Definitions Used in USGS Data Management Training Modules

Module 1: Value of Data Management

- **Data Management (DM)** – The development, execution and supervision of plans, policies, programs and practices that control, protect, deliver and enhance the value of data and information assets.[2]
- **Data Stewardship** – Taking responsibility for a set of data for the well-being of the larger organization, and operating in service to, rather than in control of, those around us.[5]
- **Metadata** – Information that describes a dataset, such that a dataset can be understood, re-used, and integrated with other datasets.[5]
- **Data Steward** – The person most knowledgeable about the resource or program who ensures that pertinent data meets any defined data standards and accurately describes the resource for which they have responsibility. Data stewards are responsible for managing data with regard to the informational needs of others.[1]

Module 2: Planning for Data Management

- **Data Management Plan (DMP)** – A formal document that outlines data management considerations before work begins and how data will be handled during and after the completion of research.[3]
- **Quality Assurance** – A set of activities to ensure that the data are generated and compiled in a way that meets the goal of the project.[4]
- **Quality Control** – Testing or other activities designed to identify problems in the data.[4] It is intended to ensure the data adhere to a defined set of quality criteria.[1]
- **Archive** – To place or store data in a data center; archiving is typically done for ensuring long-term preservation of the data and to promote discovery and use of the data.[4]
- **Data Model** – A set of diagrams and definitions that represents the enterprise data and their interrelationships in a specific and consistent way. It is a definition of the structure, rules, and constraints on data required by an enterprise to conduct all of its business functions. The data model contains entities, attributes, relationships, primary and foreign keys, and rules governing the data. Data modeling is the practice of analyzing and representing the data in a meaningful fashion, easily understood by nearly anyone in the enterprise.[1]
- **Metadata Standard** – Requirements for metadata documentation that are intended to ensure correct use and interpretation of the data by its owners and users. Different scientific communities use different sets of metadata standards; common examples are EML (Ecological Metadata language), FGDC (Federal Geographic Data Committee) standard, and ISO 19115 (International Organization for Standardization Geographic information metadata).[4]

Module 3: Best Practices for Preparing Science Data to Share

- **Provenance** – History of a data file or data set, including collection, transformations, quality control, analyses, or editing.[4]
- **File Format** – The specific organization of information in a digital computer file.[4]
- **Preservation** – A formally established, directed, and coordinated set of activities associated with maintaining records for use - either in their original form or in some other usable form.[6]

References:

[1] Chatfield, T., Ricketts, L., Selbach, R. November, 2009. Data Management for Data Stewards. Data Management Training Workshop. Bureau of Land Management (BLM).

[2] DAMA, 2009. The DAMA Guide to the Data Management Body of Knowledge (DAMA-DMBOK). New Jersey, Technics Publications, LLC; 1st Ed. 430 p.

[3] Smith, S., Tessler, S., McHale, M., Burley, T., 2012. Data Management Planning Framework (DMPf), Part III - Research Data Management Plan (RDMP) – Enterprise, version 2.2, U.S. Geological Survey.

[4] Strasser, C., Cook, R., Michener, W., Budden, A. 2012. Primer on data management: what you always wanted to know. DataONE. Accessed August 23, 2013. http://www.dataone.org/sites/all/documents/DataONE_BP_Primer_020212.pdf

[5] USGS Data Management Website. Accessed August 23, 2013. http://www.usgs.gov/datamanagement/

[6] Garlick K. 1992. Modern Archives Institute.