenced by various thematic point features, such as orien-

tation_points, geochron_points, fossil_points, and map_unit_points. These point feature classes have an optional relationship with the stations feature class. In other words, these feature classes do not require a relationship to a station, but in practice, these points typically do have this relationship. These relationships are shown with a dark blue connecting line in the ER diagram provided in appendix A.

Associated Feature Relationships

The orientation_points and cartographic_points feature classes include an assoc_field_id field that can be used to relate a feature, via a GUID, to another feature in one of several other feature classes, including map_unit_points, map_unit_lines, map_unit_polys, contacts_and_faults, structures_lines, etc. The primary use of this relationship is to associate decoration types with other mapped features. For example, a dip symbol decoration can be associated with a fold feature stored in the structure_lines feature class. These relationship fields are color-coded steel gray (without connecting lines) in the ER diagram provided in appendix A.

Relationships to the Alaska DGGS GERILA Database

Several fields in AK GeMS relate map data to values in the GERILA database. These relationship fields are color-coded cyan (without connecting lines) in the ER diagram provided in appendix A.

DATA DICTIONARY EXCEL FILE ORGANIZATION

The data dictionary Excel file is organized into the following tabs:

- **Tables List tab:** A description of the tables (feature classes and stand-alone tables) included in AK GeMS. The following columns are included in this tab:
    - º Feature Dataset: The feature dataset where the table resides.
    - º Table: The name of the table in AK GeMS.
    - º Table Description: The description of the table.
    - º Example Features: Example features that may be included in the table.
    - º Notes: General notes associated with the table.
    - º Fed GeMS Table Name: The name of the table when mapped to a Fed GeMS database.

- **Domains List tab:** A description of the coded and range domains in AK GeMS.
    - º Domain Name: The name of the domain in AK GeMS.
    - º Description: The description of the domain.
    - º Field Type: Field type using ESRI terminology.
    - º Domain Type: The type of domain, coded or range.
    - º Split Policy: The split policy set in the template database.
    - º Merge Policy: The merge policy set in the template database.
    - º Where Used: Which tables have this domain for a field.
    - º Notes: General notes associated with the domain.

- **Tabs for each thematic table:** The following columns are included in these tabs:
    - º Field Name: The name of a field.
    - º Description: The description of the field.

- º Data Type: Field type using ESRI terminology.
- º Default Value: Default value set in the template geodatabase.
- º Null Value Allowed: NULL values are allowed in the published geodatabase.
- º Multiple Values Allowed: Can multiple values occupy this field value? If so, individual values should be delimitated with a semicolon.
- º Example Values: Example values that might be found in this field.
- º Domain: The domain name with a link to the domain values.
- º Lineage: From where this field was derived. 'Fed GeMS' indicates the field maps directly to a field in the Federal GeMS Schema. 'AK GeMS' provides a field unique to AK GeMS. 'ESRI' indicates that the ESRI software adds this field.
- º Notes: General notes associated with the table.
- º When: When this field is populated.
- º Who: Who populates this field.
- º How: How this field is populated.
- º Field Name Fed GeMS: The related field name in the Fed GeMS standard.

- **Domains Coded Values tab:** The single tab with the unique values included in each coded domain.
  - º Code: The integer value of the domain. This value is used only in the multi-map database.
  - º Description: The description of the domain value. This description is used as the code value for the single map AK GeMS and Fed GeMS databases.
  - º Glossary Definition: Definition of the domain value that is in the glossary table.
  - º Notes: General notes associated with the value.

- **Domains Range Values tab:** The description of the range of values in each range domain.
  - º Field Type: Field type associated with this domain. Typically, field type of Float.
  - º Minimum: Minimum value of the range.
  - º Maximum: Maximum value of the range.

- **Annotation Domain Values tab:** The unique values included in each annotation domain.
  - º Code: The integer value of the domain. This value is used only in the multi-map database.
  - º Description: The description of the domain value. Note: this is used as the code value for the single map AK GeMS and Fed GeMS databases.
  - º Glossary Definition: Definition of the domain value. A corresponding record exists in the glossary table.
  - º Notes: General notes associated with the value.

- **Suggested Values tab:** The suggested values for select fields that do not have attribute domains.
  - º Value: The suggested value.

- **Glossary Definition:** Definition of the suggested value that is used in the glossary table.
  - º Notes: General notes associated with the value.

- **geo_material_dom tab:** The unique values of the geo_material_dom domain. This table is directly replicated from Fed GeMS. There is no integer code associated with this domain.
  - ○ Code: The string value of the domain.
  - ○ Description: The description of the domain value.
  - ○ Notes: General notes associated with the value.
- **location_conf_lookup_TABLE tab:** A scale-based lookup table used to convert location confidence values to numeric values. Only applicable to the single map production AK GeMS database.
- **orientation_conf_lookup_TABLE tab:** A lookup table used to convert orientation confidence values to numeric values. Only applicable to the single map production AK GeMS database.
- **Topology tab:** A description of topology rules suggested for use with a single map production AK GeMS database. Only applicable to the single map production AK GeMS database.
  - ○ Feature Class 1: The feature class involved in the topologic rule.
  - ○ Rule: The topological rule name.
  - ○ Feature Class 2: If applicable, the second feature class involved in the topologic rule.
  - ○ Notes: General notes associated with the rule.
  - ○ Lineage: From where this field was derived. 'Fed GeMS' indicates the rule is identified in the GeMS documentation. 'AK GeMS' indicates a rule added to AK GeMS.
- **Glossary Definitions tab:** Listing of all glossary definitions. These definitions are stored in the glossary table. The glossary definitions are either directly defined in the AK GeMS data dictionary or referenced from the AGI Glossary of Geology (American Geosciences Institute, 2020).
  - ○ Term: Term that is defined.
  - ○ Glossary Definition: The definition of the term.
  - ○ Data Source: Source of the definition.

SINGLE MAP PRODUCTION PHASES, ROLES, AND METHODS (WHO, WHEN, HOW)

The AK GeMS data dictionary provides general guidance on database production for a single map production database. Additional guidance is provided by internal DGGS documentation. General guidance on phases, roles, and methods is provided with values in three columns, identified as When, Who, and How within the data dictionary.

Phases (When)

- Prep: The beginning phase of the mapping project. The database is generated and project-specific settings, such as coordinate systems, are set in this phase. Typically, only the product_info table has values assigned in this phase.
- Fieldwork: Fieldwork associated with the mapping project.
- Interpretation: The interpretative phase of the mapping project.
- Pub Prep: The Pub Data Lead calculates database-centric fields from data populated in the interpretation phase.
- Pub: In this phase the database is prepared for publication.

- On the fly: Continuously occurring phase with on-the-fly calculations by ESRI software.
- Static: Data is static.

Roles (Who)

- Geologist: The geologist(s) editing the geodatabase while producing the geologic map.
- Pub Data Lead: The map publication data lead for the geodatabase.
- GeMS Admin: The GIS team GeMS administrator for the geodatabase.
- ESRI: The ArcMap or ArcPro software on-the-fly calculations.

Method (How)

- Manual: Manually entered value.
- Calculated: Calculated programmatically by software or script.
- Picked: Picked from an attribute domain.
- Imported: Values imported from an outside data source.
- GPS: Values generated from GPS-based collections in the field.

ACKNOWLEDGMENTS

REFERENCES

American Geosciences Institute, Glossary of Geology [website], last accessed December 14, 2020. https://www.americangeosciences.org/pubs/glossary

DGGS Staff, and Papp, K.R., 2005, Alaska GeoSurvey News - A summary of the past, present, and future of distributing Alaskan geologic data in digital format: Alaska Division of Geological & Geophysical Surveys Newsletter 2005-2, 8 p. https://doi.org/10.14509/14595

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

U.S. Geological Survey, 2020, Data Management: Data Dictionaries [website], last accessed October 26, 2020. https://www.usgs.gov/products/data-and-tools/data-management/data-dictionaries

U.S. Geological Survey National Cooperative Geologic Mapping Program, 2020, GeMS (Geologic Map Schema)—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

Walker, J.D., Geissman, J.W., Bowring, S.A., and Babcock, L.E., eds., 2018, Geologic Time Scale v. 5.0: Geological Society of America. https://www.geosociety.org/documents/gsa/timescale/timescl.pdf

APPENDIX A: AK GEMS KEY RELATIONSHIPS

AK GeMS Key Relationships

product_info
- product_id [GUID]
- product_id_dggs [LONG]
- project_id_dggs [LONG]
- citation_id_dggs [LONG]

data_sources
- data_sources_id [TEXT] w/GUID
- citation_id_dggs [LONG]
- source [Text]
- product_id [GUID]

data_sources_polys
- data_sources_polys_id [TEXT] w/GUID
- data_sources_id [TEXT] w/GUID
- product_id [GUID]

iso_value_lines
- iso_value_lines_id [TEXT] w/GUID
- data_sources [TEXT]
- product_id [TEXT] w/GUID

join

join

Legend:
- Primary Key
- 1-M relate to product_id field in product_info table
- Relates to DGGS GERILA DB
- 1-M relate to source field in data_sources table ***
- *** data_sources field can accommodate multiple entries with a semicolon delimitations. Ex Gillis, 2018; Gillis, 2019
- GUID value relating to _id field in either: map_unit, structure, or geology poly, line, or point

map_unit_polys|lines|points
- map_unit_polys_id [TEXT] w/GUID
- stations_id [TEXT] w/GUID *points only*
- map_unit [TEXT]
- dmu_guid [TEXT] w/GUID
- field_id [TEXT] *points only*
- data_sources [TEXT]
- product_id [GUID]

join

dmu
- dmu_id [TEXT] w/GUID
- map_unit [TEXT]
- geo_material [TEXT]
- geo_material_dict_id [TEXT] w/GUID
- description_source_id [TEXT]
- product_id [GUID]

join

geo_material_dict
- geo_material_dict_id [TEXT] w/GUID
- geo_material [TEXT]

overlay_polys
- overlay_polys_id [TEXT] w/GUID
- data_sources [TEXT]
- product_id [TEXT] w/GUID

structure_polys|lines
- structure_polys_id [TEXT] w/GUID
- data_sources [TEXT]
- product_id [GUID]

geologic_polys|lines|points
- geologic_polys_id [TEXT] w/GUID
- data_sources [TEXT]
- product_id [GUID]

cartographic_polys|lines|points
- cartographic_polys_id [TEXT] w/GUID
- data_sources [TEXT]
- product_id [GUID]
- assoc_feature_id [TEXT] w/GUID *points only*

stations
- stations_id [TEXT] w/GUID
- field_id [TEXT]
- field_station_id [LONG]
- map_unit_assoc [TEXT]
- data_sources [TEXT]
- product_id [TEXT] w/GUID

orientation_points
- orienation_points_id [TEXT] w/GUID
- stations_id [TEXT] w/GUID
- field_id [TEXT]
- map_unit_assoc [TEXT]
- assoc_feature_id [TEXT] w/GUID
- data_sources [TEXT]
- product_id [TEXT] w/GUID

geochron_points
- geochron_points_id [TEXT] w/GUID
- stations_id [TEXT] w/GUID
- field_id [TEXT]
- map_unit_assoc [TEXT]
- assoc_feature_id [TEXT] w/GUID
- data_sources [TEXT]
- analysis_sources [TEXT] w/GUID
- product_id [TEXT] w/GUID

fossil_points
- fossil_points_id [TEXT] w/GUID
- stations_id [TEXT] w/GUID
- field_id [TEXT]
- map_unit_assoc [TEXT]
- assoc_feature_id [TEXT] w/GUID
- field_sample_id [TEXT]
- alternate_sample_id [TEXT]
- fossil_forms_source_id [TEXT]
- fossil_age_source_id [TEXT]
- data_sources [TEXT]
- analysis_source_id [TEXT] w/GUID
- product_id [TEXT] w/GUID