Complex Citation Implementation IG Statement of Work
12 May 2025

Name: **Complex Citation Implementation Interest Group**

Introduction

The current scholarly communication landscape does not easily support the ability for an author to cite a large number of existing data and other digital objects in their manuscript in a way that enables credit for each object as well as the provenance needed for transparency and traceability. Funders commonly require accurate identification of the datasets and other digital objects used to support research findings. In some disciplines this can mean hundreds to millions of digital object citations. The Research Data Alliance (RDA) Complex Citation Working Group completed the recommendations in early 2025[1] and this Interest Group is focused on the implementation activities across the broad number of workflow actors: researchers, infrastructure providers, publishers, indexers, DOI Registry Agencies, funders, institutions, and more. The Complex Citation use cases include all disciplines. All workflow actors have a role in implementation and securing a common approach to ensure a successful path forward such that all researchers have the tools needed to include all digital object citations in their manuscript.

"Solving a Complex Problem in a Simple Way"

Within this sphere the following new vocabulary is utilised:

- **Complex Citation Object (CCO)** - a new type of Digital Object Identifier (DOI) supported by the DOI Registration Agencies that includes, as its primary content, a list of Linked Digital Objects.

- **Linked Digital Object (LDO)**- the metadata necessary to identify the digital objects that support the research, or scholarly product. This includes data, software, physical samples, images, and videos. This does not include the scholarly literature normally included in the Reference section of a paper.

[1] Agarwal, D., Ayliffe, J. O., Buck, J. J. H., Damerow, J., Stall, S., Stockhause, M., Wyborn, L., & RDA Complex Citation Working Group Members. (2025). Complex Citation WG Recommendation (Version 4). Research Data Alliance. https://doi.org/10.15497/RDA00130

User scenario(s) or use case(s) to address

The Complex Citation Working Group Recommendation, published in April 2025, includes an extensive list of use cases captured for the recommendation. Additional details can be found in the recommendation.

	Name	Type/ Alignment	Domain	Short Description
1	AmeriFlux	Citing many datasets across repositories	Earth Sciences	• Datasets are stored in a repository • Users combine 100s of datasets • Citing many datasets across repositories • Giving credit to the data providers requires the citation of 100s of datasets stored in a repository.
2	British Oceanographic Data Centre (BODC)	Citing subsets of larger datasets	Earth Sciences	• Data and samples from many cruises and other activities • Datasets organized in data collections • User downloads contain subsets from several collections • Citing subsets of a larger dataset: • Giving credit to the data providers requires the citation of the subsets/elements of the larger datasets
3	Inter-governmental Panel on Climate Change (IPCC) FAIR Guidelines	Citing subsets of larger datasets across repositories Tracing figures and other outcomes back to their origins	Earth Sciences	• Enhance the transparency of the Sixth Assessment Report: • Many figures combine input data (subsets) from multiple repositories applying scripts to create figure datasets • Tracing figures and other outcomes back to their origins: • Giving credit to the data and software providers requires the citation of 100s of datasets and data subsets across multiple repositories • Figure reproducibility requires information on input data, software, and provenance information

	Name	Type/ Alignment	Domain	Short Description
4	CLARIN Virtual Collections	Citing collections of distributed resources	Humanities, but in principle domain agnostic	• Researchers need to be able to share and cite resources that are distributed over different systems as one dataset. • Bundling references to a set of such resources with collection metadata and issuing a PID creates a Virtual Collection (VC) that can be used for sharing and citing. • Automatic processing of the VC the content is something that has been recently considered and worked on/
5	Monthly Weather Review paper using many datasets	Citing a large number of datasets in a publication	Earth Sciences	• Datasets from CMIP3, CMIP5 and CMIP6 were used in a publication • Proper citations were retrieved for all of these datasets and then inserted into the reference list • During the editing process, the list of cited data went into the supplementary PDF, which resulted in the citations of datasets or related papers not being "seen" by automated citations trackers • Although provided, DOIs for used datasets were not mentioned in the reference list; maybe due to the citation identifier @misc in the bibtex file and the way this entry type was handled by the bibtex library used for AMS journals?
6	Soil profiles data publication and reuse	multiple authors, citation, dynamic citation	Agriculture	• Single data sets are published, data collections were created and published by reusers, authors rights should be respected
7	Agricultural Data Collections	citing data collections	Agriculture	• Complex data are published as a hierarchical arrangement of tables under one DOI. Individual tables can be cited analogously to a book chapter.

	Name	Type/ Alignment	Domain	Short Description
8	NASA SciX, previously named Astrophysics Data System (ADS)	Linking data and software to the scholarly literature and capturing their citations	Astrophysics, Heliophysics, Planetary Science, Earth Science	• The SciX has implemented a workflow for capturing software citations. • Since both data and software citations are crucial for the transparency of research results and for the transmission of credit, the ADS will implement indexing of high-level data products, in particular those published by NASA Archives, and track their citations.
9	Library of spectral samples	citing portions of a larger corpus	Earth Science, but applies more broadly	• EcoSIS (https://ecosis.org) is a collection of over 200,000 spectra from laboratory and field conditions, which is useful in constructing models, such as for hyperspectral missions, such as AVIRIS, EMIT, SBG, PRISM. • These spectra have been contributed by many different researchers. We want to address reproducibility and provide credit so that someone using EcoSIS as part of the process of building models to predict ecological quantities can clearly specify which spectra from EcoSIS were used in that work.
10	GBIF: Global Biodiversity Information Facility	Citations for downloads drawing records from 10s, 100s or 1000s of individual datasets, with credit attributed to source datasets and data publishers	Life sciences with links to biodiversity	• GBIF has issued a DOI for every user download since mid-2015. • Users are repeatedly reminded and encouraged to use this DOI when citing the data. • Properly cited uses allocate credit to all contributing datasets and publishers. • Adoption of DOI-based citations as of this writing (21/7/23) now stands at 60% (avg 20 / 32 published peer-reviewed uses each week). Likely to integrate ROR to attribute uses to funders in coming months.

	Name	Type/ Alignment	Domain	Short Description
11	EarthScope: GAGE GNSS data facility	Composite and Aggregate data citations - both are citations of two or more data citations	Earth Sciences	• Aggregated GPS/GNSS Datasets - These will often be an associated group of campaign datasets or a network of stations. • A campaign example is the Mammoth/Mojave 1994 campaign - https://doi.org/10.7283/T57H1GGM, which consists of three individual primary datasets: Mammoth, Mojave, and Combined Sites). • For permanent/continuous stations, networks or sub-networks of stations may be assigned an aggregated DOI. An example is Plutons GPS Network - https://doi.org/10.7283/T5V98697. • The collection of stations aggregated does not have to be a network; in this case, the purpose of the aggregated dataset is for collecting a potentially large number of station DOIs for citing in a journal article (ie, in order to avoid citation lists containing tens or hundreds of dataset references). • Composite GPS/GNSS Datasets - A composite dataset DOI is one that is comprised of two or more subset DOIs that together make up what would normally be considered to be a single dataset. The most common example is a permanent (continuous) GPS/GNSS station where the principal investigator (author) changed at a particular point in time. • The existing network (Nucleus) stations that were adopted by GAGE/UNAVCO as part of PBO are examples. The entire dataset is one DOI and is comprised of a separate DOI for each time period with a different author or set of authors. An example is the composite DOI for the station NOMT - https://doi.org/10.7283/T5B27SN9

	Name	Type/ Alignment	Domain	Short Description
12	KBase (US Dept of Energy, Systems Biology)	citation of a research workflow with many types of raw data from several sources, data products, private/public data, physical samples, instrument output, and software	Life sciences w/ links to physical sciences	• GROW (Genome Resolved Open Watershed) - community data collection and processing effort that includes a range of data types: physical samples with IGSN, that have been converted into Metagenome, Metatranscriptome, and Metabolomic data by at least 2 US Dept of Energy user facilities (JGI - MetaG, MetaT; EMSL - MetaB). • KBase is providing a DOI for the "parent landing page", in this case the Samples Narrative, but the DOI would ideally cite all PIDs related to the samples (IGSN), data (DOI, accession numbers), funders (Funder ID/ROR), funded proposals (DOIs), etc.
13	SeaDataNet/Emodnet	Citation of subsets from up to millions of files	Marine (but could be generic)	• Marine observation data are organised in files, sometimes millions of files. In aggregators like SeaDataNet, EMODnet and others these files come together in a cloud. • The newly developed BEACON software allows users to extract subsets of these files on the fly to generate one single file - NetCDF or other - as output. It contains the reference ID to the original files in the metadata for every snippet of the original files. • How to best provide access to the citation for users of the subsetted data?
14	Physical Samples (IGSN)	Citation of individual samples in aggregated collections	Earth Science, Environmental Sciences, Soils	• May include individual physical samples from nature and field-based acquisitions, may also include synthetic or manufactured samples. • In the laboratory, a source sample can be split into multiple subsamples, or observations on multiple sites may be made in microanalytical in situ techniques.
15	Research Resource Identifier (RRID)	Physical Sample	Biological	• The RRID represents antibodies, model organisms and software projects) in the biomedical literature to

	Name	Type/ Alignment	Domain	Short Description
		Persistent Identifier		improve transparency of research methods.
16	Laboratory Analyses	Citation of Individual analyses	Earth Science, Environmental Sciences, Soils	• Laboratory analyses are often carried out in batches in the laboratory or individual analyses sometimes from multiple laboratories are aggregated into a collection. • It is important that the individual analysis, as well as the sample they were derived from can be credited.
17	Instrument PID Research Coordination Network (RCN) - Matt Mayernik	Citation of instrument used in the research	All domains	• This Research Coordination Network is developing the recommendations around assigning PIDs to instruments. In Sept 2025 they will be holding their third (and final) workshop with the goal of finalizing recommendations, and developing implementation guidance.
18	GNSS observation data - GNSS-DCAT-AP PIDs for GNSS observation files (RINEX)	Citing subsets of larger datasets	Earth Science	• Cite subsets of GNSS observation files (RINEX) from multiple GNSS stations. The subsets can refer to a specific time range. • Create a CCO where each Linked Data Object is built from the metadata of a GNSS observation file (RINEX), identified via GNSS-DCAT-AP PID, to enable citation, enhance reproducibility and give credit to data providers. • GNSS-DCAT-AP is a standardized metadata schema that has been created to address the specific characteristics of a GNSS observation file (RINEX).
19	Complex Software citations - Peter Löwe	Citing larger, complex software workflows Building provenance graphs	All domains	• CCOs that are specific to software alone, broadening out and developing the software citation sphere • Looking at transition from software provenance graphs for workflow reproducibility to final citation and credit assignment • Creating CCOs that cover inclusion of software workflows alongside other types of LDOs

Objectives

The Complex Citation Implementation IG (CCI IG) continues the work of the RDA Complex Citation WG by focusing on the coordination of the implementation and adoption of the published recommendations. The CCI IG is also responsible for maintaining and further developing these recommendations. Based on the experience of a few pilot studies, the initial tasks for the CCI IG have been identified:

- Establish CC Technical Task Group
    - Finalize the specifications of the CCO
        - Define and establish, with DataCite and the users of the CCO, a new resourceType.
        - Standardize the Complex Citation Object's (CCO) metadata.
    - Finalize the specifics of the LDOs
        - For the enabling of the right interpretation of the related LDOs, a controlled vocabulary for the LDO types needs to be defined to ensure the LDOs are correctly interpreted and credit is assigned as indicated.
    - Establish pilots for creating a CCO and sustainable hosting.
        - Identify pilots
        - Reconnect to organizations with implementations similar to Complex Citation: RO-Crate, CLARIN Virtual Collection Registry, Pangaea, BioStudies / ELIXIR, Open Research Knowledge Graph (ORKG) Reborn Statements (paper)
    - Develop technical tools/guides for creating CCOs
        - As packages to capture workflow components
        - Construction of CCOs as part of the publication process
- Establish CC Implementation Task Group
    - Implementation in a scholarly publication workflow:
        - Develop guidance for creating a CCO and citing in the scholarly publication.
        - Develop approach for scholarly publication to support an intext citation of an LDO.
        - Provide guidance on citation elements. Guidelines for the use of CCOs in publications have to be defined and included in author guidelines and the publication process, including inline references, flags in the reference list.
    - Implementation for scholarly infrastructure to enable automated attribution and credit. This includes workflow actors such as publishers (e.g., STM, CSE, SSP, Charleston, ALPSP, and other publishing associations), the editorial workflow managers (e.g., Aries, ScholarOne, Editorial Manager), the peer-review publication platform providers (e.g., Atypon, High Wire, SilverChair), DOI Registration Agencies (e.g., Crossref, DataCite), Indexers (e.g., SciX, Web of Science, Dimensions), open science metrics (e.g., OSMI), publishing standards

to include JATS (e.g., NISO), style guide providers (e.g., APA, Chicago), reference managers (e.g., Zotero), Librarian community.
    - Develop implementation materials specific to the publishing workflow actors noted.
    - Determine which scholarly infrastructure workflow actors are essential for a CCO to be functional to prioritize the work (e.g., DataCite). The other workflow actors help with sustainability, integration, and broad adoption (e.g., JATS).
    - Collaborate with the existing publication infrastructure to leverage existing processes and mechanisms.
    - Promote adoption with all workflow actors.
  - Determine the exact workflow actors that will "unpack" the CCO to ensure automated attribution and credit is enabled. This includes the right credit assignment in citation impact matrices in downstream services.
- Establish CC Adoption Task Group
  - Develop a communication plan and training materials.
  - Engage with the research community to gather feedback and support the change of data citation culture use of CCOs with drive.
  - Organize workshop events.

Participation

The Complex Citation Implementation Interest Group has identified the workflow actors needed for the implementation and adoption. Participants from each of the workflow actors participated in the development of the recommendation and have committed to continuing with the Interest Group to support development of Implementation Guidelines and adoption.

Workflow Actor	**Process**
Complex Citation Object Generator and Host (i.e. Repository, Institution)	Creation, registration, and ongoing management of Complex Citation, landing page, preservation, version control, reference support
Author team for journal article of report / Researcher	Creates the Complex Citation Object and includes it in the paper, or other scholarly document

Scholarly Publishing Infrastructure: publishers, the editorial workflow managers, peer-review publication platform providers, DOI Registration Agencies, Indexers, open science metrics, publishing standards to include JATS, style guide providers, reference managers, Librarian community.	Where the Complex Citation is in the Reference Section, described in the Availability Section, and supports in-text citation, included in the Crossref files, shared with indexers, included in citation guidance, implemented by references, communicated by librarians, and measured by research assessment entities.
Scholarly Indexers and any other workflow actors that unpack a CCO and enable direct reference to LDOs.	Ensures the Linked Digital Objects included in the Complex Citation are treated as directly referenced citations.
Consumers incl. Harvesters	Recognizes and integrates Complex Citation Objects into their schema. Supports metrics around data and data use.
DataCite	The DOI Registration Agency that supports the new DOI data type.

UN Sustainable Development Goals (SDGs)

The CCI IG aims at enabling an effective traceability of research outputs and an easy and efficient credit assignment according to best practices in science and the UNESCO Open Science recommendations. Thus, the CCI IG contributes to all SDGs by making research results generally more transparent.

Outcomes

The CCI IG will prioritize the tasks identified in the objectives section and set up the identified task groups of IG members to accomplish the individual tasks. The IG will take these outcomes up for broader discussion and communication to the adopters.

The CCI IG will focus on establishing the Complex Citation as the approach for researchers to cite a large number of digital objects.

Mechanism

The CCI IG meets twice a month in two different time zones. Task groups will organize their meetings in between and report back in the CCI IG meetings.

Timeline

Month 3: Establish task groups.
CC Technical Task Group: Required standards for CCO and LDO types have been drafted and are open for review.
CC Implementation Task Group: Determine workflow actors for CCO unpacking. The DataCite metadata extension has been submitted to DataCite's Metadata WG.

Month 6:
CC Technical task Group: First version of standards for CCO and LDO types is published. Early adoption of a few Complex Citation Object Generators and Hosts has started.
CC Implementation Task Group: The state of work is presented and discussed at the RDA plenary with the broader workflow actors.
RDA Plenary: Share the current progress and encourage additional engagement.

Month 9:
CC Implementation Task Group: Based on early adopters implementation, work with the identified workflow actors to pilot unpacking the CCO and evaluate the impact on workflow.
CC Adoption Task Group: Engage more implementation.

Month 12:
RDA Plenary: Share the current pilot implementation including all workflow actors. After review, additional adopters implement the CCO, training and capacity building initiatives are started.
CC Implementation Task Group and CC Adoption Task Group: Continue encouraging and supporting implementation.

Potential Group Members

The objectives for the CCI IG are significant and will require support of a large number of workflow actors across three task groups. We have selected 8 representing North America, Europe, and Australia. Our membership includes worldwide participation. This Interest Group will be supported by the Earth, Space, and Environmental Data CoP that represents over 10 countries.

FIRST NAME	LAST NAME	EMAIL	TITLE
Martina	Stockhause	stockhause@dkrz.de	Co-Chair
Lesley	Wyborn	lesley.wyborn@anu.edu.au	Co-Chair
Joan	Damerow	JoanDamerow@lbl.gov	Co-Chair

FIRST NAME	LAST NAME	EMAIL	TITLE
Justin	Buck	justin.buck@noc.ac.uk	Co-Chair
James	Ayliffe	jamayl@noc.ac.uk	Co-Chair
Graham	Parton	graham.parton@stfc.ac.uk	Co-Chair
Shelley	Stall	sstall@agu.org	Co-Chair
Deb	Agarwal	daagarwal@lbl.gov	Co-Chair
The Complex Citation WG Members			Members

Copyright Licence