

MAGIP Best Practice: Persistent Identifier

Michael Fashoway

2010 MAGIP Technical Session

October 28, 2010



Background

- MAGIP and Best Practices
 - One of the goals of MAGIP is to “*foster technical cooperation and promote the development of sound policy and practices that will support the efficient and effective use of GIS in Montana.*”
 - “*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.*”



Background

- 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 a large number of people.
- It must allow for the easy transfer of ideas, knowledge, and standards while remaining flexible and adaptable to changes in business needs or technology.
- A best practice is not mandatory nor is it applicable to every situation.

Background

- Timeline
 - Summer 2006 – MLIAC requested MAGIP propose a standard for a persistent unique identifier
 - Fall 2008 – a group met at the Technical Session to discuss establishing or formalizing a set of GIS best practices
 - February 2009 – MAGIP Board approved the GIS Portal Metadata Standard, Persistent Identifiers, and Spatial Reference System best practices
 - Summer 2009 – MLIAC endorsed the above best practices



Background

- Persistent Identifier Best Practice
 - Around the time the Transportation and Critical Infrastructure data models were being developed, it was recognized there was a need for a meaningful unique identifier for records in the above *federated* databases that would be *unique, persistent, traceable, and easily implemented*.
 - What does that mean?

Background

- Federated – method that supports the integration and utilization of data from multiple data providers, while retaining the primary maintenance responsibility with the provider
- Unique – relational databases require “keys”, or unique IDs, to work
 - Relationships between tables are made using keys
 - Even for databases with one table, a unique ID is good practice even when the database doesn’t require it

Background

- Persistent – for any given record, the ID will never be deleted or changed
 - This guarantees that any related tables that are built to add value will not lose the relationship
- Traceable – data from any provider can easily be tracked and identified
 - Semi-meaningful (the parts actually stand for something)
 - A form a record-level metadata
- Easily implemented!



Description

- The persistent identifier consists of three parts:
 - Entity (provider) identifier (unique identifier assigned to any data provider)
 - Dataset identifier (three characters)
 - Record (local) identifier (the unique identifier in any given dataset)
- Three parts are concatenated into a single identifier, with a period (".") delimiting each part to make it possible to parse

Description

	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

Description

- The Base Map Service Center (BMSC) maintains the table of entity (provider) identifiers, guaranteeing they are unique
- The BMSC will generate an entity identifier for any data provider
- The BMSC also maintains the table of dataset identifiers
- The record (local) identifier gives a data provider control over the assignment and maintenance of the unique identifier in their data
 - May use an existing identifier of any type, as long as it is persistent and unique

Description

- Examples:
 - A BMSC created record in the Structures Framework: **306101002.STR.848**
 - A Lewis & Clark County provided record in the Structures Framework:
99049000.STR.{4477EDAF-7681-46BE-BB7E-A73D31101A73}
← ArcGIS-maintained GlobalID
 - A GNIS provided record in the Geographic Names Framework: **1434.GNM.1677497**
← GNIS ID

Implementation

- Couple ways to implement
 - Maintain all four parts as individual fields in the attribute table
 - Advantages: makes it easier to sort and query by any of the parts; easy to calculate the full PKEY
 - Disadvantages: more fields in the attribute table; provider and dataset IDs don't usually change so there is redundant data

Implementation

Attributes of STR_Point

OBJECTID *	PKEY *	PROVIDERID	DatasetID	RECORDID	
583482	306101002.STR.151129	306101002	STRUCTURE RECORD (MONTANA)	151129	04;
583483	306101002.STR.164451	306101002	STRUCTURE RECORD (MONTANA)	164451	04;
583484	306101002.STR.104978	306101002	STRUCTURE RECORD (MONTANA)	104978	<N
583485	306101002.STR.95355	306101002	STRUCTURE RECORD (MONTANA)	95355	<N
583486	306101002.STR.110562	306101002	STRUCTURE RECORD (MONTANA)	110562	<N
583487	306101002.STR.59446	306101002	STRUCTURE RECORD (MONTANA)	59446	56;
583488	306101002.STR.199341	306101002	STRUCTURE RECORD (MONTANA)	199341	04;
583489	306101002.STR.127739	306101002	STRUCTURE RECORD (MONTANA)	127739	<N
583490	306101002.STR.93282	306101002	STRUCTURE RECORD (MONTANA)	93282	<N
583491	306101002.STR.96968	306101002	STRUCTURE RECORD (MONTANA)	96968	<N
583493	99049000.STR.{1DF49D77-5D35-4337-A8CF-DAA	99049000	STRUCTURE RECORD (MONTANA)	{1DF49D77-5D35-4337-A8	{1C
583495	306101002.STR.61955	306101002	STRUCTURE RECORD (MONTANA)	61955	56;
583497	306101002.STR.168864	306101002	STRUCTURE RECORD (MONTANA)	168864	04;
583498	306101002.STR.180944	306101002	STRUCTURE RECORD (MONTANA)	180944	<N
583499	306101002.STR.50830	306101002	STRUCTURE RECORD (MONTANA)	50830	56;
583500	99049000.STR.{014F161F-4F39-4648-BCDF-285E	99049000	STRUCTURE RECORD (MONTANA)	{014F161F-4F39-4648-BC	{01
583501	306101002.STR.134591	306101002	STRUCTURE RECORD (MONTANA)	134591	<N
583502	306101002.STR.199811	306101002	STRUCTURE RECORD (MONTANA)	199811	<N
583504	306101002.STR.82429	306101002	STRUCTURE RECORD (MONTANA)	82429	<N
583505	99049000.STR.{AF856552-370B-4DA0-8B9F-394	99049000	STRUCTURE RECORD (MONTANA)	{AF856552-370B-4DA0-8	{AF
583506	306101002.STR.121087	306101002	STRUCTURE RECORD (MONTANA)	121087	<N
583507	99089000.STR.1611	99089000	STRUCTURE RECORD (MONTANA)	1611	<N
583509	306101002.STR.100528	306101002	STRUCTURE RECORD (MONTANA)	100528	<N
583510	306101002.STR.128531	306101002	STRUCTURE RECORD (MONTANA)	128531	<N
583511	306101002.STR.140601	306101002	STRUCTURE RECORD (MONTANA)	140601	<N
583512	99013000.STR.4912	99013000	STRUCTURE RECORD (MONTANA)	4912	02;
583513	306101002.STR.59739	306101002	STRUCTURE RECORD (MONTANA)	59739	56;
583514	306101002.STR.57662	306101002	STRUCTURE RECORD (MONTANA)	57662	56;
583515	306101002.STR.130454	306101002	STRUCTURE RECORD (MONTANA)	130454	<N

Implementation

- Maintain the two parts that change, PKEY and the record (local) identifier
 - Advantages: minimized the number of extra fields needed and redundant data
 - Disadvantages: more difficult sorts / queries;

Summary

- Unique identifiers are a good idea for any database
- The MAGIP Persistent Identifier Best Practice is a proven unique identifier format and is ideal for GIS datasets that integrate data from multiple providers OR for GIS data that is contributed to federated datasets
- Questions, comments?

