U.S. Global Change Research Program

The Global Change Information System: an overview of its Data Model and Metadata Conventions

Introduction

With over 15,000 people, 9,000 organizations, over 3,000 datasets, and much more, the Global Change Information System, or GCIS, catalogs a copious amount of information and metadata. Metadata in the system include the names of USGCRP report authors and their affiliations, report figures and images, datasets, and a substantial list of publications referenced in USGCRP Reports. These are either relationally or semantically linked with one another to improve the traceability of climate information and data provenance, thereby bolstering the overall utility of the GCIS. As the prime access point to U.S federal assessments on Climate Change, reports, and tools produced by the USGCRP – including the native data underpinning these resources – the GCIS is an important resource for Global Change research.

Given this brief description of the GCIS, it is imperative that the system make use of existing and highly regarded metadata standards and conventions where possible. This document aims to describe the current GCIS data model, as well as highlight all the conventions and standards deployed in the system.

Metadata Standards and External Resources deployed in the GCIS:

The GCIS conforms with well-established metadata standards and leverages external resources. A list of these is provided below:

- **Identifiers:**
  GCIS makes use of external identifiers such as Digital Object Identifiers (DOIs), ORCIDs, ISBNs, ISSNs. ISBN resolution is handled through WorldCat. They may also utilize identifiers created for other aggregator systems, such as Data.gov identifiers. When GCIS must create its own identifier, it may utilize a Universally Unique Identifiers (UUIDs) when no human-readable identifier is reasonable.

- **Provenance:**
  GCIS utilizes W3C provenance to represent some relationships inside the GCIS. In particular, GCIS utilizes PROV-O and CiTO verbs.

- **Technology:**
  GCIS is written in Perl using the Mojolicious web framework, the Rose::DB Database interface, and many other fine modules from the CPAN. It relies on PostgreSQL for data storage.

Map inserts are supported by OpenStreetMap data with MapBox tiles and implementation through LeafletJS.

- **Data Export Formats:**
  GCIS offers a multitude of human- and machine-readable data export formats, defaulting with the HTML pages. Publication pages can be downloaded as JSON and YAML data, the semantic formats Turtle, N-Triples, JSON Triples, RDF+XML, and RDF+JSON, and visual representations in Graphviz, and SVG of the semantic mapping. Additionally, the data Array(s) behind Table Publications may be downloaded as CSV. List pages can be downloaded as JSON, YAML, or CSV formatted data.

Metadata Quality Characteristics observed in the GCIS:

The Harvard Law School characterizes metadata as "electronic fingerprints", that aid the searchability of documents or online resources[1]. Therefore, good metadata records provide users with enough information to understand and (or) locate the object that the record is about.
There is no shortage of qualities that make metadata records "good". Nevertheless, GCIS metadata records strongly align with the following quality characteristics:

- **Completeness:** Is there enough information to facilitate findability?
- **Conformance:** Do metadata records abide by established standards?
- **Credibility:** Is the metadata based on trustworthy sources?
- **Readability:** Is the metadata record both machine and human-readable?

The GCIS is designed to aid users in their search for climate science information referenced in USGCRP products. The qualities above reflect this and thus, are central to the GCIS data model.

[1] https://hls.harvard.edu/dept/its/what-is-metadata/

The GCIS Data Model

Global change information in the GCIS is structured using the GCIS data model. This data model represents relationships and entities (publications or resources) such as reports, report chapters, figures, images, tables, bibliographic entries, organizations, and people. These entities will, heretofore, be referred to as "resources", in accordance with GCIS parlance.

GCIS Resources:

The terminology below describes the resources represented in the GCIS :

Resource	Definition
Activity	A description of the process used to construct a publication. Examples include: 063fd83f-nca3-maurer-r201304-process.
Article	An independent story in a journal, book, or other published material. In GCIS, an article has a DOI (if available), like 10.1002/grl.50527, which is used to uniquely identify it.
Book	A long-form publication, generally centered around one particular topic. In GCIS, books are cataloged using UUIDs, though ISBN numbers are also listed. Examples include : Water Is for Fighting Over and Other Myths about Water in the West, 6d3eb0d6-568a-40d3-9b10-393f943038da.
Chapter	Used to separate reports into smaller, more subject-intensive sections. In GCIS, chapters feature a unique mnemonic identifier and possibly a number. Some chapters, like an appendix, do not have a number. Chapters may have figures, tables, findings, and references associated with them. Chapters can also have distinctive Contributors to their parent Report. An example is NCA4's Chapter 4 : Energy Supply, Delivery, and Demand.
Contributor	A person or organization that creates information (usually a figure or image) listed in GCIS. Analogous to Author records linked to publications. A few examples of Contributor role types can be found here.
Dataset	A collection of data pertaining to a specific subject. Datasets are considered a type of publication. In GCIS, a dataset is given a unique mnemonic identifier, such as nca3-ncep-ncar-r1.

Figure	A visual representation (e.g., a graph, map, photograph, diagram, chart, or a satellite image) referenced in the text. It may be composed of one or more images. Example: observed-us-temperature-change.
Finding	A major conclusion or discovery garnered from the scientific evidence presented in a report. Has a statement as well as phrases and bibliographic references describing the confidence level, uncertainties, and evidence for that finding. Example: global-climate-is-changing.
Image	The unique graphic used to create a figure. One or more images make a figure. Example: 1f5a3cdd-fc45-403e-bf11-d1772005b430.
Instrument	A measuring apparatus. Often associated with one or more platforms via instrument instances. Example: poseidon-2.
Instrument Instance	A platform/instrument combination. It may be associated with one or more datasets. Example: jason-1/poseidon-2.
Journals	A publication containing a collection of articles, original research, and conclusions. Generally shorter than a book, peer-reviewed, and released in multiple volumes or periodicals. In GCIS, a journal may have a print and online ISSN, and is uniquely identified within GCIS using a mnemonic identifier (such as climatic-change). Example: Biogeochemistry.
Model	A simulation or projection based on previously gathered data. Models are used to simulate climate systems and create model runs. Example: ccsm3.
Model Run	A unique piece of data ascertained by combining a model, scenario, time range, time resolution, and sequence number.
Organization	A group, such as an institution, agency, or NGO. In GCIS, organizations are uniquely identified by mnemonic identifiers. Example: us-global-change-research-program.
Person	Individuals associated with global change research. In GCIS, they are uniquely identified by numeric identifiers. Example: person/16009.
Platform	A host for information gathering tools/instruments. May host one or more instruments. Example: Terra.
Project	Initiatives led by an organization to generate models and eventually reach conclusions. Example: cmip3.

Publication	A report-specific resource (such as a chapter, figure, finding, or table), or a non-report-specific resource (such as a journal, article, image, web page, book, or dataset).
Reference	A bibliographic entry or citation pointing to a publication, such as an article, web page, or book. In a report, a reference may appear in a chapter, figure, finding or table. References are labeled with a universally unique identifier (UUID). Example: 0006123e-10a3-4501-a89c-95a7921a9c3d.
Region	An area separated by its climate, geographic makeup, or another identifying factor. Example: alaska-us
Report	An authored compilation of information on a particular subject. May contain chapters, figures, tables, findings, and references. Example: NCA3 and NCA4.
Scenario	A set of assumptions used to help understand potential future conditions. Example: rcp_2_6
Table	An organized list referenced in the text. Often embedded in a report. Example: list-us-droughts
Web Page	A single page of a greater website. In GCIS, a web page is assigned a UUID, like 26625ddf-dd19-4dd1-a35d-33c68c5b2d6e, to identify its state at a particular point in time.

Climate Science is a complex discipline. Given the amount of data generation and data processing involved, and the use of climate science data in critical decision making, traceable provenance becomes invaluable. The GCIS adopts both a relational model and a semantic model, to maintain traceable provenance for the resources it catalogs.

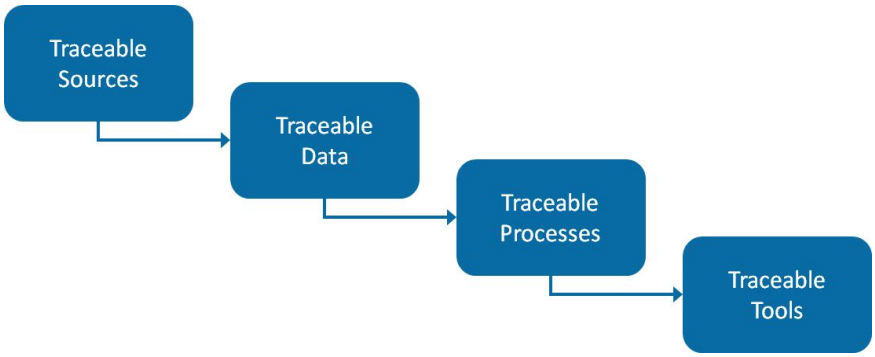

Figure 1: The GCIS "traceable accounts" model. Each step in this chain is an additional layer of reproducibility, but also a significant increase in the difficulty of capturing and curating.

Relational Model:

The relational model used by GCIS captures one-to-many, many-to-many, and many-to-one relationships between the resources listed above. Journals have many articles, reports have many figures, findings, tables, and chapters. The relationship between images and figures, is many-to-many, as is the relationship between resources or publications of any type and contributors.

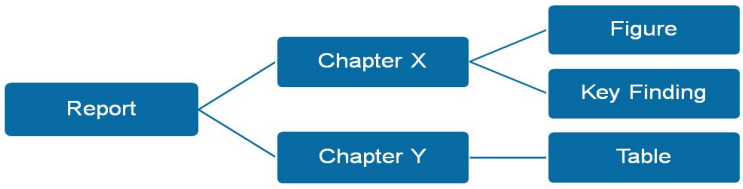

Figure 2: A representation of the GCIS relational model.

Semantic Model:

Resources may also be related to each other, and in this case, the relationship between two resources can be annotated with a term from a semantic vocabulary. Furthermore, the relationship may have an "activity" associated with it. All GCIS resources have representations in Turtle (a semantic format for storing data). The GCIS vocabulary is defined in the GCIS ontology. Many other ontologies are used including, most notably PROV. The entity-activity-agent model of PROV has been applied to the GCIS through the use of resources, activities, and contributors.

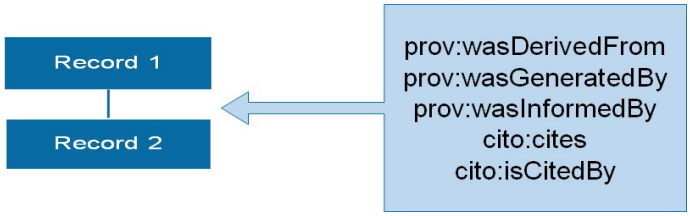

Figure 3: A representation of the GCIS semantic model.

As illustrated above, the GCIS data model leverages widely recognized conventions and standards. Almost every GCIS resource conforms to existing international standards. However, in unique instances where established conventions are non-existent, the GCIS team develops its own conventions.

The GCIS Resource Conventions

To better provide authoritative, accessible, usable, and timely information, GCIS resources need to be traceable and transparent (ref: IQA). There are several resource types in the GCIS, each with an associated field convention, provenance convention and relationship convention. These conventions could be "default", or "resource-specific".

Default Conventions:

Assigning citable, unique and persistent identifiers to information held within the GCIS greatly increases the transparency and traceability of that information as well as addressing a growing trend within the Federal research community: requiring researchers to formally cite the data and document the processes that led to a given research result (Mayernik, et al., 2012).

- **Title:**
  The Resources title (or name, if no title field) is modified to become the identifier. Spaces are replaced with "-". All letters are lowercase.

- **UUID:**
  A UUID is generated for the resource and used as the identifier. Globally unique.

- **Number:**
  The Resource type has an incrementing number as the identifier. Unique only for a given resource.

Resource Specific Conventions:

The conventions, together with the use case of each GCIS resource type, is discussed below.

① Resource Type: Activity

Appropriate Use Case: Activities are created to connect components of USGCRP reports to the source publication it was created from, in order to convey the process of creating the asset. For example, an Activity can link

- An Image back to its dataset.
- An Array back to its dataset.
- A Figure back to its Article.

Activities are not created for non-USGCRP products.

Field Conventions:

Field	Description
identifier	The format of the activity identifier going forward should follow the template: "report_identifier-UUID-activity"
methodology	The process of creating the resulting object from the input, in the author's own words and in such a way that another expert party could reproduce the output.
visualization_methodology	The process of creating the visual portion of the output object, if any and if distinguished from the main methodology.
methodology_citation	The citation to the methodology, if it has been published.
methodology_contact	The point of contact for the methodology, if any.
activity_duration	Captures the time taken in the process to get from the source object to the final one.
source_access_date	The date the parent resource was accessed.
source_modifications	A written description of modifications done to the source object.
modified_source_location	The location of the modified source, if available.
interim_artifacts	Deprecated outside of NCO assessment activities. The names of files created along the way to create the final product.
output_artifacts	Deprecated outside of NCO assessment activities. The final output filenames from the process.
computing_environment	Operating systems and versions used to perform this activity
software	Primary software (with version) used.
visualization_software	Primary visualization software (with version) used.

start_time	Time bounds used to restrict the input object. Optional, depending on applicability. If equal to end_time, indicates a temporal moment.
end_time	Time bounds used to restrict the input object. Optional, depending on applicability. If equal to start_time, indicates a temporal moment.
dataset_variables	A list of Dataset Variables applied in this activity.
spatial_extent	Spatial bounds used to restrict the input object. GeoJSON. Optional, depending on applicability.
data_usage	DEPRECATED: A description of the way in which input data were used for this activity.
notes	DEPRECATED: Other information about this activity which might be useful for traceability or reproducibility.
duration	DEPRECATED: Use activity_duration to document the time taken to perform the activity.

Provenance Conventions: None.

Relationship Conventions:

- Associated with two other publications (one parent, one child) as the process of going from one to the other.

② Resource Type: Array

Appropriate Use Case: Arrays hold the low level content of a Table entity. Arrays are created whenever a table is created.

Field Conventions:

Field	Description
identifier	Automatically created UUIDs. In the case of multiple arrays per table, we may identify the array with short, internally distinct identifiers the show which is which.
rows_in_header	Usually one, indicates how many rows to be ignored as headers, non-data.

rows	The actual data, including headers, if any, stored as a database array. Avoid special characters, if possible.

Provenance Conventions:

- **prov:wasDerivedFrom:**
  With very rare exceptions, should be used on the Array instead of the Tables. Most commonly, Arrays are connected to a Dataset via an Activity.

- **cito:cites:**
  Generally the Table object will be the thing using cito:cites. It is plausible to have a good use case for an array to use cito:cites.

Relationship Conventions:

- **table:**
  Arrays should always belong to a Table. The relationship is generally added through the Table object. Multiple arrays are added in their original order.

③ Resource Type: Article

Appropriate Use Case: The article resource type in GCIS is used to represent articles referenced in USGCRP reports. This means that each article in the GCIS is used as a reference in a USGCRP report.

Identifier Convention: An article's DOI is used as the identifier when it is available. Acceptable fall-back identifiers include PMID, PMCID, and CDC MMWR. If the article does not have any of these external identifiers, the article's title -- words (all in lowercase) separated by hyphens, and no numbers -- becomes the identifier.

Examples include:

- For an article having the doi of 10.1002/aqc.880:
  http://data.globalchange.gov/article/10.1002/aqc.880
- For an article without a DOI, but has a PubMed ID of pmid-22377962:
  https://data.globalchange.gov/article/pmid-22377962
- For an article without a doi or other acceptable external identifier:
  http://data.globalchange.gov/article/resilience-adaptability-and-transformability-in-socialecological-systems

Field Conventions:

Field	Description
Title	Title should match the exact Article title
DOI	The exact DOI, if one exists. Otherwise blank.
Year	The year the article was published, as in the year this volume of the journal came out.
Notes	Not In Use. May be used in the future for extraordinary circumstances, i.e. journal retractions.
Journal	Should refer to the GCIS Journal entity corresponding to the article's Journal. Should never be blank.
URL	Should always be provided if we don't have a DOI. Preferred to have regardless. In the case of neither a DOI nor a URL existing, may be left blank.
journal vol	Which volume of the Journal this article was published from. Sourced from crossref.org, not from the corresponding imported reference.
journal pages	Which pages of the journal this article occurs at. Preferred to be a page range. Initial page is acceptable. Sourced from crossref.org, not from the corresponding imported reference.

Provenance Conventions:

- Articles are related to USGCRP produced publications only.
- An article can be cito:isCitedBy a USGCRP publication.
- An article can cito:cites a USGCRP publication.

Relationship Conventions:

- **Contributors:**
  Persons and Organizations are connected to Articles with the following conventions:
  - Role Types
    - author is the assumed role
    - point of contact role is used for specified corresponding author
    - other roles are not used unless they are called out in the article
  - sort key should be used to match the article's order

  In cases of articles with hundreds of authors, some editors, and a point of contact, just the editors and point of contact are added.

- **Gcmd_keywords:**
  - Not to be used at this time.
- **Regions:**
  - Not to be used at this time.
- **Files:**
  - Not used in articles.

④ Resource Type: Book

Appropriate Use Case: Books are not created by us directly, but rather as children publications for references. Reports may have ISBNs, and be published as physical books, but they still better fit as a Report object over a Book object.

Field Conventions:

Field	Description
identifier	Fine as UUID. GCIS will automatically generate one from a blank identifier field
title	The book title, as given by TSU. Checked against the WorldCat.
isbn	Required if the book has one. We prefer the thirteen digit one if it exists. Should be the one that links properly to WorldCat. We do not have a preference for ISBNs of different editions (e.g. hardcover vs paperback), so long as it is correct and matches our other metadata.
year	The publication year of the edition of the book matching the ISBN. If ISBN is not known, then the publication year of the first edition is preferred.
publisher	Deprecated. Use a contributor with the role, Publisher.
number_of_pages	Imported from WorldCat. Not a priority if not available from WorldCat.
url	Official online published copy of the book is fine, if available. Use the publisher's webpage for this book. Imported from TSU or WorldCat. Research to find it otherwise.
in_library	DEPRECATED
topic	Imported from TSU or WorldCat. Not a priority to fill in otherwise.

Provenance Conventions:
Only "cito:isCitedBy", is used; created from the Reference.
Nothing is created off of the book itself.

Relationship Conventions:

- **Contributors:**
  - The publisher role is the first priority.
  - Editors and authors roles are second priority.
  - Other roles can be added as appropriate for each case
- **File:**
  - Can have a related file if available. Not a priority.
- **gcmd_keywords:**
  - not to be used for the moment. To be determined.
- **Regions:**
  - not to be used.

⑤ Resource Type: Chapter

Appropriate Chapter Use Case: Chapters are created for USGCRP Products and for external reports that are being cited in GCIS a large number of times at a chapter specificity.

Field Conventions:

Field	Description
report_identifier	the GCIS identifier of the report this resource belongs to.
title	official published chapter title. Does not include a subtitle, unless needed for uniqueness. Doesn't include "Chapter 1:" prefix. Doesn't include the Report name. See CSSR as an example of a good title convention.
identifier	title of the chapter with dashes in place of the spaces.
number	the official published chapter ID, if any. (i.e. 1, 2, 3, A, B, C, i, ii, etc). Likely blank for chapters like front matter and executive summary. Does not determine chapter ordering. See sort key.
url	The direct URL for the given chapter. Same as what the chapter DOI, if any, would resolve to. Blank if chapter does not have its own URL. Do not use the report URL.

sort key	the internal number to make the chapters sort in order.
doi	the DOI for the Chapter, if any. Blank if not. Do not use the report's DOI.

Provenance Conventions:
Chapters use cito:CitedBy and cito:Cites.
Chapters should be created with their internal citations as cito:Cites
We are generally only concerned with cito:CitedBy when it is cited by a future USGCRP product. We do not try to proactively find external citations.

Relationship Conventions:

- **Report**
  - must exist.
- **figures/findings/tables**
  - only populate these for USGCRP reports
  - If an external federal agency is creating a report and collects metadata in GCIS-compliant processes, they may populate those records.
    - See good convention exemplar **epa-multi-model-framework-for-quantitative-sectoral-impacts-analysis-2017**.
- **Contributors**
  - only populated for the USGCRP produced reports.
  - For some reports, we list all authors on the report and not per-chapter. (See CSSR)
  - Generally speaking, follow how the report links authors.
    - If they are listed on the report as a whole, list on the report.
    - If they have the authors specified by chapter, list on the chapter.
- **Files**
  - should be the downloadable PDF of the chapter, usually hosted externally to GCIS.
  - Populated for USGCRP produced reports.
  - Maybe populated for external reports.
- **gcmd_keywords**
  - not to be used for the moment. To be determined.
- **regions**
  - not to be used.

⑥ Resource Type: Contributor

Appropriate Use Case: Contributors are created on all publications that provide publishers, authors, editors, etc unless the decision is made to not specify them.

Field Conventions:

Field	Description
id	An opaque numeric identifier for this contributor.
person_id	The person (optional).
organization_identifier	The organization.
role_type_identifier	Generally speaking these should match the exact role type specified in the publication. Where reasonable, synonymous types may be used (Chief Editor / Lead Editor). If the publication gives the role in a foreign language, we would use the existing matching translation or create a new English role to match.
Point of Contact	We want to have a PoC on all publications where it is available, going forward. We are undecided on whether we will back-fill this in. May simply be the author. If an email ID is only provided for one person, they are PoC. If there is only one with a mail symbol but all have email IDs, the person with the symbol is the PoC. For third party reports we are loading into GCIS, we ask the third party who the PoC is.
Author	This is the default role type assumed if no role is listed on the publication.

Provenance Conventions: None.

Relationship Conventions:

- **Publication:**
  - Contributors are created on all publications that provide publishers, authors, editors, etc unless the decision is made to not specify them.
- **Reference:**
  - Deprecated.
  - We will not connect References to Contributors going forward, nor attempt to backfill in the connection.

⑦ Resource Type: Dataset

Appropriate Use Case: Datasets are created in GCIS when cited by Reports, Figures, Tables, etc. A dataset should be unique per version. Dataset has many extraneous or problematic fields. Create with caution.

Field Conventions:

Field	Description
identifier	an external ID is preferred if sufficiently unique. Data.gov is preferred as an external ID. DOI is also acceptable. An identifier must be unique for this version of the dataset. If no acceptable external ID, org_name + dataset_name + version. (In version, "." becomes "_"). If there is no version, org_name + dataset_name + publication_year.
name	Use the same name as the external source.
version	The external version, if any. Follow their convention.
description	Any available description from the landing page or website of the organization. May be shortened.
native_id	The external id, if any. The identifier for this dataset given by the producer or archive for the dataset.
url	A URL for a landing page. Probably can't be guaranteed unique.
doi	The DOI for this dataset, if any. Probably can't be guaranteed unique.
release_dt	The date on which this version of this dataset was released.
publication_year	The date on which this dataset was initially published, regardless of version.
description_attribution	A URL which contains the description of this dataset given in the description field. If the same as URL, duplicate URL here.

Provenance Conventions:
Datasets use prov:wasDerivedFrom, where a figure "prov:wasDerivedFrom" a dataset, through an activity.

Relationship Conventions:

- **Contributors:**
  - Datasets should have their owner organization linked to them via a contributor.
  - Normally via the data_archive role.
- **Files:**
  - Historically Datasets have sometimes had a jpg associated with them.
    Not required

⑧ Resource Type: Figure

Appropriate Use Case: Figures should be created for USGCRP Publications. They are generally created automatically via synchronization with the TSU Resources Metadata. In the case of supported non-USGCRP Publications, Figures may be created if the publishing entity provides the information.

Field Conventions:

Field	Description
identifier	The title of the figure with hyphens. May differ if the figure is renamed late in the report process. Generally, we write "US" instead of "United-States"
chapter_identifier	The associated chapter identifier inside the report, if any, associated with the figure.
title	The title of the Figure as set by TSU. In the case of Figures not from TSU, the title as it appears in the report.
caption	Any text appearing alongside the figure.
attributes	*deprecated.* Unknown use.
time_start time_end	The start and end of the temporal extent represented by this figure. We *prefer* to have these on the image rather than the figure.
lat_max lat_min lon_max lon_min	The northernmost/southernmost latitude & eastmost/westmost longitudes in the bounding box for this figure. We *prefer* to have these on the image rather than the figure.
usage_limits	Copyright restrictions describing how this figure may be used by the general public. Historically this has not been filled in properly. Needs to come from TSU, but should not be filled in from their usage limits, but rather whether we created it or not. For non-TSU figures, fill in accordingly.
submission_dt	Currently, we think this is the last update before submission. Followup needed. Final determination to be made after conversation with TSU.
create_dt	Currently, we think this is the initial creation in TSU's system. Follow-up needed. Final determination to be made after conversation with TSU.
source_citation	If this isn't an original creation, the citation to that.

	Follow-up: should this be captured in provenance?
ordinal	The internal chapter or report numbering (alphanumeric)
report_identifier	The associated report identifies the figure.
url	Direct URL to the Figure landing page in the associated report, if any.
_origination	Internal field to hold data needed by TSU for presentation of report websites. Eventually to be phased out via Activity & TSU Sync updates.

Provenance Conventions:

- **Cito:cites:**
  - Figures can cite their sources
- **cito:isCitedBy:**
  - If the Figure has been cited in a future USGCRP publication, it may be marked as isCitedBy
- **prov:wasGeneratedBy, prov:wasDerivedFrom, prov:wasInformedBy:**
  - These provenance relationships would be better placed on the Image associated with a Figure, rather than on the Figure itself.

Relationship Conventions:

- **Image:**
  - All Figures should have an Image.
  - A Figure should have an Image for each logical subpanel within the figure.
  - When a logical subdivision of the Figure is unavailable, a single backing Image is acceptable.
  - For when the TSU creates the Figures & Images, the Image subobjects are expected to have proper context and not just be cropped subsections of the Figure.
- **Report:**
  - Figures must belong to a Report or Indicator.
- **Chapter:**
  - Figures may belong to a Chapter, if the Report is broken into Chapters and the Figure is within them.
- **Contributors:**
  - All Figures should have a Point of Contact
  - Figures may have authors, editors, or other contributors depending on the associated Assessment.
- **regions:**
  - Not yet implemented.
- **Gcmd_keywords:**
  - On a case by case basis for the parent report, we may assign GCMD keywords to the Figure.

- **File:**
  - All figures should have a high resolution image file associated with it.
  - Preference for PNG for vector-like graphics and JPG for photo-like graphics.

⑨ Resource Type: Finding

Appropriate Use Case: In USGCRP Reports, every chapter has a "Key Message" - a short statement of fact with backing supporting evidence. Only used on USGCRP published reports, and only exists as a subobject.

Field Conventions:

Field	Description
identifier	Should follow the convention of: *Key-Finding-CH#-ORD* *Key-Message-CH#-ORD* The initial part of the identifier should match the naming of this object type for that report
chapter_identifier	The chapter, if any, contains this finding.
statement	The shorter, headline sentence that conveys the key information. Filled in via data provided from TSU. Data contained within is provided by the authors following guidance. HTML Formatting is supported. Callouts to References should be done via their UUIDs, and captured within <*tbib*> tags.
ordinal	The ordinal as specified in the actual report.
report_identifier	The report that contains this finding.
process	Filled in via data provided from TSU. Data contained within is provided by the authors following guidance. HTML Formatting is supported. Callouts to References should be done via their UUIDs, and captured within <*tbib*> tags.
evidence	Filled in via data provided from TSU. Data contained within is provided by the authors following guidance. HTML Formatting is supported. Callouts to References should be done via their UUIDs, and captured within <*tbib*> tags.
uncertainties	Filled in via data provided from TSU. Data contained within is provided by the authors following guidance. HTML Formatting is supported. Callouts to References should be done via their UUIDs, and captured within <*tbib*> tags.

confidence	Filled in via data provided from TSU. Data contained within is provided by the authors following guidance. HTML Formatting is supported. Callouts to References should be done via their UUIDs, and captured within <tbib> tags.
url	If there is a direct URL to the finding, add it here. Prefer not linking to the general parent page (aka the chapter).

Provenance Conventions:

- **cito:cites & cito:isCitedBy**
  - The Finding cites its references.
  - If future USGCRP products cite this Finding, it may have 'isCitedBy'.

Relationship Conventions:

- **Chapter:**
  - If it is within a report chapter, this must be filled in.
- **Report**
  - This must be filled in.

⑩ Resource Type: Image

Appropriate Use Case:

- Images should only exist underneath Figures. All Figures should have at least one Image.
- For reports done in conjunction with the TSU, the Images should be imported alongside their Figures, or filled in where needed following the pattern they use.
- For reports not done with TSU, we may create Images.
- Images for Figures with just one panel should be the full Figure (i.e. the same as the image file attached on the Figure record). Images for Figures with multiple panels should be reconstructed from their Figure to contain all relevant information (i.e. labels, legend, etc) rather than simply cropped. If no other option, cropped is better than nothing.

Field Conventions:

Field	Description
dentifier	Standard UUID identifier.
position	This is more of a nice-to-have. Not a focus for filling in.
title	Usually imported directly from TSU. Indicator Images do not have Titles. Images should incorporate the Title of their Figure in their Title, plus a relevant distinctive postfix.

description	Imported from TSU, if one exists. We do not attempt to fill this in otherwise.
attributes	Imported from TSU, if any exists. For Indicators, filled in if provided. Otherwise, not filled in.
time_start	Existing definition fine. Should match the activity that created it, if any. Blank is okay if irrelevant. If dates aren't exactly specified, they are assumed the earliest possible (i.e. 2017 is 2017-01-01, June 2017 is 2017-06-01). Same for time.
time_end	Existing definition fine. Should match the activity that created it, if any. Blank is okay if irrelevant. If dates aren't exactly specified, they are assumed the earliest possible (i.e. 2017 is 2017-01-01, June 2017 is 2017-06-01). Same for time.
lat_max	Existing definition fine. Should correlate with the activity that created it, if any. Blank is okay if irrelevant.
lat_min	Existing definition fine. Should correlate with the activity that created it, if any. Blank is okay if irrelevant.
lon_max	Existing definition fine. Should correlate with the activity that created it, if any. Blank is okay if irrelevant.
lon_min	Existing definition fine. Should correlate with the activity that created it, if any. Blank is okay if irrelevant.
usage_limits	Existing definition fine. Should come from the source of the Figure. Be certain before filling in; this should be the limits on usage for a public user, not the limits on USGCRP's use of the image.
submission_dt	Currently, we think this is the last update before submission. *Followup needed.*
create_dt	Currently, we think this is the initial creation in TSU's system. Followup needed.
url	Not used.

Provenance Conventions:

- **Cito:cites:**
  - Generally the Figure object will be the thing using cito:cites.
  - It is plausible to have a good use case for an image to use cito:cites.
- **prov:wasDerivedFrom:**
  - Image is preferred to be the child to wasDerivedFrom as opposed to the Figure.

- Many types of publication can be the parent type.

Relationship Conventions:

- **Figure**
  - Generally, Figures and Image tend to be one-to-one.
  - It is acceptable if two Figures use the same component images, they may use the same Image object.
- **Gcmd_keyword**
  - Not yet implemented, but probably better applied to Figures.
- **Region**
  - Not yet implemented.
- **File**
  - Should have a file with the type best suited to its image type. High quality images strongly preferred.

⑪ Resource Type: Instrument

Appropriate Use Case:

We do not actively create or curate instruments.

⑫ Resource Type: Instrument Instance

Appropriate Use Case:

We do not actively create or curate instrument instances.

⑬ Resource Type: Journal

Appropriate Use Case: Journal objects are created in the GCIS to improve the traceable accounts of articles cited in USGCRP reports. Therefore, every article in GCIS must have a corresponding journal where possible.

Field Conventions:

Field	Description
identifier	This is based on the title. When creating a journal, leave the identifier blank such that the system autocreates it without the articles.
title	The title according to crossref. Fallback to the title on the website or best available source.

publisher	Deprecated in favor of the Contributor relationship, role Publisher.
country	Optional. If filled in, use the country codes as they appear in our Countries table. Fill in if this improves clarity.
url	We strongly want this filled in. For defunct journals, point to the official archive website, if any. Can be left blank if no such url is available.
notes	Currently used to capture the provenance connection we can't capture semantically. No other good use. Do not use it.
print_issn	We want to have at least one of the ISSNs. If there is more than one print ISSN, use the first listed & functional ISSN.
online_issn	We want to have at least one of the ISSNs. If there is more than one online ISSN, use the first listed & functional ISSN.

Provenance Conventions: None.

Relationship Conventions:

- **Article:**
  - Journals should only be created when needed because of an Article being in GCIS.
- **Contributors:**
  - Every Journal should have a contributor Organization with role Publisher
  - If a Journal changes Publisher, and we have articles under both, we should update to have a Publisher and a 'Former Publisher'
- **Files:**
  - An exemplary image of an issue of the Journal, pulled from their website.
- **Gcmd_keywords:**
  - Not used on this object.
- **Regions:**
  - Not yet implemented.

⑭ Resource Type: Model

Appropriate Use Case:

We do not actively create or curate models.

⑮ Resource Type: Model-Run

Appropriate Use Case:

We do not actively create or curate model-runs.

⑯ Resource Type: Organization

Appropriate Use Case: Organization objects capture the affiliations of "people" in GCIS. Ideally, each "person" in GCIS should have an affiliation. Having affiliations in GCIS improves provenance tracing.

Field Conventions:

Field	Description
Identifier	Organization name separated by hyphens. To denote divisions or offices within a larger entity, begin with the larger entity. Spell out names of agencies and universities in their entirety.
name	The Organization's name. Spell out entire organization name, e.g. "National Oceanic and Atmospheric Administration". Use official name where at all possible, even where it may differ from the entry in a publication.
organization type	The type of organization. For example, Federal Organization, Research Organization, Commercial Organization, Academic Organization, and so on.
url	Direct link to the Organization's landing page.
country	The country within which the organization's primary HQ is located.
international	Flag indicating a multinational organization with HQs in multiple countries.

Provenance Conventions: None

Relationship Conventions:

- University of X at Y predecessor of University of X; (implies that former univ became latter one)
- University of X at Y branch of University System of Y; (implies former univ part of a univ system)
- University of X branch of University System of Y (implies newer univ part of that same system)
- Clearly this can be extended to offices, etc. Also as a corollary if in changing names the university changed systems, we'd clearly adjust the relationship accordingly. Wikipedia can be a huge help here.

Since we associate people and organizations through publications. We use what is mentioned in the publication even if the affiliation has since changed. For example, we keep "National Climatic Data Center" even though it is now the "National Centers for Environmental Information." Since it was NCDC at the time of the publication, we keep it as such and employ the methodology used above. As such, if a web page URL is the same as that in a previous publication but now includes the "National Centers for Environmental Information" banner, we use NCDC in-lieu of NCEI if that is what the author would have seen at the time it was cited.
One last thing - if an organization relationship changes, do not revise. Simply update with the new relationship, keeping the older one.

⑰ Resource Type: Person

Appropriate Use Case: This resource represents contributors to a particular product in GCIS (for example, an Article or Report). Each person in GCIS is represented by a unique number that serves as the identifier.

Field Conventions:

Field	Description
id	Leave as is, as this is an internal identifier used by GCIS.
first_name	Examples: Jeffrey; C. Ben; Richard A. Basically, we use the first initial and middle name, the first name, or the first name and middle initial. Note that some authors publish under variations of the same name depending on the publication. In that case, we use the following system: If an NCA3 authors, use what is used there, else: if Health Assessment author, use what is there. else: the most complete formal name used in a GCIS publication. Note that we only keep one name in GCIS (i.e., no aliases). For example, "Robert E. Wolfe" and "Robert Wolfe" cannot both exist in the database and be affiliated with the same article if they refer to the same person. This too can be changed.
middle_name	currently does not show on html, hence we use the middle name or middle initial in the "first_name" field (UPDATE upon completion of https://github.com/USGCRP/gcis/issues/298)
last_name	The person's last-name, including if it's hyphenated. Note that sometimes first and last names may be flipped in publications (common with many non-US authors)

orcid	The person's ORCiD (see http://orcid.org) If no information is available on the person's ORCiD page to definitively attribute something to that person, leave the field blank. This situation occurs very commonly when people do not leave identifying information on their ORCiD pages. Hence, it is possible that someone else with the same name created that ORCiD.Very often, people will just link their ORCiDs to their ResearcherID pages, which many times is enough to make a definite match with the article author. Enter only the sixteen digit value with hyphens, with nothing else. Also, refrain from adding other IDs like Scopus since the system will return an error if the value exceeds sixteen digits and doesn't begin with the standard "0000."
url	An "official" URL for that person. It could be a page on one's company website. It could provide any sort of information, and can even be an obituary. Could also be a wordpress account if that could be identified as an official page (i.e., created by the author). Could also be accounts on gulfbase.org. Note: many add their lab's URL in their ORCiDs. Normally, we go with the page on the lab's server about the person, not the main page for the lab, but either way works. Avoid linking to third-party pages, such as a UN or other entitty's writeup of that person in order to advertise his or her upcoming speech - this occurs quite often. It is advisable to use the most up to date website if possible. e.g. if someone has a website on one university's server from his/her graduate school days and one from a subsequent research appointment, use that from the research appointment. Think of it this way: if I wanted to advertise my official website next to my email address on a business card, what would I use beside social media websites, researchgate, academia.edu, and linkedin? Update if necessary but don't go out of one's way to locate all not working URLs.

Provenance Conventions: None

Relationship Conventions: None

⑱ Resource Type: Platform

Appropriate Use Case:

We do not actively create or curate platforms.

⑲ Resource Type: Project

Appropriate Use Case:

We do not actively create or curate projects.

⑳ Resource Type: Reference

Appropriate Use Case:

- References are created for USGCRP assessments.
- They are converted from EndNote references handled by external users (usually TSU).
- We will support other entitiy's reports in GCIS having references if they provide properly formatted EndNotes.
- References may or may not be evergreen, depending on the base type of the cited publication. For example, a published Article or Report is probably evergreen, but a website definitely isn't.

Field Conventions:

Field	Description
identifier	Must be a UUID. Generated by GCIS, but assigned by the entity controlling the EndNote file before import.
child_publication_id	After the reference's child publication has been found/created, this should link to its Publication ID.
attrs	A hash capture of the EndNote fields, as mapped by the TSU EndNote conventions, with subfields based on the Reference Type. attrs should not be edited after the initial release except to correct errors, as needed. They are meant to be correct as of the date of publication of the parent publication.

Provenance Conventions: None.

Relationship Conventions:

- **Publication:**
  - All references should have a parent publication, or they are considered Orphans without a purpose.
  - Reference child publications should be filled in as time allows with a goal of 100%. Many references may point to the same child publication, but only for non-evergreen publications or other exceptions as determined. Evergreen publications should only have one reference for which they are the child publication. Child publication types

of Article, Report, and Book can be done automatically. The rest are handled manually.

㉑ Resource Type: Region

Appropriate Use Case: The Regions resource type in GCIS represents the geographical regions described in USGCRP reports.

-

Field Conventions:

Field	Description
identifier	A simplified version of the label with dashes. In the future, we would like to change this to be a UUID to avoid too specific or broad naming.
label	Should match the most recent NCA report specified name of the region. If no region, should match the governmental name.
description	Ideally, taken from the text of the report. At a minimum, a list of the governmental boundaries contained.

Provenance Conventions:

- None

Relationship Conventions:

- **Publication**
  - Used to connect reports, chapters, figures, findings, and tables to region(s).

㉒ Resource Type: Report

Appropriate Use Case: Publications, both external publications (not published by USGCRP) and USGCRP publications, that have officially been released as reports, and that need to be uploaded to the GCIS database, are created as reports in the GCIS.

Field Conventions:

Field	Description

identifier	The title of the report with dashes in place of the spaces.
title	Title is generally imported directly from TSU's EndNote. It should match the external title.
url	Should always be provided if we don't have a DOI. Preferred to have regardless. In the case of neither a DOI nor a URL existing, may be left blank.
doi	The exact DOI, if one exists. Otherwise blank.
summary	A brief summary of the report.
publication year	The year of publication as indicated on the report's official landing page.
in library	Whether or not this report is available in the USGCRP Resources Library.
report type	We use the type as named by the distributor. Default is “report”.
frequency	Captures how often the report would be released, if the report is meant to be released periodically.
public	Indicates that the report is publicly readable.
topic	A brief free form comma-separated list of topics associated with the report.
contact note	A note about contacting someone about the report. Phrases [in brackets] in the note will become links to the contact email. Added when this is available.
contact email	A contact email address for the report. Added when this is available.

Additional Notes on Reports:

Identifier:

- If the distributor website has an official identifier, we use that.
  - US Forest Service Report "Pacific Northwest Research Station General Technical Report 255" with internal identifier PNW-GTR-855.
  - http://data.globalchange.gov/report/usfs-pnw-gtr-855
- If there is a widely accepted identifier, we use that.
  - USGCRP's Third National Climate Assessment is known as NCA3.
  - http://data.globalchange.gov/report/nca3.

- If we have to create it: Distributor abbreviation-hyphenated-title-version, if known, otherwise year.
    - This is an ACWI report called "A National Framework for Ground-Water Modeling in the U.S." which was published in 2013.
    - http://data.globalchange.gov/report/acwi-nationalframework-2013.

- If a report is an indicator, follow the title process with an added postfix of the year. Should match the published title of the Indicator, with an added prefix of **Indicator:** Only the most recent year Indicator should have a URL.

Contributors:

- Reports should have their contributors.
- When possible, contributors should have the person and the organization, but Organization only is acceptable.
- The "Sort" key should be used to make the contributor order match the report.

Provenance Conventions:

- cito:cites
- cito:isCitedBy
- prov:wasDerivedFrom
- prov:wasInformedBy

㉓ Resource Type: Scenario

Appropriate Use Case: The Scenario resource type currently captures common IPCC Climate scenarios used in climate science.

We do not actively create or curate scenarios.

㉔ Resource Type: Table

Appropriate Use Case:

- Tables are the high level representation of array data. They represent the Table objects in Reports.
- We only create them for USGCRP publications.
- We create them by hand.
- Tables may contain multiple arrays and in this case, a table caption would be used to briefly explain the relationship between the arrays.

Field Conventions:

Field	Description
identifier	Short identifier based on the title.
ordinal	The numeric position of this table within a chapter. Must start with a number, may contain numbers, letters, dashes, dots and underscores
title	The title of the table as it appears in the publication.
caption	The visible explanatory text that accompanies the array data in the publication. When there is a difference in the caption between the print and online tables, we match the online table.
url	Generally not used. Generally not used. If the Table has a proper, stand alone landing page, it should go here.

Provenance Conventions:

- **As child:**
  - prov:wasDerivedFrom
    - With very rare exceptions, not to be used on Tables, but rather their underlying Array.
  - Cito:cites
    - If a cell in the array or a statement in the caption includes a citation, the table should use this semantic connection to link to the reference.

- **As parent:**
  - cito:isCitedBy and prov:wasDerivedFrom
    - These are used when future USGCRP publications cite or derive things from the table.
    - We do not actively look for outside use of the table.

㉕ Resource Type: Web Page

Appropriate Use Case:

- Web pages are brought in via citations but Reports, Figures, Images, etc.
- Web tools are categorized as web pages.
- **Distinguishing Webpages from Datasets:**
  - After web page type references are imported, a QA should be done before production release to check if any would be better categorized as datasets.

Field Conventions:

Field	Description

identifier	fine as UUID. See default: UUID. GCIS will automatically generate one from a blank identifier field
url	the full url, not including any parameters. Parameters should be collected on either the activity or references as appropriate.
title	whatever the Reference names the URL, otherwise as stated on the web page itself, correct as of the creation of the object.
access_date - deprecated	the access date for this webpage. Should not be used, as we have a unique URL requirement. A access date should go on either the Reference object (e.g. if this is a referenced publication) or an Activity object (e.g. if this cited on a Figure)

Provenance Conventions:

- Webpages are valid to be cito:CitedBy, usually by Figures or Images.
- They do not use cito:Cites, as they are not USGCRP products.

Relationship Conventions:

- **Contributors:**
  - Web Pages often have a host, and may have authors, as appropriate.
- **File:**
  - Not used on Webpages.
- **Gcmd_keywords:**
  - Not to be used for the moment. To be determined.
- **Regions:**
  - Not to be used.