

ICONC Project Source Data Entry Specifications

Document Introduction

Critical Data Entry Specifications:

1. Each row in the 'Source Data' worksheet should represent a unique activity-organization-pair.
2. For each activity-organization pairing, a row will list activity info (identical across all organization pairings for the given activity), and organization activity (unique for each row.) Columns I to R will list information for each organization involved in an activity.
3. After completing data entry for a single activity and all of its related organizations, you should have multiple rows with identical activity data posted in columns A to H and S to Y.
4. Care must be taken with respect to spelling, spacing, and text formatting. Data is pulled directly from the Source Data spreadsheet into reports.



Omission of any of the above will result in low-quality data of limited to no use, which would have to be re-entered.

Data Inclusion Specifications:

Activities must meet two criteria to be represented in this data set:

- 1) Digital preservation/curation/stewardship must be a core component of the activity.
 - Often this will be specified on the website for the activity, if not apparent from the name itself.
- 2) The activity must be collaborative in nature.
 - Multiple organizations (i.e. institutions) must be leading, contributing, building, or participating in the production or provision of the activity.
 - If activity is centralized, i.e. one institution controls all services or governance, then it is not collaborative and would be omitted.



If you are unsure whether an activity should be included, err on the side of inclusion and check in with Christina.

Source Data Worksheet Notes:

- Yellow column headings pertain to activity data fields.
- Green column headings pertain to organization data fields.
- Orange column headings pertain to data relevant for activity-organization pairings involving an activity type of 'Digital Preservation Network'

Activity Related Field Definitions

A) Activity Name

- Spreadsheet location: Column A
- Data definition: the name of the collaborative digital preservation activity (e.g. ANADP Conference or ADPNet)
- *Data formatting notes:*
 - Use title case style and capitalization, as this data will pull directly into reports.
 - If the name is an acronym (e.g. ADPNet), follow the acronym with a hyphen and the full name of the initiative (e.g. ADPNet - Alabama Digital Preservation Network)

B) Activity Status

- Spreadsheet location: Column B
- Data definition: classification of the time-bound status of a collaborative digital preservation/curation/stewardship activity.
- *Restricted data field options: Project, Program, Other*
 - *Project:*
 - *Should be time limited, i.e. have an end date, either in the past or future.*
 - *e.g. projects tied to grant awards*
 - *Program:*
 - *Often a matured project.*
 - *Could have an end date if the program ended.*
 - *If ongoing, it should not have an end date (if so, classify as a project.)*

C) Activity Type

- Spreadsheet location: Column C
- Data definition: classification of the type of collaborative digital preservation/curation/stewardship activity.
- *Restricted data field options: Research, Digital Preservation Network, Open Repository / Digital Library, Closed Repository, Conference, Training, Service/Software, Other*
 - *Research:*
 - *Only include if digital preservation/curation/stewardship oriented.*
 - *Includes grant awards, e.g. the ETD project.*
 - *Digital Preservation Network*
 - *Must have primary purpose of preservation (not public access or something else. If focus is elsewhere, use a different activity type.)*
 - E.g. Portal to Texas History - would not be classified as Digital Preservation Network, as primary focus is access. Would be classified as Digital Library.
 - *If multiple institutions are merely contributing content (as opposed to running the network), then classify as a Digital Library/Repository.*
 - *Nodes hosting content for the network should be found on the network's website. Otherwise, enter N/A*
 - *Multiple institutions acts as nodes, hosting all of the content.*

Activity Related Field Definitions Cont.

- *Open Repository / Digital Library*
 - *Multi-institution repositories should be included if explicit emphasis on preservation/curation.*
 - *If multiple institutions are contributing content, this counts as multi-institution.*
 - *e.g. Portal to Texas History*
 - *Implicit, non-stated preservation activities are not enough to include a digital library or knowledgebase. These entries should all have an explicit focus on long-term management and preservation.*
 - *Single institution repositories should not be entered, as they are not collaborative.*
 - *Public content access (i.e. open access) is often a marker of the repositories.*
- *Closed Repository*
 - *Closed repositories do not have access as a primary goal. Instead, these are often focused on disaster recovery and or data retention.*
 - *Multi-institution repositories should be included if explicit emphasis on preservation/curation.*
 - *If multiple institutions are contributing content, this counts as multi-institution.*
 - *Single institution repositories should not be entered, as they are not collaborative.*
 - *Includes Dark Archives if multi-institutional*
 - *Single institution acts as the node, hosting all of the content.*
- *Conference*
 - *Primary focus of the conference must be digital preservation/curation/stewardship.*
 - *Cannot be included if digital preservation is only a thread, or theme.*
 - *Collaboration may be in the form of putting on the conference (organizing/planning), or be a result (e.g. CurateCamp)*
 - *Don't organizational attendees as organization/activity pairs.*
 - *For each conference, do capture:*
 - *Organizing/Program committee members & institutions:*
 - *Host/lead organization*
 - *Founding institution*
- *Training*
 - *Note: this category includes collaborative programs, initiatives (e.g. DPOE, SAA DAS), and training events (e.g. week-long institutes, pre-conference workshops). As long as they are explicitly preservation focused and collaborative.*
- *Service/Software*
 - *Must be explicitly preservation/curation/retention focused*
 - *Must be collaborative in nature (hosted, run, or produced by multiple institutions.)*

Activity Related Field Definitions Cont.

- *Standard / Policy Development*
 - *Development effort for a common standard (technical, policy, etc.)*
 - *Must be at a co-investigator level or management level to be listed as a partner. Others can be classified as participants.*
- *Data formatting notes:*
 -  ○ *If activity fits the definition of more than one activity type, enter text in the format: 'Multiple: <activity type1>, <activity type2>.'*
 - Entries in this field must be capitalized and spaced identical to one of the fixed data options to properly generate reports.

D) Activity URL

- Spreadsheet location: Column D
- Data definition: url for activity website. If activity has been suspended, a link to website referencing final project reports, description, partners, which maybe provided by LoC if project was NDIIPP supported.
- *Data formatting notes:*
 - Include <http://www> in address.
 - If entering multiple urls, separate with a comma and space. Enter multiple urls if a good summary page on the LoC's website, in addition to project page.
 - If none available: leave blank.

E) Activity Short Description

- Spreadsheet location: Column E
- Data definition: general description of activity, of eight words or less. Words should not reference information that is captured elsewhere in the spreadsheet (i.e. NDIIPP funded, geographic focus, current operational status). Consider a functional definition, which may include a focused community reference.
- *Data formatting notes:*
 - Use phrases in sentence case capitalization.
 - Omit periods.

F) Was Activity NDIIPP Funded

- Spreadsheet location: Column F
- Data definition: answer of whether NDIIPP provided direct funding to activity in grant form.
- Restricted data field options: Yes, No
- *Data formatting notes:*
 - Entries in this field must be capitalized and spaced identical to one of the fixed data options to properly generate reports.
 - *CD can make this a drop down if needed.*

Activity Related Field Definitions Cont.

G) Year Activity Started

- Spreadsheet location: Column G
- Data definition: enter first year of activity operations.
- *Data formatting notes:*
 - Enter year in four digits, *e.g.* 2013
 - If unknown, leave blank.

H) Year Activity Ended H

- Spreadsheet location: Column F
- Data definition: enter last year of activity operations.
- *Data formatting notes:*
 - Enter year in four digits, *e.g.* 2013.
 - If activity has not ended, enter 'Ongoing'.

I) Activity Geographic Focus

- Spreadsheet location: Column I
- Data definition: classification of the intended regional scope of the collaborative activity.
- Restricted data field options: International, National, Regional, State, Local, None Specified.
- *Data formatting notes:*
 - Entries in this field must be capitalized and spaced identical to one of the fixed data options to properly generate reports.
 - *CD can make this a drop down if needed.*

J) Current Host/Lead Organization for Activity

- Spreadsheet location: Column I
- Data definition: Name of the organization that is currently hosting and/or leading the activity. This should be explicitly listed on the website.
- *Data formatting notes:*
 - Capitalize as a proper noun.
 - Multiple organizations could have this designation, but it has to be documented online.
 - If unknown, leave blank.

K) Founding Institution of Activity

- Spreadsheet location: Column K
- Data definition: Name of organization that founded the activity in the first year. This needs to be explicitly stated.
- *Data formatting notes:*
 - Capitalize as a proper noun.
 - If unknown, leave blank.

Organization Related Field Definitions

L) U.S. State/Territory of Participating Organizations

- Spreadsheet location: Column L
- Data definition: name of U.S. State or Territory where participating organization is located.
- Restricted data field options: List of U.S. States and Territories
- *Data formatting notes:*
 - If participating organization is based outside of the U.S. states and territories, enter 'International'.
 - Entries in this field must be capitalized and spaced identical to one of the fixed data options to properly generate reports.
 - *CD can make this a drop down if needed.*

M) Participating Organization Name

- Spreadsheet location: Column M
- Data definition: full name of participating organization. If name is an acronym, begin the entry with the acronym followed by a hyphen and then the full name.
 - For academic institutions: individual schools, departments, centers, and institutes will be collapsed and represented solely as the university.
 - For state governments: state agencies will be represented individually. E.g. Secretary of State and State Archivist would both be included as organizations if they were involved, even if the Archivist office was nested within the Secretary of State's office.
 - For international partners: only include an international organization if they are a host or founder of the activity. International organizations acting in the other roles will not be represented in this version of the data set, given time limitations and the domestic focus of the project.
- *Data formatting notes:*
 - Capitalize as proper nouns.

N) Organization Role in Activity

- Spreadsheet location: Column N
- Data definition: classification of the organization's role(s) in the collaborative activity. If multiple, state, 'Multiple: <role 1>, <role 2>
- Restricted data field options: Founding Institution, Host/Lead, Partner, Sponsor, Participant, Developer, Member Organization
- *Data formatting notes:*
 - Entries in this field must be capitalized and spaced identical to one of the fixed data options to properly generate reports.

O) Organization Address Line 1

- Spreadsheet location: Column O
- Data definition: First line of listed address of organization, inclusive of street number and street.
- *Data formatting notes:*
 - *Enter St., Rd. and other common street abbreviations in abbreviated form.*

Organization Related Field Definitions Cont.

P) Org Address Line 2

- Spreadsheet location: Column P
- Data definition: Second line of listed address of organization, inclusive of street number and street. Only use if needed.
- *Data formatting notes:*
 - *Spell out Suite (instead of Ste.)*

Q) Org City

- Spreadsheet location: Column Q
- Data definition: City of listed address of organization.

R) Org State/Province (incl. country if international)

- Spreadsheet location: Column R
- Data definition: State or territory for listed address of organization. If organization is based outside of the U.S., enter the nation instead (e.g. Spain).
 - *Note, while this information mirrors the data in column I, it must be formatted differently to enable address Geolocation.*

S) Org Zip

- Spreadsheet location: Column S
- Data definition: Postal code for listed address of organization.

T) Organizational Contact Name

- Spreadsheet location: Column T
- Data definition: Primary contact at this organization for the given activity, if known. For activities that have ceased, enter the person of record at the given organization. If you cannot find this information, enter, "None specified".
- *Data formatting notes:*
 - *Include both first and last names.*
 - *Do not include titles.*

U) Organizational Contact Email

- Spreadsheet location: Column U
- Data definition: email address for primary contact listed in column Q, if known. If this information is not readily available from activity website, leave data blank.
- *Data formatting notes:*
 - *Use @ within the address, (e.g. Christina@unt.edu)*

Digital Preservation Network Related Field Definitions

V) If Activity is Digital Preservation Network: List Member/client/participant organizations

- Spreadsheet location: Column V
- Data definition: a listing of all participating organizations in the network. Note: all organizations within this list should also have their own rows tied to the activity.
- *Data formatting notes:*
 - *Enter each organization as a Proper Noun*
 - *Enter each new organization on a new line within the cell. Press 'Alt' and 'Enter' simultaneously to add line break within the cell. Alternatively, use notepad to create or clean the list from the web and then paste the clean list into the cell.*

W) If Activity is Digital Preservation Network: Is the network a membership organization?

- Spreadsheet location: Column W
- Data definition: classification of membership organization status. Data should only be entered for activity type pairings (rows) where activity type= 'Digital Preservation Networks'.
- Restricted data field options: Yes, No, N/A
- *Data formatting notes:*
 - *Entries in this field must be capitalized and spaced identical to one of the fixed data options to properly generate reports.*

X) If Activity is Digital Preservation Network: List node hosting organizations

- Spreadsheet location: Column X
- Data definition: a listing of all organizations hosting nodes on the network. Note: all organizations within this list should also have their own rows tied to the activity. This would be explicitly listed on the website.
- *Data formatting notes:*
 - *Enter each organization as a Proper Noun*
 - *Enter each new organization on a new line within the cell. Press 'Alt' and 'Enter' simultaneously to add line break within the cell. Alternatively, use notepad to create or clean the list from the web and then paste the clean list into the cell.*

Y) If Activity is Digital Preservation Network: Technical Administration Organization - Technical Platform

- Spreadsheet location: Column Y
- Data definition: Composite data element listing both the organization taking care of the technical administration of the network, and the technical platform being used to support the network.
- *Data formatting notes:*
 - *List both organizations and technical platforms as proper nouns, separated by a hyphen, e.g. Educopia - MetaArchive.*