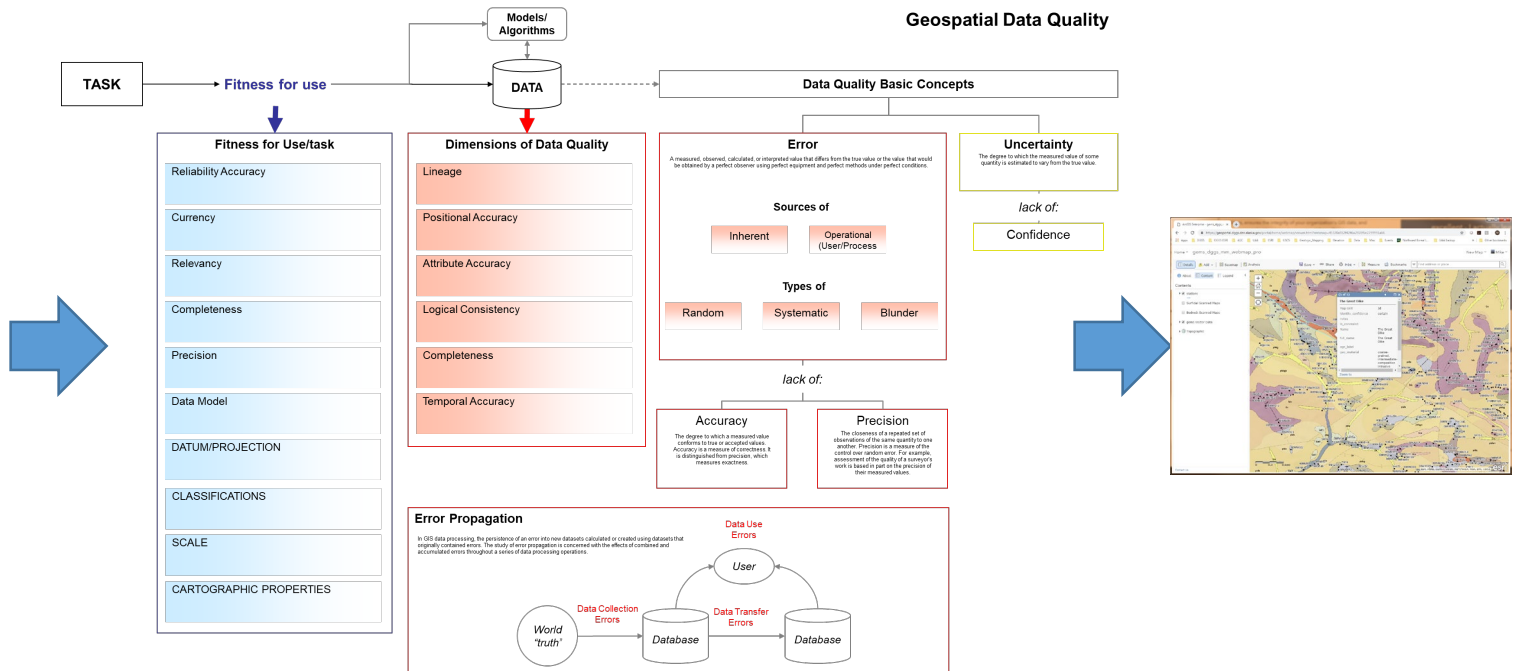


QA/QC of GeMS Data

Case Study: The Alaska DGGs Geologic Mapping System

10 Jan 2022



Mike Hendricks, Jen Athey, Amy Macpherson

Alaska Division of Geological & Geophysical Surveys
3354 College Rd, Fairbanks AK 99709

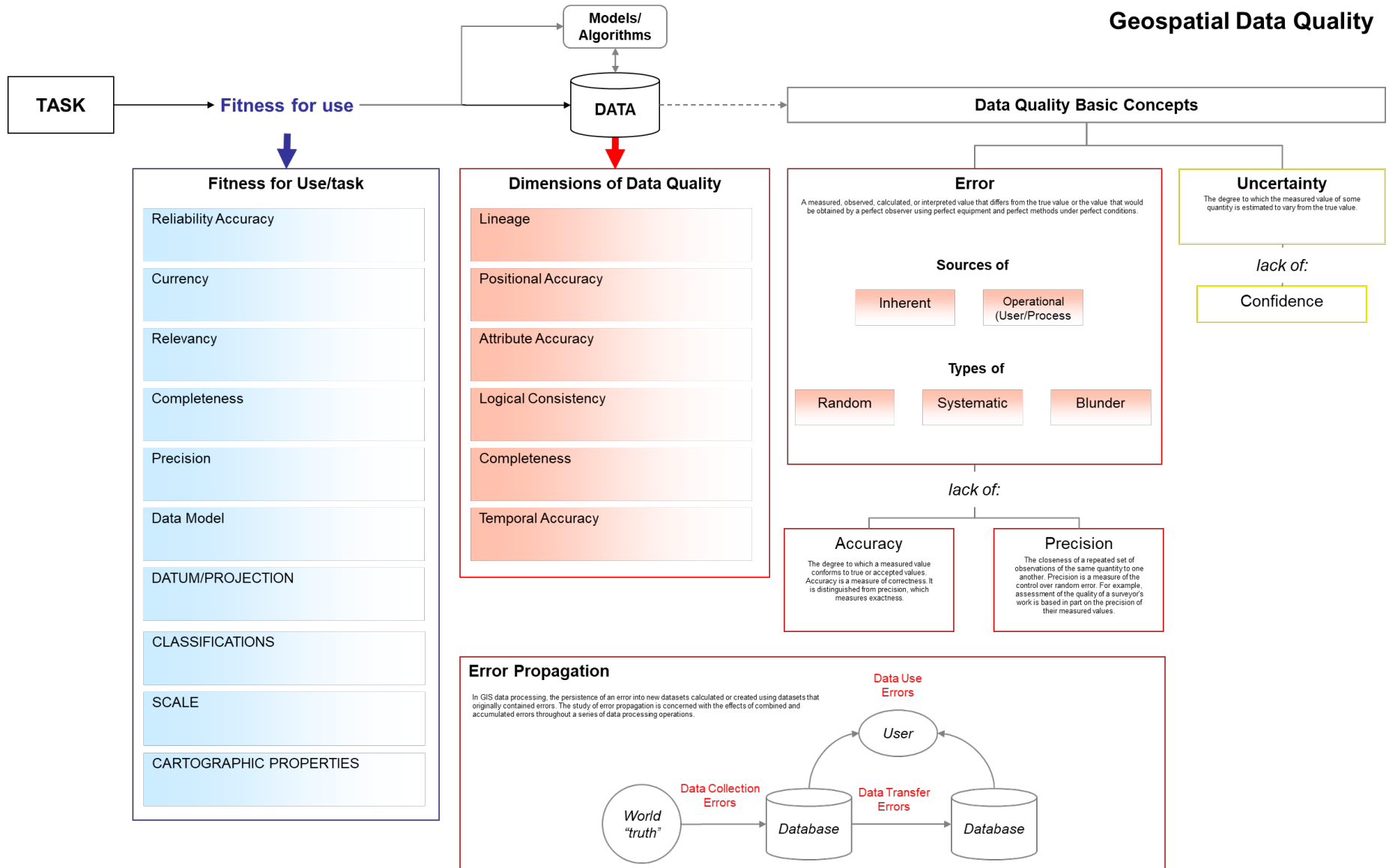


Agenda

- Geospatial Data Quality Fundamentals
- Quality Assurance (QA) & Quality Control (QC)
- QA/QC in the Alaska DGGS Geologic Mapping System
- Data Reviewer updates in ArcPro

Geospatial Data Quality Fundamentals

Geospatial Data Quality



Fitness for Use versus Data Quality

Fitness for Use: The ability of data to effectively be used for some intended purpose.

Data Quality: How faithfully the data represents the true (a) location, (b) shape, or (c) characteristics of the phenomena.

What level of data quality is required for a GeMS database to be fit for use? Not simple question

Some Causes of Data Quality Issues

Conceptualization errors

- Example: Raster Cell size issues
- Example: Vector representations of gradual change

Data Collection & Analysis Methods

- Example: GPS errors
- Example: Digitizing errors
- Example: Process model errors
- Example: Age analysis

Human Error

- Example: Blunders – typed 01.01.10 instead of 01.01.01

Temporal

- Example: Old Data

Elements of Data Quality

- **Positional Accuracy:** The accuracy of the position of features within a spatial reference system
- **Completeness:** The presence and absence of features, their attributes and relationships
- **Temporal Quality:** The quality of the temporal attributes and temporal relationships of features
- **Thematic Accuracy:** Classification correctness related to features and their attributes
- **Logical Consistency:** Adherence to logical rules of data structure, attribution, and relationships
- **Usability:** The data adhering to the user requirements for its intended use
- **Lineage:** What is the data source

Data Quality Basic Concepts

Error

A measured, observed, calculated, or interpreted value that differs from the true value or the value that would be obtained by a perfect observer using perfect equipment and perfect methods under perfect conditions.

Sources of

Inherent

Operational
(User/Process)

Types of

Random

Systematic

Blunder

lack of:

Accuracy

The degree to which a measured value conforms to true or accepted values. Accuracy is a measure of correctness. It is distinguished from precision, which measures exactness.

Precision

The closeness of a repeated set of observations of the same quantity to one another. Precision is a measure of the control over random error. For example, assessment of the quality of a surveyor's work is based in part on the precision of their measured values.

Uncertainty

The degree to which the measured value of some quantity is estimated to vary from the true value.

lack of:

Confidence

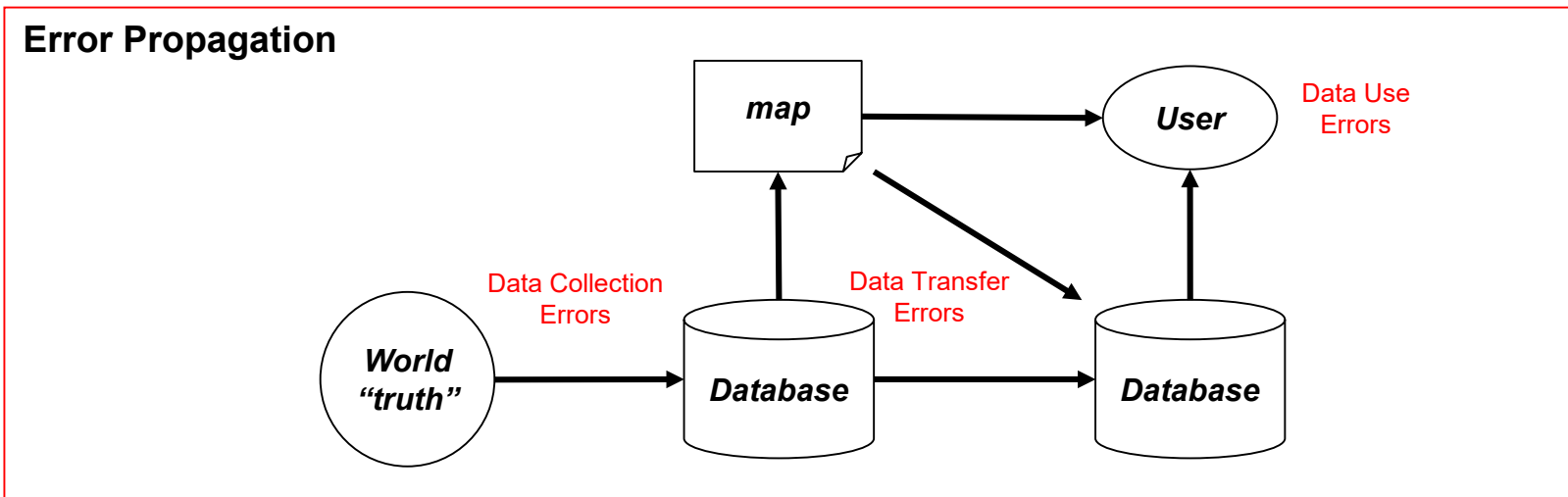
FGDC-STD-013-2006

Levels of SCIENTIFIC CONFIDENCE & LOCATIONAL ACCURACY

- “Discrete levels of scientific confidence and location accuracy have been developed to use as a terminology that can clearly yet concisely communicating the identity, existence, locatability, and position of geologic map features.”
- “These level are directly derived from, or are closely associated with, the feature attributes.”

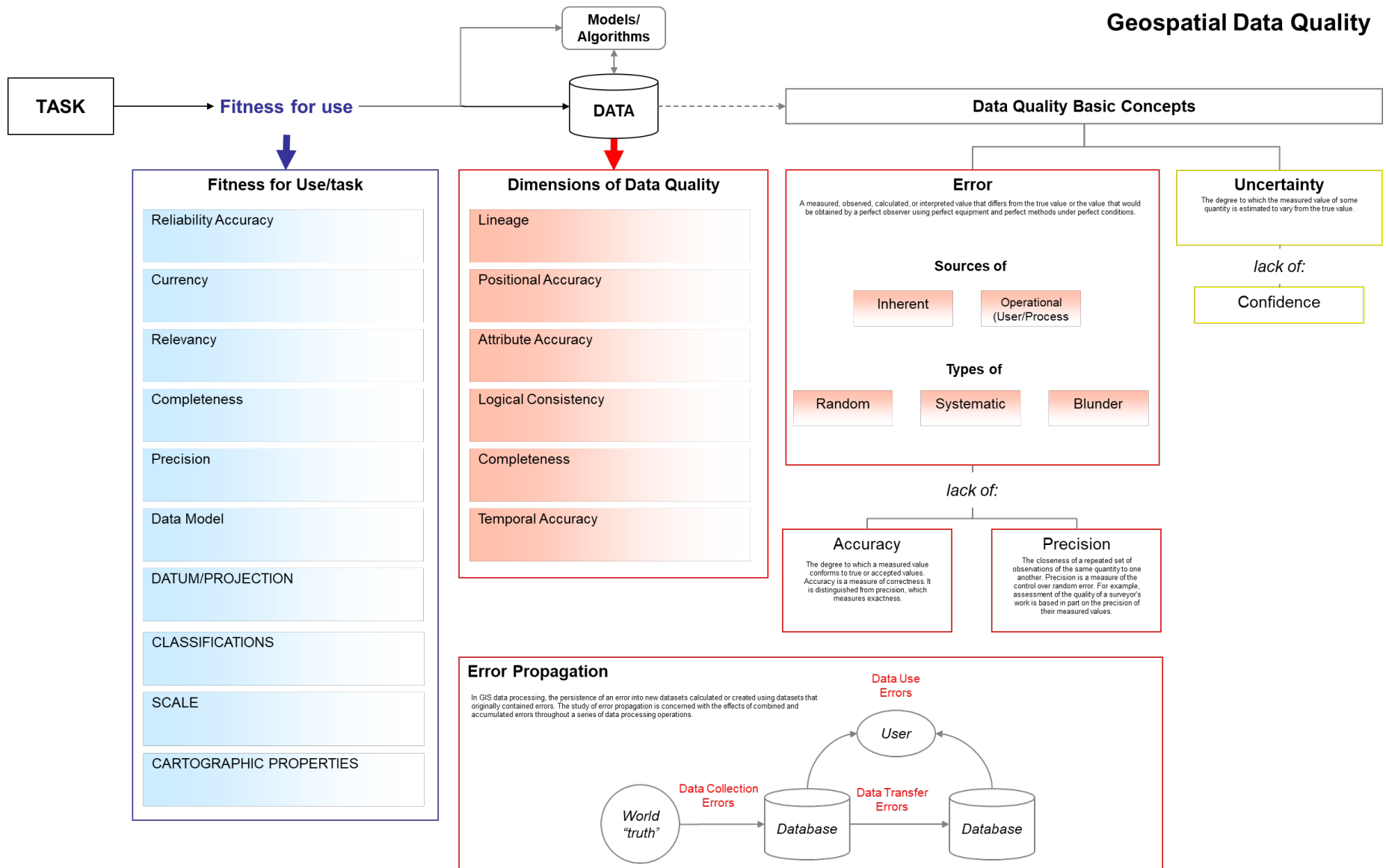
Error Propagation

In GIS data processing, the persistence of an error into new datasets calculated or created using datasets & maps that originally contained errors. *The study of error propagation is concerned with the effects of combined and accumulated errors throughout a series of data processing operations.*



Geospatial Data Quality

Geospatial Data Quality



Agenda

- Geospatial Data Quality Fundamentals
- **Quality Assurance (QA) & Quality Control (QC)**
- QA/QC in the Alaska DGGs Geologic Mapping System
- Data Reviewer

QA versus QC

- Quality Assurance – Processes or methods to help prevent errors from being introduced in the data.
 - Data models – GeMS, AK GeMS
 - Domains
 - Attribute Rules, etc. (ArcPro only)
 - Feature Templates
 - Editing Tools
 - Established Procedures, Documentation, and Training
- Quality Control – Processes or tools to identify errors that are already in the data.
 - Visual
 - GP Tools
 - Data Reviewer
 - USGS GeMS Validation Tool(s)
 - Custom Python Tools

Develop & Document a QA/QC Workflow

- Plan
 - Get organizational buy in
 - Understand resources are required
- Develop a QA/QC Plan
 - Purpose and Scope
 - Roles and Responsibilities
 - Testing Environment
 - Requirements and Acceptance Criteria
 - QC Workflow, Process and Tools
- Metadata Integration

Implementing a Quality Assurance plan

- Develop and use a stable well-documented Data Model
 - Domains
 - Attribute Rules (Pro only)
- Establish Feature Editing Templates
- Consider creating custom editing tools
- Develop and “publish” data generation guide (documentation)
- Training

Implementing a Quality Control plan

- Manual Data Validation
 - Establish and Use a Check list
 - Visual
 - Geoprocessing Tools
- Automated Data Validation
 - Custom Scripts &/or models
 - Data Reviewer
 - GeMS validation tool
- When to perform QC
 - On-demand
 - Periodically
 - New/update Data

Agenda

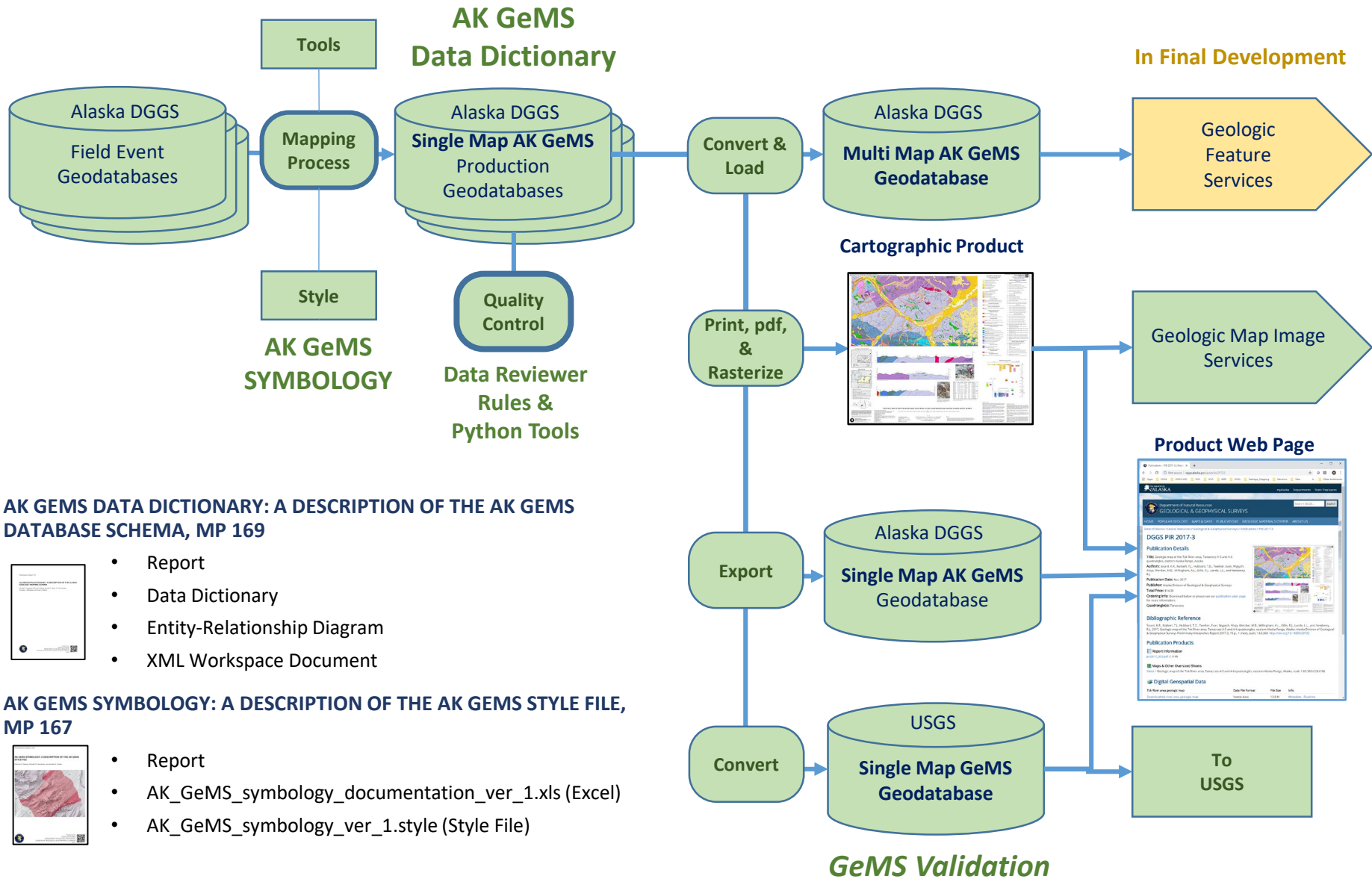
- Geospatial Data Quality Fundamentals
- Quality Assurance (QA) & Quality Control (QC)
- **QA/QC in the Alaska DGGS Geologic Mapping System**
- Data Reviewer

Some DMT AK DGGS Related Presentations

- **DMT 2021: Documentation of Alaska's geologic-GIS data management and delivery system**
By Jennifer Athey, Mike Hendricks, and Patricia Ekberg
- **DMT 2021: Symbolizing a GeMS geodatabase**
By Mike Hendricks, Patricia Ekberg, Jennifer Athey, and Amy Macpherson
- **DMT 2021: Digital review of GeMS-based products**
By Mike Hendricks, Jennifer Athey, and Amy Macpherson
- **DMT 2020-Lite: Customizing the GeMS Toolbox for local requirements**
By Mike Hendricks
- **DMT 2020: Lessons From Converting Alaska Digital Geologic Maps to the USGS Geologic Map Schema (GeMS)**
By Chris Wyatt, Mike Hendricks, Jennifer Athey, and Patricia Ekberg
- **DMT 2020: Alaska-GeMS Multi-map Database Schema Changes from the Federal GeMS Standard**
By Jennifer Athey, Mike Hendricks, Patricia Ekberg, and Chris Wyatt
- **DMT 2020: Improving Quality Control of a GeMS Database with ESRI's Data Reviewer**
By Mike Hendricks, Jennifer Athey, Patricia Ekberg, and Chris Wyatt

Alaska DGGS Geologic Mapping System Components

Organizational Procedures



AK GEMS DATA DICTIONARY: A DESCRIPTION OF THE AK GEMS DATABASE SCHEMA, MP 169



- Report
- Data Dictionary
- Entity-Relationship Diagram
- XML Workspace Document

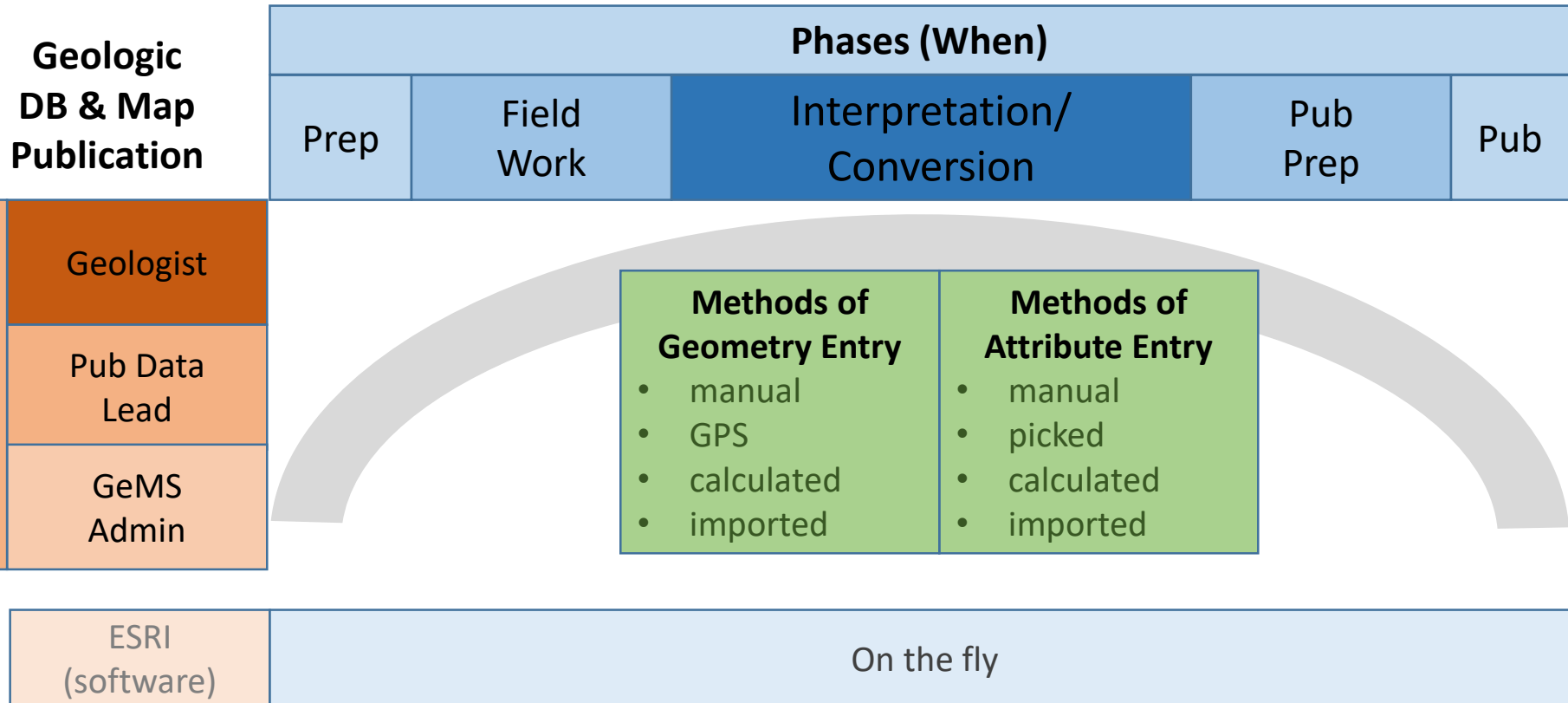
AK GEMS SYMBOLOGY: A DESCRIPTION OF THE AK GEMS STYLE FILE, MP 167



- Report
- AK_GeMS_symbology_documentation_ver_1.xls (Excel)
- AK_GeMS_symbology_ver_1.style (Style File)

GeMS Validation

AK GeMS Phases, Roles, and Methods

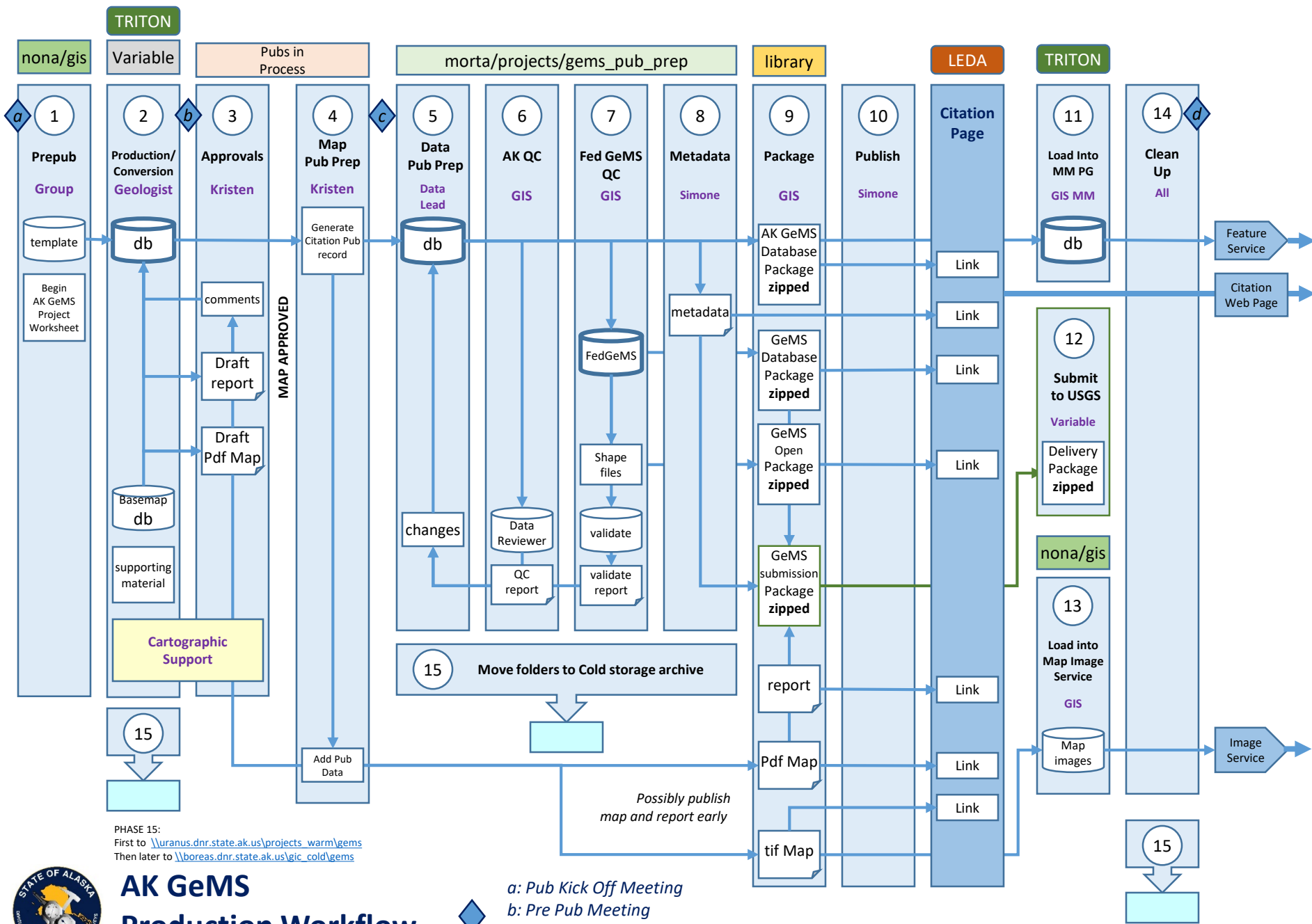


Example: map_unit_polys Documentation from Data Dictionary

Field	When	Who	How	Why	default value	Domain	data type
OBJECTID	on the fly	esri	calculated	software			OBJECTID
Shape	on the fly	esri	calculated	software			Geometry
map_unit_polys_id	pub prep	pub data lead	calculated				string: 50
layer	Interpretation	geologist	calculated		0		Long Integer
category	pub prep	pub data lead	calculated			map_unit_polys_category_dom	string: 50
type	Interpretation	geologist				map_unit_polys_type_dom	string: 254
symbol	pub prep	variable	calculated				string: 254
label	pub prep	pub data lead	calculated				string: 50
map_unit	interpretation	geologist	manual				string: 10
identity_confidence	interpretation	geologist	manual			identity_confidence_dom	string: 50
is_concealed	interpretation	geologist	manual			is_concealed_dom	string: 10
modifier	interpretation	geologist	manual	allows special characteristic of a unit, without making a new unit in the DMU		no domain, but suggested values of: hornfels, migmatite, mylonite, outcrop, shear, non, other, unspecified, unknown	
notes	interpretation	geologist	manual				string: 4096
data_sources_method	interpretation	geologist	manual			data_sources_method_dom	Long Integer
data_sources_id	interpretation	geologist	manual				string: 50
product_id	pub prep	pub data lead	calculated				string: 50
draw_policy	Interpretation	Geologist	manual	Controls if the feature is drawn on the map product	yes	draw_policy_dom	string: 10
distribution_policy	pub prep	pub data lead	manual		1	distribution_policy_dom	Long Integer

AK GeMS Production Workflow

- To improve QA/QC we developed and published a 15 phase workflow document
- Identifies:
 - Workflow
 - Responsibilities
 - Location of Data
 - Production Meetings
 - Products



PHASE 15:
 First to [\\uranus.dnr.state.ak.us/projects_warm/gems](https://uranus.dnr.state.ak.us/projects_warm/gems)
 Then later to [\\boreas.dnr.state.ak.us/gis_cold/gems](https://boreas.dnr.state.ak.us/gis_cold/gems)



AK GeMS Production Workflow 01 Sep 2021



a: Pub Kick Off Meeting
 b: Pre Pub Meeting
 c: Pub Data Handoff Meeting
 d: Post Pub Meeting



We are using MS Teams Planner App to track GeMS work

The screenshot displays the MS Teams Planner app interface for a project titled "DNR GeMS (ECC) Team / General". The interface is organized into columns representing different stages of the workflow, each with a "Add task" button at the top. The columns are:























- 1-2. In Production/In Conversion**: Contains tasks like "AK GeMS 1.1 New Mapping" (assigned to oimok) and "Early GeMS CONVERSION" (assigned to Kasatochi).
- 3. Map Approvals**: Contains "AK GeMS 1.1 New Mapping" (assigned to Styr).
- 4-5. Map & Data Pub Prep**: Contains "AK GeMS New Mapping" (assigned to Road Atlas) and "AK GeMS 1.1 CONVERSION" (assigned to livengood).
- 6 & 7. Quality Control (Mike)**: Contains "AK GeMS 1.1 CONVERSION" (assigned to DB Conversion).
- 8. Metadata (Simone)**: Contains "AK GeMS 1.1 New Mapping" (assigned to NE Tanacross) and "AK GeMS 1.1 CONVERSION" (assigned to caribou).
- 9. Packaging (Mike)**: No tasks are currently visible in this column.
- 10. Publish (Simone)**: Contains "AK GeMS 1.1 CONVERSION" (assigned to tanana_b1).
- 11. Load into MM DB (Pedro)**: No tasks are currently visible in this column.

Each task card provides details such as the task name, status (e.g., "Grant Deliverable", "New Mapping", "CONVERSION"), assignee, and due date. The interface also includes navigation options like "Group by Bucket", "Filter", "List", "Board", "Charts", and "Schedule".

QA/QC focused workflow phases

- ② Phase 2: Production (QA Focused) – Geologist/Map Maker
 - AK GeMS Schema, Style file
 - Topologies
 - Some Editing Templates
 - Custom Tools/Scripts (Limited)
- ③ Phase 3: Approvals (QC Focused) – Map & Report Centric
 - Manual/Visual
 - Peer Review
 - Internal Approvals
- ④ ⑤ Phases 4-5: Map & Data Pub Prep (QA Focused) – Data Lead/GIS
 - AK GeMS Schema, Style file
 - AK GeMS Custom Tools/Scripts
- ⑥ ⑦ Phases 6-7: AKGeMS/GeMS QC (QC Focused) – Data Centric
 - Semi-automated
 - Topologies
 - Data Reviewer
 - GeMS Validation Tool

AK GeMS Custom Tools/Scripts

- [-]  GeMS_production_tools_python_ArcMap.pyt
 - [-]  Phase 2 Production/Conversion
 -  Change Source Value
 -  Disable Editor Tracking
 -  Enable Editor Tracking
 - [-]  Phase 5 Data Pub Prep
 -  Build Pub Prep Folder
 -  Remove Empty Tables
 -  Remove Leading and Trailing Spaces from ALL string fields in DB
 -  Update Feature Extents
 -  Update geo_material_dict_id field in the DMU
 -  Update Location Confidence Fields
 -  Update map unit dmu_guid value
 -  Update map unit dmu_guid value for cross section
 -  Update map_unit_assoc field value
 -  Update Product id
 - [-]  Phase 6 QC
 -  Check orientation_confidence_meters field Values
 - [-]  Phase 9 Packaging
 -  Remove Editor Tracking Fields
 -  Save all layers in map
 -  Update Distribution Policy

Issues in our QA/QC implementation

- Need to improve QA/QC tools/checks to validate symbology with attributes & vice versa
- Need to complete transition to ArcPro for all operations
 - particularly data reviewer QC
 - Style file ArcPro
 - Tools all transitioned
- Need a few additional data reviewer rules
- Need a few additional python tools
- Improve documentation
 - Data Producer Guide Needed
 - Data Conversion Guide needs to be updated
- Continue training of all levels of personnel
 - Weekly GIS Tips and Tricks Class ongoing for 5+ years
 - Weekly GeMS meeting
- AK GeMS Schema/Template Updates (publish version 1.2 ~ summer 2022)
 - Update Domains
 - Attribute Rules
 - Contingent Values
 - New tables/feature classes? (very limited)

Agenda

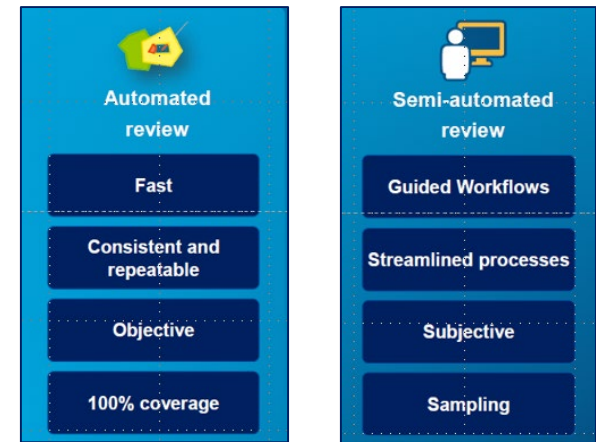
- Geospatial Data Quality Fundamentals
- Quality Assurance (QA) & Quality Control (QC)
- QA/QC in the Alaska DGGs Geologic Mapping System
- **Data Reviewer**

What is ArcGIS Data Reviewer?

Data Quality Management for ArcGIS

Provides

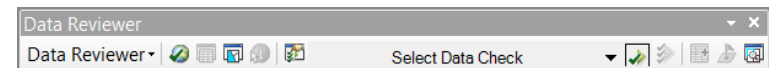
- **Rule-based automated validation**
- Semi-automated workflows
- **Track** and report quality



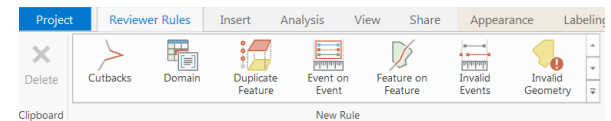
Execution Mechanisms

- Desktop ArcMap & **ArcGIS Pro**
- Models & python scripts
- Workflow manager
- As a Service via ArcGIS Server

Requires Installation of the Extension in ArcMap



Integrated into ArcPro



Our focus at this time in Bold

Data Reviewer Checks

~ 40 configurable data check to choose from

Methods

- Adhoc/on the fly validation
- Batch validation

Categories of Data Reviewer Checks

- Attribute Checks
- Event Checks
- Feature Integrity Checks
- Polygon Checks
- Polyline Checks
- Spatial Relationship Checks
- Z value checks

The screenshot displays the 'Data Reviewer for ArcGIS Pro Checks' interface, organized into several columns and rows of check categories. Each category includes a title, a brief description, and a small illustrative diagram. The categories shown are:

- EVENT CHECKS:** Check an Event, Invalid Events.
- POLYGON CHECKS:** Evaluate Polygon Perimeter and Area, Invalid Hole Features, Multiple Polygons, Polygon Overlap/Gap or Sliver, Polygon Ring Closed, Polygon Sliver, Unnecessary Polygon Boundaries.
- POLYLINE CHECKS:** Evaluate Polyline Length, Find Gaps, Multipartity, Multiple Line, Duplicates, Polyline or Path Closed on Self, Unnecessary Nodes.
- ATTRIBUTE CHECKS:** Domain, Query Attributes (Domain), Regular Expression, Subarea, Table to Table Attributes, Unique ID.
- FEATURE INTEGRITY CHECKS:** Collapsing, Duplicate Vertices, Evaluate Edges, Evaluate Part Count, Evaluate Vertex Count, Evaluate Z Values, Invalid Geometry, Nonlinear Segment, Self-intersecting, Snapping.
- GEODATABASE CHECKS:** Relationship, Topology Rules.
- SPATIAL RELATIONSHIP CHECKS:** Adjacent Vertices Elevation, Composite, Difference Z or Intersection, Duplicate Feature.

At the bottom right, there is a 'LEGEND' section with three items: 'REVIEWER BATCH JOBS', 'REVIEWER RULES IN A GEODATABASE', and 'REVIEWER RULES IN A MAP'. The Esri logo and 'THE SCIENCE OF INSIGHT' tagline are visible at the bottom right corner.

Our Development Process

2020-2021

- ESRI Online Course
- In the Spring of 2020 and Winter of 2021 DGGs Used Dedicated ESRI Advantage Program Credits to develop Data Reviewer Rules and Procedures
- Developed List of Rules
- Built and Testing Rule Files with Data Reviewer in ArcMap (.rbj)
- Developed Organizational Procedures

2022

- ArcPro 2.9 now includes vast majority of rules from ArcMap
- Spring 2022 use ESRI Advantage Program Credits for transition

Current Procedure to Check a Map Database

- Used near the END of the production process
- Build a separate dedicated data reviewer database (for results) for each map (as opposed to a single data reviewer database for the organization)
 - Questions: How should this database be archived?
- Run by the GIS Staff (for now)

Some of the AK DGGS Rules List






















Attribute Centric

- All values must meet database domain constraints
- The symbol code for a map unit must match that same map unit's symbol in the description_of_map_units table
- The feature label must match the correct value in the identity_confidence field and vice versa
- Contacts_and Faults features must only be split when key attributes change
- Map_units_polys features must only be split when key attributes change

Geometric Centric

- All features must have valid geometries
- Line features must not self intersect
- Contacts and Faults and Map Units must be single part features
- Curved segments for lines and polygons are not allowed
- Point feature classes that reference a station feature must be collated with that feature in the stations feature class
- Contracts and Faults must be on the boundary of map unit polygons
- Line features must be longer than 10 meters
- Polygon features must be larger than 100 square meters

Rule Files

Name	Size
 _AK_GeMS_All_Rules.rbj	640 KB
 AK_GeMS_check_data_sources_poly_has_related_data_source_record.rbj	11 KB
 AK_GeMS_check_if_category_is_other_then_type_is_other.rbj	92 KB
 AK_GeMS_check_map_unit_assoc_equal_intersecting_map_unit.rbj	36 KB
 AK_GeMS_check_map_unit_symbol_to_dmu.rbj	27 KB
 AK_GeMS_check_product_id_values.rbj	125 KB
 AK_GeMS_check_type_value_includes_category.rbj	97 KB
 AK_GeMS_Contact_and_fault_must_be_on_map_unit_poly_boundary.rbj	10 KB
 AK_GeMS_domain_checks.rbj	98 KB
 AK_GeMS_invalid_geometry.rbj	88 KB
 AK_GeMS_non_linear_check.rbj	57 KB
 AK_GeMS_other_value_has_notes.rbj	101 KB
 AK_GeMS_point_not_colocated_with_station.rbj	29 KB
 AK_GeMS_polyline_closes_on_self_check.rbj	29 KB
 AK_GeMS_segment_length_greater_than_1km_check.rbj	7 KB
 AK_GeMS_short_lines.rbj	41 KB
 AK_GeMS_single_part_check.rbj	19 KB
 AK_GeMS_small_polygons.rbj	34 KB
 AK_GeMS_uncertain_identity_confidence_has_question_mark_label.rbj	18 KB
 AK_GeMS_unnecessary_node_check.rbj	7 KB
 AK_GeMS_unnecessary_polygon_boundary_check.rbj	8 KB

\\nona.dnr.state.ak.us\gis\standards\GeMS\Data_Reviewer

Some references

- ESIRUC 2019 - ESRI Best Practices: QA/QC for Your GIS Data
<https://www.esri.com/content/dam/esrisites/en-us/about/events/media/UC-2019/technical-workshops/tw-6391-1016.pdf>
- ISO 19157:2013 Geographic information — Data quality
<https://www.iso.org/standard/32575.html>
- FGDC Digital Cartographic Standard for Geologic Map Symbolization
https://ngmdb.usgs.gov/fgdc_gds/geolsymstd/download.php
- GeMS (Geologic Map Schema)—A Standard Format for the Digital Publication of Geologic Maps
<https://pubs.er.usgs.gov/publication/tm11B10>
- Introduction to attribute rules <https://pro.arcgis.com/en/pro-app/latest/help/data/geodatabases/overview/an-overview-of-attribute-rules.htm>