

# Spatial Data Storage/Data Discovery Terms Definitions

December 17, 2012 (author: Diane Papineau)

Document Status: Final Draft

## General

***Spatial Data Collection***—all the spatial data that Montana State Library (MSL) stores and preserves that, by definition, fit MSL’s Collection Development Policy. The collection may include multiple copies of data in different forms (coverage, shapefile, etc.) stored in the [Dark Archive](#) as well as duplicates of the data files stored in the [Accessible Archive](#) and [Active Store](#) locations for patron discovery and access.

***Clearinghouse***—a group of *public* data discovery tools and resources, including the Montana GIS Portal, the MSL Data List, web mapping applications, web mapping services, webpages offering data download, as well as staff time for manual packaging of large datasets for patrons.

## Data Storage Repositories

***Dark Archive***—a physical location on MSL’s SAN storing archival copies of all data MSL accepts into its data collection. The Dark Archive serves as the archival copy of MSL spatial data that is made accessible through the Clearinghouse. The Dark Archive will not *permanently* store spatial data that is archived by other organizations such as the National Agricultural Imagery Program (NAIP) files and wetlands data; however, it will archive that data for a defined period of time to serve as restore source if needed.

*Note:* MSL proposes that at some point the Dark Archive will also store preservation copies of MSL electronic maps (state publications) and MSL GIS map projects.

***Security Dark Archive***—a physical location in an off-site data center that stores an exact copy of the Dark Archive. The integrity of the data in the Security Dark Archive will be integrity checked using the same processes applied to the Dark Archive at MSL. The Security Dark Archive and the Dark Archive at MSL will each serve as the restore source for each other.

**Accessible Archive**—a physical location offering superseded data made easily accessible to patrons via the Internet. This is older data that is not of interest to most patrons, but is still valuable to many, justifying a dedicated volume and access tools for data download (i.e. superseded Cadastral, superseded Land Cover, etc.).

*Note:* The storage location and services supporting this Accessible Archive have not been determined or defined as of November 2012.

**Active Store**—a group of MSL storage locations holding active data (primarily the most current data) made available in different forms for online, self-service access (exception: large datasets such as NAIP will be manually provided from the Active Store). Data in Active Store locations is data that will be used most frequently by MSL staff and library patrons (internal partners and the public). Different storage devices and tools will serve Active Store data:

- For internal desktop use (GeoInfo staff and MSL partners) in an SDE database
- For web map services and map applications, likely in file geodatabases
- As downloadable files
- On patron portable hard drives for large dataset manual delivery

## **Data Structure**

**Stand-alone Data**—data that functions entirely on its own without introducing ambiguity and is offered for use without dependency. Standalone data can be offered in different forms, including as a geodatabase as long the geodatabase stores only one feature class

**Series Data**—data that changes regularly creating superseded data, which is documented via parent and children metadata records. The changes can be either attribute changes or changes in geometry/landscape. Some series data will change in expected intervals (MSDI); other series data changes when unpredictable landscape changes occur and are mapped (beetle kill extents, forest fire boundaries); other data can be assembled into a series by MSL because of the advantage offered to patrons (i.e. all non-NAIP aerial imagery of a city).

Reference: <http://www2.archivists.org/glossary/terms/s/serial> (Society of American Archivists)

**Collection Data**—a group of stand-alone datasets or feature classes that fit any of these descriptions, documented via parent and children metadata records (possibly also grandparent/parent/child records):

- Data that comprises a related regional collection (i.e. Yellowstone River, Flathead Basin mapping)
- Data created as part of one defined project
- Individual data files for download derived out of one larger dataset that is not suitable for patron download or use as supplied (i.e. Flathead LiDAR; NAIP?)
- Data offered by an agency via their own clearinghouse, online or offline (i.e. FWP's data list, City of Bozeman's data list; however, MSL's Data List and our Accessible Archive would not be considered a collection because the "parent" collection metadata record would be too large and it would continue to grow)
- A geodatabase comprised of more than one feature class.

Reference: <http://www2.archivists.org/glossary/terms/c/collection> (Society of American Archivists)

## **Data Packages** (applying OAIS to MSL workflow)

**Working Data**—data that is currently being prepared by MSL staff. Working data may be submissions toward a statewide MSDI dataset or it may be data that requires little or no modification before becoming publishable data. Working data may reside in an editing SDE environment, but would not reside in an SDE environment for published data. Working data is equivalent to the OAIS “Submission Information Package” or SIP.

**Published Data Package**—data, with complete metadata record, that is made available for internal MSL staff and partner use (i.e. via SDE) as well as via public Clearinghouse tools. Published data can be offered in a variety of forms (shapefile, geodatabase, etc.) and via different vehicles (web map service, zip file, SDE layer, web map application). A Published Data Package is equivalent to the OAIS “Dissemination Information Package” or DIP.

**Archived Data Package**—data that is stored in an archival repository. This could be a package in the Dark Archive or a package in the Accessible Archive. An Archived Data Package is equivalent to the OAIS “Archival Information Package” (AIP).

## **Data Status**

**Superseded Data**—data that has been *supplanted* by a newer version of the same data theme in a series (i.e. MSDI). The supplanting dataset may contain records that have been added, removed, or edited. Changes in the data series through time represent changes in real-world features that the data represents. However, changes may also include data corrections as the data series matures. Superseded data will remain available to patrons via self-serve access. Some data in a series may only be Dark Archived via a defined snapshot interval.

**Snapshot Data**—series data that is captured at specified intervals during the year for the purpose of archiving. The snapshot interval is defined specific to the data being captured and may vary from theme to theme. Snapshot intervals are defined in the [MSDI Publishing Standard](#) (document currently in development and very rough).

**Obsolete Data**—data that has been *replaced* by a better dataset that represents the same region and theme. The new data is recognized as more accurate or more up-to-date, it may not conform to the same schema as the obsolete data, and it may be from a different source. Obsoleted data may not remain available to patrons via self-serve access; however, it will remain in the Dark Archive in perpetuity as a record of data once served and it will be made available to patrons upon request.

DRAFT