



Supplemental F

National Heritage Solution L16PA00033

DATA CONVERSION DOCUMENT

Version 2.0

**United States Department of the Interior
Bureau of Land Management
National Operations Center
Division of Resource Services
Denver Federal Center
Denver, Colorado 80225**



Document Information

General Information

Project Name	National Heritage Solution, Phase 2		
Prepared By	Maggie Satterfield	Document Version	2.0
Title	Senior Analyst	Document Version Date	03/27/2020
Reviewed By	Maggie Satterfield	Review Date	03/27/2020

History

Ver. No.	Ver. Date	Revised By	Description	Filename
1.0	12/18/2018	Nicole Wigston	Initial document, including details for Idaho & Oregon	BLM NHS Data Management Plan v1
1.1	12/19/2018	Maggie Satterfield	Additional material to script logic	BLM NHS Data Management Plan v1.1
1.2	4/4/2019	Maggie Satterfield	Overview information; state-specific details; scoring system implemented	BLM NHS Data Management Plan v1.2
1.3	4/26/2019	Maggie Satterfield	State-specific details; scoring scale updated	BLM NHS Data Management Plan v1.3
1.3.1	4/30/2019	Maggie Satterfield	Full review and update of state-specific details; scoring system details updated	BLM NHS Data Management Plan v1.3.1
1.3.1	5/1/2019	Nicole Wigston	Edited for clarity	BLM NHS Data Management Plan v1.3.1
1.4	5/27/2019	Maggie Satterfield	State-specific details	BLM NHS Data Conversion Document v1.4
1.5	6/13/2019	Maggie Satterfield	Add details to Global section, review existing content for accuracy	BLM NHS Data Conversion Document v1.5
1.5.1	6/14/19	Nicole Wigston	Editorial	BLM NHS Data Conversion Document v1.5.1
2.0	3/27/2020	Maggie Satterfield	Updated all state details, appendices, and toolbox overview for 2019 ETL cycle.	BLM NHS Data Conversion Document v2.0

Distribution List

From	Date	Contact Information
Nicole Wigston	12/21/18, Version 1.1	nwigston@quantumspatial.com
Nicole Wigston	4/08/19, Version 1.2	
Nicole Wigston	4/29/19, Version 1.3	



Nicole Wigston	5/1/19, Version 1.3.1	
Maggie Satterfield	5/28/19, Version 1.4	msatterfield@quantumspatial.com
Maggie Satterfield	3/27/20, Version 2.0	

To	Date	Version	Contact Information
Kirk Halford and/or Katy Carpenter	12/21/19	1.1	fhalford@blm.gov
Kirk Halford and Amy Fehringer-Schopp	4/15/19	1.2	
Kirk Halford and Amy Fehringer-Schopp	4/29/19	1.3	
Kirk Halford and Amy Fehringer-Schopp	5/1/19	1.3.1	
Kirk Halford	5/28/19	1.4	
Kirk Halford and Amy Fehringer-Schopp	6/14/19	1.5.1	
Kirk Halford	3/27/20	2.0	



Purpose of Data Conversion Document

This document describes the Extract-Transform-Load (ETL) processes and approach specific to each contributing State. A version of this document will accompany each ETL tool release, corresponding to the implementation, as the logic is developed for each State. The final version of this document will include any ongoing maintenance requirements and procedures.

Table of Contents

Document Information	2
General Information	2
History	2
Distribution List	2
1. Global Overview	8
1.1 State Scores of Delivered Data and Recommendations	8
2. NCRIMS ETL Process	10
2.1 NCRIMS ETL Script Tool	10
2.2 Required Directory Structure	11
2.2.1 BLM Surface Management Areas	12
2.3 Validation Processes	12
2.3.1 Geometry Validation	12
2.3.2 Duplicate Identifiers	13
2.3.3 Orphaned Domain Values	13
2.3.4 Data Migration Logs	14
2.4 Use of Crosswalks	14
2.4.1 Schema Crosswalk	14
2.4.2 Domain Crosswalk	15
3. Quality Control Methods	16
4. Compiling States to a National Dataset	17
4.1 Moving Forward	17
5. 2019 State Observations and Recommendations	18
5.1 Alaska	18
<i>Recommendations for Future Deliverables (2)</i>	19
5.2 Arizona	20



<i>Recommendations for Future Deliverables (5)</i>	20
5.3 California	22
<i>Recommendations for Future Deliverables (5)</i>	23
5.4 Colorado	25
<i>Recommendations for Future Deliverables (2)</i>	25
5.5 Idaho	26
<i>Recommendations for Future Deliverable (4)</i>	26
5.6 Montana	28
<i>Recommendations for Future Deliverable (4)</i>	28
5.7 Nevada	30
<i>Recommendations for Future Deliverable (4)</i>	30
5.8 New Mexico	32
<i>Recommendations for Future Deliverable (4)</i>	32
5.9 Oregon	34
<i>Recommendations for Future Deliverable (4)</i>	34
5.10 Utah	36
<i>Recommendations for Future Deliverable (5)</i>	36
5.11 Wyoming	38
<i>Recommendations for Future Deliverables (5)</i>	38
APPENDIX A: 2019 State Toolsets and Workflows	40
Alaska	40
General Workflow and Tool Parameters	40
Arizona	42
General Workflow and Tool Parameters	42
California	44
General Workflow and Tool Parameters	44
Colorado	46
General Workflow and Tool Parameters	46
Idaho	48
General Workflow and Tool Parameters	48
Montana	49
General Workflow and Tool Parameters	49
Nevada	51
General Workflow and Tool Parameters	51



New Mexico	54
General Workflow and Tool Parameters	54
Oregon	57
General Workflow and Tool Parameters	57
Utah	59
General Workflow and Tool Parameters	59
Wyoming	61
General Workflow and Tool Parameters	61
APPENDIX B: 2019 Data Directory Structure	65
APPENDIX C: 2018 – 2019 Data Resolutions	66
Alaska 2018 Score and Recommendations	66
<i>Recommendations for Future Deliverables</i>	66
<i>2019 Resolution</i>	68
Arizona 2018 Score and Recommendations	68
<i>Recommendations for Future Deliverables</i>	68
<i>2019 Resolution</i>	69
California 2018 Score and Recommendations	69
<i>Recommendations for Future Deliverables</i>	70
<i>2019 Resolution</i>	71
Colorado 2018 Score and Recommendations	71
<i>Recommendations for Future Deliverables</i>	71
<i>2019 Resolution</i>	72
Idaho 2018 Score and Recommendations	72
<i>Recommendations for Future Deliverable</i>	73
<i>2019 Resolution</i>	74
Montana 2018 Score and Recommendations	74
<i>Recommendations for Future Deliverable:</i>	74
<i>2019 Resolution</i>	76
Nevada 2018 Score and Recommendations	76
<i>Recommendations for Future Deliverable</i>	76
<i>2019 Resolution</i>	77
New Mexico 2018 Score and Recommendations	78
<i>Recommendations for Future Deliverable</i>	78
<i>2019 Resolution</i>	80



Oregon 2018 Score and Recommendations	80
<i>Recommendations for Future Deliverable</i>	80
<i>2019 Resolution</i>	82
Utah 2018 Score and Recommendations	82
<i>Recommendations for Future Deliverables</i>	82
<i>2019 Resolution</i>	83
Wyoming 2018 Score and Recommendations	83
<i>Recommendations for Future Deliverables</i>	83
<i>2019 Resolution</i>	85



1. Global Overview

The Cultural Resource Management geospatial dataset schema was developed to standardize state-level cultural resource data. The purpose of this dataset is to provide “seamless information for high-level landscape analysis” (CRM Implementation Guidelines, 2018). This analysis is accessible through the BLM National Heritage web map application.

The purpose of the Extract, Transform, and Load (ETL) processes is to migrate data from state-level schemas to the CRM National schema. This migration is targeted to occur annually (September). After migration, the states’ data must then be aggregated to the one National CRM dataset.

The aggregated dataset is to be published as a feature service visible within the BLM National Heritage web map application. This dataset includes five fields from the CRM Standard Schema that are critical to the Sensitivity Analysis reporting tool within the NCRIMS web map application. These five key fields are:

1. Investigation Class - INVSTGTN_CL
2. Resource Temporal Cultural Assignment - RSRCE_TMPRL_CLTRL_ASGNMNT
3. Resource Primary Property Class - RSRCE_PRMRY_PPRTY_CL
4. Resource Primary Category Name - RSRCE_PRMRY_CTGRY_NM
5. Resource NRHP Eligibility Status - RSRCE_NRHP_ELGBLE_STTS

1.1 State Scores of Delivered Data and Recommendations

The state deliverables were rated using numerical values -1, 0, and 1. The purpose of this system is to provide a quantifiable status intended to guide future data releases and improve consistency and compatibility of state deliverables to the CRM standard. This compatibility and consistency is important to the efficacy of ETL process in producing a more accurate National Dataset. This rating system applies to all states.

The deliverables were rated against 3 categories - Organization, Data Input, and Maintenance. Any state deliverables with a negative Total Score should be considered a priority by the BLM for improving compatibility with the CRM standard. Categories with a negative total score contain components which posed critical challenges for an automated ETL process of the corresponding state’s data. Scores may range from -3 to 3, with 3 being the best score.

The Organization category contains 2 criteria:

1. The data was delivered in a file geodatabase. The use of shapefiles is not recommended for deliveries due to storage constraints. It is recommended that if tables are included in the data release, they be included inside the file geodatabase. A file geodatabase provides a consistent delivery mechanism for state data and is considered a best practice as it is optimized and scalable (Childs, 2009).¹

¹ Childs, Colin. The Top Nine Reasons to Use a File Geodatabase. ArcUser (2009). Retrieved from: <https://www.esri.com/news/arcuser/0309/files/9reasons.pdf>

2. The file geodatabase contained one feature class for Resource features and one feature class for Investigation features. Some states may store data according to geometry. Before execution of the ETL, these geometries must be transformed and merged into a single, usable geometry for the national database. Additionally, other states may store feature information in multiple feature classes based on status. This introduces multiple schemas which need to be in agreeance before the automate ETL is executed.

The Data Input category contains 4 criteria:

1. The data did not contain redundant fields. Redundant fields store the same information and can cause confusion during migration, especially if discrepancies exist between the crosswalk and delivered data.
2. The data contained a date field or an alternative field to track the most recent visit to a site. A date field or a system of tracking when the record was input or when the site was visited is recommended, especially if legacy records exist in the data. This helps narrow down the records to the most recent available.
3. The data was complete across the CRM fields. It is recommended that states maintain data fields which exist in the CRM schema. States that do so are considered complete across the CRM fields. This improves the effectiveness of the National Dataset.
4. The data contained normalized/ consistent values. Normalized values are extremely important for an effective ETL. Normalized values are considered values which fall within a domain or range of values. For example, consider a NRHP eligibility field. If the field contains multiple versions of the value “No” (i.e. “No”, “nO”, “n0”) then the values are not consistent. Likewise, if multiple values exist for ineligibility (i.e. “No”, “Not Eligible”, “Ineligible”) then the field is not normalized.

The Maintenance category also contains 4 criteria:

1. The data did not contain legacy or duplicate records. Legacy records are considered as any record of a previous visit to the site or survey which has not been removed from the data delivery. Duplicate records are considered records which appear in the dataset more than once and contain the same information and geometry.
2. The legacy or duplicate records were able to be queried and easily removed (if they exist). Often times, without a tracking field, it is difficult to determine if the record is multipart, legacy, or truly duplicate. This is especially true if two records with the same identifier are located on top of one another.
3. Each record in the delivered dataset contained an identifier, generally the SHPO ID. This allows a user to contact the SHPO office with a reference in case more information is needed. If a SHPO ID is not available for the record, then an Agency ID should be available. These identifiers also assist the BLM in creating and maintaining the CRM ID fields. Features without an identifier will not be included in the National Dataset.
4. Features did not exhibit geometry or topology issues. Severe issues cause errors in geoprocessing during the ETL processes. These issues also may negatively impact the analytical and reporting services in the web application.

The category score is the sum of the criteria ratings. The Total Score is the sum of the category scores. Any criteria met received a rating of 1. Any criteria that was not met but did not pose a critical issue for execution of the ETL received a rating of 0. Any criteria which is not met and

posed critical issues for execution of the ETL received a rating of -1. Ratings and corresponding recommendations are included in Section 5: State Details. For collective view of the scores, see Section 5.

2. NCRIMS ETL Process

Each State contributed unique datasets that were migrated to the National CRM standard as part of this Task. Crosswalks were used to map the input state schema to the output CRM dataset and a script tool was written to perform the migration. The ArcGIS ArcToolbox titled, NCRIMS Data Migration, contains a single tool applicable to each state. The tool includes options to configure the schema crosswalk, configure the domain crosswalks, and perform the migration.

This tool is intended to be run in ArcMap with an ArcGIS Desktop Standard license or higher for Windows OS. All crosswalks produced as output from this tool are intended to be populated in ArcMap. Additionally, this tool references the Python 2.7 library included with the ArcGIS Desktop installation. It does not call any third-party libraries and runs in the background of an ArcMap session. Last, the script corresponds to a specific folder structure which currently mimics the structure used by the BLM.



The script corresponds to a specific folder structure which currently mimics the structure used by the BLM. ***Any changes to this structure may result in errors.***

2.1 NCRIMS ETL Script Tool

Seven input parameters exist for this tool (Image 1.)

1. Process Type – (*string*) a dropdown containing three values (Configure schema crosswalk, Configure domain crosswalks, Migrate Values) which designate the procedure to be executed on the input layers.
2. Did the state require a domain crosswalk? – (*Boolean*) a checkbox which should be utilized with the Migrate Values procedure if a domain crosswalk was configured for the state deliverable. This generally is relevant to states which are not using the CRM schema.
3. Config GDB – (*workspace*) this parameter is auto-populated and points to the crosswalk_config.gdb schema in the ETLtool folder (see Section 2.2 for more information). This is required for creating the schema and domain crosswalks.
4. Original Delivery Folder – (*folder*) the folder on the BLM network that houses the data delivery. See Section 2.2 for more information.
5. Input Investigations – (*feature class*) the investigation feature class(es) included in the delivery. This parameter is not required for successful execution as not all states maintain both an investigation and resource feature class.
6. Input Resources – (*feature class*) the resource feature class(es) included in the delivery. This parameter is not required for successful execution as not all states maintain both an investigation and resource feature class.



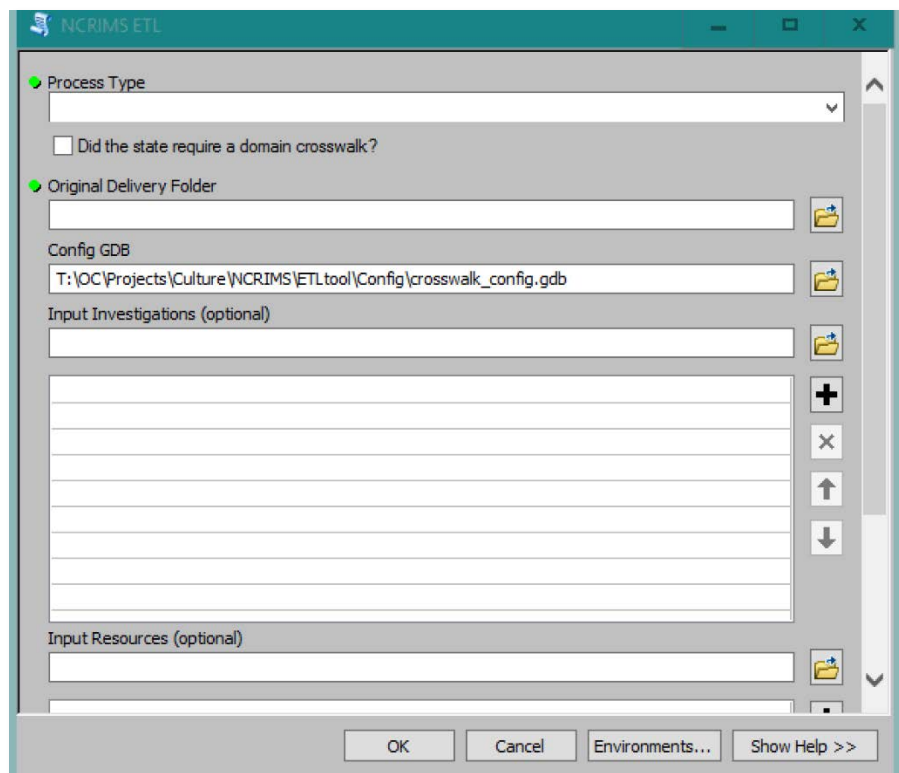


Image 1: NCRIMS ETL tool user interface

2.2 Required Directory Structure

Design, development, and testing of the ArcToolbox Scripts used file geodatabases. Before execution of the scripts, ensure state delivered data is stored in a single file geodatabase. Any actions taken to ensure the data was in an appropriate format before execution can be found in Appendix A.

Each data delivery should be stored in a separate directory inside of a state folder, titled as the abbreviation. For example, the 2019 Idaho delivery was stored inside a folder called “OriginalDelivery”. The “OriginalDelivery” folder was stored inside of a root folder titled “ID”. A “DataMigration” folder and “Staging” folder was also created within the root folder. The root-child directory structure along with the abbreviated state title of the root folder was critical to the ETL. This complete file structure is shown in Appendix B.

The ETL requires the BLM Surface Management Areas (SMA) in order to calculate the BLM Acres of each feature. The path to this data was coded in relation to the script path, thus it is critical that this data be stored similarly if the tool is to be executed in a separate environment. Additionally, geoprocessing was performed on the SMA layer to create a buffered dataset. This buffered dataset was utilized for states which delivered all features, regardless of which federal lands the feature intersected.

For a complete overview of these structures, see Appendix B.

2.2.1 BLM Surface Management Areas

The BLM Surface Management Areas (SMA) were required for the 2019 ETL process. This feature class also underwent geoprocessing before the completion of the ETL. The script referenced the output names included in the steps below. Changing folder, geodatabase, and outputs names will result in errors.

1. Download the BLM SMA polygons from:
<https://landscape.blm.gov/geoportal/catalog/search/resource/details.page?uuid=%7B2A8B8906-7711-4AF7-9510-C6C7FD991177%7D>
2. Extract the downloaded zip folder and copy the geodatabase to this location:
 ...EGIS\OC\Projects\Culture\NCRIMS\ETLtool\BLM_SMA
3. Ensure the file geodatabase is named *sma_wm.gdb*. If it is not, rename it. Delete all feature classes except for *SurfaceMgtAgy_BLM*.
4. In ArcMap, run the Repair Geometry tool on the *SurfaceMgtAgy_BLM* feature class.
5. In ArcMap, use the Data Management > Projections and Transformations > Project tool to reproject the BLM SMA feature class. The parameters are below.
 - a. Input Dataset or Feature Class: *sma_wm.gdb\SurfaceMgtAgy_BLM*
 - b. Output Dataset or Feature Class: *sma_wm.gdb\SurfaceMgtAgy_BLM_projected*
 - c. Output Coordinate System: *North_America_Albers_Equal_Area_Conic (WKID: 102008)*
 - d. Geographic Transformation: *Leave default*
6. In ArcMap, use the Analysis > Proximity > Buffer tool to buffer the projected SMA (output from step 4) 100 meters. Name the output *SurfaceMgtAgy_BLM_100m_buffer*



Changing folder, geodatabase, and outputs names in this workflow will result in errors.

2.3 Validation Processes

The NCRIMS ETL includes three validations with corresponding output logs to assist the user with data quality control. These validations include a geometry validation, duplicate identifier validation, and orphaned domains validation. The output logs were placed in the DataMigration folder automatically. See Appendix B for the location of the DataMigration folder.

2.3.1 Geometry Validation

This validation outputs a csv file logging the OID and geometric issue of the feature. This logging process is fully automated and occurs on every state delivery. This validation logs any features with:

1. Short segment—Some segments are shorter than allowed by the system units of the spatial reference associated with the geometry.
2. Null geometry—The feature has no geometry or nothing in the SHAPE field.
3. Incorrect ring ordering—The polygon is topologically simple, but its rings may not be oriented correctly (outer rings clockwise; inner rings counterclockwise).
4. Incorrect segment orientation—Individual segments are not consistently oriented. The to point of segment *i* should be incident on the from point of segment *i+1*.
5. Self intersections—A polygon must not intersect itself.



6. Unclosed rings—The last segment in a ring must have its to-point incident on the from-point of the first segment.
7. Empty parts—The geometry has multiple parts and one of them is empty (has no geometry).
8. Duplicate vertex—The geometry has two or more vertices with identical coordinates.
9. Mismatched attributes—The z- or m-coordinate of a line segment's endpoint does not match the z- or m-coordinate of the coincident endpoint on the next segment.
10. Discontinuous parts—One of the geometry's parts is made up of disconnected or discontinuous parts.
11. Empty Z values—The geometry has one or more vertices with an empty z-value (NaN, for example).
12. Bad envelope—The envelope does not match the coordinate extent of the geometry.
13. Bad dataset extent—The extent property of the dataset does not contain all of the features in the dataset. For this problem the FEATURE_ID will be -1.²

If any of the above issues are detected, the script returns a warning message alerting the user the geometry problems were detected. Please note, features with null geometries will not be migrated into the National Dataset. These features pose significant risk to the functionality of geoprocessing services.

2.3.2 Duplicate Identifiers

This validation outputs a csv which logs any SHPO Identifiers (SHPO IDs) that have occurred in the data more than once. If duplicated SHPO IDs are detected, a warning message appears alerting the user. Upon receiving the warning, these duplicate IDs were reviewed to determine if they are truly duplicate, legacy records, or multipart records.

2.3.3 Orphaned Domain Values

This validation outputs a csv which logs values that do not abide by the domain-restricted values (orphaned domain values). This validation applies only to the CRM fields that are domain-restricted. A list of these fields is below.

- INVSTGTN_CL
- INVSTGTN_AUTH
- RSRCE_TMPRL_CLTRL_ASGNMNT
- RSRCE_PRMRY_PRPTY_CL
- RSRCE_PRMRY_CTGRY_NM
- RSRCE_NRHP_ELGBLE_STTS
- RSRCE_NRHP_ELGBLE_CRTRA
- RSRCE_NRHP_ELGBLE_AUTH_NM
- RSRCE_CNDTN_ASSMNT
- RSRCE_CLCTN_PRFRM_STTS
- RSRCE_SPTL_CLCTN_MTHD

If orphaned domains are detected, a warning message appears alerting the user. Upon receiving the warning, the values were reviewed then queried and corrected in the final dataset.

² ArcGIS Pro Tool Reference- Check Geometry. ESRI (2018). Retrieved from <https://pro.arcgis.com/en/pro-app/tool-reference/data-management/check-geometry.htm>

2.3.4 Data Migration Logs

In addition to validations, the script also produces logs which record information pertaining to the migration processes. Such logs include:

1. Schema Crosswalks – This folder stores geodatabases containing the configured schema crosswalk tables.
2. Domain Crosswalks – This folder stores geodatabases containing the configured domain crosswalk tables.
3. Error Logs – This folder stores a date-stamped text file. The text file contains any error messages encountered during the migration process. Error messages cause the script to exit and no longer execute.

2.4 Use of Crosswalks

2.4.1 Schema Crosswalk

The crosswalks received from each state, typically in Microsoft Excel (.xls) format, were referenced by the script user. A configuration method was created to guide the crosswalk of values from state fields to CRM standard fields. This configuration method appropriately matches state fields to CRM fields.

The configuration method references a file geodatabase which was stored in the parent directory of the NCRIMS script tool. This geodatabase consisted of two tables - a table containing the CRM Resource fields and a table containing the CRM Investigation fields. During the 2019 cycle, these were edited by selecting the appropriate state field name in the dropdown. If more than one feature class was required for input to a CRM feature class, separate columns were created to store the schema crosswalk information. If more than one field was required for population of a CRM field, the field names were typed in the cell, separated by a semi-colon (Image 2). More information about how these were used in each state ETL can be found in Appendix A.

Table				
INVSTGTNcrosswalk				
OBJECTID *	NCRIMS_standard	Archaeo_Inventory	In_Process_Inventory	Rural_Archit_Inventory
1	CRM_INVSTGTN_ID	<Null>	<Null>	<Null>
2	INVSTGTN_AGCY_ID	<Null>	LEAD_AGENCY_NUMBER	LEAD_AGENCY
3	INVSTGTN_SHPO_ID	AUTHOR	REPORT_NUMBER	REPORT_NUMBER
4	INVSTGTN_CMPLT_MONTH_YR	CLASS_II_ACREAGE	REPORT_DATE	REPORT_DATE
5	INVSTGTN_DATE	CLASS_II_ACREAGE	FIELDWORK_END	SURVEY_YEAR
6	INVSTGTN_LEAD_BLM_ADMIN_ST	COMMENTS	<Null>	<Null>
7	INVSTGTN_TITLE	DATE_DIGITIZED	TITLE	TITLE
8	INVSTGTN_AUTH	DIGITIZED_BY	INVSTGTN_ATHRTY	INVSTGTN_ATHRTY
9	INVSTGTN_CL	FIELDWORK_END	CLASS_II_ACREAGE;CLASS_III_ACREAGE	SURVEY_TYPE
10	INVSTGTN_PRFRM_PARTY_NM	FIELDWORK_START	LEAD_AGENCY	LEAD_AGENCY
11	INVSTGTN_NEPA_ID	FORM_SOURCE	INVSTGTN_NEPA_ID	INVSTGTN_NEPA_ID
12	INVSTGTN_DATA_SRCE	GlobalID	<Null>	<Null>
13	INVSTGTN_CMT	INVSTGTN_ATHRTY	<Null>	COMMENTS
14	ADMIN_ST	INVSTGTN_NEPA_ID	<Null>	<Null>
15	GIS_ACRES	LEAD_AGENCY	<Null>	<Null>
16	BLM_ACRES	LEAD_AGENCY_NUMBER	<Null>	<Null>
17	CREATE_DATE	LINEAR_INVENTORY	<Null>	<Null>
18	CREATE_BY	NO_FINDS	<Null>	<Null>
		NSM_NUMBER	<Null>	<Null>
		OBJECTID	<Null>	<Null>
		OTHER_ACREAGE		
		PROONENT		
		REPORT_DATE		
		REPORT_NUMBER		
		REPORT_NUMBER_2		
		REPORT_YEAR		
		SHAPE		
		SHAPE_Area		
		SHAPE_Length		
		SHAPE_STArea__		
		SHAPE_STLength		

Image 2. An example of a schema crosswalk that required multiple feature classes as input and multiple field names.

2.4.2 Domain Crosswalk

Similar to the Schema Crosswalk tool, an additional configuration method was created to control the migration of values to domain values. This is generally the second step in the ETL process for most states. More information about how these were used in each state ETL can be found in Appendix A.

The configuration method references file geodatabases which store frequency tables with an additional field called NCRIMS_Value. During the 2019 cycle, the NCRIMS_Value field was edited by typing the appropriate CRM value in the empty cell. If more than one state field was utilized in the schema crosswalk, a column existed for each field (Image 3). The script referenced these tables to control the crosswalk of state values to CRM domain values. After these tables were populated, the Migrate Values procedure was executed.

RSRCE_NRHP_ELGBLE_STTS						
	OBJECTID *	FREQUENCY	FIELD_NRHP	LEAD_NRHP	SHPO_NRHP	NCRIMS_Value
▶	1	32	<Null>	<Null>	<Null>	Undetermined
	2	1	<Null>	<Null>	No Info	Undetermined
	3	1	<Null>	<Null>	Unevaluated	Undetermined
	4	1	<Null>	Ineligible	Ineligible	No
	5	1	<Null>	Ineligible	Unevaluated	No
	6	129	Eligible	<Null>	<Null>	Yes
	7	40	Eligible	<Null>	Default	Yes
	8	28	Eligible	<Null>	Eligible	Yes
	9	31	Eligible	<Null>	No Info	Yes
	10	27	Eligible	<Null>	Unevaluated	Yes
	11	58	Eligible	Eligible	<Null>	Yes
	12	9	Eligible	Eligible	Default	Yes
	13	108	Eligible	Eligible	Eligible	Yes
	14	2	Eligible	Eligible	No Info	Yes
	15	2	Eligible	Eligible	Not Reviewed for Sec. 106	Yes

Image 3. An example of a domain crosswalk where multiple fields were utilized in the schema crosswalk

3. Quality Control Methods

During the 2019 cycle, processes were conducted to check the data quality of the National Dataset. These processes (listed below) were conducted either before or after the state/SHPO data was added to National Dataset. Additionally, the percentage of complete information provided by each state/SHPO was calculated.

1. Verify no features were lost in the migration – This was accomplished by ensuring the number of features in the state’s staging dataset matched the number of features in the feature class(es) used as output. In some cases, a feature had null geometry. These feature OIDs were logged in the geometry validation log and were not included in the staging dataset. See Section 4 for information on the staging datasets.
2. Verify the data in each staging field is logical – If a field such as Investigation Title showed dates, this was not considered a logical match for the crosswalk. This was reconciled by viewing the delivered data and choosing a field which captured the Investigation Titles more accurately.
3. Orphaned domain values are reconciled – The NCRIMS script tool warned and logged any values under a domain-restricted field which did not agree with domain coded values. These were reconciled using Select by Attributes and Field Calculator in ArcMap.
4. Geometric issues are reconciled – The NCRIMS script tool warned and logged and feature OIDs which had geometry that did not abide by ESRI’s recommended best practices for feature geometry (see Section 2.3.1). These were reconciled by running the Repair Geometry tool on the feature class(es) in question. See Appendix A for details on this process as it applied to each state.
5. Duplicate IDs are reconciled – In most cases, this was not easily accomplished for various reasons, hence this quality check was unable to be fulfilled. However, these duplicate IDs were logged for each state in the DataMigration folder and reconciled

where possible. See Section 2.3.2 for information on this validation process and Appendix A for details on this process as it applied to each state.

6. All features in the national dataset contain an identifier – This was accomplished by deleting any feature in the National Dataset that did not have a SHPO or Agency Identifier.
7. Percent default values – After the state/SHPO data was complied to a National Dataset, an additional metric was logged that quantified to what extent the state/SHPO data did not fulfill the information for the CRM standard. This metric was quantified in a percentage of values which were “Unknown”, “Not Specified” or “Not Applicable” in any CRM fields which were domain-restricted. These metrics are shown in Table 1.

	AK	AZ	CA	CO	ID	MT	NM	NV	OR	UT	WY
Investigation Authority		27.33	99.53	6.84	99.35	100.00	100.00	100.00	100.00	100.00	100.00
Investigation Class		43.07	0.00	2.57	3.68	98.98	100.00	0.98	42.28	0.10	0.00
Resource Temporal Cultural Assignment	10.87	40.77	22.40	4.96	2.28	52.52	27.20	26.49	14.04	39.04	12.60
Resource Primary Property Class	0.00	0.00	22.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Resource Primary Category Name	94.93	100.00	40.24	48.83	26.37	7.27	44.82	65.00	100.00	23.44	100.00
Resource NRHP Eligibility Status	97.92	59.45	78.65	21.12	38.69	77.60	100.00	53.27	80.52	32.91	29.08
Resource NRHP Eligibility Criteria	99.22	73.29	9.01	99.20	94.44	96.95	32.53	91.92	100.00	98.66	100.00
Resource NRHP Eligibility Authority Name	97.92	50.36	78.58	7.20	1.52	100.00	32.53	53.19	3.41	0.00	72.15
Resource Condition Assessment	96.18	60.64	98.22	47.41	30.22	100.00	42.60	79.85	100.00	31.17	100.00
Resource Collection Performed Status	100.00	65.66	58.62	82.12	2.68	100.00	93.22	21.31	0.00	96.78	0.37
Resource Spatial Collection Method	96.43	34.41	19.22	100.00	31.26	100.00	100.00	6.74	100.00	0.10	100.00

Table 1. Percent of default values used for each state's data per the CRM fields which are domain-restricted.

4. Compiling States to a National Dataset

The state/SHPO data were aggregated into a National Dataset after completion of the NCRIMS script tool. This aggregation was performed manually after QC using the Append tool in ArcMap. Additional processing was performed on the National Dataset after the aggregation. Details of the additional processing can be found in Appendix A.

Prior to aggregation, all data was stored in a staging file geodatabase titled “Staging.gdb”. These geodatabases stored all delivered state/SHPO data in the CRM standard. Some states delivered all data regardless of the location of feature in relation to federal agencies lands other than the BLM. These data for these states remained in staging, however only the features which intersected the BLM SMA 100-meter buffer were placed in the National Dataset. This was accomplished using the Select by Location tool in ArcMap prior to appending. Details on this process as it applied to each state can be found in Appendix A.

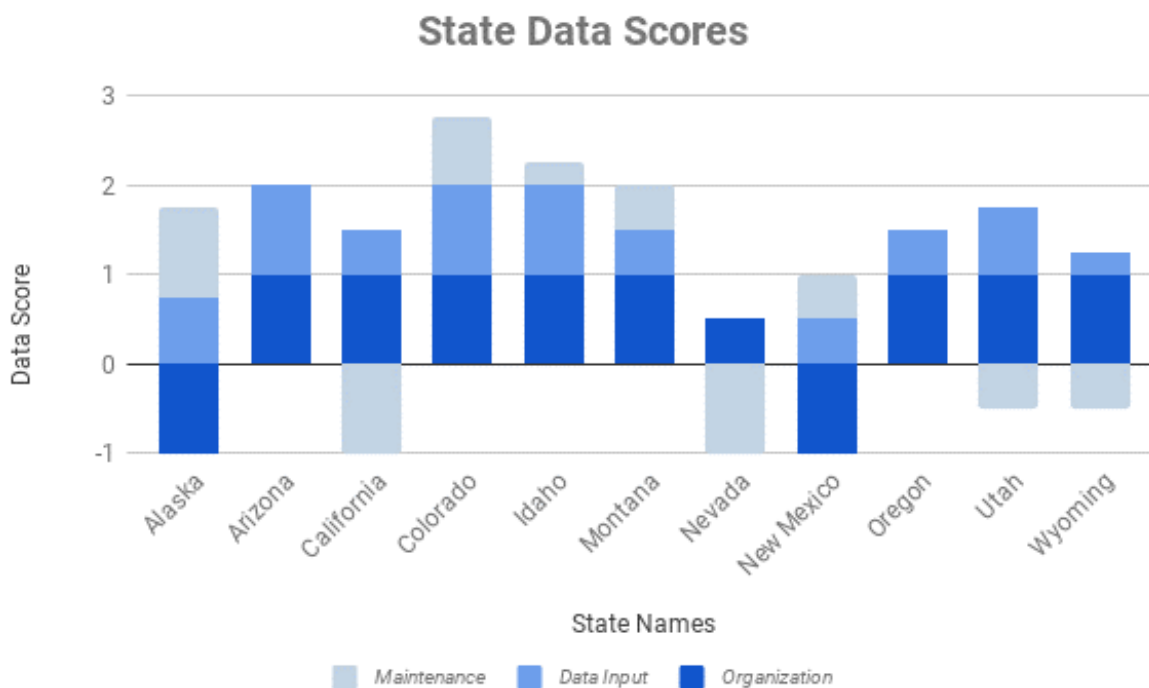
4.1 Moving Forward

Moving forward with the standard, the NOC may consider the following to ensure data is retained in the future.

1. CRM Identifiers – A plan should be developed for ensuring these persist. Cooperation from State and SHPO may be required for ensuring each record in the delivered dataset contains an identifier. Moreover, if the State or SHPO intends to change or update their current system for assigning or creating these identifiers, the State or SHPO should inform the NOC of this change and both entities should work together to develop a plan for migrating previous CRM IDs to the newly delivered data.
2. Create Date/ By fields – Similarly to the CRM Identifiers, a plan should be developed for ensuring these persist, if persistence is desired. It is recommended in order to decipher between current and previous data for future cycles.

5. 2019 State Observations and Recommendations

It is expected that each state delivery will be unique. Observations specific to each states' 2019 deliverable are described below along with the total data score and recommendations for the 2020 deliverable. These scores are summarized in the graph below. Please note, these recommendations are intended to streamline the national ETL process and provide guidance for the states on creating compatibility with the national standard.



5.1 Alaska

Alaska participated in the 2019 NCRIMS update cycle and delivered data in January of 2020. The deliverable contained data for one region, so the 2018 data was used in addition to the 2019

deliverable in the National Dataset. The targeted acquisition date is September, annually. The status on Alaska's work to improve the compatibility of deliverables was summarized in their June 2019 Update document. Details on resolutions observed in the 2019 deliverable based on the 2018 recommendations are provided in Appendix C. Although it is known Alaska is continuously working on compatibility improvements, the following section quantifies the compatibility of the 2019 deliverable and provides recommendations for improvement.

Recommendations for Future Deliverables (2)

Total Score: 0.75

A. Organization: -1 (-2 out of 2)

- a. The SHPO/ State data was delivered in a file geodatabase: -1
Criteria description: The data was delivered in a file geodatabase. The use of shapefiles is not recommended for deliveries due to storage constraints. It is recommended that if tables are included in the data release, they be included inside the file geodatabase. A file geodatabase provides a consistent delivery mechanism for state data and is considered a best practice as it is optimized and scalable.
- b. Two feature classes exist in the data deliverable: -1
Criteria description: The file geodatabase contained one feature class for Resource features and one feature class for Investigation features. Some states may store data according to geometry. Before execution of the ETL, these geometries must be transformed and merged into a single, usable geometry for the national database. Additionally, other states may store feature information in multiple feature classes based on status. This introduces multiple schemas which need to be in agreeance before the automate ETL is executed.

Alaska delivered 3 shapefiles representing different geometries for resource data. The attribute data for these shapefiles was delivered in 3 separate tables. Manual processes were performed prior to the ETL to reformat the deliverable for the NCRIMS script tool. These processes are described in detail in Appendix A.

Recommendation: Perform the reformatting processes described in Appendix A: Alaska prior to the 2020 delivery.

B. Data Input: 0.75 (3 out of 4)

- a. Redundant fields do not exist in the delivered data: 1
Criteria description: The data did not contain redundant fields. Redundant fields store the same information and can cause confusion during migration, especially if discrepancies exist between the crosswalk and delivered data.
- b. Records in the data are tracked: 1
Criteria description: The data contained a date field or an alternative field to track the most recent visit to a site. A date field or a system of tracking when the record was input or when the site was visited is recommended, especially if legacy records exist in the data. This helps narrow down the records to the most recent available. If reconciliation of legacy records is not required, the state/SHPO will receive a 1 for this criteria.
- c. The data includes a field for each CRM Standard field: 0
Criteria description: The data was complete across the CRM fields. It is recommended that states maintain data fields which exist in the CRM schema. States that do so are considered complete across the CRM fields. This improves the effectiveness of the National Dataset.
- d. The delivered data has normalized and/or consistent values: 1
Criteria description: The data contained normalized/ consistent values. Normalized values are extremely important for an effective ETL. Normalized values are considered values which fall within a domain or range of values. For example, consider a NRHP eligibility field. If the field

contains multiple versions of the value “No” (i.e. “No”, “nO”, “n0”) then the values are not consistent. Likewise, if multiple values exist for ineligibility (i.e. “No”, “Not Eligible”, “Ineligible”) then the field is not normalized.

Alaska provided data for all CRM Resource fields except for Primary Property Class and Collection Performed Status.

Recommendation: Add and maintain the RSRCE_PRMRY_PRPTY_CL and RSRCE_PRFRM_STTS fields or supply data to populate these fields in the 2020 delivery.

C. Maintenance: 1 (4 out of 4)

- a. Legacy and duplicate records do not exist: 1
Criteria description: The data did not contain legacy or duplicate records. Legacy records are considered as any record of a previous visit to the site or survey which has not been removed from the data delivery. Duplicate records are considered records which appear in the dataset more than once and contain the same information and geometry.
- b. Legacy or duplicate records can be quickly reconciled: 1
Criteria description: The legacy or duplicate records were able to be queried and easily removed (if they exist). Often times, without a tracking field, it is difficult to determine if the record is multipart, legacy, or truly duplicate. This is especially true if two records with the same identifier are located on top of one another.
- c. Each record contains an identifier: 1
Criteria description: Each record in the delivered dataset contained an identifier, generally the SHPO ID. This allows a user to contact the SHPO office with a reference in case more information is needed. If a SHPO ID is not available for the record, then an Agency ID should be available. These identifiers also assist the BLM in creating and maintaining the CRM ID fields. Features without an identifier will not be included in the National Dataset.
- d. Features do not exhibit geometry or topology issues: 1
Criteria description: Features did not exhibit geometry or topology issues. Severe issues cause errors in geoprocessing during the ETL processes. These issues also may negatively impact the analytical and reporting services in the web application.

No critical issues were observed regarding data maintenance.

5.2 Arizona

Arizona participated in the 2019 NCRIMS update cycle and delivered data in October of 2019. The targeted acquisition date is September, annually. Details on resolutions observed in the 2019 deliverable based on the 2018 recommendations are provided in Appendix C. The following section quantifies the compatibility of the 2019 deliverable and provides recommendations for improvement.

Recommendations for Future Deliverables (5)

Total Score: 2

D. Organization: 1 (2 out of 2)

- a. The SHPO/ State data was delivered in a file geodatabase: 1
Criteria description: The data was delivered in a file geodatabase. The use of shapefiles is not recommended for deliveries due to storage constraints. It is recommended that if tables are included in the data release, they be included inside the file geodatabase. A file geodatabase

provides a consistent delivery mechanism for state data and is considered a best practice as it is optimized and scalable.

b. Two feature classes exist in the data deliverable: 1

Criteria description: The file geodatabase contained one feature class for Resource features and one feature class for Investigation features. Some states may store data according to geometry. Before execution of the ETL, these geometries must be transformed and merged into a single, usable geometry for the national database. Additionally, other states may store feature information in multiple feature classes based on status. This introduces multiple schemas which need to be in agreement before the automate ETL is executed.

Manual procedures were performed to update the schema of the deliverable prior to executing the NCRIMS script tool. These procedures are documented in Appendix A: Arizona.

Recommendation: Change the field name RSRCE_PRMRY_CAT_NM to RSRCE_PRMRY_CTGRY_NM.

B. Data Input: 1 (4 out of 4)

a. Redundant fields do not exist in the delivered data: 1

Criteria description: The data did not contain redundant fields. Redundant fields store the same information and can cause confusion during migration, especially if discrepancies exist between the crosswalk and delivered data.

b. Records in the data are tracked: 1

Criteria description: The data contained a date field or an alternative field to track the most recent visit to a site. A date field or a system of tracking when the record was input or when the site was visited is recommended, especially if legacy records exist in the data. This helps narrow down the records to the most recent available. If reconciliation of legacy records is not required, the state/SHPO will receive a 1 for this criteria.

c. The data includes a field for each CRM Standard field: 1

Criteria description: The data was complete across the CRM fields. It is recommended that states maintain data fields which exist in the CRM schema. States that do so are considered complete across the CRM fields. This improves the effectiveness of the National Dataset.

d. The delivered data has normalized and/or consistent values: 1

Criteria description: The data contained normalized/ consistent values. Normalized values are extremely important for an effective ETL. Normalized values are considered values which fall within a domain or range of values. For example, consider a NRHP eligibility field. If the field contains multiple versions of the value "No" (i.e. "No", "nO", "n0") then the values are not consistent. Likewise, if multiple values exist for ineligibility (i.e. "No", "Not Eligible", "Ineligible") then the field is not normalized.

Manual processing was required to correct the domain values which did not comply with the CRM standard. A list of these domain values can be found in Appendix A.

Recommendation: Update the values under domain-restricted fields to comply with the CRM standard.

C. Maintenance: 0 (0 out of 4)

a. Legacy and duplicate records do not exist: 1

Criteria description: The data did not contain legacy or duplicate records. Legacy records are considered as any record of a previous visit to the site or survey which has not been removed from the data delivery. Duplicate records are considered records which appear in the dataset more than once and contain the same information and geometry.

b. Legacy or duplicate records can be quickly reconciled: 1



Criteria description: The legacy or duplicate records were able to be queried and easily removed (if they exist). Often times, without a tracking field, it is difficult to determine if the record is multipart, legacy, or truly duplicate. This is especially true if two records with the same identifier are located on top of one another.

- c. Each record contains an identifier: -1

Criteria description: Each record in the delivered dataset contained an identifier, generally the SHPO ID. This allows a user to contact the SHPO office with a reference in case more information is needed. If a SHPO ID is not available for the record, then an Agency ID should be available. These identifiers also assist the BLM in creating and maintaining the CRM ID fields. Features without an identifier will not be included in the National Dataset.

- d. Features do not exhibit geometry or topology issues: -1

Criteria description: Features did not exhibit geometry or topology issues. Severe issues cause errors in geoprocessing during the ETL processes. These issues also may negatively impact the analytical and reporting services in the web application.

1 Investigation SHPO ID and 8 Resource SHPO IDs occurred in the data more than once. However, most of these appeared to be multipart features or contained different Agency identifiers. A complete list of SHPO IDs which occurred more than once in the data can found under the DataMigration folder. See Section 2.3.3 for more information regarding this folder.

Recommendation: Request and review these SHPO identifiers to confirm these features are not legacy nor duplicated. If legacy or duplicated records exist, remove them from the 2020 deliverable. If these features are multipart, perform dissolves, where possible. It is understood some multipart features contain different attribute information.

3,911 Investigation records and 13,220 Resource records did not contain an assigned SHPO ID. Of the 3,911 Investigation records, 153 did not have an assigned Agency identifier. Of the 13,220 Resource records, 3,350 did not have an assigned Agency identifier.

Recommendation: Request and review these records. Assign a unique identifier for these in the 2020 deliverable.

18 features contained NULL geometry. In other words, these features did not have any geometric information for the GIS and thus, could not be projected or undergo geoprocessing. As a result, these 18 features were not migrated to the National Dataset and will not appear in the web application. See Appendix B for a list of these features' OIDs.

Recommendation: Request and review these feature's OID's. Add geometry to these features via digitizing or another form of drawing these boundaries in GIS.

5.3 California

California participated in the 2019 NCRIMS update cycle and delivered data in January of 2020. The targeted acquisition date is September, annually. Details on resolutions observed in the 2019 deliverable based on the 2018 recommendations are provided in Appendix C. The following section quantifies the compatibility of the 2019 deliverable and provides recommendations for improvement.



Recommendations for Future Deliverables (5)

Total Score: 0.5

A. Organization: 1 (2 out of 2)

- a. The SHPO/ State data was delivered in a file geodatabase: 1
Criteria description: The data was delivered in a file geodatabase. The use of shapefiles is not recommended for deliveries due to storage constraints. It is recommended that if tables are included in the data release, they be included inside the file geodatabase. A file geodatabase provides a consistent delivery mechanism for state data and is considered a best practice as it is optimized and scalable.
- b. Two feature classes exist in the data deliverable: 1
Criteria description: The file geodatabase contained one feature class for Resource features and one feature class for Investigation features. Some states may store data according to geometry. Before execution of the ETL, these geometries must be transformed and merged into a single, usable geometry for the national database. Additionally, other states may store feature information in multiple feature classes based on status. This introduces multiple schemas which need to be in agreeance before the automate ETL is executed.

California delivered a file geodatabase with related tables. Manual processing was performed to format the delivered data to make it compatible with the NCRIMS script tool. Details on this process can be found in Appendix A: California.

Recommendation: Perform the reformatting processes described in Appendix A prior to the 2020 delivery.

B. Data Input: 0.5 (2 out of 4)

- a. Redundant fields do not exist in the delivered data: 1
Criteria description: The data did not contain redundant fields. Redundant fields store the same information and can cause confusion during migration, especially if discrepancies exist between the crosswalk and delivered data.
- b. Records in the data are tracked: 0
Criteria description: The data contained a date field or an alternative field to track the most recent visit to a site. A date field or a system of tracking when the record was input or when the site was visited is recommended, especially if legacy records exist in the data. This helps narrow down the records to the most recent available. If reconciliation of legacy records is not required, the state/SHPO will receive a 1 for this criteria.
- c. The data includes a field for each CRM Standard field: 0
Criteria description: The data was complete across the CRM fields. It is recommended that states maintain data fields which exist in the CRM schema. States that do so are considered complete across the CRM fields. This improves the effectiveness of the National Dataset.
- d. The delivered data has normalized and/or consistent values: 1
Criteria description: The data contained normalized/ consistent values. Normalized values are extremely important for an effective ETL. Normalized values are considered values which fall within a domain or range of values. For example, consider a NRHP eligibility field. If the field contains multiple versions of the value "No" (i.e. "No", "nO", "n0") then the values are not consistent. Likewise, if multiple values exist for ineligibility (i.e. "No", "Not Eligible", "Ineligible") then the field is not normalized.

California provided data for all CRM fields except for the Resource NRHP Eligibility Criteria and Resource Condition Assessment.



Recommendation: Add and maintain the RSRCE_NRHP_ELGBLE_CRTRA and RSRCE_CNDTN_ASSMNT fields or supply data to populate these fields in the 2020 delivery.

C. Maintenance: -1 (-4 out of 4)

- a. Legacy and duplicate records do not exist: -1

Criteria description: The data did not contain legacy or duplicate records. Legacy records are considered as any record of a previous visit to the site or survey which has not been removed from the data delivery. Duplicate records are considered records which appear in the dataset more than once and contain the same information and geometry.

- b. Legacy or duplicate records can be quickly reconciled: -1

Criteria description: The legacy or duplicate records were able to be queried and easily removed (if they exist). Often times, without a tracking field, it is difficult to determine if the record is multipart, legacy, or truly duplicate. This is especially true if two records with the same identifier are located on top of one another.

- c. Each record contains an identifier: -1

Criteria description: Each record in the delivered dataset contained an identifier, generally the SHPO ID. This allows a user to contact the SHPO office with a reference in case more information is needed. If a SHPO ID is not available for the record, then an Agency ID should be available. These identifiers also assist the BLM in creating and maintaining the CRM ID fields. Features without an identifier will not be included in the National Dataset.

- d. Features do not exhibit geometry or topology issues: -1

Criteria description: Features did not exhibit geometry or topology issues. Severe issues cause errors in geoprocessing during the ETL processes. These issues also may negatively impact the analytical and reporting services in the web application.

81 Investigation SHPO ID and 645 Resource SHPO IDs occurred in the data more than once. These appeared to be legacy records. Because the date field was not fully populated, these were not resolved. A complete list of SHPO IDs which occurred more than once in the data can found under the DataMigration folder. See Section 2.3.2 for more information regarding this folder.

Recommendation: Request and review these SHPO identifiers to confirm these features are not legacy nor duplicated. Ensure the date field is fully populated and, if legacy or duplicated records exist, remove them from the 2020 deliverable. If these features are multipart, perform dissolves, where possible. It is understood some multipart features contain different attribute information.

8,504 Investigation records and 13,960 Resource records did not contain an assigned SHPO ID. Of the 8,504 Investigation records, 1,131 did not have an assigned Agency identifier. Of the 13,960 Resource records, 8,193 did not have an assigned Agency identifier.

Recommendation: Request and review these records. Assign a unique identifier for these in the 2020 deliverable.

5 features contained NULL geometry. In other words, these features did not have any geometric information for the GIS and thus, could not be projected or undergo geoprocessing. As a result, these 5 features were not migrated to the National Dataset and will not appear in the web application. See Appendix A for a list of these features' OIDs. Additionally, the delivered feature classes contained bad spatial extents.

Recommendation: Request and review these feature's OID's. Add geometry to these features via digitizing or another form of drawing these boundaries in GIS. Run the Add Spatial Index tool in ArcMap prior to the 2020 delivery.

5.4 Colorado

Colorado participated in the 2019 NCRIMS update cycle and delivered data and the state ETL script tool in November of 2019. The targeted acquisition date is September, annually. Details on resolutions observed in the 2019 deliverable based on the 2018 recommendations are provided in Appendix C. The following section quantifies the compatibility of the 2019 deliverable and provides recommendations for improvement.

Recommendations for Future Deliverables (2)

Total Score: 2.75

A. Organization: 1 (2 out of 2)

- a. The SHPO/ State data was delivered in a file geodatabase: 1
Criteria description: The data was delivered in a file geodatabase. The use of shapefiles is not recommended for deliveries due to storage constraints. It is recommended that if tables are included in the data release, they be included inside the file geodatabase. A file geodatabase provides a consistent delivery mechanism for state data and is considered a best practice as it is optimized and scalable.
- b. Two feature classes exist in the data deliverable: 1
Criteria description: The file geodatabase contained one feature class for Resource features and one feature class for Investigation features. Some states may store data according to geometry. Before execution of the ETL, these geometries must be transformed and merged into a single, usable geometry for the national database. Additionally, other states may store feature information in multiple feature classes based on status. This introduces multiple schemas which need to be in agreeance before the automate ETL is executed.

No critical issues were observed with the organizational structure of the delivered data.

B. Data Input: 1 (4 out of 4)

- a. Redundant fields do not exist in the delivered data: 1
Criteria description: The data did not contain redundant fields. Redundant fields store the same information and can cause confusion during migration, especially if discrepancies exist between the crosswalk and delivered data.
- b. Records in the data are tracked: 1
Criteria description: The data contained a date field or an alternative field to track the most recent visit to a site. A date field or a system of tracking when the record was input or when the site was visited is recommended, especially if legacy records exist in the data. This helps narrow down the records to the most recent available. If reconciliation of legacy records is not required, the state/SHPO will receive a 1 for this criteria.
- c. The data includes a field for each CRM Standard field: 1
Criteria description: The data was complete across the CRM fields. It is recommended that states maintain data fields which exist in the CRM schema. States that do so are considered complete across the CRM fields. This improves the effectiveness of the National Dataset.
- d. The delivered data has normalized and/or consistent values: 1
Criteria description: The data contained normalized/ consistent values. Normalized values are extremely important for an effective ETL. Normalized values are considered values which fall within a domain or range of values. For example, consider a NRHP eligibility field. If the field contains multiple versions of the value "No" (i.e. "No", "nO", "n0") then the values are not

consistent. Likewise, if multiple values exist for ineligibility (i.e. “No”, “Not Eligible”, “Ineligible”) then the field is not normalized.

Manual processing was required to correct the domain values which did not comply with the CRM standard. A list of these domain values can be found in Appendix A.

Recommendation: Update the values under domain-restricted fields to comply with the CRM standard.

C. Maintenance: 0.75 (3 out of 4)

a. Legacy and duplicate records do not exist: 1

Criteria description: The data did not contain legacy or duplicate records. Legacy records are considered as any record of a previous visit to the site or survey which has not been removed from the data delivery. Duplicate records are considered records which appear in the dataset more than once and contain the same information and geometry.

b. Legacy or duplicate records can be quickly reconciled: 1

Criteria description: The legacy or duplicate records were able to be queried and easily removed (if they exist). Often times, without a tracking field, it is difficult to determine if the record is multipart, legacy, or truly duplicate. This is especially true if two records with the same identifier are located on top of one another.

c. Each record contains an identifier: 1

Criteria description: Each record in the delivered dataset contained an identifier, generally the SHPO ID. This allows a user to contact the SHPO office with a reference in case more information is needed. If a SHPO ID is not available for the record, then an Agency ID should be available. These identifiers also assist the BLM in creating and maintaining the CRM ID fields. Features without an identifier will not be included in the National Dataset.

d. Features do not exhibit geometry or topology issues: 0

Criteria description: Features did not exhibit geometry or topology issues. Severe issues cause errors in geoprocessing during the ETL processes. These issues also may negatively impact the analytical and reporting services in the web application.

The Colorado deliverable contained resource and investigation features with self intersections.

Recommendation: Run the Repair Geometry tool in ArcMap prior to the 2020 delivery.

5.5 Idaho

Idaho participated in the 2019 NCRIMS update cycle and delivered data in November of 2019. The targeted acquisition date is September, annually. Details on resolutions observed in the 2019 deliverable based on the 2018 recommendations are provided in Appendix C. The following section quantifies the compatibility of the 2019 deliverable and provides recommendations for improvement.

Recommendations for Future Deliverable (4)

Total Score: 2.25

A. Organization: 1 (2 out of 2)

a. The SHPO/ State data was delivered in a file geodatabase: 1

Criteria description: The data was delivered in a file geodatabase. The use of shapefiles is not recommended for deliveries due to storage constraints. It is recommended that if tables are included in the data release, they be included inside the file geodatabase. A file geodatabase

provides a consistent delivery mechanism for state data and is considered a best practice as it is optimized and scalable.

b. Two feature classes exist in the data deliverable: 1

Criteria description: The file geodatabase contained one feature class for Resource features and one feature class for Investigation features. Some states may store data according to geometry. Before execution of the ETL, these geometries must be transformed and merged into a single, usable geometry for the national database. Additionally, other states may store feature information in multiple feature classes based on status. This introduces multiple schemas which need to be in agreement before the automate ETL is executed.

No critical issues were observed regarding the organizational structure of the deliverable.

B. Data Input: 1 (4 out of 4)

a. Redundant fields do not exist in the delivered data: 1

Criteria description: The data did not contain redundant fields. Redundant fields store the same information and can cause confusion during migration, especially if discrepancies exist between the crosswalk and delivered data.

b. Records in the data are tracked: 1

Criteria description: The data contained a date field or an alternative field to track the most recent visit to a site. A date field or a system of tracking when the record was input or when the site was visited is recommended, especially if legacy records exist in the data. This helps narrow down the records to the most recent available. If reconciliation of legacy records is not required, the state/SHPO will receive a 1 for this criteria.

c. The data includes a field for each CRM Standard field: 1

Criteria description: The data was complete across the CRM fields. It is recommended that states maintain data fields which exist in the CRM schema. States that do so are considered complete across the CRM fields. This improves the effectiveness of the National Dataset.

d. The delivered data has normalized and/or consistent values: 1

Criteria description: The data contained normalized/ consistent values. Normalized values are extremely important for an effective ETL. Normalized values are considered values which fall within a domain or range of values. For example, consider a NRHP eligibility field. If the field contains multiple versions of the value "No" (i.e. "No", "nO", "n0") then the values are not consistent. Likewise, if multiple values exist for ineligibility (i.e. "No", "Not Eligible", "Ineligible") then the field is not normalized.

Although Idaho has implemented the CRM standard, manual processing was required to correct the domain values which did not comply with the CRM standard. A list of these domain values can be found in Appendix A.

Recommendation: Update the values under domain-restricted fields to comply with the CRM standard.

C. Maintenance: 0.25 (1 out of 4)

a. Legacy and duplicate records do not exist: 1

Criteria description: The data did not contain legacy or duplicate records. Legacy records are considered as any record of a previous visit to the site or survey which has not been removed from the data delivery. Duplicate records are considered records which appear in the dataset more than once and contain the same information and geometry.

b. Legacy or duplicate records can be quickly reconciled: 1

Criteria description: The legacy or duplicate records were able to be queried and easily removed (if they exist). Often times, without a tracking field, it is difficult to determine if the record is multipart, legacy, or truly duplicate. This is especially true if two records with the same identifier are located on top of one another.



- c. Each record contains an identifier: -1

Criteria description: Each record in the delivered dataset contained an identifier, generally the SHPO ID. This allows a user to contact the SHPO office with a reference in case more information is needed. If a SHPO ID is not available for the record, then an Agency ID should be available. These identifiers also assist the BLM in creating and maintaining the CRM ID fields. Features without an identifier will not be included in the National Dataset.

- d. Features do not exhibit geometry or topology issues: 0

Criteria description: Features did not exhibit geometry or topology issues. Severe issues cause errors in geoprocessing during the ETL processes. These issues also may negatively impact the analytical and reporting services in the web application.

911 Investigation SHPO IDs and 33 Resource SHPO IDs occurred in the data more than once. However, most of these appeared to be multipart features or contained different Agency identifiers. A complete list of SHPO IDs which occurred more than once in the data can found under the DataMigration folder.

Recommendation: Request and review these SHPO identifiers to confirm these features are not legacy nor duplicated. Remove any legacy or duplicate records prior to the 2020 delivery. If these features are multipart, perform dissolves, where possible. It is understood some multipart features contain different attribute information.

629 Investigation records and 241 Resource records did not contain an assigned SHPO ID. Of the 629 Investigation records, 7 did not have an assigned Agency identifier. Of the 241 Resource records, 19 did not have an assigned Agency identifier.

Recommendation: Request and review these records. Assign a unique identifier for these in the 2020 deliverable.

Last, the Idaho deliverable contained 88 Resource features with self-intersections.

Recommendation: Run the Repair Geometry tool in ArcMap prior to the 2020 delivery.

5.6 Montana

Montana participated in the 2019 NCRIMS update cycle and delivered data in December of 2019. The targeted acquisition date is September, annually. Details on resolutions observed in the 2019 deliverable based on the 2018 recommendations are provided in Appendix C. The following section quantifies the compatibility of the 2019 deliverable and provides recommendations for improvement.

Recommendations for Future Deliverable (4)

Total Score: 2

A. Organization: 1 (2 out of 2)

- a. The SHPO/ State data was delivered in a file geodatabase: 1

Criteria description: The data was delivered in a file geodatabase. The use of shapefiles is not recommended for deliveries due to storage constraints. It is recommended that if tables are included in the data release, they be included inside the file geodatabase. A file geodatabase provides a consistent delivery mechanism for state data and is considered a best practice as it is optimized and scalable.

- b. Two feature classes exist in the data deliverable: 1



Criteria description: The file geodatabase contained one feature class for Resource features and one feature class for Investigation features. Some states may store data according to geometry. Before execution of the ETL, these geometries must be transformed and merged into a single, usable geometry for the national database. Additionally, other states may store feature information in multiple feature classes based on status. This introduces multiple schemas which need to be in agreeance before the automate ETL is executed.

No critical issues were observed regarding the organizational structure of the deliverable.

B. Data Input: 0.5 (2 out of 4)

- a. Redundant fields do not exist in the delivered data: 1
Criteria description: The data did not contain redundant fields. Redundant fields store the same information and can cause confusion during migration, especially if discrepancies exist between the crosswalk and delivered data.
- b. Records in the data are tracked: 1
Criteria description: The data contained a date field or an alternative field to track the most recent visit to a site. A date field or a system of tracking when the record was input or when the site was visited is recommended, especially if legacy records exist in the data. This helps narrow down the records to the most recent available. If reconciliation of legacy records is not required, the state/SHPO will receive a 1 for this criteria.
- c. The data includes a field for each CRM Standard field: -1
Criteria description: The data was complete across the CRM fields. It is recommended that states maintain data fields which exist in the CRM schema. States that do so are considered complete across the CRM fields. This improves the effectiveness of the National Dataset.
- d. The delivered data has normalized and/or consistent values: 1
Criteria description: The data contained normalized/ consistent values. Normalized values are extremely important for an effective ETL. Normalized values are considered values which fall within a domain or range of values. For example, consider a NRHP eligibility field. If the field contains multiple versions of the value "No" (i.e. "No", "nO", "n0") then the values are not consistent. Likewise, if multiple values exist for ineligibility (i.e. "No", "Not Eligible", "Ineligible") then the field is not normalized.

Montana provided data for all CRM fields except for the Investigation Authority, Investigation NEPA ID, Resource NRHP Eligibility Authority Name, Resource Condition Assessment, Resource Collection Performed Status, and Resource Spatial Collection Method.

Recommendation: Add and maintain the INVSTGTN_AUTH, INVSTGTN_NEPA_ID, RSRCE_NRHP_ELGBLE_AUTH_NM, RSRCE_CNDTN_ASSMNT, RSRCE_CLCTN_PRFRM_STTS, and RSRCE_SPTL_CLCTN_MTHD fields or supply data to populate these in the 2020 delivery.

C. Maintenance: 0.5 (2 out of 4)

- a. Legacy and duplicate records do not exist: 1
Criteria description: The data did not contain legacy or duplicate records. Legacy records are considered as any record of a previous visit to the site or survey which has not been removed from the data delivery. Duplicate records are considered records which appear in the dataset more than once and contain the same information and geometry.
- b. Legacy or duplicate records can be quickly reconciled: 1
Criteria description: The legacy or duplicate records were able to be queried and easily removed (if they exist). Often times, without a tracking field, it is difficult to determine if the record is

multipart, legacy, or truly duplicate. This is especially true if two records with the same identifier are located on top of one another.

c. Each record contains an identifier: 0

Criteria description: Each record in the delivered dataset contained an identifier, generally the SHPO ID. This allows a user to contact the SHPO office with a reference in case more information is needed. If a SHPO ID is not available for the record, then an Agency ID should be available. These identifiers also assist the BLM in creating and maintaining the CRM ID fields. Features without an identifier will not be included in the National Dataset.

d. Features do not exhibit geometry or topology issues: 0

Criteria description: Features did not exhibit geometry or topology issues. Severe issues cause errors in geoprocessing during the ETL processes. These issues also may negatively impact the analytical and reporting services in the web application.

9 Investigation SHPO IDs and 2 Resource SHPO IDs occurred in the data more than once. However, most of these appeared to be multipart features. A complete list of SHPO IDs which occurred more than once in the data can found under the DataMigration folder.

Recommendation: Request the list of SHPO IDs and dissolve these. Dissolve any additional multipart features in the 2020 deliverable.

563 Investigation records did not contain an assigned SHPO ID nor an assigned Agency identifier.

Recommendation: Request and review these records. Assign a unique identifier for these in the 2020 deliverable.

Last, the Idaho deliverable contained 60 Resource features and 127 Investigation features with self-intersections.

Recommendation: Run the Repair Geometry tool in ArcMap prior to the 2020 delivery.

5.7 Nevada

Nevada participated in the 2019 NCRIMS update cycle and delivered data in December of 2019. The targeted acquisition date is September, annually. Details on resolutions observed in the 2019 deliverable based on the 2018 recommendations are provided in Appendix C. Although Nevada plans to implement the CRM standard for 2020, the following section quantifies the compatibility of the 2019 deliverable and provides recommendations for improvement.

Recommendations for Future Deliverable (4)

Total Score: -0.5

A. Organization: 0.5 (1 out of 2)

a. The SHPO/ State data was delivered in a file geodatabase: 1

Criteria description: The data was delivered in a file geodatabase. The use of shapefiles is not recommended for deliveries due to storage constraints. It is recommended that if tables are included in the data release, they be included inside the file geodatabase. A file geodatabase provides a consistent delivery mechanism for state data and is considered a best practice as it is optimized and scalable.

b. Two feature classes exist in the data deliverable: 0

Criteria description: The file geodatabase contained one feature class for Resource features and one feature class for Investigation features. Some states may store data according to geometry.



Before execution of the ETL, these geometries must be transformed and merged into a single, usable geometry for the national database. Additionally, other states may store feature information in multiple feature classes based on status. This introduces multiple schemas which need to be in agreement before the automate ETL is executed.

The geodatabase included 7 feature classes. Of the 7 feature classes, 5 of them underwent the ETL process for migration to the national standard. Because the NCRIMS script tool can consume multiple feature classes as input, no recommendation is provided. See Appendix A for the 2019 Nevada ETL workflow details.

B. Data Input: 0 (0 out of 4)

- a. Redundant fields do not exist in the delivered data: 0

Criteria description: The data did not contain redundant fields. Redundant fields store the same information and can cause confusion during migration, especially if discrepancies exist between the crosswalk and delivered data.

- b. Records in the data are tracked: 0

Criteria description: The data contained a date field or an alternative field to track the most recent visit to a site. A date field or a system of tracking when the record was input or when the site was visited is recommended, especially if legacy records exist in the data. This helps narrow down the records to the most recent available. If reconciliation of legacy records is not required, the state/SHPO will receive a 1 for this criteria.

- c. The data includes a field for each CRM Standard field: 1

Criteria description: The data was complete across the CRM fields. It is recommended that states maintain data fields which exist in the CRM schema. States that do so are considered complete across the CRM fields. This improves the effectiveness of the National Dataset.

- d. The delivered data has normalized and/or consistent values: -1

Criteria description: The data contained normalized/ consistent values. Normalized values are extremely important for an effective ETL. Normalized values are considered values which fall within a domain or range of values. For example, consider a NRHP eligibility field. If the field contains multiple versions of the value "No" (i.e. "No", "nO", "n0") then the values are not consistent. Likewise, if multiple values exist for ineligibility (i.e. "No", "Not Eligible", "Ineligible") then the field is not normalized.

Because the delivery of the data spanned multiple feature classes, redundant fields existed. However, this was not a critical issue for the ETL scripting.

Fields crosswalked to domain-restricted CRM fields did not contain normalized values. Most of these values translated to a CRM domain value without input from the state except for the DESCRIPTION fields in Archeo_Sites and In_Process_Sites. These values totaled over 11,000 unique combinations, hence input was needed from the state to finalize and complete the domain crosswalk.

Recommendation: Standardize values in the fields that will be crosswalked to domain-restricted CRM fields. Target 1,000 or less unique combinations.

C. Maintenance: -1 (-4 out of 4)

- a. Legacy and duplicate records do not exist: -1

Criteria description: The data did not contain legacy or duplicate records. Legacy records are considered as any record of a previous visit to the site or survey which has not been removed from the data delivery. Duplicate records are considered records which appear in the dataset more than once and contain the same information and geometry.

- b. Legacy or duplicate records can be quickly reconciled: -1



Criteria description: The legacy or duplicate records were able to be queried and easily removed (if they exist). Often times, without a tracking field, it is difficult to determine if the record is multipart, legacy, or truly duplicate. This is especially true if two records with the same identifier are located on top of one another.

- c. Each record contains an identifier: -1

Criteria description: Each record in the delivered dataset contained an identifier, generally the SHPO ID. This allows a user to contact the SHPO office with a reference in case more information is needed. If a SHPO ID is not available for the record, then an Agency ID should be available. These identifiers also assist the BLM in creating and maintaining the CRM ID fields. Features without an identifier will not be included in the National Dataset.

- d. Features do not exhibit geometry or topology issues: -1

Criteria description: Features did not exhibit geometry or topology issues. Severe issues cause errors in geoprocessing during the ETL processes. These issues also may negatively impact the analytical and reporting services in the web application.

6,903 Investigation SHPO IDs and 16,211 Resource SHPO IDs occurred in the data more than once. Some of these features appeared to be legacy records but were unable to be resolved because the date field was not fully populated. Others were true duplicates or multipart features.

Recommendation: Request and review these SHPO identifiers. Ensure the date field is fully populated and remove legacy and duplicated records from the 2020 deliverable. If these features are multipart, perform dissolves, where possible. It is understood some multipart features contain different attribute information.

26 Investigation features and 1,093 Resource features did not contain a SHPO identifiers. Of the 26 Investigation features, 1 did not contain an Agency identifier. Of the 1,093 Resource features, 66 did not contain an Agency identifier.

Recommendation: Request and review these records. Assign a unique identifier for these in the 2020 deliverable.

Last, all delivered datasets had bad spatial extents. This caused the ETL tool to error out during geoprocessing operations such as clip and geometry calculations

Recommendation: Run the Add Spatial Index tool in ArcMap prior to the 2020 delivery.

5.8 New Mexico

New Mexico participated in the 2019 NCRIMS update cycle and delivered data in January of 2020. The targeted acquisition date is September, annually. Details on resolutions observed in the 2019 deliverable based on the 2018 recommendations are provided in Appendix C. The following section quantifies the compatibility of the 2019 deliverable and provides recommendations for improvement.

Recommendations for Future Deliverable (4)

Total Score: 0

A. Organization: -1 (-2 out of 2)

- a. The SHPO/ State data was delivered in a file geodatabase: -1

Criteria description: The data was delivered in a file geodatabase. The use of shapefiles is not recommended for deliveries due to storage constraints. It is recommended that if tables are

included in the data release, they be included inside the file geodatabase. A file geodatabase provides a consistent delivery mechanism for state data and is considered a best practice as it is optimized and scalable.

b. Two feature classes exist in the data deliverable: -1

Criteria description: The file geodatabase contained one feature class for Resource features and one feature class for Investigation features. Some states may store data according to geometry. Before execution of the ETL, these geometries must be transformed and merged into a single, usable geometry for the national database. Additionally, other states may store feature information in multiple feature classes based on status. This introduces multiple schemas which need to be in agreeance before the automate ETL is executed.

New Mexico delivered shapefiles and corresponding DBF tables for the 2019 cycle. Manual processing was required to reformat the data prior to the ETL. This workflow can be found in Appendix A.

Recommendation: Perform the preprocessing workflow prior to the 2020 delivery.

B. Data Input: 0.5 (2 out of 4)

a. Redundant fields do not exist in the delivered data: 1

Criteria description: The data did not contain redundant fields. Redundant fields store the same information and can cause confusion during migration, especially if discrepancies exist between the crosswalk and delivered data.

b. Records in the data are tracked: 1

Criteria description: The data contained a date field or an alternative field to track the most recent visit to a site. A date field or a system of tracking when the record was input or when the site was visited is recommended, especially if legacy records exist in the data. This helps narrow down the records to the most recent available. If reconciliation of legacy records is not required, the state/SHPO will receive a 1 for this criterium.

c. The data includes a field for each CRM Standard field: -1

Criteria description: The data was complete across the CRM fields. It is recommended that states maintain data fields which exist in the CRM schema. States that do so are considered complete across the CRM fields. This improves the effectiveness of the National Dataset.

d. The delivered data has normalized and/or consistent values: 1

Criteria description: The data contained normalized/ consistent values. Normalized values are extremely important for an effective ETL. Normalized values are considered values which fall within a domain or range of values. For example, consider a NRHP eligibility field. If the field contains multiple versions of the value "No" (i.e. "No", "nO", "n0") then the values are not consistent. Likewise, if multiple values exist for ineligibility (i.e. "No", "Not Eligible", "Ineligible") then the field is not normalized.

The 2019 delivery did not contain data for the following fields:

INVSTGTN_AUTH

INVSTGTN_CL

RSRCE_NRHP_ELGBLE_STTS

RSRCE_PRMRY_PRPTY_CL

RSRCE_SPTL_CLCTN_MTHD

Recommendation: Add and maintain the missing fields or supply data to populate these fields in the 2020 deliverable.

C. Maintenance: 0.5 (2 out of 4)

a. Legacy and duplicate records do not exist: 0



Criteria description: The data did not contain legacy or duplicate records. Legacy records are considered as any record of a previous visit to the site or survey which has not been removed from the data delivery. Duplicate records are considered records which appear in the dataset more than once and contain the same information and geometry.

b. Legacy or duplicate records can be quickly reconciled: 1

Criteria description: The legacy or duplicate records were able to be queried and easily removed (if they exist). Often times, without a tracking field, it is difficult to determine if the record is multipart, legacy, or truly duplicate. This is especially true if two records with the same identifier are located on top of one another.

c. Each record contains an identifier: 1

Criteria description: Each record in the delivered dataset contained an identifier, generally the SHPO ID. This allows a user to contact the SHPO office with a reference in case more information is needed. If a SHPO ID is not available for the record, then an Agency ID should be available. These identifiers also assist the BLM in creating and maintaining the CRM ID fields. Features without an identifier will not be included in the National Dataset.

d. Features do not exhibit geometry or topology issues: 0

Criteria description: Features did not exhibit geometry or topology issues. Severe issues cause errors in geoprocessing during the ETL processes. These issues also may negatively impact the analytical and reporting services in the web application.

Legacy and/or duplicate records existed in the table but were reconciled during preprocessing.

Recommendation: New Mexico should review this process and QC outputs to ensure confidence in the methodology.

Both Resource and Investigation features included features with self-intersections. The ID's for these features can be found in the DataMigration folder.

Recommendation: Perform a Repair Geometry on the data prior to delivery.

5.9 Oregon

Oregon participated in the 2019 NCRIMS update cycle and delivered data in October of 2019. The targeted acquisition date is September, annually. Details on resolutions observed in the 2019 deliverable based on the 2018 recommendations are provided in Appendix C. The following section quantifies the compatibility of the 2019 deliverable and provides recommendations for improvement.

Recommendations for Future Deliverable (4)

Total Score: 1.5

A. Organization: 1 (2 out of 2)

a. The SHPO/ State data was delivered in a file geodatabase: 1

Criteria description: The data was delivered in a file geodatabase. The use of shapefiles is not recommended for deliveries due to storage constraints. It is recommended that if tables are included in the data release, they be included inside the file geodatabase. A file geodatabase provides a consistent delivery mechanism for state data and is considered a best practice as it is optimized and scalable.

b. Two feature classes exist in the data deliverable: 1

Criteria description: The file geodatabase contained one feature class for Resource features and one feature class for Investigation features. Some states may store data according to geometry. Before execution of the ETL, these geometries must be transformed and merged into a single,

usable geometry for the national database. Additionally, other states may store feature information in multiple feature classes based on status. This introduces multiple schemas which need to be in agreeance before the automate ETL is executed.

No critical issues were observed regarding the organizational structure of the deliverable.

B. Data Input: 0.5 (2 out of 4)

- a. Redundant fields do not exist in the delivered data: 1
Criteria description: The data did not contain redundant fields. Redundant fields store the same information and can cause confusion during migration, especially if discrepancies exist between the crosswalk and delivered data.
- b. Records in the data are tracked: 1
Criteria description: The data contained a date field or an alternative field to track the most recent visit to a site. A date field or a system of tracking when the record was input or when the site was visited is recommended, especially if legacy records exist in the data. This helps narrow down the records to the most recent available. If reconciliation of legacy records is not required, the state/SHPO will receive a 1 for this criterium.
- c. The data includes a field for each CRM Standard field: 0
Criteria description: The data was complete across the CRM fields. It is recommended that states maintain data fields which exist in the CRM schema. States that do so are considered complete across the CRM fields. This improves the effectiveness of the National Dataset.
- d. The delivered data has normalized and/or consistent values: 0
Criteria description: The data contained normalized/ consistent values. Normalized values are extremely important for an effective ETL. Normalized values are considered values which fall within a domain or range of values. For example, consider a NRHP eligibility field. If the field contains multiple versions of the value "No" (i.e. "No", "nO", "n0") then the values are not consistent. Likewise, if multiple values exist for ineligibility (i.e. "No", "Not Eligible", "Ineligible") then the field is not normalized.

The 2019 deliverable included data for all CRM fields except for the Investigation Authority and the Resource Spatial Collection Method.

Recommendation: Add the INVSTGTN_AUTH and RSRCE_SPTL_CLCTN_MTHD fields.

The fields crosswalked to domain-restricted CRM fields contained mostly normalized values except for values in the Resource Primary Category Name field. This field contained 953 different values, hence the default value "Unknown" was utilized in the 2019 National Dataset.

Recommendation: Normalize the values for the RSRCE_PRMRY_CTGRY_NM field for the 2020 delivery.

C. Maintenance: 0 (0 out of 4)

- a. Legacy and duplicate records do not exist: -1
Criteria description: The data did not contain legacy or duplicate records. Legacy records are considered as any record of a previous visit to the site or survey which has not been removed from the data delivery. Duplicate records are considered records which appear in the dataset more than once and contain the same information and geometry.
- b. Legacy or duplicate records can be quickly reconciled: 0
Criteria description: The legacy or duplicate records were able to be queried and easily removed (if they exist). Often times, without a tracking field, it is difficult to determine if the record is



multipart, legacy, or truly duplicate. This is especially true if two records with the same identifier are located on top of one another.

c. Each record contains an identifier: 1

Criteria description: Each record in the delivered dataset contained an identifier, generally the SHPO ID. This allows a user to contact the SHPO office with a reference in case more information is needed. If a SHPO ID is not available for the record, then an Agency ID should be available. These identifiers also assist the BLM in creating and maintaining the CRM ID fields. Features without an identifier will not be included in the National Dataset.

d. Features do not exhibit geometry or topology issues: 0

Criteria description: Features did not exhibit geometry or topology issues. Severe issues cause errors in geoprocessing during the ETL processes. These issues also may negatively impact the analytical and reporting services in the web application.

21 Investigation SHPO IDs and 33 Resource SHPO IDs occurred in the data more than once. Most of these appeared to be multipart features and legacy records. A complete list of SHPO IDs which occurred more than once in the data can found under the DataMigration folder.

Recommendation: Request and review these SHPO identifiers. Ensure the date field is fully populated and remove legacy and duplicated records from the 2020 deliverable. If these features are multipart, perform dissolves, where possible. It is understood some multipart features contain different attribute information.

5,899 Investigation records and 9,487 Resource records did not contain an assigned SHPO ID, however each of these records did have an assigned Agency ID. There is no recommendation for action.

The Oregon deliverable contained 2,285 Investigation features and 3 Resource features with self-intersections.

Recommendation: Run the Repair Geometry tool in ArcMap prior to the 2020 delivery.

5.10 Utah

Utah participated in the 2019 NCRIMS update cycle and delivered data in November of 2019. The targeted acquisition date is September, annually. Details on resolutions observed in the 2019 deliverable based on the 2018 recommendations are provided in Appendix C. The following section quantifies the compatibility of the 2019 deliverable and provides recommendations for improvement.

Recommendations for Future Deliverable (5)

Total Score: 1.25

A. Organization: 1 (2 out of 2)

a. The SHPO/ State data was delivered in a file geodatabase: 1

Criteria description: The data was delivered in a file geodatabase. The use of shapefiles is not recommended for deliveries due to storage constraints. It is recommended that if tables are included in the data release, they be included inside the file geodatabase. A file geodatabase provides a consistent delivery mechanism for state data and is considered a best practice as it is optimized and scalable.

b. Two feature classes exist in the data deliverable: 1



Criteria description: The file geodatabase contained one feature class for Resource features and one feature class for Investigation features. Some states may store data according to geometry. Before execution of the ETL, these geometries must be transformed and merged into a single, usable geometry for the national database. Additionally, other states may store feature information in multiple feature classes based on status. This introduces multiple schemas which need to be in agreeance before the automate ETL is executed.

Manual reformatting occurred before the NCRIMS script tool was executed. The values under the delivered RSRCE_TMPRL_CLTRL_ASGMNT were actually CRM values for the RSRCE_PRMRY_CTGRY_NM field. See Appendix A: Utah for details on the reformatting steps.

Recommendation: Perform the reformatting process described in Appendix A: Utah prior to the 2020 delivery.

B. Data Input: 0.75 (3 out of 4)

- a. Redundant fields do not exist in the delivered data: 1

Criteria description: The data did not contain redundant fields. Redundant fields store the same information and can cause confusion during migration, especially if discrepancies exist between the crosswalk and delivered data.

- b. Records in the data are tracked: 1

Criteria description: The data contained a date field or an alternative field to track the most recent visit to a site. A date field or a system of tracking when the record was input or when the site was visited is recommended, especially if legacy records exist in the data. This helps narrow down the records to the most recent available. If reconciliation of legacy records is not required, the state/SHPO will receive a 1 for this criterium.

- c. The data includes a field for each CRM Standard field: 0

Criteria description: The data was complete across the CRM fields. It is recommended that states maintain data fields which exist in the CRM schema. States that do so are considered complete across the CRM fields. This improves the effectiveness of the National Dataset.

- d. The delivered data has normalized and/or consistent values: 1

Criteria description: The data contained normalized/ consistent values. Normalized values are extremely important for an effective ETL. Normalized values are considered values which fall within a domain or range of values. For example, consider a NRHP eligibility field. If the field contains multiple versions of the value "No" (i.e. "No", "nO", "n0") then the values are not consistent. Likewise, if multiple values exist for ineligibility (i.e. "No", "Not Eligible", "Ineligible") then the field is not normalized.

The 2019 deliverable included data for all CRM fields except for the Agency Investigation Unique Identifier, Investigation Authority and the Resource Agency Unique Identifier.

Recommendation: Add the IVSTGTN_AGCY_ID, INVSTGTN_AUTH and RSRCE_AGCY_ID fields.

C. Maintenance: -0.5 (-2 out of 4)

- a. Legacy and duplicate records do not exist: -1

Criteria description: The data did not contain legacy or duplicate records. Legacy records are considered as any record of a previous visit to the site or survey which has not been removed from the data delivery. Duplicate records are considered records which appear in the dataset more than once and contain the same information and geometry.

- b. Legacy or duplicate records can be quickly reconciled: 1

Criteria description: The legacy or duplicate records were able to be queried and easily removed (if they exist). Often times, without a tracking field, it is difficult to determine if the record is multipart, legacy, or truly duplicate. This is especially true if two records with the same identifier are located on top of one another.

- c. Each record contains an identifier: -1

Criteria description: Each record in the delivered dataset contained an identifier, generally the SHPO ID. This allows a user to contact the SHPO office with a reference in case more information is needed. If a SHPO ID is not available for the record, then an Agency ID should be available. These identifiers also assist the BLM in creating and maintaining the CRM ID fields. Features without an identifier will not be included in the National Dataset.

- d. Features do not exhibit geometry or topology issues: -1

Criteria description: Features did not exhibit geometry or topology issues. Severe issues cause errors in geoprocessing during the ETL processes. These issues also may negatively impact the analytical and reporting services in the web application.

1,099 Investigation SHPO IDs and 1,099 Resource SHPO IDs occurred in the data more than once. Most of these appeared to be multipart feature. A complete list of SHPO IDs which occurred more than once in the data can found under the DataMigration folder.

Recommendation: Request and review these SHPO identifiers. If these features are multipart, perform dissolves, where possible. It is understood some multipart features contain different attribute information.

174 Investigation records did not have an assigned SHPO ID nor Agency ID.

Recommendation: Request and review these records. Assign a unique identifier for these in the 2020 deliverable.

The delivered data contained bad spatial extents. Additionally, 1 feature in the Investigation feature class contained invalid geometry and was not included in the National Dataset. These caused errors during the ETL process.

Recommendation: Run the Add Spatial Index and Repair Geometry tools in ArcMap prior to the 2020 delivery.

5.11 Wyoming

Wyoming participated in the 2019 NCRIMS update cycle and delivered data in October of 2019. The targeted acquisition date is September, annually. Details on resolutions observed in the 2019 deliverable based on the 2018 recommendations are provided in Appendix C. The following section quantifies the compatibility of the 2019 deliverable and provides recommendations for improvement.

Recommendations for Future Deliverables (5)

Total Score: 0.75

A. Organization: 1 (2 out of 2)

- a. The SHPO/ State data was delivered in a file geodatabase: 1

Criteria description: The data was delivered in a file geodatabase. The use of shapefiles is not recommended for deliveries due to storage constraints. It is recommended that if tables are included in the data release, they be included inside the file geodatabase. A file geodatabase provides a consistent delivery mechanism for state data and is considered a best practice as it is optimized and scalable.

- b. Two feature classes exist in the data deliverable: 1

Criteria description: The file geodatabase contained one feature class for Resource features and one feature class for Investigation features. Some states may store data according to geometry. Before execution of the ETL, these geometries must be transformed and merged into a single, usable geometry for the national database. Additionally, other states may store feature information in multiple feature classes based on status. This introduces multiple schemas which need to be in agreement before the automate ETL is executed.

No critical issues were observed regarding how the data was organized.

B. Data Input: 0.25 (1 out of 4)

- a. Redundant fields do not exist in the delivered data: 1

Criteria description: The data did not contain redundant fields. Redundant fields store the same information and can cause confusion during migration, especially if discrepancies exist between the crosswalk and delivered data.

- b. Records in the data are tracked: 1

Criteria description: The data contained a date field or an alternative field to track the most recent visit to a site. A date field or a system of tracking when the record was input or when the site was visited is recommended, especially if legacy records exist in the data. This helps narrow down the records to the most recent available. If reconciliation of legacy records is not required, the state/SHPO will receive a 1 for this criterion.

- c. The data includes a field for each CRM Standard field: 0

Criteria description: The data was complete across the CRM fields. It is recommended that states maintain data fields which exist in the CRM schema. States that do so are considered complete across the CRM fields. This improves the effectiveness of the National Dataset.

- d. The delivered data has normalized and/or consistent values: -1

Criteria description: The data contained normalized/ consistent values. Normalized values are extremely important for an effective ETL. Normalized values are considered values which fall within a domain or range of values. For example, consider a NRHP eligibility field. If the field contains multiple versions of the value "No" (i.e. "No", "nO", "n0") then the values are not consistent. Likewise, if multiple values exist for ineligibility (i.e. "No", "Not Eligible", "Ineligible") then the field is not normalized.

The Investigation_BLM feature class included all required CRM fields except for Investigation Authority and Investigation NEPA Identifier. The Site_BLM feature class maintained all required CRM fields.

Recommendation: Add INVSTGTN_AUTH and INVSTGTN_NEPA_ID fields.

Fields crosswalked to domain-restricted CRM fields contained mostly normalized values except for values in the Primary Category Name (RSRCE_PRMRY_CTGRY_NM) and Spatial Collection Method (RSRCE_SPTL_CLCTN_MTHD) fields. These fields contained over 3,000 different values combined, hence the default value "Unknown" was utilized in the 2019 National Dataset.

Recommendation: Normalize these values in these fields for the 2020 delivery.

Target 1,000 unique combinations.

C. Maintenance: -0.5 (-2 out of 4)

- a. Legacy and duplicate records do not exist: -1

Criteria description: The data did not contain legacy or duplicate records. Legacy records are considered as any record of a previous visit to the site or survey which has not been removed from

the data delivery. Duplicate records are considered records which appear in the dataset more than once and contain the same information and geometry.

b. **Legacy or duplicate records can be quickly reconciled: 1**

Criteria description: The legacy or duplicate records were able to be queried and easily removed (if they exist). Often times, without a tracking field, it is difficult to determine if the record is multipart, legacy, or truly duplicate. This is especially true if two records with the same identifier are located on top of one another.

c. **Each record contains an identifier: -1**

Criteria description: Each record in the delivered dataset contained an identifier, generally the SHPO ID. This allows a user to contact the SHPO office with a reference in case more information is needed. If a SHPO ID is not available for the record, then an Agency ID should be available. These identifiers also assist the BLM in creating and maintaining the CRM ID fields. Features without an identifier will not be included in the National Dataset.

d. **Features do not exhibit geometry or topology issues: -1**

Criteria description: Features did not exhibit geometry or topology issues. Severe issues cause errors in geoprocessing during the ETL processes. These issues also may negatively impact the analytical and reporting services in the web application.

530 Investigation SHPO IDs and 1,114 Resource SHPO IDs occurred in the data more than once. Most of these appeared to be multipart features and legacy records.

Additionally, only the most recent record should be included for the 2020 deliverable. A complete list of SHPO IDs which occurred more than once in the data can found under the DataMigration folder.

Recommendation: Use the Dissolve tool in ArcMap to create multipart features as described in Appendix A.

152,744 Investigation records and 25 Resource records did not contain an assigned SHPO ID. Of the 152,744 Investigation records without a SHPO ID, 45,879 did not contain an Agency Identifier. Of the 25 Resource records, 23 did not have an Agency Identifier.

Recommendation: Request and review these records. Assign a unique identifier for these in the 2020 deliverable.

The Wyoming deliverable contained 67,759 Investigation features and 3,376 Resource features with self-intersections. 1 Investigation feature contained invalid geometry and was unable to be added to the National Dataset. See Appendix A for the OID and WYCRONUM of the skipped feature.

Recommendation: Run the Repair Geometry tool in ArcMap prior to the 2020 delivery.

APPENDIX A: 2019 State Toolsets and Workflows

The following describes the general workflow used as part the NCRIMS ETL procedure for each state.

Alaska

General Workflow and Tool Parameters

1. Open the *AK\OriginalDelivery* folder in ArcMap. Create a new file geodatabase in this folder titled *AK_final2019.gdb*.



2. Use the Project tool in ArcMap to reproject the 3 delivered shapefiles. See the parameters below. Repeat this for each shapefile.
 - a. Input Dataset or Feature Class: *<shapefile>*
 - b. Input Coordinate System: *Leave default.*
 - c. Output Dataset or Feature Class:
AK\OriginalDelivery\AK_final2019.gdb\<shapefile name>
 - d. Output Coordinate System: *NAD_1983_Alaska_Albers*
 - e. Geographic Transformation: *Leave default*
 - f. Preserve shape: *Uncheck*
 - g. Maximum Offset Deviation: *Leave default*
3. In ArcMap, buffer the projected line and point feature classes in the *AK_final2019.gdb*. Use a buffer distance of 15-meters and name the output *<feature class name>_buffered*.
4. Use the Merge tool to merge the buffered and polygon feature classes. Name the output *ROOT_AHRS_spatial_merged*.
5. Join the 3 delivered tables to the *ROOT_AHRS_spatial_merged* feature class and export it as *ROOT_AHRS_spatial_final*.
6. In ArcMap, open the NCRIMS Data Migration toolbox and click on NCRIMS ETL. This tool performs the ETL process for migrating the state-delivered data into a NCRIMS staging dataset. Input the following parameters:
 - a. Input Process: *Migrate Values*
 - b. Did the state require a domain crosswalk?: *Leave unchecked.*
 - c. Did the state require a clip to the SMA buffer?: *Leave unchecked.*
 - d. Original Delivery Folder: *Data\2019_Data\AK\OriginalDelivery*
 - e. Config GDB: *ETLtool\Config\crosswalk_config.gdb*
 - f. Investigation Input(s): *Leave empty*
 - g. Resource Input(s): *Data\2019_Data\AK\OriginalDelivery\AK_final2019.gdb\ROOT_AHRS_spatial_final*
7. Open the schema crosswalks under the DataMigration folder. In ArcMap, expand the folders and drag in the geodatabase tables. These tables are used to control the crosswalk of state or SHPO fields to CRM fields. Click the empty cell in the table and select the appropriate match for each CRM field. If a match for the CRM field does not exist, leave the value Null. If the CRM field requires concatenation of two fields, use the Field Calculator to type in the field names separating them by a semi-colon. Do not use spaces.
8. In ArcMap, open the NCRIMS Data Migration toolbox and click on Configure Crosswalk a second time. Input the following parameters:
 - a. Input Process: *Configure domain crosswalk*
 - b. Original Delivery Folder: *Data\2019_Data\AK\OriginalDelivery*
 - c. Config GDB: *ETLtool\Config\crosswalk_config.gdb*
 - d. Investigation Input(s): *Leave empty*



- e. Resource Input(s): *Data\2019_Data\AK\OriginalDelivery\AK_final2019.gdb\ROOT_AHRS_spatial_final*
9. Open the domain crosswalks under the DataMigration folder. A folder is present for each delivered feature class. In ArcMap, expand the folders and drag in the geodatabase tables. These tables are used to control the crosswalk of state or SHPO values to CRM domain values. Open each table and populate the NCRIMS_Value field with the appropriate corresponding domain value for each state/SHPO value. If the state/SHPO value is Null, leave the NCRIMS_Value Null.
10. In ArcMap, open the NCRIMS Data Migration toolbox and click on Migrate Values. This tool crosswalk the state/SHPO values to the CRM staging geodatabase. Input the following parameters:
 - a. Input Process: *Migrate Values*
 - b. Did the state require a domain crosswalk?: *Check.*
 - c. Did the state require a clip to the SMA buffer?: *Leave unchecked.*
 - d. Original Delivery Folder: *Data\2019_Data\AK\OriginalDelivery*
 - e. Config GDB: *ETLtool\Config\crosswalk_config.gdb*
 - f. Investigation Input(s): *Leave empty*
 - g. Resource Input(s): *Data\2019_Data\AK\OriginalDelivery\AK_final2019.gdb\ROOT_AHRS_spatial_final*
11. QC the CRM_Resources dataset and append it to the TargetSchema dataset using the Data Management > General > Append tool in ArcMap. See below for the appropriate parameters.
 - a. Input Datasets: *Data\2019_Data\AZ\Staging\Staging.gdb\CRM_Resources*
 - b. Target Dataset: *Data\2019_Data\TargetSchema\National_CRM.gdb\CRM_Resources*
 - c. Schema Type: *NO TEST*

Arizona

General Workflow and Tool Parameters

1. Open the AZ\OriginalDelivery folder in Windows File Explorer. Make a copy of the AZ_CRM_10119.gdb and rename it AZ_CRM_10119_backup.gdb.
2. Open the AZ\OriginalDelivery\AZ_CRM_10119.gdb\CRM_Resources attribute table. Add a field called “RSRCE_PRMRY_CTGRY_NM” of text type and length 30. Using Field Calculator, set the new field equal to RSRCE_PRMRY_CAT_NM. Click OK.
3. In ArcMap, open the NCRIMS Data Migration toolbox and click on NCRIMS ETL. This tool performs the ETL process for migrating the state-delivered data into a NCRIMS staging dataset. Input the following parameters:
 - a. Input Process: *Migrate Values*
 - b. Did the state require a domain crosswalk?: *Leave unchecked.*
 - c. Did the state require a clip to the SMA buffer?: *Leave unchecked.*
 - d. Original Delivery Folder: *Data\2019_Data\AZ\OriginalDelivery*



- e. Config GDB: *ETLtool\Config\crosswalk_config.gdb*
 - f. Investigation Input(s):
Data\2019_Data\AZ\OriginalDelivery\AZ_CRM_10119.gdb\CRM_Investigations
 - g. Resource Input(s):
Data\2019_Data\AZ\OriginalDelivery\AZ_CRM_10119.gdb\CRM_Resources
4. In ArcMap, use the Select by Attributes button to locate and select the orphaned domains values below. Once selected, use the Field Calculator to change the original value to the CRM value shown to the right. Orphaned domain values are any values included in the delivered dataset that do not abide with the CRM domain values. The values will thus need to be edited to comply with the CRM standard schema. These may also be located under *DataMigration\orphaned_domains*
- a. Feature class: CRM_Investigations
 - i. *AZSite – Other*
 - ii. *Section 107 – Section 106*
 - iii. *unknown – Unknown*
 - iv. *ClassII and Class III – Class II and Class III*
 - v. *Class iII Inventory – Class III Inventory*
 - vi. *Collections/Non-Field Study – Collections Non-Field Study*
 - vii. *FLPMA – Other*
 - b. Feature class: CRM_Resources
 - i. *Protohistoric – Ethnohistoric*
 - ii. *BLM – AZ BLM*
 - iii. *Historic Native American – Ethnohistoric*
 - iv. *Not Available – NA*
 - v. *AZSite – Other*
 - vi. *no – No*
 - vii. *Digitized –USGS2400 – Digitized - USGS2400*
 - viii. *<Null> - Unknown, NA, Not Specified (The default value depending on the assigned domain)*
 - ix. *Consultant Recommendation – Consultant*
 - x. *SHPO Concurrence – SHPO*
 - xi. *GPS – GPS (*Remove trailing space in original value)*
 - xii. *Eligible (D Only) – Eligible D Only*
 - xiii. *HIstoric – Historic*
 - xiv. *Eligible ((A,B,and/or C) and D) – Eligible ((A, B and/or C) and D)*
 - xv. *BLM-Legacy – AZ BLM*
 - xvi. *Not Applicable – NA*
 - xvii. *SHPO Determination – SHPO*
 - xviii. *Imminently Threatened – Threatened*
5. Review the output tables from the geometry validation process found under *DataMigration\check_geometry_log*. The following list shows topology errors critical to geoprocessing in the ETL. The steps proceeding the dataset names and errors were used to reconcile the topology errors. These processes were completed manually. For more information on type of topology errors, see the ESRI documentation here:



<http://desktop.arcgis.com/en/arcmap/10.3/tools/data-management-toolbox/check-geometry.htm>

- a. CRM_Investigations – *null geometry (OID: 858)*
 - b. CRM_Resources – *null geometry (OIDs: 941, 942, 943, 944, 945, 4180, 5395, 5396, 5698, 5699, 5826, 5827, 5846, 5847, 5857, 5882, 5883)*
 - i. No manual reconciliation required. The ETL script ignores these features; these features were not migrated into the CRM standard dataset. See Section 2.3.2 for more information.
6. QC the CRM_Resources dataset and append it to the TargetSchema dataset using the Data Management > General > Append tool in ArcMap. See below for the appropriate parameters.
 - a. Input Datasets: *Data\2019_Data\AZ\Staging\Staging.gdb\CRM_Resources*
 - b. Target Dataset: *Data\2019_Data\TargetSchema\National_CRM.gdb\CRM_Resources*
 - c. Schema Type: *NO TEST*
7. QC the CRM_Investigations dataset and append it to the TargetSchema dataset using the Data Management > General > Append tool in ArcMap. See below for the appropriate parameters.
 - a. Input Datasets: *Data\2019_Data\AZ\Staging\Staging.gdb\CRM_Investigations*
 - b. Target Dataset: *Data\2019_Data\TargetSchema\National_CRM.gdb\CRM_Investigations*
 - c. Schema Type: *NO TEST*

California

General Workflow and Tool Parameters

1. Open the *CA\OriginalDelivery* folder in Windows File Explorer. Make a copy of the *CaCRM01022020.gdb* and rename it *CaCRM01022020_backup.gdb*.
2. Join the *CaCRM01022020_backup.gdb\CRM_InvProtocol_tbl* table to the *CaCRM01022020_backup.gdb\CRM_Investigations\inv_inproc_proxy* feature class on the *PK_GUID* field.
3. Export the joined feature class as *inv_inproc_proxy_joined* to the *CaCRM01022020_backup.gdb*
4. Join the *CaCRM01022020_backup.gdb\CRM_ResAttribute_tbl* table to the *CaCRM01022020_backup.gdb\CRM_Investigations\res_inproc_proxy* feature class on the *PK_GUID* field.
5. Export the joined feature class as *res_inroc_prox_joined* to the *CaCRM01022020_backup.gdb*



6. In ArcMap, open the NCRIMS Data Migration toolbox and click on NCRIMS ETL. This tool performs the ETL process for migrating the state-delivered data into a NCRIMS staging dataset. Input the following parameters:
 - a. Input Process: *Configure schema crosswalk*
 - b. Did the state require a domain crosswalk?: *Leave unchecked.*
 - c. Did the state require a clip to the SMA buffer?: *Leave unchecked.*
 - d. Original Delivery Folder: *Data\2019_Data\CA\OriginalDelivery*
 - e. Config GDB: *ETLtool\Config\crosswalk_config.gdb*
 - f. Investigation Input(s): *Data\2019_Data\CA\OriginalDelivery\CaCRM01022020_backup.gdb\inv_inproc_proxy*
 - g. Resource Input(s): *Data\2019_Data\CA\OriginalDelivery\CaCRM01022020_backup.gdb\res_inproc_proxy*
7. Open the schema crosswalks under the DataMigration folder. In ArcMap, expand the folders and drag in the geodatabase tables. These tables are used to control the crosswalk of state or SHPO fields to CRM fields. Click the empty cell in the table and select the appropriate match for each CRM field. If a match for the CRM field does not exist, leave the value Null. If the CRM field requires concatenation of two fields, use the Field Calculator to type in the field names separating them by a semi-colon. Do not use spaces.
8. In ArcMap, open the NCRIMS Data Migration toolbox and click on NCRIMS ETL a second time. Input the following parameters:
 - a. Input Process: *Configure domain crosswalk*
 - b. Original Delivery Folder: *Data\2019_Data\CA\OriginalDelivery*
 - c. Config GDB: *ETLtool\Config\crosswalk_config.gdb*
 - d. Investigation Input(s): *Data\2019_Data\CA\OriginalDelivery\CaCRM01022020_backup.gdb\inv_inproc_proxy*
 - h. Resource Input(s): *Data\2019_Data\CA\OriginalDelivery\CaCRM01022020_backup.gdb\res_inproc_proxy*
9. Review the output tables from the geometry validation process. These are found under DataMigration\check_geometry_log. The following list shows topology errors critical to geoprocessing in the ETL. The steps proceeding the dataset names and errors were used to reconcile the topology errors. These processes were completed manually. For more information on type of topology errors, see the ESRI documentation here: <http://desktop.arcgis.com/en/arcmap/10.3/tools/data-management-toolbox/check-geometry.htm>
 - a. Inv_inproc_proxy – *self-intersections*
 - b. Inv_inproc_proxy – *bad dataset extent*
 - c. Res_inproc_proxy – *self-intersections*
 - d. Res_inproc_proxy – *bad dataset extent*
 - i. In ArcMap, open the tool Repair Geometry under the Data Management > Features.
 - ii. Drag the corresponding staging feature class copy into the Input Features parameter. These copies are found under the Staging folder > Staging.gdb.
 - iii. Open the Add Spatial Index tool under Data Management > Indexes.



- iv. Drag the corresponding staging feature class copy into the Input Features parameter. These copies are found under the Staging folder > Staging.gdb.
 - v. Click OK. Repeat the above steps for each listed feature class's copy.
10. Open the domain crosswalks under the DataMigration folder. A folder is present for each delivered feature class. In ArcMap, expand the folders and drag in the geodatabase tables. These tables are used to control the crosswalk of state or SHPO values to CRM domain values. Open each table and populate the NCRIMS_Value field with the appropriate corresponding domain value for each state/SHPO value. If the state/SHPO value is Null, leave the NCRIMS_Value Null.
11. In ArcMap, open the NCRIMS Data Migration toolbox and click on NCRIMS ETL. This tool crosswalks the state/SHPO values to the CRM staging geodatabase. Input the following parameters:
- a. Input Process: *Migrate Values*
 - b. Did the state require a domain crosswalk?: *Check*.
 - c. Did the state require a clip to the SMA buffer?: *Leave unchecked*.
 - d. Original Delivery Folder: *Data\2019_Data\CA\OriginalDelivery*
 - e. Config GDB: *ETLtool\Config\crosswalk_config.gdb*
 - f. Investigation Input(s): *2019_Data\CA\OriginalDelivery\CaCRM01022020_backup.gdb\inv_inproc_proxy*
 - g. Resource Input(s): *2019_Data\CA\OriginalDelivery\CaCRM01022020_backup.gdb\res_inproc_proxy*
12. QC the CRM_Resources dataset and append it to the TargetSchema dataset using the Data Management > General > Append tool in ArcMap. See below for the appropriate parameters.
- a. Input Datasets: *Data\2019_Data\CA\Staging\Staging.gdb\CRM_Resources*
 - b. Target Dataset: *Data\2019_Data\TargetSchema\National_CRM.gdb\CRM_Resources*
 - c. Schema Type: *NO TEST*
13. QC the CRM_Investigations dataset and append it to the TargetSchema dataset using the Data Management > General > Append tool in ArcMap. See below for the appropriate parameters.
- a. Input Datasets: *Data\2019_Data\CA\Staging\Staging.gdb\CRM_Investigations*
 - b. Target Dataset: *Data\2019_Data\TargetSchema\National_CRM.gdb\CRM_Investigations*
 - c. Schema Type: *NO TEST*

Colorado

General Workflow and Tool Parameters

- 1. In 2019_Data directory, create the following folders under the CO directory.
 - a. DataMigration
 - b. Staging

2. Under the DataMigration folder, create a subfolder called check_geometry_log.
3. Under the Staging folder, create a file geodatabase called Staging.gdb
4. Copy the CRM_Investigations and CRM_Resources feature classes in the BLM_CO_NCRIMS_Crosswalk_1911151050_1.gdb to the Staging.gdb. Rename the CRM_Investigations copy to CRM_Investigations_copy. Rename the CRM_Resources copy to CRM_Resources_copy.
5. In ArcMap, open the Check Geometry tool under Data Management > Features. Input the following parameters.
 - a. In features: *2019_Data\CO\Staging\Staging.gdb\CRM_Investigations_copy*
 - b. Out table:
2019_Data\CO\DataMigration\check_geometry_log\CRM_Investigations_copy_11252019.csv
6. Repeat step with the following parameters
 - a. In features: *2019_Data\CO\Staging\Staging.gdb\CRM_Resources_copy*
 - b. Out table:
2019_Data\CO\DataMigration\check_geometry_log\CRM_Resources_copy_11252019.csv
7. Review the output tables from the check geometry tool. The following list shows topology errors critical to geoprocessing in the ETL. The steps proceeding the dataset names and errors were used to reconcile the topology errors. These processes were completed manually. For more information on type of topology errors, see the ESRI documentation here: <http://desktop.arcgis.com/en/arcmap/10.3/tools/data-management-toolbox/check-geometry.htm>
 - a. CRM_Investigations_copy – *self-intersections*
 - b. CRM_Resources_copy – *self-intersections*
 - i. In ArcMap, open the tool Repair Geometry under the Data Management > Features.
 - ii. Drag the corresponding staging feature class copy into the Input Features parameter. These copies are found under the Staging folder > Staging.gdb.
8. Using Select by Attributes, review the values under CRM domain-restricted fields. The following values were discovered and changed to the corresponding value to the right. These values were revised using the Field Calculator in ArcMap.
 - a. RSRCE_PRMRY_CTGRY_NM:
 - i. NULL – *Unknown*
 - ii. Temporary camp – *Temporary Camp*
 - iii. Water Storage and Conveyance – *Wtr Strg Cnvync*
 - b. RSRCE_NRHP_ELGBL_STTS:
 - i. Unknown – *Undetermined*
 - c. RSRCE_NRHP_CRTRA:
 - i. NULL – *Not Specified*

- d. RSRCE_NRHP_AUTH_NM:
 - i. Keeper of the National Register – *Keeper of the Natl Register*
 - e. RSRCE_CNDTN_ASSMNT:
 - i. Imminently Threatened – *Threatened*
 - ii. Not Applicable – *NA*
9. Append the CRM_Resources copy it to the TargetSchema dataset using the Data Management > General > Append tool in ArcMap. See below for the appropriate parameters.
- a. Input Datasets: *Data\2019_Data\CO\Staging\Staging.gdb\CRM_Resources_copy*
 - b. Target Dataset: *Data\2019_Data\TargetSchema\National_CRM.gdb\CRM_Resources*
 - c. Schema Type: *NO TEST*
10. Append the CRM_Investigations copy to the TargetSchema dataset using the Data Management > General > Append tool in ArcMap. See below for the appropriate parameters.
- a. Input Datasets: *Data\2019_Data\CO\Staging\Staging.gdb\CRM_Investigations_copy*
 - b. Target Dataset: *Data\2019_Data\TargetSchema\National_CRM.gdb\CRM_Investigations*
 - c. Schema Type: *NO TEST*

Idaho

General Workflow and Tool Parameters

1. Open the *ID\OriginalDelivery* folder in Windows File Explorer. Make a copy of the *ID_National_CRM_2019.gdb* and rename it *ID_National_CRM_2019_backup.gdb*.
2. In ArcMap, open the NCRIMS Data Migration toolbox and click on NCRIMS ETL. This tool performs the ETL process for migrating the state-delivered data into a NCRIMS staging dataset. Input the following parameters:
 - a. Input Process: *Migrate Values*
 - b. Did the state require a domain crosswalk?: *Leave unchecked.*
 - c. Did the state require a clip to the SMA buffer?: *Leave unchecked.*
 - d. Original Delivery Folder: *Data\2019_Data\ID\OriginalDelivery*
 - e. Config GDB: *ETLtool\Config\crosswalk_config.gdb*
 - f. Investigation Input(s): *Data\2019_Data\ID\OriginalDelivery\ID_National_CRM_2019.gdb\ID_CRM_Investigations*
 - g. Resource Input(s): *Data\2019_Data\ID\OriginalDelivery\ID_National_CRM_2019.gdb\ID_CRM_Resources*
3. In ArcMap, use the Select by Attributes button to locate and select the orphaned domains values below. Once selected, use the Field Calculator to change the original value to the CRM value shown to the right. Orphaned domain values are any values included in the

delivered dataset that do not abide with the CRM domain values. The values will thus need to be edited to comply with the CRM standard schema. These may also be located under *DataMigration\orphaned_domains*

- a. Feature class: CRM_Investigations
 - i. *BLM – ID BLM*
 - ii. *2018-1244 – ID BLM*
 - iii. *IPCO – ID BLM*
 - iv. *Class II – Class II Inventory*
 - v. *Class III – Class III Inventory*
 - b. Feature class: CRM_Resources
 - i. *GPS – GPS (*Remove trailing space in original value)*
 - ii. *unevaluated – Undetermined*
 - iii. *Ineligible – No*
 - iv. *Eligible – Yes*
 - v. *Good – Stable*
4. QC the CRM_Resources dataset and append it to the TargetSchema dataset using the Data Management > General > Append tool in ArcMap. See below for the appropriate parameters.
- a. Input Datasets: *Data\2019_Data\ID\Staging\Staging.gdb\CRM_Resources*
 - b. Target Dataset:
Data\2019_Data\TargetSchema\National_CRM.gdb\CRM_Resources
 - c. Schema Type: *NO TEST*
5. QC the CRM_Investigations dataset and append it to the TargetSchema dataset using the Data Management > General > Append tool in ArcMap. See below for the appropriate parameters.
- a. Input Datasets: *Data\2019_Data\ID\Staging\Staging.gdb\CRM_Investigations*
 - b. Target Dataset:
Data\2019_Data\TargetSchema\National_CRM.gdb\CRM_Investigations
 - c. Schema Type: *NO TEST*

Montana

General Workflow and Tool Parameters

1. In ArcMap, open the NCRIMS Data Migration toolbox and click on Configure Crosswalk. This tool generates crosswalks for the user to map the state/SHPO schemas to the CRM schema or to map the state/SHPO values to the CRM domain values. Input the following parameters:
 - a. Input Process: *Configure schema crosswalk*
 - b. Original Delivery Folder: *Data\2019_Data\MT\OriginalDelivery*
 - c. Config GDB: *ETLtool\Config\crosswalk_config.gdb*
 - d. Investigation Input(s):
2019_Data\MT\OriginalDelivery\MT_SHPO_2020_gdb_121619\MT_SHPO_2020.gdb\Survey_Poly

- e. Resource Input(s):
 2019_Data\MT\OriginalDelivery\MT_SHPO_2020_gdb_121619\MT_SHPO_2020.gdb\Site_Poly
2. Open the schema crosswalks under the DataMigration folder. In ArcMap, expand the folders and drag in the geodatabase tables. These tables are used to control the crosswalk of state or SHPO fields to CRM fields. Click the empty cell in the table and select the appropriate match for each CRM field. If a match for the CRM field does not exist, leave the value Null. If the CRM field requires concatenation of two fields, use the Field Calculator to type in the field names separating them by a semi-colon. Do not use spaces.
3. In ArcMap, open the NCRIMS Data Migration toolbox and click on Configure Crosswalk a second time. Input the following parameters:
 - a. Input Process: *Configure domain crosswalk*
 - b. Original Delivery Folder: *Data\2019_Data\MT\OriginalDelivery*
 - c. Config GDB: *ETLtool\Config\crosswalk_config.gdb*
 - d. Investigation Input(s):
 2019_Data\MT\OriginalDelivery\MT_SHPO_2020_gdb_121619\MT_SHPO_2020.gdb\Survey_Poly
 - e. Resource Input(s):
 2019_Data\MT\OriginalDelivery\MT_SHPO_2020_gdb_121619\MT_SHPO_2020.gdb\Site_Poly
4. Review the output tables from the geometry validation process. These are found under DataMigration\check_geometry_log. The following list shows topology errors critical to geoprocessing in the ETL. The steps proceeding the dataset names and errors were used to reconcile the topology errors. These processes were completed manually. For more information on type of topology errors, see the ESRI documentation here:
<http://desktop.arcgis.com/en/arcmap/10.3/tools/data-management-toolbox/check-geometry.htm>
 - a. Survey_Poly – *self-intersections*
 - b. Site_Poly – *self-intersections, incorrect ring ordering*
 - i. In ArcMap, open the tool Repair Geometry under the Data Management > Features.
 - ii. Drag the corresponding staging feature class copy into the Input Features parameter. These copies are found under the Staging folder > Staging.gdb.
5. Open the domain crosswalks under the DataMigration folder. A folder is present for each delivered feature class. In ArcMap, expand the folders and drag in the geodatabase tables. These tables are used to control the crosswalk of state or SHPO values to CRM domain values. Open each table and populate the NCRIMS_Value field with the appropriate corresponding domain value for each state/SHPO value. If the state/SHPO value is Null, leave the NCRIMS_Value Null.
6. In ArcMap, open the NCRIMS Data Migration toolbox and click on Migrate Values. This tool crosswalk the state/SHPO values to the CRM staging geodatabase. Input the following parameters:

- a. Input Process: *Migrate Values*
 - b. Did the state require a domain crosswalk?: *Check.*
 - c. Did the state require a clip to the SMA buffer?: *Leave unchecked.*
 - d. Original Delivery Folder: *Data\2019_Data\MT\OriginalDelivery*
 - e. Config GDB: *ETLtool\Config\crosswalk_config.gdb*
 - f. Investigation Input(s):
2019_Data\MT\OriginalDelivery\MT_SHPO_2020_gdb_121619\MT_SHPO_2020.gdb\Survey_Poly
 - g. Resource Input(s):
2019_Data\MT\OriginalDelivery\MT_SHPO_2020_gdb_121619\MT_SHPO_2020.gdb\Site_Poly
7. QC the CRM_Resources dataset and append it to the TargetSchema dataset using the Data Management > General > Append tool in ArcMap. See below for the appropriate parameters.
 - a. Input Datasets: *Data\2019_Data\MT\Staging\Staging.gdb\CRM_Resources*
 - b. Target Dataset:
Data\2019_Data\TargetSchema\National_CRM.gdb\CRM_Resources
 - c. Schema Type: *NO TEST*
 8. QC the CRM_Investigations dataset and append it to the TargetSchema dataset using the Data Management > General > Append tool in ArcMap. See below for the appropriate parameters.
 - a. Input Datasets: *Data\2019_Data\MT\Staging\Staging.gdb\CRM_Investigations*
 - b. Target Dataset:
Data\2019_Data\TargetSchema\National_CRM.gdb\CRM_Investigations
 - c. Schema Type: *NO TEST*

Nevada

General Workflow and Tool Parameters

6. In ArcMap, open the NCRIMS Data Migration toolbox and click on Configure Crosswalk. This tool generates crosswalks for the user to map the state/SHPO schemas to the CRM schema or to map the state/SHPO values to the CRM domain values. Input the following parameters:
 - a. Input Process: *Configure schema crosswalk*
 - b. Original Delivery Folder: *Data\2019_Data\NV\OriginalDelivery*
 - c. Config GDB: *ETLtool\Config\crosswalk_config.gdb*
 - d. Investigation Input(s):
Data\2019_Data\NV\OriginalDelivery\NVCRI_92419.gdb\Archeo_Inventory,
Data\2019_Data\NV\OriginalDelivery\NVCRI_92419.gdb\In_Process_Inventory,
Data\2019_Data\NV\OriginalDelivery\NVCRI_92419.gdb\Rural_Archit_Inventor
y,
 - e. Resource Input(s):
Data\2019_Data\NV\OriginalDelivery\NVCRI_92419.gdb\Archeo_Sites,
Data\2019_Data\NV\OriginalDelivery\NVCRI_92419.gdb\In_Process_Sites,
Data\2019_Data\NV\OriginalDelivery\NVCRI_92419.gdb\Rural_Archit_Sites,

Data\2019_Data\NV\OriginalDelivery\NVCRI_92419.gdb\National_Register

7. Open the schema crosswalks under the DataMigration folder. In ArcMap, expand the folders and drag in the geodatabase tables. These tables are used to control the crosswalk of state or SHPO fields to CRM fields. Click the empty cell in the table and select the appropriate match for each CRM field. If a match for the CRM field does not exist, leave the value Null. If the CRM field requires concatenation of two fields, use the Field Calculator to type in the field names separating them by a semi-colon. Do not use spaces.
8. In ArcMap, open the NCRIMS Data Migration toolbox and click on Configure Crosswalk a second time. Input the following parameters:
 - a. Input Process: *Configure domain crosswalk*
 - b. Original Delivery Folder: *Data\2019_Data\NV\OriginalDelivery*
 - c. Config GDB: *ETLtool\Config\crosswalk_config.gdb*
 - d. Investigation Input(s):
Data\2019_Data\NV\OriginalDelivery\NVCRI_92419.gdb\Archeo_Inventory,
Data\2019_Data\NV\OriginalDelivery\NVCRI_92419.gdb\In_Process_Inventory,
Data\2019_Data\NV\OriginalDelivery\NVCRI_92419.gdb\Rural_Archit_Inventor
y,
 - e. Resource Input(s):
Data\2019_Data\NV\OriginalDelivery\NVCRI_92419.gdb\Archeo_Sites,
Data\2019_Data\NV\OriginalDelivery\NVCRI_92419.gdb\In_Process_Sites,
Data\2019_Data\NV\OriginalDelivery\NVCRI_92419.gdb\Rural_Archit_Sites,
Data\2019_Data\NV\OriginalDelivery\NVCRI_92419.gdb\National_Register
9. Review the output tables from the geometry validation process. These are found under DataMigration\check_geometry_log. The following list shows topology errors critical to geoprocessing in the ETL. The steps proceeding the dataset names and errors were used to reconcile the topology errors. These processes were completed manually. For more information on type of topology errors, see the ESRI documentation here:
<http://desktop.arcgis.com/en/arcmap/10.3/tools/data-management-toolbox/check-geometry.htm>
 - a. Archeo_Inventory – *bad dataset extent*
 - b. Archeo_Sites – *bad dataset extent*
 - c. In_Process_Inventory – *bad dataset extent*
 - d. In_Process_Sites – *bad dataset extent*
 - e. National_Register – *bad dataset extent*
 - f. Rural_Archit_Inventory – *bad dataset extent*
 - g. Rural_Archit_Resources – *bad dataset extent*
 - i. In ArcMap, open the tool Add Spatial Index under the Data Management > Indexes.
 - ii. Drag the corresponding staging feature class copy into the Input Features parameter. These copies are found under the Staging folder > Staging.gdb.
 - iii. Click OK. Repeat the above steps for each listed feature class's copy.
10. Open the domain crosswalks under the DataMigration folder. A folder is present for each delivered feature class. In ArcMap, expand the folders and drag in the geodatabase tables.



These tables are used to control the crosswalk of state or SHPO values to CRM domain values. Open each table and populate the NCRIMS_Value field with the appropriate corresponding domain value for each state/SHPO value. If the state/SHPO value is Null, leave the NCRIMS_Value Null.

7. In ArcMap, open the NCRIMS Data Migration toolbox and click on Migrate Values. This tool crosswalk the state/SHPO values to the CRM staging geodatabase. Input the following parameters:
 - a. Input Process: *Migrate Values*
 - b. Did the state require a domain crosswalk?: *Check*.
 - c. Did the state require a clip to the SMA buffer?: *Leave unchecked*.
 - d. Original Delivery Folder: *Data\2019_Data\NV\OriginalDelivery*
 - e. Config GDB: *ETLtool\Config\crosswalk_config.gdb*
 - f. Investigation Input(s):
Data\2019_Data\NV\OriginalDelivery\NVCRI_92419.gdb\Archeo_Inventory,
Data\2019_Data\NV\OriginalDelivery\NVCRI_92419.gdb\In_Process_Inventory,
Data\2019_Data\NV\OriginalDelivery\NVCRI_92419.gdb\Rural_Archit_Inventor
y,
 - g. Resource Input(s):
Data\2019_Data\NV\OriginalDelivery\NVCRI_92419.gdb\Archeo_Sites,
Data\2019_Data\NV\OriginalDelivery\NVCRI_92419.gdb\In_Process_Sites,
Data\2019_Data\NV\OriginalDelivery\NVCRI_92419.gdb\Rural_Archit_Sites,
Data\2019_Data\NV\OriginalDelivery\NVCRI_92419.gdb\National_Register
9. In ArcMap, create a backup copy of the CRM_Investigations feature class. Then, open the Delete Identical tool under Data Management > General. Use the parameters below to execute the tool and minimize any detected duplicate features.
 - a. Input Dataset: *CRM_Investigations*
 - b. Field(s): *Select All*
 - c. XY Tolerance: *Leave default or blank*
 - d. Z Tolerance: *Leave default or blank*
10. In ArcMap, create a backup copy of the CRM_Resources feature class. Then, open the Delete Identical tool under Data Management > General. Use the parameters below to execute the tool and minimize any detected duplicate features.
 - a. Input Dataset: *CRM_Resources*
 - b. Field(s): *Select All*
 - c. XY Tolerance: *Leave default or blank*
 - d. Z Tolerance: *Leave default or blank*
11. QC the CRM_Resources dataset and append it to the TargetSchema dataset using the Data Management > General > Append tool in ArcMap. See below for the appropriate parameters.
 - a. Input Datasets: *Data\2019_Data\NV\Staging\Staging.gdb\CRM_Resources*
 - b. Target Dataset:
Data\2019_Data\TargetSchema\National_CRM.gdb\CRM_Resources
 - c. Schema Type: *NO TEST*



12. QC the CRM_Investigations dataset and append it to the TargetSchema dataset using the Data Management > General > Append tool in ArcMap. See below for the appropriate parameters.
 - a. Input Datasets: *Data\2019_Data\NV\Staging\Staging.gdb\CRM_Investigations*
 - b. Target Dataset:
Data\2019_Data\TargetSchema\National_CRM.gdb\CRM_Investigations
 - c. Schema Type: *NO TEST*
13. Use Select by Location to remove additional features in the National Dataset's CRM_Investigations and CRM_Resources that do not intersect BLM Surface Management Areas 100-meter buffer.

New Mexico

General Workflow and Tool Parameters

1. Create a file geodatabase in the *2019_Data\NM\OriginalDelivery* folder. Name the geodatabase *NM_2019Final.gdb*
2. Import the delivered *Sitebounds.shp* and *Surveys.shp* into the geodatabase created in step 1.
3. Dissolve the *Surveys* feature class using the Dissolve tool under Data Management > Generalization.

Parameters:

 - a. Input Features: *NM_2019Final.gdb\Surveys*
 - b. Output Feature Class: *NM_2019Final.gdb\Surveys_dissolved*
 - c. Dissolve_Fields: *NMCRIS_NUM, SREPORTTIT, PERF_AG, SCOMMENTS, LEAD_NUM1, LEAD_NUM2*
 - d. *Check the box* next to Create multipart features.
4. Import the *ACTIV.dbf* table into the *NM_Final2019.gdb*. Delete identical rows in the imported *ACTIV* table using the Delete Identical tool under Data Management > General > Delete Identical.

Parameters:

 - a. Input Dataset: *NM_2019Final.gdb\ACTIV*
 - b. Field(s): *NMCRIS_NUM, DFIELDENDDD*
5. Dissolve the *Sitebounds* feature class using the Dissolve tool under Data Management > Generalization.

Parameters:

 - a. Input Dataset: *NM_2019Final.gdb\Sitebounds*
 - b. Output Feature Class: *NM_2019Final.gdb\Sitebounds_dissolved*
 - c. Dissolve_Fields: *ArmsArchID, ARCH_OCCUP*
 - d. *Check the box* next to Create multipart features



6. Import the *ALIAS.dbf* table into the *NM_2019Final.gdb*. Add the following field to the imported table.
Parameters:
 - a. Field Name: *Alias_2*
 - b. Field Type: *Text*
 - c. Field Length: *2000*
7. Concatenate the row values using the Concatenate Row Values under the BLM CRM Data Migration toolbox > Additional Tools.
Parameters:
 - a. Input: *NM_2019Final.gdb\ALIAS*
 - b. Case Field: *LANUM*
 - c. Read from Field: *ALIAS*
 - d. Copy to Field: *Alias_2*
 - e. Delimiter: “, ”
8. Delete identical rows in the *ALIAS* table using the Delete Identical tool under Data Management > General > Delete Identical. Parameters:
 - a. Input Dataset: *NM_2019Final.gdb\ALIAS*
 - b. Field(s): *LANUM, Alias_2*
9. Delete the field titled *ALIAS*.
10. Add a field to the *ALIAS* table using the following parameters:
 - a. Field Name: *ALIAS*
 - b. Field Type: *Text*
 - c. Field Length: *255*
11. Right click on the newly added *ALIAS* field. Select Field Calculator and set the *ALIAS* field to equal *Alias_2*. Ignore any truncation warnings.
12. Import the *COMP.dbf* into the *NM_2019Final.gdb* and delete identical rows on the imported *COMP* table using the tool from steps 2 and 6.
Parameters:
 - a. Input Dataset: *NM_2019Final.gdb\COMP*
 - b. Field(s): *LANUM, COMPTYPE, REMARKS*
13. Add a field called *COMPPTYPE_2* and a field called *REMARKS_2*, both of type Text and length *2000*. Make a copy of the imported *COMP* table in the *NM_2019Final.gdb* and name it *COMP_1*
14. Repeat steps 7 – 11 for the *COMPTYPE* field in the imported *COMP* table.
15. Repeat steps 7 – 11 for the *REMARKS* field in the *COMP_1* table.



16. Join the *COMP_1* table to the *COMP* table on the *LANUM* field and use the Field Calculator to set *COMP.REMARKS_2 = COMP_1.REMARKS_2*
17. Use Field Calculator a second time to strip leading “, “ characters from the *COMPTYPE_2* and *REMARKS_2* fields. Use the following syntax:
 - a. `!<field name>!lstrip(", ")`
18. Import the *DOE.dbf* into the *NM_2019Final.gdb* and delete identical rows on the imported *DOE* table using the tool from steps 2 and 6.

Parameters:

 - a. Input Dataset: *NM_2019Final.gdb\DOE*
 - b. Field(s): *LANUM, SNREVAL, SAUTHORITY, BCRIT_A, BCRIT_B, BCRIT_C, BCRIT_D*
19. Use the Data Management > General > Sort tool to sort the records in the imported *DOE* table. See the parameters below.
 - a. Input Dataset: *NM_2019Final.gdb\DOE*
 - b. Output Dataset: *NM_2019Final.gdb\DOE_sorted*
 - c. Field(s): *LANUM, DEVELDATE*
20. Add a field called *queryField* of type Long Integer to the *DOE_sorted* table. Open the python window in ArcMap and paste the following code.


```
>>> _id_set = set([_row[0] for _row in arcpy.da.SearchCursor("DOE_sorted",
    "LANUM")])
>>> with arcpy.da.UpdateCursor("DOE_sorted", ["LANUM", "queryField"]) as
    _cursor:
...     for _row in _cursor:
...         if _row[0] in _id_set:
...             _row[1] = 9999
...             _cursor.updateRow(_row)
...             _id_set.remove(_row[0])
```
21. Use a definition query on the *DOE_sorted* table to show only NULL *queryField* values.
22. Open Data Management > Tables > Delete Rows. Use the *DOE_sorted* table as the input.
23. Use the Data Management > General > Sort tool to sort the records in the imported *VISIT* table. See the parameters below.
 - a. Input Dataset: *NM_2019Final.gdb\VISIT*
 - b. Output Dataset: *NM_2019Final.gdb\VISIT_sorted*
 - c. Field(s): *LANUM, VISIT_NUM (Descending)*
24. Add a field called *queryField* of type Long Integer to the *VISIT_sorted* table. Open the python window in ArcMap and paste the following code.


```
>>> _id_set = set([_row[0] for _row in
    arcpy.da.SearchCursor("VISIT_sorted", "LANUM")])
>>> with arcpy.da.UpdateCursor("VISIT_sorted", ["LANUM", "queryField"]) as
    _cursor:
...     for _row in _cursor:
```



```

...         if _row[0] in _id_set:
...             _row[1] = 9999
...             _cursor.updateRow(_row)
...             _id_set.remove(_row[0])

```

25. Use a definition query on the *VISIT_sorted* table to show only NULL *queryField* values.
26. Open Data Management > Tables > Delete Rows. Use the *VISIT_sorted* table as the input.
27. Join the *ACTIV* table to the *Surveys_dissolved* feature class on the *NMCRIS_NUM* field and export the joined feature class to the *NM_2019Final.gdb* as *Surveys_final*.
28. Join the following tables to the *Sitebounds_dissolved* feature class on the *ArmArchID* field and the field listed next to the table name.
 - a. ALIAS: LANUM
 - b. COMP: LANUM
 - c. DOE_sorted: LANUM
 - d. VISIT_sorted: LANUM
29. Export the joined feature class to the *NM_2019Final.gdb* as *Sitebounds_final*
30. Follow the workflow under Montana for *Surveys_final* and *Sitebounds_final* to append the data to the National Dataset.

Oregon

General Workflow and Tool Parameters

1. In ArcMap, open the NCRIMS Data Migration toolbox and click on Configure Crosswalk. This tool generates crosswalks for the user to map the state/SHPO schemas to the CRM schema or to map the state/SHPO values to the CRM domain values. Input the following parameters:
 - a. Input Process: *Configure schema crosswalk*
 - b. Original Delivery Folder: *Data\2019_Data\OR\OriginalDelivery*
 - c. Config GDB: *ETLtool\Config\crosswalk_config.gdb*
 - d. Investigation Input(s):
Data\2019_Data\OR\OriginalDelivery\Cultural.gdb\CULT_INVEST_POLY
 - e. Resource Input(s):
Data\2019_Data\OR\OriginalDelivery\Cultural.gdb\CULT_RESOURCE_POLY
2. Open the schema crosswalks under the DataMigration folder. In ArcMap, expand the folders and drag in the geodatabase tables. These tables are used to control the crosswalk of state or SHPO fields to CRM fields. Click the empty cell in the table and select the appropriate match for each CRM field. If a match for the CRM field does not exist, leave the value Null. If the CRM field requires concatenation of two fields, use the Field Calculator to type in the field names separating them by a semi-colon. Do not use spaces.

3. In ArcMap, open the NCRIMS Data Migration toolbox and click on Configure Crosswalk a second time. Input the following parameters:
 - a. Input Process: *Configure domain crosswalk*
 - b. Original Delivery Folder: *Data\2019_Data\OR\OriginalDelivery*
 - c. Config GDB: *ETLtool\Config\crosswalk_config.gdb*
 - d. Investigation Input(s):
Data\2019_Data\OR\OriginalDelivery\Cultural.gdb\CULT_INVEST_POLY
 - e. Resource Input(s):
Data\2019_Data\OR\OriginalDelivery\Cultural.gdb\CULT_RESOURCE_POLY
4. Review the output tables from the geometry validation process. These are found under DataMigration\check_geometry_log. The following list shows topology errors critical to geoprocessing in the ETL. The steps proceeding the dataset names and errors were used to reconcile the topology errors. These processes were completed manually. For more information on type of topology errors, see the ESRI documentation here:
<http://desktop.arcgis.com/en/arcmap/10.3/tools/data-management-toolbox/check-geometry.htm>
 - a. CULT_INVEST_POLY – *self-intersections*
 - b. CULT_RESOURCE_POLY – *self-intersections*
 - i. In ArcMap, open the tool Repair Geometry under the Data Management > Features.
 - ii. Drag the corresponding staging feature class copy into the Input Features parameter. These copies are found under the Staging folder > Staging.gdb.
 - iii. Click OK. Repeat the above steps for each listed feature class's copy.
11. Open the domain crosswalks under the DataMigration folder. A folder is present for each delivered feature class. In ArcMap, expand the folders and drag in the geodatabase tables. These tables are used to control the crosswalk of state or SHPO values to CRM domain values. Open each table and populate the NCRIMS_Value field with the appropriate corresponding domain value for each state/SHPO value. If the state/SHPO value is Null, leave the NCRIMS_Value Null.
6. In ArcMap, open the NCRIMS Data Migration toolbox and click on Migrate Values. This tool crosswalk the state/SHPO values to the CRM staging geodatabase. Input the following parameters:
 - a. Input Process: *Migrate Values*
 - b. Did the state require a domain crosswalk?: *Check*.
 - c. Did the state require a clip to the SMA buffer?: *Leave unchecked*.
 - d. Original Delivery Folder: *Data\2019_Data\OR\OriginalDelivery*
 - e. Config GDB: *ETLtool\Config\crosswalk_config.gdb*
 - f. Investigation Input(s):
Data\2019_Data\OR\OriginalDelivery\Cultural.gdb\CULT_INVEST_POLY
 - g. Resource Input(s):
Data\2019_Data\OR\OriginalDelivery\Cultural.gdb\INVEST_RESOURCE_POLY

7. In ArcMap, use the Select by Attributes button to locate and select the orphaned domains values below. Once selected, use the Field Calculator to change the original value to the CRM value shown to the right. Orphaned domain values are any values included in the delivered dataset that do not abide with the CRM domain values. The values will thus need to be edited to comply with the CRM standard schema. These may also be located under *DataMigration\orphaned_domains*
 - a. Feature class: CRM_Resources
 - i. Agency – Managing Agency
8. QC the CRM_Resources dataset and append it to the TargetSchema dataset using the Data Management > General > Append tool in ArcMap. See below for the appropriate parameters.
 - a. Input Datasets: *Data\2019_Data\OR\Staging\Staging.gdb\CRM_Resources*
 - b. Target Dataset:
Data\2019_Data\TargetSchema\National_CRM.gdb\CRM_Resources
 - c. Schema Type: *NO TEST*
9. QC the CRM_Investigations dataset and append it to the TargetSchema dataset using the Data Management > General > Append tool in ArcMap. See below for the appropriate parameters.
 - a. Input Datasets: *Data\2019_Data\OR\Staging\Staging.gdb\CRM_Investigations*
 - b. Target Dataset:
Data\2019_Data\TargetSchema\National_CRM.gdb\CRM_Investigations
 - c. Schema Type: *NO TEST*

Utah

General Workflow and Tool Parameters

1. Open the *UT\OriginalDelivery* folder in Windows File Explorer. Make a copy of the *NCRMS_Utah_20191113.gdb* and rename it *NCRMS_Utah_20191113_backup.gdb*.
2. Open the
UT\OriginalDelivery\NCRMS_Utah_20191113\NCRMS_Utah_20191113\NCRMS_Utah_20191113.gdb\Resources_2019113 attribute table. Use the Data Management > Fields > Alter Field to change the name if the *RSRCE_TMPRL_CLTRL_ASGMNT* field to *RSRCE_PRMRY_CRGRY_NM*.
3. In ArcMap, open the NCRIMS Data Migration toolbox and click on NCRIMS ETL. This tool generates crosswalks for the user to map the state/SHPO schemas to the CRM schema or to map the state/SHPO values to the CRM domain values. Input the following parameters:
 - a. Input Process: *Configure schema crosswalk*
 - b. Original Delivery Folder: *Data\2019_Data\UT\OriginalDelivery*
 - c. Config GDB: *ETLtool\Config\crosswalk_config.gdb*
 - d. Investigation Input(s):
Data\2019_Data\UT\OriginalDelivery\NCRMS_Utah_20191113\NCRMS_Utah_20191113\NCRMS_Utah_20191113.gdb\Investigations_2019113



- e. Resource Input(s):
Data\2019_Data\UT\OriginalDelivery\NCRMS_Utah_20191113\NCRMS_Utah_20191113\NCRMS_Utah_20191113.gdb\Resources_2019113
4. Open the schema crosswalks under the DataMigration folder. In ArcMap, expand the folders and drag in the geodatabase tables. These tables are used to control the crosswalk of state or SHPO fields to CRM fields. Click the empty cell in the table and select the appropriate match for each CRM field. If a match for the CRM field does not exist, leave the value Null. If the CRM field requires concatenation of two fields, use the Field Calculator to type in the field names separating them by a semi-colon. Do not use spaces.
5. In ArcMap, open the NCRIMS Data Migration toolbox and click on NCRIMS ETL a second time. Input the following parameters:
 - a. Input Process: *Configure domain crosswalk*
 - b. Original Delivery Folder: *Data\2019_Data\UT\OriginalDelivery*
 - c. Config GDB: *ETLtool\Config\crosswalk_config.gdb*
 - d. Investigation Input(s):
Data\2019_Data\UT\OriginalDelivery\NCRMS_Utah_20191113\NCRMS_Utah_20191113\NCRMS_Utah_20191113.gdb\Investigations_2019113
 - e. Resource Input(s):
Data\2019_Data\UT\OriginalDelivery\NCRMS_Utah_20191113\NCRMS_Utah_20191113\NCRMS_Utah_20191113.gdb\Resources_2019113
6. Review the output tables from the geometry validation process. These are found under DataMigration\check_geometry_log. The following list shows topology errors critical to geoprocessing in the ETL. The steps proceeding the dataset names and errors were used to reconcile the topology errors. These processes were completed manually. For more information on type of topology errors, see the ESRI documentation here:
<http://desktop.arcgis.com/en/arcmap/10.3/tools/data-management-toolbox/check-geometry.htm>
 - a. Investigations_20191113 – *bad extent*
 - i. In ArcMap open the Recalculate Feature Class Extent under Data Management.
 - ii. Use the Staging Investigations_20191113_copy feature class as the input layer.
 - iii. Click OK.
 - b. Investigations_20191113 – *self-intersections*
 - c. Investigations_20191113 – *invalid topology (StateProj: U09UY0025) *This feature was not migrated to the National Dataset.*
 - d. Resources_20191113 – *self-intersections*
 - i. In ArcMap, open the tool Repair Geometry under the Data Management > Features.
 - ii. Drag the Resources_20191113 copy into the Input Features parameter. These copies are found under the Staging folder > Staging.gdb.
 - iii. Click OK.

7. Open the domain crosswalks under the DataMigration folder. A folder is present for each delivered feature class. In ArcMap, expand the folders and drag in the geodatabase tables. These tables are used to control the crosswalk of state or SHPO values to CRM domain values. Open each table and populate the NCRIMS_Value field with the appropriate corresponding domain value for each state/SHPO value. If the state/SHPO value is Null, leave the NCRIMS_Value Null.
8. In ArcMap, open the NCRIMS Data Migration toolbox and click on NCRIMS ETL. This tool crosswalks the state/SHPO values to the CRM staging geodatabase. Input the following parameters:
 - a. Input Process: *Migrate Values*
 - b. Did the state require a domain crosswalk?: *Check*.
 - c. Did the state require a clip to the SMA buffer?: *Leave unchecked*.
 - d. Original Delivery Folder: *Data\2019_Data\UT\OriginalDelivery*
 - e. Config GDB: *ETLtool\Config\crosswalk_config.gdb*
 - f. Investigation Input(s):
Data\2019_Data\UT\OriginalDelivery\NCRMS_Utah_20191113\NCRMS_Utah_20191113\NCRMS_Utah_20191113.gdb\Investigations_2019113
 - g. Resource Input(s):
Data\2019_Data\UT\OriginalDelivery\NCRMS_Utah_20191113\NCRMS_Utah_20191113\NCRMS_Utah_20191113.gdb\Resources_2019113
9. Use Select by Location to select the staging resource features that intersect BLM Surface Management Areas 100-meter buffer.
10. Append the selected records to the TargetSchema dataset using the Data Management > General > Append tool in ArcMap. See below for the appropriate parameters.
 - a. Input Datasets: *Data\2019_Data\UT\Staging\Staging.gdb\CRM_Resources*
 - b. Target Dataset:
Data\2019_Data\TargetSchema\National_CRM.gdb\CRM_Resources
 - c. Schema Type: *NO TEST*
11. Use Select by Location to select the staging investigation features that intersect BLM Surface Management Areas 100-meter buffer.
12. Append the selected records to the TargetSchema dataset using the Data Management > General > Append tool in ArcMap. See below for the appropriate parameters.
 - a. Input Datasets: *Data\2019_Data\UT\Staging\Staging.gdb\CRM_Investigations*
 - b. Target Dataset:
Data\2019_Data\TargetSchema\National_CRM.gdb\CRM_Investigations
 - c. Schema Type: *NO TEST*

Wyoming

General Workflow and Tool Parameters

1. In ArcMap, open the NCRIMS Data Migration toolbox and click on NCRIMS ETL. This tool generates crosswalks for the user to map the state/SHPO schemas to the CRM

schema or to map the state/SHPO values to the CRM domain values. Input the following parameters:

- a. Input Process: *Configure schema crosswalk*
 - b. Original Delivery Folder: *Data\2019_Data\WY\OriginalDelivery*
 - c. Config GDB: *ETLtool\Config\crosswalk_config.gdb*
 - d. Investigation Input(s):
Data\2019_Data\WY\OriginalDelivery\WY_NCRIMS_20191007.gdb\Investigations_BLM
 - e. Resource Input(s):
Data\2019_Data\WY\OriginalDelivery\WY_NCRIMS_20191007.gdb\Sites_BLM
2. Open the schema crosswalks under the DataMigration folder. In ArcMap, expand the folders and drag in the geodatabase tables. These tables are used to control the crosswalk of state or SHPO fields to CRM fields. Click the empty cell in the table and select the appropriate match for each CRM field. If a match for the CRM field does not exist, leave the value Null. If the CRM field requires concatenation of two fields, use the Field Calculator to type in the field names separating them by a semi-colon. Do not use spaces.
 3. In ArcMap, open the NCRIMS Data Migration toolbox and click on NCRIMS ETL a second time. Input the following parameters:
 - a. Input Process: *Configure domain crosswalk*
 - b. Original Delivery Folder: *Data\2019_Data\WY\OriginalDelivery*
 - c. Config GDB: *ETLtool\Config\crosswalk_config.gdb*
 - d. Investigation Input(s):
Data\2019_Data\WY\OriginalDelivery\WY_NCRIMS_20191007.gdb\Investigations_BLM
 - e. Resource Input(s):
Data\2019_Data\WY\OriginalDelivery\WY_NCRIMS_20191007.gdb\Sites_BLM
 4. Review the output tables from the geometry validation process. These are found under DataMigration\check_geometry_log. The following list shows topology errors critical to geoprocessing in the ETL. The steps proceeding the dataset names and errors were used to reconcile the topology errors. These processes were completed manually. For more information on type of topology errors, see the ESRI documentation here:
<http://desktop.arcgis.com/en/arcmap/10.3/tools/data-management-toolbox/check-geometry.htm>
 - a. Investigations_BLM – *self-intersections*
 - b. Investigations_BLM – *invalid topology (OID: 132287/ WYCRONUM: 11-225-2)*
**This feature was not migrated to the National Dataset.*
 - c. Sites_BLM – *self-intersections*
 - i. In ArcMap, open the tool Repair Geometry under the Data Management > Features.
 - ii. Drag the corresponding staging feature class copy into the Input Features parameter. These copies are found under the Staging folder > Staging.gdb.
 - iii. Click OK. Repeat the above steps for each listed feature class's copy.

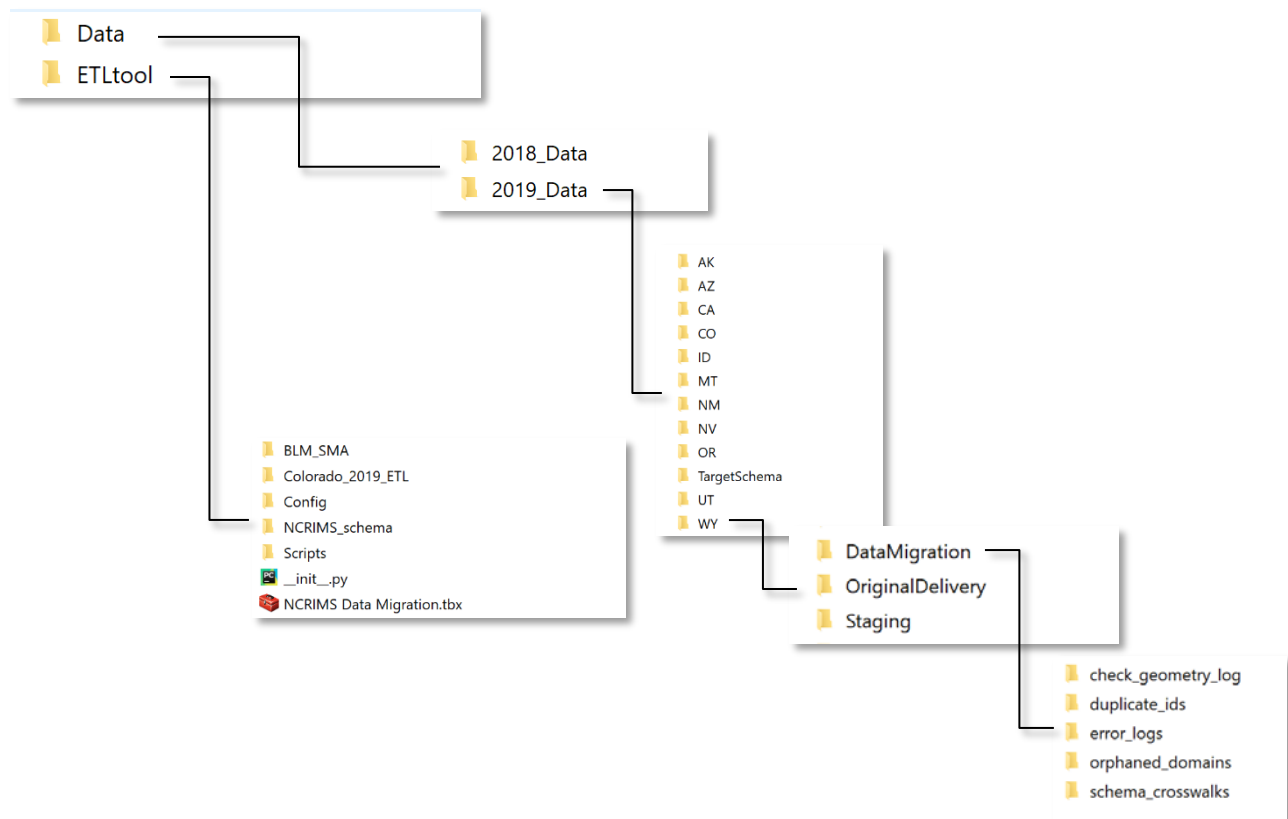
12. Open the domain crosswalks under the DataMigration folder. A folder is present for each delivered feature class. In ArcMap, expand the folders and drag in the geodatabase tables. These tables are used to control the crosswalk of state or SHPO values to CRM domain values. Open each table and populate the NCRIMS_Value field with the appropriate corresponding domain value for each state/SHPO value. If the state/SHPO value is Null, leave the NCRIMS_Value Null.
12. In ArcMap, open the NCRIMS Data Migration toolbox and click on NCRIMS ETL. This tool crosswalks the state/SHPO values to the CRM staging geodatabase. Input the following parameters:
 - a. Input Process: *Migrate Values*
 - b. Did the state require a domain crosswalk?: *Check*.
 - c. Did the state require a clip to the SMA buffer?: *Leave unchecked*.
 - d. Original Delivery Folder: *Data\2019_Data\WY\OriginalDelivery*
 - e. Config GDB: *ETLtool\Config\crosswalk_config.gdb*
 - f. Investigation Input(s):
Data\2019_Data\WY\OriginalDelivery\WY_NCRIMS_20191007.gdb\Investigations_BLM
 - h. Resource Input(s):
Data\2019_Data\WY\OriginalDelivery\WY_NCRIMS_20191007.gdb\Sites_BLM
8. Open the Staging geodatabase (*Data\2019_Data\WY\Staging\Staging.gdb*). Delete the feature classes who's name trail with *_clipped* and *_projected*
9. Drag the *Staging.gdb\Sites_BLM_copy* feature class into the ArcMap data frame. Use Selection > Select by Attributes to select all features with a RSRCE_SHPO_ID that is NULL or 0.
10. Use Select by Attributes a second time to select all features with a RSRCE_AGCY_ID that is NULL from the existing selected records.
11. Right click on the *Sites_BLM_copy* layer in the Table of Contents. Click Selection > Create layer from selected features.
12. Open the *Sites_BLM_copy* attribute table and reverse the selection.
13. Use Field Calculator to set the GIS_Acres and BLM_Acres values to 0.
14. Use Data Management > Generalization > Dissolve to convert the single-part polygons to multi-part polygons. Use the parameters below.
 - a. Input features: *Sites_BLM_copy*
 - b. Output feature class:
Data\2019_Data\WY\Staging\Staging.gdb\Sites_BLM_copy_dissolved
 - c. Dissolve field(s): *all RSRCE_* fields and GIS_Acres, BLM_Acres, CREATE_DATE, CREATE_BY*
 - d. Statistics field(s): *Leave blank*
 - e. Create multipart features: *Check*



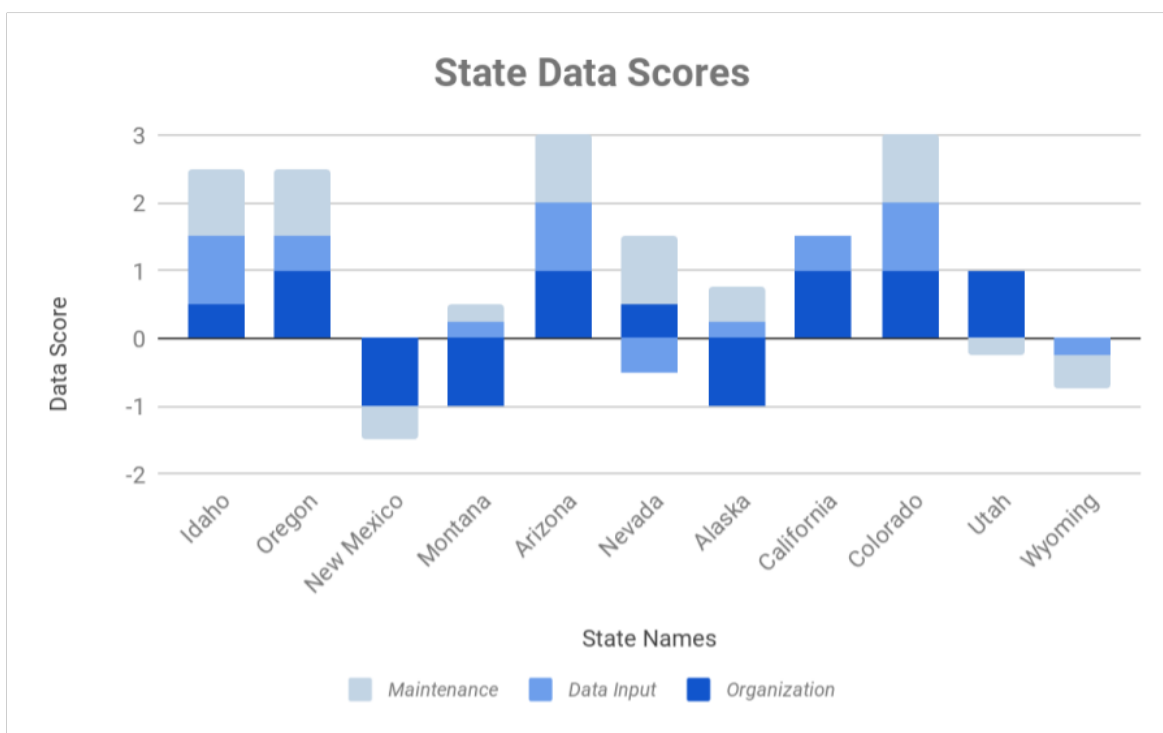
- f. Unsplit lines: *Uncheck*
15. Repeat steps 9 through 14 for the *Staging\Investigations_BLM_copy* feature class. Use all *INVTGTN_** fields, *BLM_Acres*, *GIS_Acres*, *CREATE_DATE*, *CREATE_BY*, *WYCRONUM* in the Dissolve tool (step 14.)
16. Delete all features in the *Staging\Staging.gdb\CRM_Resources* and *Staging\Staging.gdb\CRM_Investigations* feature classes.
17. Rename *Staging.gdb*_copy* feature classes to *Staging.gdb*_old* and *Staging.gdb*_dissolved* feature classes to *Staging.gdb*_copy*.
18. Add a field called TEMP_ID (type – long) to the *_copy* feature classes. Set this field equal to the OBJECTID using the Field Calculator.
19. Make sure the data frame is the North America Albers Equal Area Conic coordinate system (projected). Open the *Staging\Staging.gdb\Sites_BLM_copy* feature class attribute table. Right click on the GIS_Acres field and select Calculate Geometry. Use the coordinate system of the data frame and calculate acres.
20. Use Data Management > Analysis > Extract > Clip to clip the *_copy* feature class to the BLM SMA boundary. The parameters are below.
 - a. Input features: *Data\2019_Data\WY\Staging\Staging.gdb\Sites_BLM_copy*
 - b. Clip features: *ETLtool\BLM_SMA\sma_wm.gdb\SurfaceMgtAgy_BLM_projected*
 - c. Output feature class:
Data\2019_Data\WY\Staging\Staging.gdb\Sites_BLM_copy_clipped
 - d. XY Tolerance: *Leave blank*
21. Make sure the data frame is the North America Albers Equal Area Conic coordinate system (projected). Open the *Staging\Staging.gdb\Sites_BLM_copy_clipped* feature class attribute table. Right click on the BLM_Acres field and select Calculate Geometry. Use the coordinate system of the data frame and calculate acres.
22. Join the *Sites_BLM_copy_clipped* to the *Sites_BLM_copy* feature class on the TEMP_ID and OBJECTID fields.
23. Start an edit session and, using Field Calculator, set the *Sites_BLM_copy.BLM_Acres* field equal to *Sites_BLM_copy_clipped.BLM_Acres*.
24. Repeat steps 18 through 24 for the *Investigations_BLM_copy* feature class.
25. Using Data Management > General > Append, append the records in *Sites_BLM_copy* to *Staging.gdb\CRM_Resources* (Schema Type: NO TEST).

26. Using Data Management > General > Append, append the records in - *Investigations_BLM_copy* to *Staging.gdb\CRM_Investigations* (Schema Type: NO TEST).
27. QC the CRM_Resources dataset and append it to the TargetSchema dataset using the Data Management > General > Append tool in ArcMap. See below for the appropriate parameters.
 - a. Input Datasets: *Data\2019_Data\WY\Staging\Staging.gdb\CRM_Resources*
 - b. Target Dataset: *Data\2019_Data\TargetSchema\National_CRM.gdb\CRM_Resources*
 - c. Schema Type: *NO TEST*
28. QC the CRM_Investigations dataset and append it to the TargetSchema dataset using the Data Management > General > Append tool in ArcMap. See below for the appropriate parameters.
 - a. Input Datasets: *Data\2019_Data\WY\Staging\Staging.gdb\CRM_Investigations*
 - b. Target Dataset: *Data\2019_Data\TargetSchema\National_CRM.gdb\CRM_Investigations*
 - c. Schema Type: *NO TEST*

APPENDIX B: 2019 Data Directory Structure



APPENDIX C: 2018 – 2019 Data Resolutions



Alaska 2018 Score and Recommendations

Alaska delivered four Resource layers in shapefile format. Steps were taken to reformat the data in preparation for the execution of the ETL script tool. The shapefiles were imported into a file geodatabase prior to execution of the ETL script. This state currently does not maintain geospatial Investigation data.

This state's automated transform process differs from other states in that the originally delivered geodatabase includes three separate geometries for the Resources data. These layers were buffered 15 meters for geometry that is not polygon, then merged to create a single Resources layer. The originally delivered feature classes were exported to the schema_test2018.gdb for future validation. The data delivery did not include a one-to-one match with the CRM standard. Thus, the configure field map method was used to control the migration process. The domain crosswalk method was also used to migrate the RSRCE_TMPRL_CLTRL_ASGNMNT values to the appropriate domain values. See Appendix B for information on the Alaska ETL workflow and tool inputs.

Recommendations for Future Deliverables

Total Score: -0.25

A. Organization: -1 (-2 out of 2)

- a. The SHPO/ State data was delivered in a file geodatabase: -1

Criteria description: A file geodatabase ensures all data is delivered as it was captured and without any field name truncation.

- b. Two feature classes exist in the data deliverable: -1

Criteria description: One feature class exists containing cultural Resources data and one feature class exists containing cultural Investigations data. This reduces geoprocessing of different geometries and ensures consistency and simplicity in the data deliverables

For future deliveries, Alaska should deliver data in a file geodatabase format. Shapefiles exhibit storage constraints, resulting in truncated field names which do not reflect the original name. This causes discrepancies between the crosswalk spreadsheet and the delivered data. Alaska should also consider merging and storing all Resource data in one feature class. By doing so, this process can be incorporated into the script allowing for the removal of Step 1 from the workflow (see Appendix B - Alaska). Although more than one feature class was delivered, the feature classes shared a similar schema and did not critically inhibit the development ETL process.

B. Data Input: 0.25 (1 out of 4)

- a. Redundant fields do not exist in the delivered data: 0

Criteria description: The data deliverable does not contain multiple fields holding the same information. This causes confusion in determining the appropriate field for the ETL if the crosswalk is not current with the data deliverable.

- b. Records in the data are tracked: 1

Criteria description: The data contains a field that indicates a visitation date or some other form of tracking. This enables differentiation of current records from legacy records, in the event duplicate records exist in the dataset.

- c. The data includes a field for each CRM Standard field: -1

Criteria description: The delivered data includes enough information to populate all fields of the CRM standard. This ensures all CRM data is captured in the National Dataset.

- d. The delivered data has normalized and/or consistent values: 1

Criteria description: The values within each CRM-required field fall within a domain. The spelling of the values is consistent throughout the dataset. This improves the efficacy of automation and allows for easily codable crosswalks.

No critical issues were observed regarding the data input and values. The values for most fields appeared to be normalized. This normalization should be confirmed for the next delivery for all fields, especially Resource Temporal Cultural Assignment. Upon confirmation, these values can be incorporated into the script for further automation so that Step 3 does not need to be performed (see Appendix B - Alaska). It is recommended that Alaska maintain any missing CRM fields to ensure their data is reflected accurately and wholly in the National Dataset. Last, although redundant fields existed in the shapefiles, they did not pose any critical issues for ETL development. However, to reduce the risk of potential confusion, Alaska should remove any redundant fields in future deliverables.

C. Maintenance: 0.5 (2 out of 4)

- a. Legacy records do not exist: 1

Criteria description: The data deliverable does not contain previously recorded records in addition to current or recent records. This results in multiple records per feature which inhibits the efficacy of automation.

- b. Legacy records can be queried: 1



Criteria description: If the data deliverable does contain legacy records, then the most recent or most current record can be easily queried for migration to the National Dataset.

c. **Duplicate records do not exist: 0**

Criteria description: The data deliverable does not contain duplicated records for each feature. Duplicated records include the same information for each field and inhibit the efficacy of automation.

d. **Duplicate records can be easily removed: 0**

Criteria description: If the data deliverable does contain duplicate records, then the most current record can be easily queried for migration to the National Dataset.

No critical issues were observed regarding how the data was maintained. The feature classes did not include legacy records. Some features with matching AHRSSNO values were present in the point, line, and/or polygon feature classes, producing duplicate records in the staging dataset. These records were not resolved and remain duplicated. It is recommended that Alaska convert all Resource data to matching geometry and resolve the duplicated records for the next deliverable.

2019 Resolution

- A. Organization – No resolution was observed regarding the Organization category.
- B. Data Input – More CRM fields were added in the 2019 deliverable.
- C. Maintenance – No duplicate records were observed in the 2019 deliverable.

Arizona 2018 Score and Recommendations

Arizona delivered a file geodatabase with a match to the legacy CRM standard. Arizona maintains all Resource data in one feature class and all Investigation data in a separate feature class. No additional steps were required for data preparation.

The data delivery did not include a one-to-one match with the current CRM standard. However, the configuration method was not used to control the migration process. This was hard-coded in the script. Alternatively, the domain crosswalk method was used to migrate values to domain values. The Arizona dataset is included in the schema validation method. Details on the workflow can be found in Appendix B.

Recommendations for Future Deliverables

Total Score: 3

A. Organization: 1 (2 out of 2)

- a. The SHPO/ State data was delivered in a file geodatabase: 1

Criteria description: A file geodatabase ensures all data is delivered as it was captured and without any field name truncation

- b. Two feature classes exist in the data deliverable: 1

Criteria description: One feature class exists containing cultural resources data and one feature class exists containing cultural Investigations data. This reduces geoprocessing of different geometries and ensures consistency and simplicity in the data deliverables.

No critical issues were observed with the organizational structure of the delivered data.

B. Data Input: 1 (4 out of 4)

- a. Redundant fields do not exist in the delivered data: 1



Criteria description: The data deliverable does not contain multiple fields holding the same information. This causes confusion in determining the appropriate field for the ETL if the crosswalk is not current with the data deliverable.

b. Records in the data are tracked: 1

Criteria description: The data contains a field that indicates a visitation date or some other form of tracking. This enables differentiation of current records from legacy records, in the event duplicate records exist in the dataset.

c. The data includes a field for each CRM Standard field: 1

Criteria description: The delivered data includes enough information to populate all fields of the CRM standard. This ensures all CRM data is captured in the National Dataset.

d. The delivered data has normalized and/or consistent values: 1

Criteria description: The values within each CRM-required field fall within a domain. The spelling of the values is consistent throughout the dataset. This improves the efficacy of automation and allows for easily codable crosswalks.

No critical issues were observed regarding the data input and values. It is recommended that Arizona update their CRM schema to match the new, final 2018 CRM schema. Additionally, Arizona should ensure the domain *codes* are used in their dataset and that they match the codes used in the CRM schema.

C. Maintenance: 1 (4 out of 4)

a. Legacy records do not exist: 1

Criteria description: The data deliverable does not contain previously recorded records in addition to current or recent records. This results in multiple records per feature which inhibits the efficacy of automation.

b. Legacy records can be queried: 1

Criteria description: If the data deliverable does contain legacy records, then the most recent or most current record can be easily queried for migration to the National Dataset.

c. Duplicate records do not exist: 1

Criteria description: The data deliverable does not contain duplicated records for each feature. Duplicated records include the same information for each field and inhibit the efficacy of automation.

d. Duplicate records can be easily removed: 1

Criteria description: If the data deliverable does contain duplicate records, then the most current record can be easily queried for migration to the National Dataset.

No critical issues were observed regarding the maintenance of the data.

2019 Resolution

D. Organization – No resolution was needed regarding the Organization category.

E. Data Input – No resolution was needed regarding the Data Input category.

F. Maintenance – No resolution was needed regarding the Maintenance category.

California 2018 Score and Recommendations

California delivered a file geodatabase with relationship classes and a one-to-many relationship structure. California maintains all Resource data in one feature class and all Investigation data in a separate feature class; however, additional logic was required to extract information from the related tables. Steps were taken to reformat the data before the execution of the ETL script tool. These steps are saved as a ModelBuilder tool in the California toolset. See Appendix B for information on the California ETL workflow and tool inputs.

The data delivery did not include a one-to-one match with the CRM standard. The configuration method was used to control the migration process. Additionally, the domain crosswalk method was used to migrate values to domain values. The California dataset did not undergo schema validation due to the related tables included in the data deliverable. The related tables required the use of query tables which did not maintain the complete field schema in the output. The PK_GUID field was used to join the related tables to the feature classes. This field was also used as a means of unique identification of each feature class for geoprocessing. Any features which lack a PK_GUID were not migrated to the standard dataset. Four features in the res_inproc_proxy dataset did not have a PK_GUID assigned. Seven features in the inv_inproc_proxy dataset did not have a PK_GUID assigned.

Recommendations for Future Deliverables

Total Score: 1.5

D. Organization: 1 (2 out of 2)

- a. The SHPO/ State data was delivered in a file geodatabase: 1
Criteria description: A file geodatabase ensures all data is delivered as it was captured and without any field name truncation
- b. Two feature classes exist in the data deliverable: 1
Criteria description: One feature class exists containing cultural Resources data and one feature class exists containing cultural Investigations data. This reduces geoprocessing of different geometries and ensures consistency and simplicity in the data deliverables.

No critical issues were observed with the organizational structure of the delivered data. It is recommended that California extract values from the related tables and store them in the appropriate feature class for the next deliverable.

E. Data Input: 0.5 (2 out of 4)

- a. Redundant fields do not exist in the delivered data: 1
Criteria description: The data deliverable does not contain multiple fields holding the same information. This causes confusion in determining the appropriate field for the ETL if the crosswalk is not current with the data deliverable.
- b. Records in the data are tracked: 0
Criteria description: The data contains a field that indicates a visitation date or some other form of tracking. This enables differentiation of current records from legacy records, in the event duplicate records exist in the dataset.
- c. The data includes a field for each CRM Standard field: 0
Criteria description: The delivered data includes enough information to populate all fields of the CRM standard. This ensures all CRM data is captured in the National Dataset.
- d. The delivered data has normalized and/or consistent values: 1
Criteria description: The values within each CRM-required field fall within a domain. The spelling of the values is consistent throughout the dataset. This improves the efficacy of automation and allows for easily codable crosswalks.

No critical issues were observed regarding the data input and values. A date field exists in each feature class and is useful for tracking; however, this field is not populated for all records. Most fields in the delivered appear to have normalized and consistent values.

F. Maintenance: 0 (0 out of 4)

- a. Legacy records do not exist: -1
Criteria description: The data deliverable does not contain previously recorded records in addition to current or recent records. This results in multiple records per feature which inhibits the efficacy of automation.
- b. Legacy records can be queried: -1
Criteria description: If the data deliverable does contain legacy records, then the most recent or most current record can be easily queried for migration to the National Dataset.
- c. Duplicate records do not exist: 1
Criteria description: The data deliverable does not contain duplicated records for each feature. Duplicated records include the same information for each field and inhibit the efficacy of automation.
- d. Duplicate records can be easily removed: 1
Criteria description: If the data deliverable does contain duplicate records, then the most current record can be easily queried for migration to the National Dataset.

Legacy records existed in the related tables. There was not a method provided to determine which record is most recent, nor was a tracking field present. Fields were concatenated by row and crosswalked to the standard schema. It is recommended that California include a field which can be queried to extract the most recent record, or remove legacy records from the tables. No duplicate records existed in the feature classes or tables.

2019 Resolution

- A. Organization – No resolution was observed regarding the Organization category.
- B. Data Input – No resolution was observed regarding the Data Input category.
- C. Maintenance – No resolution was observed regarding the Maintenance category.

Colorado 2018 Score and Recommendations

Colorado delivered a file geodatabase with a one-to-one match to the CRM schema. No additional steps were required for data preparation.

The data delivery included slight discrepancies in the delivered field properties. A method was created to consume the migrated data and alter the field properties. Details on the workflow can be found in Appendix B.

Recommendations for Future Deliverables

Total Score: 3

A. Organization: 1 (2 out of 2)

- a. The SHPO/ State data was delivered in a file geodatabase: 1
Criteria description: A file geodatabase ensures all data is delivered as it was captured and without any field name truncation
- b. Two feature classes exist in the data deliverable: 1
Criteria description: One feature class exists containing cultural resources data and one feature class exists containing cultural Investigations data. This reduces geoprocessing of different geometries and ensures consistency and simplicity in the data deliverables.

No critical issues were observed with the organizational structure of the delivered data.

B. Data Input: 1 (4 out of 4)

- a. Redundant fields do not exist in the delivered data: 1
Criteria description: The data deliverable does not contain multiple fields holding the same information. This causes confusion in determining the appropriate field for the ETL if the crosswalk is not current with the data deliverable.
- b. Records in the data are tracked: 1
Criteria description: The data contains a field that indicates a visitation date or some other form of tracking. This enables differentiation of current records from legacy records, in the event duplicate records exist in the dataset.
- c. The data includes a field for each CRM Standard field: 1
Criteria description: The delivered data includes enough information to populate all fields of the CRM standard. This ensures all CRM data is captured in the National Dataset.
- d. The delivered data has normalized and/or consistent values: 1
Criteria description: The values within each CRM-required field fall within a domain. The spelling of the values is consistent throughout the dataset. This improves the efficacy of automation and allows for easily codable crosswalks.

No critical issues were observed regarding the data input and values.

C. Maintenance: 1 (4 out of 4)

- a. Legacy records do not exist: 1
Criteria description: The data deliverable does not contain previously recorded records in addition to current or recent records. This results in multiple records per feature which inhibits the efficacy of automation.
- b. Legacy records can be queried: 1
Criteria description: If the data deliverable does contain legacy records, then the most recent or most current record can be easily queried for migration to the National Dataset.
- c. Duplicate records do not exist: 1
Criteria description: The data deliverable does not contain duplicated records for each feature. Duplicated records include the same information for each field and inhibit the efficacy of automation.
- d. Duplicate records can be easily removed: 1
Criteria description: If the data deliverable does contain duplicate records, then the most current record can be easily queried for migration to the National Dataset.

No critical issues were observed regarding the maintenance of the data.

2019 Resolution

- A. Organization – No resolution was needed regarding the Organization category.
- B. Data Input – No resolution was needed regarding the Data Input category.
- C. Maintenance – No resolution was needed regarding the Maintenance category.

Idaho 2018 Score and Recommendations

Idaho delivered two file geodatabases with data migrated to the CRM standard. Because the data was contained in two separate file geodatabases and across multiple geometries, steps were taken before the script tool's execution to reformat the delivered data.

This state's automated transform process differs from other states in that the originally delivered geodatabase includes three separate geometries for the Resources data. These layers were buffered 15 meters for geometry that is not polygon, then merged to create a single Resources

layer. The originally delivered feature classes were exported to the schema_test2018.gdb for future validation. All three Resource feature classes and the Investigation feature class participate in the schema validation for Idaho. During the geometry validation of the script, one feature class in the Resources layer included NULL geometry. This feature was not reflected in the output National Dataset. A list of features with geometric or topological errors is included in the DataMigration folder. Details specific to the Idaho workflow are included Appendix B - Idaho.

Recommendations for Future Deliverable

Total Score: 2.5

A. Organization: 0.5 (1 out of 2)

- a. The SHPO/ State data was delivered in a file geodatabase: 1
Criteria description: A file geodatabase ensures all data is delivered as it was captured and without any field name truncation.
- b. Two feature classes exist in the data deliverable: 0
Criteria description: One feature class exists containing cultural resources data and one feature class exists containing cultural Investigations data. This reduces geoprocessing of different geometries and ensures consistency and simplicity in the data deliverables.

For future deliveries, Idaho should consider storing all Resource data in one feature class and all Investigation data in a separate feature class.

B. Data Input: 1 (4 out of 4)

- a. Redundant fields do not exist in the delivered data: 1
Criteria description: The data deliverable does not contain multiple fields holding the same information. This causes confusion in determining the appropriate field for the ETL if the crosswalk is not current with the data deliverable.
- b. Records in the data are tracked: 1
Criteria description: The data contains a field that indicates a visitation date or some other form of tracking. This enables differentiation of current records from legacy records, in the event duplicate records exist in the dataset.
- c. The data includes a field for each CRM Standard field: 1
Criteria description: The delivered data includes enough information to populate all fields of the CRM standard. This ensures all CRM data is captured in the National Dataset.
- d. The delivered data has normalized and/or consistent values: 1
Criteria description: The values within each CRM-required field fall within a domain. The spelling of the values is consistent throughout the dataset. This improves the efficacy of automation and allows for easily codable crosswalks.

No critical issues were observed regarding data input.

C. Maintenance: 1 (4 out of 4)

- a. Legacy records do not exist: 1
Criteria description: The data deliverable does not contain previously recorded records in addition to current or recent records. This results in multiple records per feature which inhibits the efficacy of automation.
- b. Legacy records can be queried: 1
Criteria description: If the data deliverable does contain legacy records, then the most recent or most current record can be easily queried for migration to the National Dataset.

- c. Duplicate records do not exist: 1

Criteria description: The data deliverable does not contain duplicated records for each feature. Duplicated records include the same information for each field and inhibit the efficacy of automation.

- d. Duplicate records can be easily removed: 1

Criteria description: If the data deliverable does contain duplicate records, then the most current record can be easily queried for migration to the National Dataset.

No critical issues were observed regarding how the data was maintained.

2019 Resolution

- D. Organization – For the 2019 data delivery, Idaho compiled all Resource feature classes into a single, polygon-type feature class.
- E. Data Input – No resolution was needed regarding the Data Input category.
- F. Maintenance – No resolution was needed regarding the Maintenance category.

Montana 2018 Score and Recommendations

Montana delivered three shapefiles and two Excel spreadsheets. Steps were taken to reformat the data in preparation for the execution of the ETL script tool. The Excel tables and shapefiles were imported into a file geodatabase to improve consistency in inputs across the states.

Because the data delivery did not include a one-to-one match with the CRM standard, the configure field map method was used to control the migration process. The configuration method uses the *Input Tables* parameter to add additional fields for each table that contains data needed for the migration. The domain crosswalk method was also used to migrate field values to the appropriate domain values. Refer to Appendix C - Montana for the schema crosswalk.

Additionally, Montana's data deliverable included multiple records for each feature. In the shapefiles, the duplicate records were actually multi-part features, but a dissolve on these features was not performed to improve processing of the tool. In the tables, the multiple records were legacy records. Additional logic was added to the script to extract the most recent record from the tables. Because of the challenges the format of the deliverable posed, Montana was exempted from the schema validation. Details specific to the Montana workflow are included Appendix B - Montana.

Recommendations for Future Deliverable:

Total Score: -0.5

A. Organization: -1 (-2 out of 2)

- a. The SHPO/ State data was delivered in a file geodatabase: -1

Criteria description: A file geodatabase ensures all data is delivered as it was captured and without any field name truncation.

- b. Two feature classes exist in the data deliverable: -1

Criteria description: One feature class exists containing cultural Resources data and one feature class exists containing cultural Investigations data. This reduces geoprocessing of different geometries and ensures consistency and simplicity in the data deliverables

For future deliverables, it is recommended that Montana deliver data in a file geodatabase. Shapefiles exhibit storage constraints, resulting in truncated field names which do not reflect the original name. This causes discrepancies between the crosswalk

spreadsheet and the delivered data. Additionally, it is recommended that Montana import any tabular data into the file geodatabase prior to delivery. Upon doing so, ensure all text “Null” values are converted to NULL and all NULL values are removed from date-type fields. It is recommended that Montana perform joins between the feature classes and tables prior to delivery; however, it is not critical to this state’s ETL.

B. Data Input: 0.25 (1 out of 4)

- a. Redundant fields do not exist in the delivered data: 1
Criteria description: The data deliverable does not contain multiple fields holding the same information. This causes confusion in determining the appropriate field for the ETL if the crosswalk is not current with the data deliverable.
- b. Records in the data are tracked: 1
Criteria description: The data contains a field that indicates a visitation date or some other form of tracking. This enables differentiation of current records from legacy records, in the event duplicate records exist in the dataset.
- c. The data includes a field for each CRM Standard field: -1
Criteria description: The delivered data includes enough information to populate all fields of the CRM standard. This ensures all CRM data is captured in the National Dataset.
- d. The delivered data has normalized and/or consistent values: 0
Criteria description: The values within each CRM-required field fall within a domain. The spelling of the values is consistent throughout the dataset. This improves the efficacy of automation and allows for easily codable crosswalks.

Montana should add and maintain any missing fields required by the national standard to ensure all migrated data is complete and accurate in the National Dataset. Montana data appeared to be largely normalized with the exception of the Resource Primary Category field. It is recommended that Montana enact methods such as domains to control the input of data relevant to this CRM field. If used, domains need to be included in the crosswalk spreadsheet and be current with the data deliverable.

C. Maintenance: 0.25 (1 out of 4)

- a. Legacy records do not exist: -1
Criteria description: The data deliverable does not contain previously recorded records in addition to current or recent records. This results in multiple records per feature which inhibits the efficacy of automation.
- b. Legacy records can be queried: 0
Criteria description: If the data deliverable does contain legacy records, then the most recent or most current record can be easily queried for migration to the National Dataset.
- c. Duplicate records do not exist: 1
Criteria description: The data deliverable does not contain duplicated records for each feature. Duplicated records include the same information for each field and inhibit the efficacy of automation.
- d. Duplicate records can be easily removed: 1
Criteria description: If the data deliverable does contain duplicate records, then the most current record can be easily queried for migration to the National Dataset.

Montana delivered legacy records of site and surveys. As a result, additional logic was required in the ETL script to account for the extraction of the most recent record. A method for identifying the most recent record was not explicitly stated or included with the data delivery. As a result, date fields were used to determine the most recent record. It

is recommended that Montana extract the most recent record from the tables for future data delivery. This can be achieved by enacting methods that allow these records to be easily queried. Additionally, it is recommended that Montana perform a dissolve on feature classes to reduce the number of records included in the deliverable. This will improve performance of the ETL.

2019 Resolution

- A. Organization – The 2019 data was delivered in a file geodatabase with one resource feature class and one investigation feature class.
- B. Data Input – The values were largely normalized. No resolution was observed regarding additional fields or data to populate all CRM fields.
- C. Maintenance – Minimal or no duplicate/ legacy records were encountered.

Nevada 2018 Score and Recommendations

Nevada delivered a file geodatabase with multiple Resource and Investigation layers. In an effort to maintain consistency across workflows, steps were taken to reformat the data in preparation for the execution of the ETL script tool. The required Resource and Investigations layers were extracted from the delivered geodatabase and imported into a separate file geodatabase.

Because the data delivery did not include a one-to-one match with the CRM standard, the configure field map method was used to control the schema migration process. The domain crosswalk method was also used to migrate field values to the appropriate domain values. The originally delivered feature classes were exported to the schema_test2018.gdb for future validation. All three Resource feature classes and two Investigation feature classes participate in the schema validation for Nevada. See Appendix B for information on the Nevada ETL workflow and tool inputs.

Recommendations for Future Deliverable

Total Score: 1

A. Organization: 0.5 (1 out of 2)

- a. The SHPO/ State data was delivered in a file geodatabase: 1
Criteria description: A file geodatabase ensures all data is delivered as it was captured and without any field name truncation.
- b. Two feature classes exist in the data deliverable: 0
Criteria description: One feature class exists containing cultural Resources data and one feature class exists containing cultural Investigations data. This reduces geoprocessing of different geometries and ensures consistency and simplicity in the data deliverables

For future deliveries, Nevada should consider storing all Resource data in one feature class and all Investigation data in a separate feature class.

B. Data Input: -0.5 (-2 out of 4)

- a. Redundant fields do not exist in the delivered data: 0
Criteria description: The data deliverable does not contain multiple fields holding the same information. This causes confusion in determining the appropriate field for the ETL if the crosswalk is not current with the data deliverable.
- b. Records in the data are tracked: 0

Criteria description: The data contains a field that indicates a visitation date or some other form of tracking. This enables differentiation of current records from legacy records, in the event duplicate records exist in the dataset.

- c. The data includes a field for each CRM Standard field: -1

Criteria description: The delivered data includes enough information to populate all fields of the CRM standard. This ensures all CRM data is captured in the National Dataset.

- d. The delivered data has normalized and/or consistent values: -1

Criteria description: The values within each CRM-required field fall within a domain. The spelling of the values is consistent throughout the dataset. This improves the efficacy of automation and allows for easily codable crosswalks.

Because the delivery of the data spanned multiple feature classes, redundant fields existed. However, this was not a critical issue for the ETL scripting. Due to multiple varieties of syntax and spelling for identical values in the delivered data, it is recommended that Nevada enact methods which enforce control over data input or collection. This would simplify the domain crosswalk process and ensure all relevant values are reflected in the national standard GDB. Such methods include creating and adding domains to CRM-equivalent fields. This would normalize the data and improve the flow of the ETL tool. If used, *domains need to be included in the crosswalk spreadsheet and be current with the data deliverable*. In other words, if the state changes these domains after completing the crosswalk spreadsheet, then the state should update the spreadsheet to match the data before the deliverable. Nevada should add and maintain any missing fields required by the national standard to ensure all migrated data is complete and accurate in the National Dataset.

C. Maintenance: 1 (4 out of 4)

- a. Legacy records do not exist: 1

Criteria description: The data deliverable does not contain previously recorded records in addition to current or recent records. This results in multiple records per feature which inhibits the efficacy of automation.

- b. Legacy records can be queried: 1

Criteria description: If the data deliverable does contain legacy records, then the most recent or most current record can be easily queried for migration to the National Dataset.

- c. Duplicate records do not exist: 1

Criteria description: The data deliverable does not contain duplicated records for each feature. Duplicated records include the same information for each field and inhibit the efficacy of automation.

- d. Duplicate records can be easily removed: 1

Criteria description: If the data deliverable does contain duplicate records, then the most current record can be easily queried for migration to the National Dataset.

No critical issues were observed regarding how the data was maintained. The feature classes did not include legacy records or duplicated records.

2019 Resolution

D. Organization – No resolution was observed regarding the delivery of the data.

E. Data Input – The following CRM fields were added for future maintenance:

- a. INVSTGTN_ATHRTY,
- b. INVSTGTN_NEPA_ID,
- c. RSRCE_LAST_RCRD_DT



- F. Maintenance – No resolution was needed regarding how the Nevada data delivery was maintained.

Additional comments: Nevada plans to implement the CRM standard for future deliveries after their receipt of the 2019 migrated dataset.

New Mexico 2018 Score and Recommendations

New Mexico delivered two shapefiles and several standalone DBFs representing related tables at a one-to-many match. Steps were taken to reformat the data in preparation for the execution of the ETL script tool. A file geodatabase was created to store the delivered shapefiles and DBF tables.

Because the data delivery did not include a one-to-one match with the CRM standard, the configure field map method was used to control the schema crosswalk process. The configuration method uses the *Input Tables* parameter to add additional fields for each table that contains data needed for the migration. The crosswalk is shown in Appendix C - New Mexico. The domain crosswalk method was also used to crosswalk field values into the appropriate domain values.

Additionally, New Mexico's data deliverable included multiple records for each feature. This resulted in a relatively large-sized deliverable, which critically affected the efficiency of the ETL process for this state, making automation a challenging goal. As a result, a manual procedure was performed and recorded to reconcile the number of records and improve performance of the ETL. Additional logic was added to the script to extract the most recent record after the manual reconciliation process. Due to the manual reconciliation that was required before the tool execution, New Mexico was exempted from the schema validation. Details specific to the New Mexico workflow are included Appendix B - New Mexico.

Recommendations for Future Deliverable

Total Score: -1.5

A. Organization: -1 (-2 out of 2)

- a. The SHPO/ State data was delivered in a file geodatabase: -1
Criteria description: A file geodatabase ensures all data is delivered as it was captured and without any field name truncation.
- b. Two feature classes exist in the data deliverable: -1
Criteria description: One feature class exists containing cultural Resources data and one feature class exists containing cultural Investigations data. This reduces geoprocessing of different geometries and ensures consistency and simplicity in the data deliverables

For future deliveries, New Mexico should deliver data in a file geodatabase format. Shapefiles and DBF tables exhibit storage constraints. Field names are generally truncated and do not reflect the original name. Shapefiles do not maintain Shape_length and Shape_area fields, making it more challenging to determine which record should be maintained in the National Dataset. Last, shapefiles introduce potential discrepancies between the crosswalk spreadsheet and the delivered data. It is recommended that New Mexico perform appropriate joins prior to delivery to include the tabular data in the attribute table of feature classes.

B. Data Input: 0 (0 out of 4)

- a. Redundant fields do not exist in the delivered data: -1
Criteria description: The data deliverable does not contain multiple fields holding the same information. This causes confusion in determining the appropriate field for the ETL if the crosswalk is not current with the data deliverable.
- b. Records in the data are tracked: 0
Criteria description: The data contains a field that indicates a visitation date or some other form of tracking. This enables differentiation of current records from legacy records, in the event duplicate records exist in the dataset.
- c. The data includes a field for each CRM Standard field: 0
Criteria description: The delivered data includes enough information to populate all fields of the CRM standard. This ensures all CRM data is captured in the National Dataset.
- d. The delivered data has normalized and/or consistent values: 1
Criteria description: The values within each CRM-required field fall within a domain. The spelling of the values is consistent throughout the dataset. This improves the efficacy of automation and allows for easily codable crosswalks.

Redundant fields existed across the tables and feature classes. This caused confusion in determining which fields could be used for tracking and which fields appropriately captured CRM standard data. This also allowed for possible scenarios of conflicting data. A tracking field existed in the required tables, but not the shapefiles. Methods were successfully implemented to retrieve the most current record from specific tables. More information on these methods can be found in Appendix B - New Mexico. It is recommended that New Mexico add and maintain any missing fields required by the national standard to ensure all migrated data is complete and accurate in the National Dataset. Second, it recommended that New Mexico ensure any date fields or fields indicating the most recent records are complete, and remove redundant fields from tables and feature classes. New Mexico data appeared to be largely normalized with the exception of the Resource Primary Category field.

C. Maintenance: -0.5 (-2 out of 4)

- a. Legacy records do not exist: -1
Criteria description: The data deliverable does not contain previously recorded records in addition to current or recent records. This results in multiple records per feature which inhibits the efficacy of automation.
- b. Legacy records can be queried: 0
Criteria description: If the data deliverable does contain legacy records, then the most recent or most current record can be easily queried for migration to the National Dataset.
- c. Duplicate records do not exist: -1
Criteria description: The data deliverable does not contain duplicated records for each feature. Duplicated records include the same information for each field and inhibit the efficacy of automation.
- d. Duplicate records can be easily removed: 0
Criteria description: If the data deliverable does contain duplicate records, then the most current record can be easily queried for migration to the National Dataset.

New Mexico delivered shapefiles with many to many relationships to records in the tables. Most of the records in the tables were legacy records and some were duplicates. However, because a consistent tracking field nor unique key was not included in the



shapefiles and tables, it was challenging to determine which record in the shapefile was paired with which record in the table. As a result, significant manual clean up was required before the execution of the ETL workflow. Additional logic was also required in the ETL script to account for the extraction of the most recent record. Methods for identifying the most recent record were not explicitly stated or included with the data delivery. As a result, date fields and number of visit fields were used to determine the most recent record. More information on these methods and manual cleanup processes can be found in Appendix B - New Mexico. It is recommended that New Mexico extract the most recent record from the tables and corresponding layers for future data delivery and include a primary/foreign key which will link each table record to a record in the shapefile. Additionally, New Mexico should manually reconcile records in the shapefiles that exist more than once. Often these records in shapefiles include different attribute values, but do not include fields to indicate which record is most recent or shape fields to indicate a multi-part feature.

2019 Resolution

- A. Organization – No resolution was observed regarding the delivery of the data.
- B. Data Input – RSRCE_PRMRY_CTGRY_NM data was standardized to match the CRM schema.
- C. Maintenance – Legacy records were encountered in the 2019 deliverable but were reconciled.

Oregon 2018 Score and Recommendations

Oregon delivered a file geodatabase which included a one-to-one match to the CRM standard. Steps were taken to reformat the data in preparation for the execution of the ETL script tool. The data was delivered in a folder titled OCRIS20180926.gdb with additional feature classes that did not contain Investigation or Resource data. For workflow consistency across states, an empty file geodatabase was created under the OriginalDelivery folder to store the Resource and Investigation feature classes.

Because the field names within the Investigations and Resources layers did not match those of the CRM standard, the configure field map method was used to control the schema crosswalk process. The crosswalk is shown in Appendix C - Oregon. The domain crosswalk method was also used to crosswalk field values into the appropriate domain values. Both the Resource (CULT_RESOURCE_POLY) and Investigation (CULT_INVEST_POLY) feature classes participate in the schema validation for Oregon. Details specific to the Oregon workflow are included Appendix B - Oregon.

Recommendations for Future Deliverable

Total Score: 2.5

A. Organization: 1 (2 out of 2)

- a. The SHPO/ State data was delivered in a file geodatabase: 1
Criteria description: A file geodatabase ensures all data is delivered as it was captured and without any field name truncation.
- b. Two feature classes exist in the data deliverable: 1

Criteria description: One feature class exists containing cultural Resources data and one feature class exists containing cultural Investigations data. This reduces geoprocessing of different geometries and ensures consistency and simplicity in the data deliverables.

No critical issues were observed regarding the organizational structure of the delivered data.

B. Data Input: 0.5 (2 out of 4)

- a. Redundant fields do not exist in the delivered data: 1
Criteria description: The data deliverable does not contain multiple fields holding the same information. This causes confusion in determining the appropriate field for the ETL if the crosswalk is not current with the data deliverable.
- b. Records in the data are tracked: 1
Criteria description: The data contains a field that indicates a visitation date or some other form of tracking. This enables differentiation of current records from legacy records, in the event duplicate records exist in the dataset.
- c. The data includes a field for each CRM Standard field: 0
Criteria description: The delivered data includes enough information to populate all fields of the CRM standard. This ensures all CRM data is captured in the National Dataset.
- d. The delivered data has normalized and/or consistent values: 0
Criteria description: The values within each CRM-required field fall within a domain. The spelling of the values is consistent throughout the dataset. This improves the efficacy of automation and allows for easily codable crosswalks.

No critical issues were observed regarding data input. The values for most fields appeared to be normalized except for values in the Resource Primary Category field and Resource Temporal Cultural Assignment. Discrepancies between the crosswalk spreadsheet and delivered data existed. It is recommended that the state enact methods such as utilizing domains on the above fields and *ensure the crosswalk spreadsheet is current with the data deliverable*. This normalization should be confirmed for the next delivery for all fields. Both datasets capture data for each CRM fields except for the Resource Spatial Collection Method field.

C. Maintenance: 1 (4 out of 4)

- a. Legacy records do not exist: 1
Criteria description: The data deliverable does not contain previously recorded records in addition to current or recent records. This results in multiple records per feature which inhibits the efficacy of automation.
- b. Legacy records can be queried: 1
Criteria description: If the data deliverable does contain legacy records, then the most recent or most current record can be easily queried for migration to the National Dataset.
- c. Duplicate records do not exist: 1
Criteria description: The data deliverable does not contain duplicated records for each feature. Duplicated records include the same information for each field and inhibit the efficacy of automation.
- d. Duplicate records can be easily removed: 1
Criteria description: If the data deliverable does contain duplicate records, then the most current record can be easily queried for migration to the National Dataset.

No critical issues were observed regarding how the data was maintained.



2019 Resolution

- A. Organization – No resolution was needed regarding the delivery of the data.
- B. Data Input – No resolution was observed regarding data input in the 2019 deliverable.
- C. Maintenance – No resolution was needed regarding how the Oregon data delivery was maintained.

Utah 2018 Score and Recommendations

Utah delivered a file geodatabase with a one-to-many relationship structure. Utah maintains all Resource data in one feature class and all Investigation data in a separate feature class; however, additional logic was required to extract information from the tables. The tables delivered were text and data file types. As a result, steps were taken to reformat the data before the execution of the ETL script tool. These steps are saved as a script tool in the Utah tool set. See Appendix B for information on the Utah ETL workflow and tool inputs.

The data delivery did not include a one-to-one match with the CRM standard. The configuration method was used to control the migration process. Additionally, the domain crosswalk method was used to migrate values to domain values. The tables required the use of query tables which did not maintain the complete field schema in the output. Legacy records existed across the tables, forcing the addition of extra logic to extract the most recent record. The Utah dataset did not undergo schema validation due to the tables included in the data deliverable.

Recommendations for Future Deliverables**Total Score: 0.75****A. Organization: 1 (2 out of 2)**

- a. The SHPO/ State data was delivered in a file geodatabase: 1
Criteria description: A file geodatabase ensures all data is delivered as it was captured and without any field name truncation
- b. Two feature classes exist in the data deliverable: 1
Criteria description: One feature class exists containing cultural resources data and one feature class exists containing cultural Investigations data. This reduces geoprocessing of different geometries and ensures consistency and simplicity in the data deliverables.

It is recommended that Utah extract values from the relevant tables and store them in the appropriate feature class for the next deliverable. It is recommended that, if Utah continue to deliver tables, the tables are in a consistent file type.

B. Data Input: 0 (0 out of 4)

- a. Redundant fields do not exist in the delivered data: 0
Criteria description: The data deliverable does not contain multiple fields holding the same information. This causes confusion in determining the appropriate field for the ETL if the crosswalk is not current with the data deliverable.
- b. Records in the data are tracked: 1
Criteria description: The data contains a field that indicates a visitation date or some other form of tracking. This enables differentiation of current records from legacy records, in the event duplicate records exist in the dataset.
- c. The data includes a field for each CRM Standard field: -1
Criteria description: The delivered data includes enough information to populate all fields of the CRM standard. This ensures all CRM data is captured in the National Dataset.

- d. The delivered data has normalized and/or consistent values: 0
Criteria description: The values within each CRM-required field fall within a domain. The spelling of the values is consistent throughout the dataset. This improves the efficacy of automation and allows for easily codable crosswalks.

It is recommended that Utah maintain a single field for each CRM field by concatenating any required field values prior to delivery. The requirement to concatenate fields via scripting poses challenges related to character length that must be uniquely handled for each instance. The values looked to be largely normalized, but lacked consistency. Consistency can be guaranteed with the use of domain values. It is recommended that Utah create and enable these on fields required by the CRM schema.

C. Maintenance: **-0.25** (-1 out of 4)

- a. Legacy records do not exist: -1
Criteria description: The data deliverable does not contain previously recorded records in addition to current or recent records. This results in multiple records per feature which inhibits the efficacy of automation.
- b. Legacy records can be queried: 1
Criteria description: If the data deliverable does contain legacy records, then the most recent or most current record can be easily queried for migration to the National Dataset.
- c. Duplicate records do not exist: -1
Criteria description: The data deliverable does not contain duplicated records for each feature. Duplicated records include the same information for each field and inhibit the efficacy of automation.
- d. Duplicate records can be easily removed: 0
Criteria description: If the data deliverable does contain duplicate records, then the most current record can be easily queried for migration to the National Dataset.

Legacy records existed in the related tables. However, date fields were present which allowed for the extraction of the most recent record. Duplicate records also existed in tables but were handled by the method to extract the most recent record. More information on this method can be found in Appendix B. It is recommended that Utah deliver a single record per feature in the future.

2019 Resolution

- A. Organization – Utah delivered two feature classes for the 2019 deliverable and reconciled tabular data per the 2018 recommendation.
- B. Data Input – The 2019 deliverable contained more fields for input into the CRM standard.
- C. Maintenance – No resolution was observed.

Wyoming 2018 Score and Recommendations

Wyoming delivered a file geodatabase with multiple feature classes and accompanying tables. Due to discrepancies between the delivered data and crosswalk document, it was unknown if the data was able to be migrated to the standard. This state's migration was handled manually and data from 2018 was used in the aggregate for Wyoming.

Recommendations for Future Deliverables

Total Score: -0.75

A. Organization: 0 (2 out of 2)

- a. The SHPO/ State data was delivered in a file geodatabase: 1
Criteria description: A file geodatabase ensures all data is delivered as it was captured and without any field name truncation
- b. Two feature classes exist in the data deliverable: -1
Criteria description: One feature class exists containing cultural resources data and one feature class exists containing cultural Investigations data. This reduces geoprocessing of different geometries and ensures consistency and simplicity in the data deliverables.

In the future, it is recommended that Wyoming deliver all Resource data in one feature class and all Investigation data in a separate feature class within a file geodatabase.

B. Data Input: -0.25 (-1 out of 4)

- a. Redundant fields do not exist in the delivered data: 0
Criteria description: The data deliverable does not contain multiple fields holding the same information. This causes confusion in determining the appropriate field for the ETL if the crosswalk is not current with the data deliverable.
- b. Records in the data are tracked: 0
Criteria description: The data contains a field that indicates a visitation date or some other form of tracking. This enables differentiation of current records from legacy records, in the event duplicate records exist in the dataset.
- c. The data includes a field for each CRM Standard field: -1
Criteria description: The delivered data includes enough information to populate all fields of the CRM standard. This ensures all CRM data is captured in the National Dataset.
- d. The delivered data has normalized and/or consistent values: 0
Criteria description: The values within each CRM-required field fall within a domain. The spelling of the values is consistent throughout the dataset. This improves the efficacy of automation and allows for easily codable crosswalks.

Due to the data delivery, it was not known if redundant fields existed, if the fields were tracked, nor if the data values were consistent. It was determined that a field for each CRM standard field did not exist.

C. Maintenance: -0.5 (-2 out of 4)

- a. Legacy records do not exist: 0
Criteria description: The data deliverable does not contain previously recorded records in addition to current or recent records. This results in multiple records per feature which inhibits the efficacy of automation.
- b. Legacy records can be queried: -1
Criteria description: If the data deliverable does contain legacy records, then the most recent or most current record can be easily queried for migration to the National Dataset.
- c. Duplicate records do not exist: 0
Criteria description: The data deliverable does not contain duplicated records for each feature. Duplicated records include the same information for each field and inhibit the efficacy of automation.
- d. Duplicate records can be easily removed: -1
Criteria description: If the data deliverable does contain duplicate records, then the most current record can be easily queried for migration to the National Dataset.

Due to the data delivery, it was not known if legacy or duplicate records existed. These fields could not be queried.



2019 Resolution

- A. Organization – Wyoming included a single file geodatabase for the 2019 deliverable. The geodatabase contained 1 resource feature class and 1 investigation feature class.
- B. Data Input – No resolution was observed regarding data input in the 2019 deliverable.
- C. Maintenance – Wyoming added a WYCRO field to assist in resolving any legacy records or duplicate records.