Notes on Data Identifiers from:

Duerr, R. E., Downs, R. R., Tilmes, C., Barkstrom, B., Lenhardt, W. C., Glassy, J., … Slaughter, P. (2011). On the utility of identification schemes for digital earth science data : an assessment and recommendations, 139–160. <http://doi.org/10.1007/s12145-011-0083-6>

**Reasons why you need a unique identifier for a dataset:**

1. provide a user with simple capabilities for for identifying data of interest. You need to be able to provide provenance and contextual information.
2. help users find and access data regardless of where the data currently reside. Data may move from repository to repository or within a repository and it still needs to be uniquely discoverable and identifiable.
3. facilitate management of the data over time - repositories change and may improve and become more complex over time. The unique identifier needs to be able to remain consistent regardless of these changes.
4. facilitate data citation in publication.

**Identifiers are ideally:**

1. "generated at the time the object is created
2. placed within the object itself
3. referenced within related material about the object, such as an associated metadata file"

**Notes on things to consider with individual identification schemes:**

Scientifically unique identifiers: if you are able to determine whether or not the contents of data set have been modified “under format transactions or content rearrangement."

ARK: Archival Resource Key

     - created by CDL, connects a user to an object, its metadata, and a statement by its provider.

     - ARKs are cited like any other URL

     - location invariant not location independent

     - unusable as scientifically unique identifiers

     - structure: [http://NMAH/]ark:/NAAN/Name[Qualifier]

     - NAAN = Name assigning authority numbers NMAH = Name mapping authority host

DOI: Digital Object Identifier

     - consists of the DOI names which are Uniform Resource Identifiers (URIs)

     - prefixes are assigned to organizations by a registry service

     - have been widely adopted by publishers and begins with "DOI"

     - DataCite promotes the use of DOIs which helps the case for scientific datasets

     - interoperable with existing numbering schemes, like ISBNs.

XRI: Extensible Resource Identifier

     - I and D (identifier and descriptor)

     - allows for semantic symbols

     - do not have widespread acceptance in the publisher community

     - unusable as scientifically unique identifier

Handle System

     - distributed by the US Military

     - relatively common to be used in a data citation

Life Science Unique Identifier

     - started through I3C

     - HTTP URIs can perform a very similar task - not sure LSUIs are necessary

PURL: Persistent Uniform Resource Locator

     - both a URL and a URI

     - developed by OCLC

URI/URN/URL:

     - use domain names to locate resources

     - URLs is preferred to URNs

UUID: Universally Unique Identifier

     - 32 hexadecimal digits that are in 5 hyphenated groups

     - opaque and non-human readable

Data Citation:

Force11 = data citation principals

* You make is easier for others to identify and acknowledge it!
* Make sure you include enough information that the next user can locate the datacite
  + Creator
  + Year of pub
  + Title
  + Publisher
  + Identifier (persistent ID)
* Two additional ones: version, resource type (dataset, database, map, soundfile, etc.)
* Always include a date of download
* Social Sciences: accession number
* Unpublished data source? The citation principal still applies! - reference it as a private communication.

Additional Resources:

Ball, A., Duke, M. (18 October 2012). How to cite datasets and link to publications [How-to guides]. Digital Curation Centre. Last updated 20 June 2012. Retrieved from [www.dcc.ac.uk/resources/how-guides/cite-datasets](http://www.dcc.ac.uk/resources/how-guides/cite-datasets)

CODATA-ICSTI Task Group on Data Citation Standards and Practices (2013). 'Out of cite, out of mind: The current state of practice, policy, and technology for the citation of data' in Data Science Journal. Vol. 12, p. CIDCR1-CIDCR75. Retrieved from [dx.doi.org/10.2481/dsj.OSOM13-043](http://dx.doi.org/10.2481/dsj.OSOM13-043)

Data Citation Synthesis Group (2014). Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA: FORCE11. Retrieved from [www.force11.org/group/joint-declaration-data-citation-principles-final](http://www.force11.org/group/joint-declaration-data-citation-principles-final)

DataCite. (2009). Why cite data? Retrieved from [www.datacite.org/whycitedata](http://www.datacite.org/whycitedata)

UK Data Archive. Create and manage data: documenting your data. Retrieved from <http://www.data-archive.ac.uk/create-manage/document>

UK Data Service. Citing data. Retrieved from [ukdataservice.ac.uk/use-data/citing-data.aspx](http://www.force11.org/group/joint-declaration-data-citation-principles-final)

University of Michigan Libguide: <http://libguides.lib.msu.edu/c.php?g=96245&p=626239>

UCSC: How to cite data: <http://guides.library.ucsc.edu/citedata>