



Best Practices & Standards: *Persistent Identifier*

Date of Adoption: February 19, 2009

Introduction:

The Montana Association of Geographic Information Professionals (MAGIP) is a non-profit professional association that works to foster technical cooperation and promote the development of sound policy and practices that will support the efficient and effective use of geographic information systems in Montana. As such, the Technical Committee recognizes that a set of formalized best practices will not only encourage the use of standards but will make it easier to share data and save GIS practitioners work and time.

A best practice can be defined as the most efficient and effective way of achieving a goal or objective, improving business results through repeatable procedures that have proven themselves over time for large numbers of people. For a best practice to be effective, it must allow for the easy transfer of ideas, knowledge, and standards, while at the same time remaining flexible and adaptable to change as business needs and technology change.

It is important to note that the adoption of a best practice by MAGIP is not a mandate for its use, nor is a best practice applicable to every situation. Only in the instance where a best practice is recognized as a standard (Metadata in the Montana GIS Portal, for example) is it required and enforced. By committing to use a best practice, however, an individual acknowledges the benefits to one's organization and to the Montana GIS community as a whole.

Overview:

The rural nature of Montana warrants a federated database approach. No single jurisdiction is likely to own and maintain all of the geographic data required. Montana defines a federated approach as a method that supports the integration and utilization of data from multiple data providers, while retaining the primary maintenance responsibility with the provider. Most of Montana's framework data model efforts have adopted a federated approach that leverages multiple data providers and thus multiple data maintainers.

Database record identification is important in a federated environment that relies on disconnected and distributed data maintenance. Montana data modeling efforts have adopted a standard approach for identifiers in federated datasets which requires uniqueness, persistence, traceability, and feasibility. Traditionally, most organizations have defined their own identification system with little regard to uniqueness or consistent generation of identifiers across multiple organizations. The adoption of a standard data exchange mechanism utilizing persistent unique identifiers with centralized integration of data allows data providers to access data from other providers. Data from other providers can easily be tracked and identified, and centralized integration means that data from other providers is readily integrated with their own holdings.

Montana also considers federated data to be transactional. Records in federated tables are only retired, never deleted. Applications that utilize federated data rely on the existence (persistence) of a given record. Many domain-specific applications also build additional tables that add value to an existing federated table. The use of a unique persistent identifier and retention of all records guarantees that entities building third party tables will retain their investment. An important aspect of the unique identifier is that it be structured to provide as much flexibility as possible by allowing data providers to use any existing identifiers as a portion of the unique identifier while still maintaining a standard that data integrators and consumers may rely on. This reduces the chance a data provider will need to modify their database schemas or workflows in order to adopt the standard. In this case, it is the responsibility of the data provider to ensure the uniqueness and persistence of any existing identifier they wish to incorporate.

Description:

Montana chose to implement a semi-meaningful unique identifier that consists of an entity (provider) identifier that is unique statewide, followed by a three-character dataset identifier, followed by a unique local identifier (see Table 1).

Table 1: Components of the Montana persistent unique identifier

	Field	Description	Data Type	Length
1	Entity (Provider) identifier	Entity identifier from the Montana Standard Table of Entity Identifiers for the provider of this record. This field cannot be altered once assigned, and cannot be null. <entity>	String	25
2	Dataset identifier	The 3-character dataset identifier assigned to this table from the Montana Standard Coded Domain for Dataset Identifiers. This field cannot be altered once assigned, and cannot be null. <dataset>	String	3
3	Record (Local) Identifier	Unique persistent record identifier as assigned by the provider of this record. This field cannot be altered once assigned, and cannot be null. <record>	String	38
4	PKEY	Unique persistent identifier in the form: “<entity>.<dataset>.<record>”. Created by concatenating the above fields. This field cannot be altered once assigned, and cannot be null.	String	68

The components of the identifier have the following characteristics of note:

- The [Montana Base Map Service Center](#) (BMSC) maintains and distributes the database of unique entity (provider) identifiers, and guarantees that identifiers are unique statewide.
- The BMSC can generate a unique identifier for any legal entity. The Montana Standard Table of Entities can carry an identifier for any organization created in law or by legal action.
- The Montana Standard Table of Entities is transactional. Identifiers are never deleted when an entity becomes obsolete. The table maintains a parent-child hierarchy to retain references to organizations that may be retired or reorganized.
- The BMSC maintains an enterprise table of dataset identifiers.
- The record (local) identifier gives a data provider local control over the assignment and maintenance of record identifiers. Data providers may cross-reference the record (local) identifier with an existing identifier of any type that may be already in use, provided that it is persistent and unique.
- For new datasets, it is recommended that record (local) identifiers be long integer data type unless there is a need for alpha-numeric values.
- All three parts are concatenated into a single string field of the form “<entity>.<dataset>.<record>”. In the concatenation, the root elements are delimited with a period for ease in parsing and to improve readability.

Implementation:

Organizations may implement the unique identifier in many ways. The following examples demonstrate three possible methods an organization may use to implement and manage the unique identifier.

1. An organization may wish to add all four unique identifier field parts to a dataset and manage each part individually. In this case, as a new record is added to the database, the three parts are populated then concatenated to populate the PKEY. This may be done one at a time or in batch.

2. Given that for a single dataset, in most situations the entity and dataset identifier parts will be the same for every record, an organization may choose to not maintain these fields to their database. In this case, only the record (local) identifier and PKEY fields are added to the dataset. Record identifiers will be populated for new records as in the above case. Instead of basing the PKEY on the concatenation of the three parts (fields), the entity and dataset identifiers will be added as a single string part such as: "00000000.STR." & [RecordID].
3. Alternatively, an organization may choose to use an existing identifier they already maintain in place of the record (local) identifier. In this case, the only field that would need to be added is the PKEY which could be maintained in the same way as above, though an organization may choose to add the additional parts.

Organizations interested in implementing the unique identifier should contact the [Montana Base Map Service Center](#) for their organization's entity identifier and for more information on dataset identifiers.

References:

[Montana Base Map Service Center](#)

Document History:

November 04, 2008	Original Document Posted by Michael Fashoway
January 22, 2009	Changes Proposed to Data Types and Field Length of Entity and Record Identifiers
February 04, 2009	Comments Incorporated by Michael Fashoway
February 19, 2009	Adopted by MAGIP Board of Directors
May 04, 2010	Documentation Standardized by MAGIP Technical Committee