

## Research Data Curation Activities Worksheet

Please indicate the data curation activities that you or a third party (e.g., a campus service, or an external service) perform for your data and your level of satisfaction with the results. N/A = Not Applicable

---

**Risk Management:** The process of reviewing data for known risks such as confidentiality issues inherent to human subjects data, sensitive information (e.g., sexual histories, credit card information) or data regulated by law (e.g. HIPAA, FERPA) and taking actions to reject or facilitate remediation (e.g., de-identification services) when necessary.

<b>Does this happen for your data?</b>	Yes	No	I Don't Know	N/A
<b>If Yes, are you satisfied with the results?</b>	Yes	No	Somewhat	

Comments:

---

**File Inventory or Manifest:** Data files are inspected and the number, file types (extensions), and file sizes of the data are understood and documented. Any missing, duplicate, or corrupt (e.g., unable to open) files are discovered.

<b>Does this happen for your data?</b>	Yes	No	I Don't Know	N/A
<b>If Yes, are you satisfied with the results?</b>	Yes	Somewhat	No	

Comments:

---

**Documentation:** Collect and maintain all information necessary to use and understand the data. Documentation may be structured (e.g., a code book) or unstructured (e.g., a plain text "Readme" file).

<b>Does this happen for your data?</b>	Yes	No	I Don't Know	N/A
<b>If Yes, are you satisfied with the results?</b>	Yes	No	Somewhat	

Comments:

---

**Data Citation:** Recommend a bibliographic citation for the data to enable appropriate attribution by third-party users.

<b>Does this happen for your data?</b>	Yes	No	I Don't Know	N/A
<b>If Yes, are you satisfied with the results?</b>	Yes	No	Somewhat	

Comments:

---

**Metadata:** Information about a data set that is structured (often in machine-readable format) for purposes of search and retrieval. Metadata elements include basic information (e.g. title, author, date created, etc.) and/or domain-specific elements (e.g., spatial coverage, time periods).

<b>Does this happen for your data?</b>	Yes	No	I Don't Know	N/A
--	-----	----	--------------	-----

If Yes, are you satisfied with the results?      Yes      No      Somewhat

Comments:

**Quality Assurance:** Ensure that all documentation and metadata are comprehensive and complete. Open and run the data files, inspect the contents in order to validate, clean, and/or enhance data for future use; look for missing documentation about codes used (e.g., the significance of “null” and “blank” values, or unclear acronyms).

Does this happen for your data?      Yes      No      I Don't Know      N/A

If Yes, are you satisfied with the results?      Yes      No      Somewhat

Comments:

**Secure Storage:** Data files are properly stored in a well-configured (in terms of hardware and software) storage environment that is routinely backed-up and physically protected. Perform routine fixity checks (to detect degradation or loss) and provide recovery services as needed.

Does this happen for your data?      Yes      No      I Don't Know      N/A

If Yes, are you satisfied with the results?      Yes      No      Somewhat

Comments:

**Persistent Identifier:** Provide a stable URL (such as a Digital Object Identifier or DOI) for a data set to ensure stable web accessibility for consistent citation and long-term discoverability.

Does this happen for your data?      Yes      No      I Don't Know      N/A

If Yes, are you satisfied with the results?      Yes      No      Somewhat

Comments:

**File Format Transformations:** Transform files into open, non-proprietary file formats that broaden the potential for long-term reuse and ensure that additional preservation actions might be taken in the future.

Does this happen for your data?      Yes      No      I Don't Know      N/A

If Yes, are you satisfied with the results?      Yes      No      Somewhat

Comments:

---

**Contextualize:** Use metadata to link the data set to related publications, dissertations, and/or projects that provide added context to how the data were generated and why.

<b>Does this happen for your data?</b>	Yes	No	I Don't Know	N/A
--	-----	----	--------------	-----

<b>If Yes, are you satisfied with the results?</b>	Yes	No	Somewhat	
--	-----	----	----------	--

Comments:

---

**Validation:** Generate and validate file checksums (e.g., test if a digital file has changed at the bit level) and ensure that file types match their extensions.

<b>Does this happen for your data?</b>	Yes	No	I Don't Know	N/A
--	-----	----	--------------	-----

<b>If Yes, are you satisfied with the results?</b>	Yes	No	Somewhat	
--	-----	----	----------	--

Comments:

---

**Versioning:** Provide mechanisms to manage new versions of the data overtime. Includes tracking any metadata describing the version history and any changes made for each version.

<b>Does this happen for your data?</b>	Yes	No	I Don't Know	N/A
--	-----	----	--------------	-----

<b>If Yes, are you satisfied with the results?</b>	Yes	No	Somewhat	
--	-----	----	----------	--

Comments:

---

**Embargo:** Restrict or mediate access to a data set, usually for a set period of time. In some cases an embargo may be used to protect any knowledge that the data exist.

<b>Does this happen for your data?</b>	Yes	No	I Don't Know	N/A
--	-----	----	--------------	-----

<b>If Yes, are you satisfied with the results?</b>	Yes	No	Somewhat	
--	-----	----	----------	--

Comments:

---

**Code review:** Run and validate computer code (e.g., look for missing files and/or errors) in order to find mistakes overlooked in the initial development phase, improving the overall quality.

<b>Does this happen for your data?</b>	Yes	No	I Don't Know	N/A
--	-----	----	--------------	-----

<b>If Yes, are you satisfied with the results?</b>	Yes	No	Somewhat	
--	-----	----	----------	--

Comments:

---

**File Audit:** Periodic review of the digital integrity of the data files and taking action when needed to protect data from digital erosion (e.g., bitrot) and/or hardware failure.

<b>Does this happen for your data?</b>	Yes	No	I Don't Know	N/A
<b>If Yes, are you satisfied with the results?</b>	Yes	No	Somewhat	

Comments:

---

**Use Analytics:** Monitor and record how often data are viewed, requested, and/or downloaded.

<b>Does this happen for your data?</b>	Yes	No	I Don't Know	N/A
<b>If Yes, are you satisfied with the results?</b>	Yes	No	Somewhat	

Comments:

---

**Metadata Brokerage:** Active dissemination of a data set's metadata to search and discovery services (e.g., article databases, catalogs, web-based indexes) for federated discovery.

<b>Does this happen for your data?</b>	Yes	No	I Don't Know	N/A
<b>If Yes, are you satisfied with the results?</b>	Yes	No	Somewhat	

Comments:

---

**De-identification:** Redacting or removing personally identifiable or protected information (e.g., sensitive geographic locations) from a dataset prior to sharing with third-parties.

<b>Does this happen for your data?</b>	Yes	No	I Don't Know	N/A
<b>If Yes, are you satisfied with the results?</b>	Yes	No	Somewhat	

Comments:

---

**Terms of Use:** Provide information to users of a data set that outline the requirements or conditions for use (e.g., a Creative Commons License).

<b>Does this happen for your data?</b>	Yes	No	I Don't Know	N/A
<b>If Yes, are you satisfied with the results?</b>	Yes	No	Somewhat	

Comments:

---

**Discovery Services:** Services that incorporate machine-based search and retrieval functionality that help users identify what data exist, where the data are located, and how can they be accessed (e.g., full-text indexing or web optimization).

<b>Does this happen for your data?</b>	Yes	No	I Don't Know	N/A
<b>If Yes, are you satisfied with the results?</b>	Yes	No	Somewhat	

Comments:

Note: Other, more repository specific activities (swap out as desired...try to keep worksheet to 4 pages printed front and back)

Chain of custody: Intentional recording of provenance metadata of data files (e.g., metadata about who created the file, when it was last edited, etc.) in order to preserve file authenticity.

Software Registry: Maintain copies of modern and obsolete versions of software (and any relevant code libraries) so that data may be opened/used overtime.

Emulation: Provide legacy system configurations in modern equipment in order to ensure long-term usability of data. (E.g., arcade games emulated on modern web-browsers)

Peer-review: The review of a data set by an expert with similar credentials and subject knowledge as the data creator for the purposes of validating the soundness and trustworthiness of the file contents.

Succession Planning: Planning for contingency, and/or escrow arrangements, in the case that the repository (or other entity responsible) ceases to operate or the institution substantially changes its scope.

File download: Allow access to the data materials by authorized third parties.

Contact Information: Keep up-to-date contact information for the data authors and/or the contact persons in order to facilitate connection with third-party users. Often involves managing ephemeral information that will change over time.

Data Visualization: The presentation of pictorial and/or graphical representations of a data set used to identify patterns, detect errors, and/or demonstrate the extent of a data set to third party users.

Full-Text Indexing: Enhance the data for discovery purposes by generating search-engine-optimized formats of the text inherent to the data.

Arrangement and Description: The re-organization of files (e.g., new folder directory structure) in a dataset that may also involve the creation of new file names, file descriptions, and the recording of technical metadata inherent to the files (e.g., date last modified).

Selection: The result of a successful appraisal. The data are determined appropriate for acceptance and ingest into the repository according to local collection policy and practice.

Repository Certification: The technical and administrative capacities of the repository undergo review through a transparent and well-documented process by a trusted third-party accreditation body (e.g., TRAC, or Data Seal of Approval).

Transcoding: With audio and video files, detect technical metadata (min resolution, audio/video codec) and encode files in ways that optimize reuse and long-term preservation actions. (E.g, Convert QuickTime files to MPEG4).

Authentication: The process of confirming the identity of a person, generally the depositor, who is contributing data to the data repository. (e.g., password authentication or authorization via digital signature). Used for tracking provenance of the data files.

Deposit agreement: The certification by the data author (or depositor) that the data conform to all policies and conditions (e.g., do not violate any legal restrictions placed on the data) and are fit for deposit into the repository. A deposit agreement may also include rights transfer to the repository for ongoing stewardship.

Conversion (Analog): In effort to increase the usability of a data set, the information is transferred into digital file formats (e.g., analog data keyed into a database). Note: digital conversion is also used to convert "fixed" data (e.g., PDF formats) into machine-readable formats.

Indexing: Verify all metadata provided by the author and crosswalk to descriptive and administrative metadata compliant with a standard format for repository interoperability.

Curation Log: A written record of any changes made to the data during the curation process and by whom. File is often preserved as part of the overall record.

Cease Data Curation: Plan for any contingencies that will ultimately terminate access to the data. For example, providing tombstones or metadata records for data that have been deselected and removed from stewardship.

File renaming: To rename files in a dataset, often to standardize and/or reflect important metadata.