Complex Citation Working Group

Final Recommendations

DOI: 10.15497/RDA00130

Authors: Deb Agarwal, James O. Ayliffe, Justin J. H. Buck, Joan Damerow, Graham Parton, Shelley Stall, Martina Stockhause, Lesley Wyborn

Published: 30 April 2025

Version: 4

Abstract: 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[1] 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 is composed of multinational, multi-discipline members from across the broad spectrum of workflow actors composing scholarly communications. RDA provides this unique platform to allow for challenging problems, such as this, to be addressed.

The working group members reviewed the current options possible to address the challenge and concluded that a new method was necessary. This Complex Citation recommendation includes the detailed problem statement, recommendations, key decisions made, roles and responsibilities, and details needed for all disciplines to consider this approach for their needs. All workflow actors have a role in implementation and securing a common approach to ensure a successful path forward such that data and other digital objects are cited across the research community, realizing the original ambitions and goals of citation for research transparency and credit for creators.

Impact: When the number of data and other digital objects to be cited is larger than what a journal will accept in the Reference Section, typically 40 or more, the citations are commonly pushed to the supplementary information, where they are not indexed, nor machine-actionable. This is contrary to the Joint Declaration on Data Citation that nearly all publishers have endorsed, requiring data to be cited in the Reference Section of the paper allowing for machine actionable links that go to the research findings. Scholarly publishers do not include content in the supplementary information as part of the scientific record. Therefore, the needed traceability and transparency are not included in the scientific record. This means:

1. Creators of these digital objects do not get attribution and credit for their contribution to the scholarly literature
2. Funders cannot measure use, impact and derived value
3. Machine-actionable transparency is not possible.
4. The supplement of the paper has a high probability of not being maintained by the publisher.

The existing tools and mechanisms to avoid the issue of having "too many" citations in papers have limitations in their use, and none allow for both automatic attribution and traceability.

The working group proposes a new approach to address the challenges in Complex Citations from the perspective of both transparent research and enabling credit for the many roles in the creation of research data, physical samples, and other research digital objects.

The findings of the Complex Citations Working Group have produced key requirements (R1 - R10) for Complex Citation Objects (CCOs) to achieve our goals. In summary:

1. CCOs capture enough detail to ensure proper credit, traceability, and transparency of cited materials (R1), supporting machine-actionable attribution for each referenced object (R2).

2. CCOs do not accrue credit themselves but simply list data and digital identifiers that require citation tracking (R3).

3. CCOs are stable, identifiable, versioned, resolvable, and persistent (R4, R5).

4. CCOs use standardized structures, limited to two PID graph levels, with a strong preference to utilize persistent identifiers (R6, R6.1, R7).

5. CCOs remain open, accessible, and flexible for various use cases, with an open license, and sufficient metadata (R8-R10).

The full recommendations are presented in subsequent sections of this document with the underpinning background material and use cases used in the development of requirements. Additionally, this document collates the roles and responsibilities of the Complex Citation Workflow Actors necessary for the CCOs to be used in practice.

The planned next step for this RDA output is to create a new working group to test, iterate, implement and promote adoption for Complex Citations into global working practice. If successful, this will lead to a fundamental improvement in the way data and other digital objects are cited across the research community, realizing the original ambitions and goals of citation for research transparency and credit for creators.

Contribution to United Nations SDGs: The Complex Citation is intended to support researchers that require a large number of digital objects for their work. Commonly, complex research such as what is required for the SDGs, needs many datasets and digital objects. The Complex Citation realizes the original ambitions and goals of citation for research transparency and credit for creators and contributors working on the SDGs.

Language: English

RDA webpage: https://www.rd-alliance.org/groups/data-granularity-wg/activity/

Citation and Download: 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

Complex Citation Working Group Recommendation 30 April 2025

Authors (alphabetical order)

Deb Agarwal, Lawrence Berkeley National Laboratory, 0000-0001-5045-2396
James O. Ayliffe, National Oceanography Centre, 0000-0002-7363-5029
Justin, J. H. Buck, National Oceanography Centre, 0000-0002-3587-4889
Joan Damerow, Lawrence Berkeley National Laboratory, 0000-0003-2601-5043
Graham Parton, Science and Technology Facilities Council, 0000-0003-4157-0352
Shelley Stall, American Geophysical Union, 0000-0003-2926-8353
Martina Stockhause, German Climate Computing Center, 0000-0001-6636-4972
Lesley Wyborn, Australian National University, 0000-0001-5976-4943

Contributors:

Complex Citation Working Group Members

November 2024 Draft Recommendation Reviewers:

Jonathan Blythe, Bureau of Ocean Energy Management, 0000-0003-2931-7100
Rorie Edmunds, DataCite
Paula Martinez, Australian Research Data Commons, 0000-0002-8990-1985
Anna Miglio, Royal Observatory of Belgium, 0000-0001-7316-4952
Helle Pedersen, Université Grenoble Alpes, 0000-0003-2936-8047
Julien Colomb, Technische Universität Berlin, 0000-0002-3127-5520

DOI: https://doi.org/10.15497/RDA00130

Complex Citation Recommendation

DOI: https://doi.org/10.15497/RDA00130

Executive Overview

The key challenges of the 21st century are complex, convergent, and transdisciplinary. Research data must be well-documented, complying with the FAIR Principles, and supported by infrastructure that integrates the mechanisms for transparency, traceability, and attribution. These same key challenges require data that are large, multidiscipline, and multinational, representing hundreds, if not thousands of people responsible for physical samples, instruments, models, code, configuration, quality checks, and the curation.

By making critical research data usable and transparent, we, in turn, must give attribution in our publications and scientific reports, connecting these contributions to the research findings and providing the detailed understanding of the data that can allow for transparency and critical consideration, specifically for AI readiness and knowledge graph usage. We must enable the creators, and other contributors for data, digital objects, hardware, and physical samples to receive credit[1] and better track the continued and growing value of a digital asset born from its original funding.

When the number of data and other digital objects[2] to be cited is larger than what a journal will accept in the Reference Section, typically 40 or more, the citations are commonly pushed to the supplementary information, where they are not indexed, nor machine-actionable. This is contrary to the Joint Declaration on Data Citation[3] that nearly all publishers have endorsed, requiring data to be cited in the Reference Section of the paper allowing for machine actionable links that go to the research findings. Scholarly publishers do not include content in the supplementary information as part of the scientific record. Therefore, the needed traceability and transparency are not included in the scientific record. This means:

1. Creators of these digital objects do not get attribution and credit for their contribution to the scholarly literature
2. Funders cannot measure use, impact and derived value
3. Machine-actionable transparency is not possible.
4. The supplement of the paper has a high probability of not being maintained by the publisher.

The existing tools and mechanisms to avoid the issue of having “too many” citations in papers have limitations in their use, and none allow for both automatic attribution and traceability.

The work to develop Complex Citation is motivated by key open research and open data policy drivers at national and international levels. These policies include the European Open Science

[1] Parsons, M. A., D. S. Katz, M. Langseth, H. Ramapriyan, and S. Ramdeen (2022), Credit where credit is due, Eos, 103, https://doi.org/10.1029/2022EO220239

[2] Digital objects include data, software, images, video, hardware, and physical sample information.

[3] Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA: FORCE11; 2014 https://doi.org/10.25490/a97f-egyk

Policy[4], the US Office of Science and Technology (OSTP) Policy Holdren[5] and Nelson[6] memos, the Canadian Tri-agency Research Data Management Policy[7], the UKRI open science policy[8] , the OECD Recommendation of the Council concerning Access to Research Data from Public Funding[9],and the UNESCO Recommendation on Open Science[10]. When successful, the outcomes of this working group through enabling unambiguous complex data citations and enabling credit to creators of different types of research objects, will contribute to the success of global data policies, including the 2023 Intergovernmental Oceanographic Commission (IOC) data policy[11], the UN Decade for Sustainable Ocean Development Challenge 8 on a Digital Representation of the Ocean[12], and the recommendations of the Intergovernmental Panel on Climate Change (IPCC)[13].

This RDA recommendation on Complex Citations brings together representatives from across the data citation ecosystem including data repositories, physical sample repositories, hardware and instrument manufacturers, core facilities[14,15], PID registries, citation indexing authorities, academic journals, and high impact user communities, including the US National Institute for Health (NIH) and the IPCC. The working group proposes a new approach to address the challenges in Complex Citations from the perspective of both transparent research and enabling

[4] European open science policy https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en

[5] Holdren Memo (2013): Increasing Access to the Results of Federally Funded Scientific Research, https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

[6] Nelson Memo (2022): Ensuring Free, Immediate, and Equitable Access to Federally Funded Research, https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-access-Memo.pdf

[7] Tri-agency Research Data Management Policy https://science.gc.ca/site/science/en/interagency-research-funding/policies-and-guidelines/research-data-management/tri-agency-research-data-management-policy

[8] UKRI open science policy https://www.ukri.org/what-we-do/supporting-healthy-research-and-innovation-culture/open-research/#:~:text=Open%20access,-UKRI%20published%20its&text=The%20policy%20aims%20to%20make,your%20research%20publications%20open%20access

[9] OECD Recommendation of the Council concerning Access to Research Data from Public Funding (OECD/LEGAL/0347) https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0347

[10] (2021). UNESCO Recommendation on Open Science. UNESCO. https://doi.org/10.54677/mnmh8546

[11] IOC data policy 2023 - has a line on attribution https://iode.org/resources/ioc-data-policy-and-terms-of-use-2023/

[12] UN Decade - digital representation of the ocean challenge https://unesdoc.unesco.org/ark:/48223/pf0000390123?posInSet=8&queryId=d61fb5dd-1226-4649-a508-7dd636d457fe

[13] Intergovernmental Panel on Climate Change. (2023). TG-Data Recommendations for AR7 (1.0). Zenodo. https://doi.org/10.5281/zenodo.10059282

[14] The authors are providing a definition of “Core Facility” to not be confused with geologic core samples. Definition: Core facilities are centralized shared research resources that provide access to instruments, technologies, services, and in many cases expert consultation and training to researchers. Accessed: https://research.mit.edu/research-resources/core-facilities-and-service-centers

[15] Jürgens, A., Tedeschi, G., D’Errico, G., Kilian, K., Zawadzki, K., Daniel, O., … Helm-Petersen, N. (2024). Navigating the frontier: research infrastructures, core facilities and a new paradigm at European Universities. *Cogent Education*, *11*(1). https://doi.org/10.1080/2331186X.2024.2365613

credit for the many roles in the creation of research data, physical samples, and other research digital objects.

The findings of the Complex Citations Working Group have produced key requirements (R1 - R10) for Complex Citation Objects (CCOs) to achieve our goals. In summary:

1. CCOs capture enough detail to ensure proper credit, traceability, and transparency of cited materials (R1), supporting machine-actionable attribution for each referenced object (R2).
2. CCOs do not accrue credit themselves but simply list data and digital identifiers that require citation tracking (R3).
3. CCOs are stable, identifiable, versioned, resolvable, and persistent (R4, R5).
4. CCOs use standardized structures, limited to two PID graph levels, with a strong preference to utilize persistent identifiers (R6, R6.1, R7).
5. CCOs remain open, accessible, and flexible for various use cases, with an open license, and sufficient metadata (R8-R10).

The full recommendations are presented in subsequent sections of this document with the underpinning background material and use cases used in the development of requirements. Additionally, this document collates the roles and responsibilities of the Complex Citation Workflow Actors necessary for the CCOs to be used in practice.

The planned next step for this RDA output is to create a new working group to test, iterate, implement and promote adoption for Complex Citations into global working practice. If successful, this will lead to a fundamental improvement in the way data and other digital objects are cited across the research community, realizing the original ambitions and goals of citation for research transparency and credit for creators.

Overview (Abstract)

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[16] 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 is composed of multinational, multi-discipline members from across the broad spectrum of workflow actors composing scholarly communications. RDA provides this unique platform to allow for challenging problems, such as this, to be addressed.

The working group members reviewed the current options possible to address the challenge and concluded that a new method was necessary. This Complex Citation recommendation includes the detailed problem statement, recommendations, key decisions made, roles and responsibilities, and details needed for all disciplines to consider this approach for their needs. 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.

New Vocabulary:

- **Complex Citation Object (CCO)** - a new type of Digital Object Identifier (DOI)[17] supported by the DOI Registration Agencies that includes, as its primary content, a list of Linked Digital Objects.

- **Linked Digital Object**[18] - 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.

[16] Digital objects include data, software, images, video, and physical sample information.
[17] The authors of this document note that based on the current state of technology, and the level of adoption of PIDs, that the Digital Object Identifier (DOI) is the only viable PID type presently.
[18] The authors of this document note the possible confusion with an existing term “Linked Data Object”. Only the use within Complex Citations is being considered.

Introduction of Complex Citation Problem

Academic publishers require that data underpinning the research within a manuscript need to be cited in an efficient way, ideally within a limited number of citations, ideally a single citation. Furthermore, the citations need to enable reproducibility of results derived from the underpinning data while also enabling credit for creators and contributors of those data.

The complexity in creating such a citation is not new. The expedition of HMS Challenger[19] during the 1870's generated scientific results spanning a 50-volume, 29,500-page report that took 23 years to compile along with a curated collection of physical samples. Such results are readily citable by volume and page for individual results, but how does one readily aggregate a subset of results into a single citation?

The digital age has enabled research to generate datasets with millions of granules (such as collections of images, files, samples, etc) and with the advent of Open Science, data granules are frequently aggregated from across many datasets or observational campaigns. Thus, the challenge in generating the single citation for publications and/or underpinning granules has continued to grow.

Over the last decade new community practices have emerged including FAIR[20] that specifies how research data should be shared with to the community, TRUST[21] underpinning trustworthy research data repositories and accreditation of repositories, CARE[22] specifying how data should be served for, or about, indigenous communities that enables ethical collaboration, community governance, and exploration of research results by indigenous communities, and the new United Nations Convention on the Law of the Sea (UNCLOS)[23] resolution that requires cite-ability of physical samples taken in international waters. These developments have both set frameworks to help with Complex Citation challenges and generated additional requirements to be addressed. In parallel to these developments, the move towards Open Science has merely increased the urgency in solving the Complex Citation challenge. New applications for data including predictive computational models, citizen science, AI and ML applications, and virtual research environments further make the case for an efficient way for researchers to accurately cite any digital object underpinning results for both reproducible research and to enable credit for creators and contributors of those digital objects especially for open data.

Once a digital object becomes citable, citations need to be discoverable by data creators; the importance of enabling credit to originators and funders should not be underestimated. An

[19] HMS Challenger, HMS is Her Majesty's Ship
[20] Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). https://doi.org/10.1038/sdata.2016.18
[21] Lin, D., Crabtree, J., Dillo, I. *et al.* The TRUST Principles for digital repositories. *Sci Data* **7**, 144 (2020). https://doi.org/10.1038/s41597-020-0486-7
[22] Stephanie Russo Carroll et al.(2020). The CARE Principles for Indigenous Data Governance. DSJ. https://doi.org/10.5334/dsj-2020-043
[23] United Nations General Assembly. (2024). Resolution 78/272 [A/RES/78/272]. https://documents.un.org/doc/undoc/gen/n24/117/55/pdf/n2411755.pdf

example of a community that needs to readily determine data usage metrics is the Argo community[24]. Argo is a global network of autonomous ocean observing platforms dependent on contributions by 20+ participating nations for the last 25 years. Argo data are open at the time of collection to meet law-of-the-sea obligations and to facilitate research collaboration. At the time of writing, Argo data have been cited 6447 times with the process of determining citations being manual and time consuming.

Thus, our vision for a Complex Citation is:

To develop an approach for aggregated citation in scholarly literature, enabling transparent and reproducible research while giving proper attribution to creators and contributors. This approach encompasses various digital objects, including data, software, physical samples, images, audio, video, and other supporting digital resources.

As described, this challenge is not new with significant related developments already in existence. The next section will move on to look at these and their limitations in achieving our vision for complex data and physical sample citations.

Discussion - Why a New Approach is Needed

In 2021, the initial group reviewed and considered the existing approaches. These are listed below in Table 1. The outcome of this analysis was that a new or updated approach is needed, which:

Enables attribution for creators/contributors of the linked digital objects: This implies that the preservation repository includes metadata for each creator and contributor, preferably a persistent identifier, to make attribution possible.

Supports provenance: This provides the relationship of the digital objects to both the scholarly item and other related digital objects. As defined by W3C, Provenance is information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness.[25]

Table 1 Overview of existing approaches

Existing approach	**Pros**	**Cons**	**Why the approach is <u>not suitable</u> for a Complex Citation Object**
Aggregate data supporting research	Easy access to data.	This moves the data from its preservation	**Enables Attribution:** Attribution to data

[24] https://argo.ucsd.edu/data/acknowledging-argo/#:~:text=To%20acknowledge%20Argo%2C%20please%20use,ocean%2Dops.org
[25] https://www.w3.org/TR/prov-overview/

Existing approach	Pros	Cons	Why the approach is <u>not suitable</u> for a Complex Citation Object
into one package assigned with one persistent identifier.		platform to an aggregated platform. Difficulty in scaling to large and many datasets. Does not support more than one type of usage license. Not commonly done for all digital object types.	creators is not possible as the original data are not being cited. **Supports Provenance**: Provenance metadata in addition to updates from data managers are disconnected from this copy of data.
Supplementary information of a manuscript.	Easy to do.	The Supplementary Information is not indexed. Journals have size limitations. Supplementary information is not included in the definition of the scientific record, and commonly not preserved over time. Journals that are signatories of the Joint Declaration on Data Citation and TOP Guidelines do not allow digital objects to be placed in the supplementary information.	**Enables Attribution:** The supplement is not considered as part of the scholarly record. Attribution of the digital object is not possible. **Supports Provenance**: Provenance is not supported. The supplementary information commonly does not include provenance information or the infrastructure necessary to make it accessible.
DataCite DOI – Collection Type - to include those that remain in their original preservation	Already in place, easy to generate within specific disciplines.	Automated attribution is not enabled for the creators of the digital objects listed in the	**Enables Attribution:** Attribution to data creators is not practiced at this time. There is additional credit value in

Existing approach	Pros	Cons	Why the approach is <u>not suitable</u> for a Complex Citation Object
hosting platform as well as those where a copy is moved to a collection site.		collection.	the collection itself which is not intended for the Complex Citation. **Supports Provenance**: Provenance metadata in addition to updates from data managers are disconnected when the collection copies data from its original source.
Data Paper publication as the primary citation for the data.	Established for scientific publications and credit assignment. Provides a deep understanding of how the dataset is created. A cited Data Paper is included in the algorithm for the impact and productivity metrics for a researcher.	The Data Paper is commonly not versioned and disconnected from datasets that are updated over time. The determination of which authors to include, commonly severely limits the full list of creators and is not equitable.	**Enables Attribution:** Attribution assignment is limited to the authors of the Data Paper. Data Paper authors may or may not be the same as the creators of the datasets. **Supports Provenance**: The Data Paper does not include metadata for dataset provenance.

Short History of the Complex Citation Community

The community building and effort to develop the recommendations presented in this document spans a sustained effort over four years (2021 - 2024) and will be briefly summarized to give context to the outcomes presented later.

Each year around 25,000 geoscience researchers and practitioners attend the American Geophysical Union (AGU) meeting, providing a regular touch point with the community to discuss a wide range of topics impacting the scientific endeavor. Within this setting, the ‘Why is Citing Data Still Hard?’ Virtual Town Hall took place in December 2020, examining continuing and emerging challenges around data citation and enabling credit. The topic of citing “many” datasets in a peer-reviewed journal was introduced with significant support from over 100 people in attendance. In 2019, AGU journals introduced a mandatory data citation policy and at

the time of this Town Hall still struggled with cultural norms around citing data in general, and the challenge of citing many datasets and other types of digital objects was still a problem. The Make Data Count[26] initiative from DataCite and other efforts within the Earth, space, and environmental sciences, including the Coalition for Publishing Data in the Earth and Space Sciences[27] (COPDESS) community has made significant progress on citing data and other digital objects.

Following the Town Hall, a determined group of participants convened three follow up working sessions[28] over the next two years. Significant progress was made in use case collection. This in turn helps in building interest across the international research ecosystem. In 2023, RDA approved the Complex Citation Working Group. To date, we have over 100 members representing the different audiences in the scholarly communication workflow and across multiple disciplines.

Key Decisions that Led to the Complex Citation Recommendation

- 2022: Drafting of the initial Complex Citation schema.
- 2023: The final recommendation needs to be multidiscipline, and broadly accepted internationally.
- 2024: The existing DOI types would not work. The Complex Citation needs a new resource type.
- November 2024: Formal recommendation as a result of the RDA Complex Citation Working Group (this document).

Overview of Complex Citation Working Group Recommendations

The initial set of use cases in conjunction with the workflow actors involved across the data publication lifecycle, led to ten requirements for the Complex Citation Object and how it is included in the existing publication lifecycle. It is through these requirements that the goals of the working group are met. At the same time, the recommendations needs to be simple enough to be implemented by the workflow actors as the first step towards a more general solution of the problem.

In the following subsections, we articulate:

1. Initial Collection of Use Cases
2. Requirements for Complex Citation Object
3. Complex Citation Object Workflow Actors: Roles and Responsibilities
4. Current State of Technology
5. Exemplar Use Cases
6. Discussion on the Challenges Yet to be Addressed
7. Discussion of Risks

[26] https://makedatacount.org/learn-about-us/
[27] https://copdess.org
[28] https://data.agu.org/DataCitationCoP/

8. Next Steps

Initial Collection of Use Cases

Following initial discussions and workshops, the initial set of use cases were collected across communities seen in the Appendix to better understand the problems and requirements for digital object citation, traceability, and transparency. A subset of use cases were then mapped against community requirements to form a broad baseline across user needs, as shown in Table 2. A subset of these use cases were further developed as proof-of-concepts and termed exemplar use cases (see below). These exemplar use cases further define what a Complex Citation Object needed to be and provide the basis on which attribution, traceability, and transparency of data and other digital objects could in turn be implemented.

The selected use cases mapped with the Complex Citation Object requirements include:

1. Intergovernmental Panel on Climate Change (IPCC): The data and software used to create individual figures in the reports and provide provenance and traceability to enhance the transparency of the results.
2. Ameriflux: The data captured from different types of instruments on the vast in-situ network of towers providing attribution and provenance to the current research teams, tower, and instrument.
3. British Oceanographic Data Centre (BODC): The objects that underpin an originators “dataset” i.e. different types of instruments and data granules used in oceanographic research.
4. Physical Samples: The digital information about individual physical samples such as with an IGSN, or an RRID.
5. Geochemistry (Geochem): The individual chemical analysis data from analyzing a rock, soil, or atmosphere sample.

Table 2 Requirements of selected initial use cases.

	Use Cases				
Requirements	IPCC	Ameriflux	BODC	Physical Samples	Geochem
Requirement 1 (citing subsets of larger datasets)	**X**	**X**	**X**	**X**	**X**
Requirement 2 (Underpinning data for graphs, chapters)	**X**				
Enable credit *	**X**	**X**	**X**	**X**	**X**
Citing larger group of datasets	**X**	**X**		**X**	**X**
Compression of knowledge			**X**	**X**	

graph levels					
User base beyond academia	**X**		**X**	**X**	**X**
Traceability (historical connections between data, objects and citation outputs)	**X**		**X**	**X**	**X**

***Enabling credit:** to enable credit is to ensure that the information needed by organizations that measure attribution and credit for contributions to the scientific record have what they need for digital objects. This includes information, with preference to persistent identifiers, for digital objects (Crossref DOI for publications, DataCite DOI for datasets and more), hosting repository, creators (ORCID), funder (Funder ID), Grant (Grant ID, DOI for awards), organizations/affiliations (ROR)

Requirements for Complex Citation Objects

There are 10 requirements for a Complex Citation Object. The first four use the term "must" to ensure common methods for implementation and application. The remaining requirements use the term "should" to encourage adopters and implementers to follow these as well. It is possible that in specific circumstances, a requirement identified as "should" might not apply.

R1: Each Complex Citation Object must capture a sufficient level of (meta)data granularity to ensure the credit and provenance to the specific cited material is included.

R2: Each Complex Citation Object must enable the mechanism of automated attribution and credit of each of the individual objects that are referenced.

R3: Each Complex Citation Object must be a clearly demarcated object that is not a primary output and should not accrue credit itself.

> Note: There can be cases, such as with **physical samples and laboratory analytical data**, where a new dataset or collection contains numerous primary objects, along with new valuable data never published before, such that the aggregated dataset should also receive credit and provenance tracking over time. The dataset should include the component objects as related identifiers with appropriate relationship types in dataset metadata. In these cases, a new PID can be assigned to the aggregated dataset, which would be included in the Complex Citation Object. In this way, both the aggregated dataset and components get credit and provenance tracking when the Complex Citation Object is cited over time.

R4: Each Complex Citation Object must be identifiable, referenceable, remain stable and resolvable through use of a suitable PID mechanism. As such they may not be deleted. The reference to the linked digital objects of a Complex Citation Object cannot be changed; any

reference update to the content would require a new version to be issued (i.e., a new Complex Citation Object).

R5: Each Complex Citation Object should employ a suitable versioning mechanism to support supersedence.

R6: Each Complex Citation Object should adhere to a standardized structure, agnostic of any particular provider, utilizing linked data defined elements to ensure interoperability and machine actionability.

> R6.1 Each Complex Citation Object should not provide elements for graph beyond two layers of connection[29] (e.g., including a Complex Citation Object as a Linked Digital Object)

R7: Each Complex Citation Object should primarily support the use of Persistent Identifiers (PIDs) for the Linked Digital Objects within a Complex Citation where possible, but permit non PID-items to be provided via a stable resolvable reference, e.g. URLs.

R8: Each Complex Citation Object should be sufficiently flexible to allow easy adoption by providers of Complex Citation Objects for a wide range of use-cases.

R9: Each Complex Citation Object should be provided, such that its metadata and content are always as open and accessible as possible/reasonable.

> R9.1: The Complex Citation Object license applies only to itself and does not transfer to any of the Linked Digital Objects.

R10: Each Complex Citation Object should have sufficient provenance metadata for its creation.

Complex Citation Workflow Actors: Roles and Responsibilities

The citation workflow that enables automated attribution and credit for the creators/contributors of digital objects includes a broad number of workflow actors (Table 3):

Table 3 Workflow actors/roles and their responsibilities

Workflow Actor / Role	Responsibility
Digital object creator/contributors	*This is a prerequisite in order to have the Linked Digital Object content ready to include in the Complex Citation Object.* When preserving the digital object, prepare community-accepted metadata. Select a preservation repository that supports making the digital object as open and FAIR as possible, provides a persistent identifier, and manages the preserved digital object.

[29] Number of layers for the knowledge graph is due to the state of technology which is expanded later

Workflow Actor / Role	Responsibility
Author team* for journal article of report	Has a document and needs an approach for referencing the many digital objects that support their findings/figures. Creates the Complex Citation Object to include all digital objects used in their research. Place the citation in the Reference Section of the paper and describe it in the Availability Statement. In-text citations will point to an individual Linked Digital Object in the Complex Citation Object.
Complex Citation Object Generator	Tool that provides a person the means to create, identify Linked Digital Objects, and register a Complex Citation Object.
Complex Citation Object Host (i.e. Repository)	Organization/entity that manages the Complex Citation Object Generator; has a relationship with a DOI Registry to assign persistent identifiers to a Complex Citation Object; commits to preserving a Complex Citation as required of a preservation repository. Data repositories can consider whether some of their existing objects (e.g. datasets with multiple components, such as samples, and collections of datasets) are suitable for Complex Citations.
Journal / Publisher - to include their third-party providers	Supports the use of Complex Citation Objects and allow/enable in-text citation in the text of the paper of individual Linked Digital Object.
DOI Registration Agencies	Registration Agencies provide services to people or organizations who need to identify and track the things that matter to them. Their work involves allocating DOI prefixes, registering DOI names, providing a metadata schema associated with each DOI record, and tracking related identifiers and use/credit. E.g., Crossref, Datacite.
Scholarly Indexer	Scholarly Indexer offers high-quality indexing services to authors, editors, and publishers of all kinds, with a specific emphasis on academic publishing. E.g., Web of Science, PubMed.
Consumers	Implements use of Complex Citation Objects and enables attribution and value of citing the Linked Digital Objects. E.g., Institutions values created datasets and other digital objects as part of a researcher's contributions to the scientific record to include those cited in a Complex Citation Object.

*Author team is using the Linked Digital Objects in the Complex Citation Object, not assumed to be the original creator of the Complex Citation Object

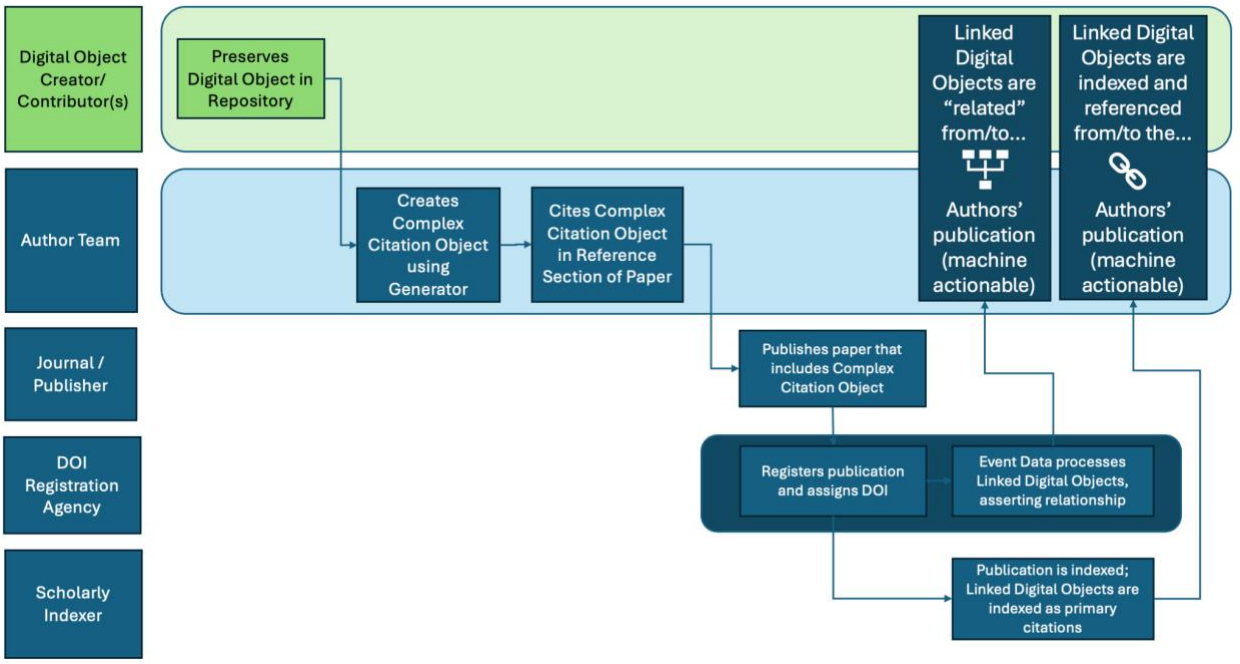

Figure 1: Visualization of the citation process. The green color indicates assumed parts outside of the Complex Citation Object recommendation

Current State of Technology

1. Natural Environmental Research Council (NERC) in the UK
   Citation is centered around the creation of DOIs on datasets. The implementation of DOIs across the NERC Environmental Data Service (EDS) remains close to the core requirements of earlier DataCite schema. The DataCite schema has evolved and more recently NERC EDS DOIs are being enhanced through the uptake of DataCite schema developments i.e. relatedIdentifier. Including these developments in conjunction with PIDs that identify data granules and sensors across the EDS will give the NERC EDS tools and technologies to support Complex Citations from high level datasets. This update to technology and development is still in the early stages.

2. Zenodo approach - IPCC
   Zenodo provides a stable and sustainable infrastructure (CERN long-term commitment) based on DataCite DOIs. It offers an API for creating and maintaining records, which supports the initiation and maintenance of Complex Citation Objects containing large numbers of Linked Digital Objects. Small adjustments will be required to accommodate the required changes to DataCite's metadata schema (new resourceTypeGeneral=ComplexCitationObject) and the metadata-only nature of a Complex Citation Object. Additionally, Zenodo supports versioning[30], including with relevant DOIs, which provides a pathway to further evolve the Complex Citation Objects in due course.

[30] https://support.zenodo.org/help/en-gb/1-upload-deposit/97-what-is-doi-versioning

3. Scholarly publishing in general and specific to AGU
   The scholarly publishing technical ecosystem is varied across journals with industry standards managed by NISO (https://www.niso.org/). Most publishers have similar workflow steps that begin with receiving the manuscript from the author team through to publishing the version of record with a DOI on a digital platform. Some publishers continue to publish paper journals. Citation for digital objects is beginning to take hold in the last 5-10 years in response to funder policy requiring that the data supporting research findings be shared in an open and FAIR manner. The percentage of journals that have author guidelines and policies that align to these policies is consistently increasing with each discipline area working at different paces reflecting their community starting point and culture. CHORUS tracks the publisher data (https://www.chorusaccess.org/resources/chorus-for-publishers/publisher-data-availability-policies-index/) and software (https://www.chorusaccess.org/resources/software-citation-policies-index/) availability and citation policies. International Association of Scientific, Technical & Medical Publishers (STM), one of several publisher societies, stood up their Research Data Program (https://www.stm-assoc.org/research-data-program/) in 2020. It has a core group of publishers sharing their percentage of data citations and openly encouraging other publishers to self-disclose their percentages. There is no automated way to generate these numbers and even with well- intentioned publisher participants, the dashboard remains publicly unpublished mostly due to the inconsistent frequency of the data being provided.

   AGU, as a society publisher, is an anomaly. They have recently completed a process-improvement effort to ensure that [nearly] all published papers have a data citation. In 2024 they publicly reported that 90% of the papers published that year have a data citation. What is not possible to know, in an automated query, is if the data are in the best possible repository for that type of research, if the data are as FAIR as possible, and if all the data sources used have been cited. Software/code citation is mandatory in their journal highlighting modeling research and selected papers in other journals where the research is about software being used. Other digital object types are not mandatory to cite.

4. Scholarly Indexers
   The indexer community is large with niche providers serving smaller communities, and a few leading providers such as Web of Science, PubMed, Google Scholar, and with a recent expansion, the Astronomical Data System (ADS).

   Most indexers do not index data and/or digital object citations. A notable exception is the Web of Science, which provides data citation indexing as an “add on” service to their current offerings. They use a combination of automated tools and human review to identify data citations. ADS, through their work with the American Astronomical Society has primarily indexed software via the Astrophysics Source Code Library (ASCL)

ascl.net and recently added in data citations. Through funding from NASA, they have expanded the number of journals they index to include all of Earth and environmental sciences in addition to planets, space, and heliophysics.

The indexers involved with the Complex Citation project have requested that the Linked Digital Objects identified in a Complex Citation Object not include a Complex Citation Object. In other words, a CCO can not include a CCO as one of its Linked Digital Objects. The team has included this request in the requirements. When discussing the Complex Citation as part of a knowledge graph, we say that it is limited to the number of levels. This is directly connected to the indexer imposed restriction.

Exemplar Use Cases

Three use cases were used in pilot studies. These are described followed by a gap analysis comparing them to the requirements.

1. BODC Use Case Pilot

The UK's Natural Environment Research Council sought to develop the idea of a Complex Citation through the UK Department for Science and Technology (DSIT). It funded the "BOOST-EDS" project based on the exemplar data collection use cases. Focusing specifically on a use case from the British Oceanographic Data Centre (BODC), as well as tracking the IPCC figure data use case (see below), the project identified connection nodes, natural points where a relation could be identified which could inform a knowledge graph. The approach was to view the nodes or granules that can form a "dataset" and the relationships that link to each node or granule i.e. sensorID of data collected are related to the data registered with a DOI. Using the exemplar use cases it was noted that often the number of levels that could form a citable entity often spanned 5 or more levels highlighting a need to flatten the graph, to meet indexer requirements. The design phase of the project then offered a means by which a more complex graph could be flattened and still provide credit to the source, along with traceability of the Complex Citation Object, in Figure 2. Each individual node should be able to provide reference citation based on the relationships related to the node supporting the provenance metadata of the schema defined. A PID, such as a Handle, was identified for the BODC use case nodes due to the lower cost and fewer metadata requirements for different data granules but relations may be connected to a range of links in the object such as sensor PIDs. Due to the relationships and connections of the node it should be possible to go back to the source data and provide appropriate citation for each data granule, meaning each node can form a part of a Complex Citation. This could also be expanded so a complex Citation Object could handle many nodes allowing for multiple granules to be cited.

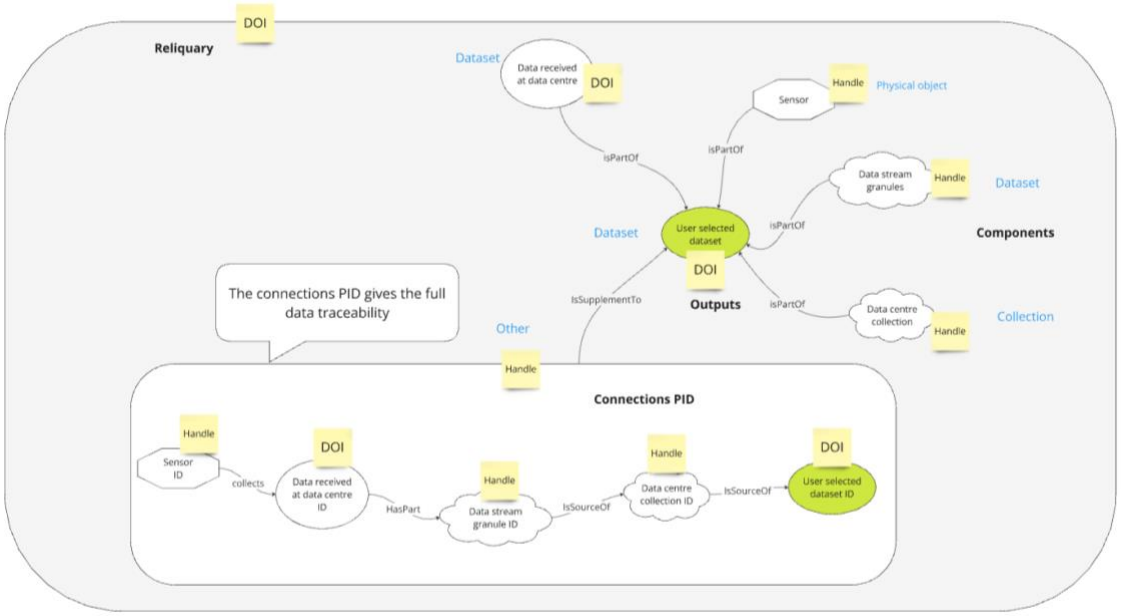

Figure 2: BODC Use Case

2. *IPCC Use Case pilot*

The pilot of the FAIR implementation into the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC WGI AR6) looked to address Complex Citations using existing technologies and services that already feed the indexer services. Complex Citation Objects were published for figure generation based on CMIP6 (Coupled Model Intercomparison Project Phase 6) data in a retrofitting approach. The information on CMIP6 input data usage in WGI AR6 figures was collected and stored at the repository responsible for archiving the CMIP6 data instead of being provided by the figure data archive. Through enhancing the Zenodo/DataCite DOI services by human-readable (csv) and machine-actionable (JSON-LD) provenance information, it could be demonstrated how a Complex Citation can provide appropriate linkages between report, different types of data (input and figure data) and enable basic figure reproducibility. Fundamentally, though, this had similar elements to the NERC pilot in that there was both the flattened list of related objects held by the record's metadata whilst also holding the richer knowledge graph within the archived files. Utilization of existing service and the associated Data Cite DOI structure was a strength of this pilot, but at this stage hits issues around the object types available at this time. This pilot is revised for AR7 due to workflow consolidation, enhanced and automated provenance creation and analysis and the content of the Complex Citation Object and the chosen service will be revisited. Details on the Complex Citation within the IPCC FAIR approach are provided in the Appendix.

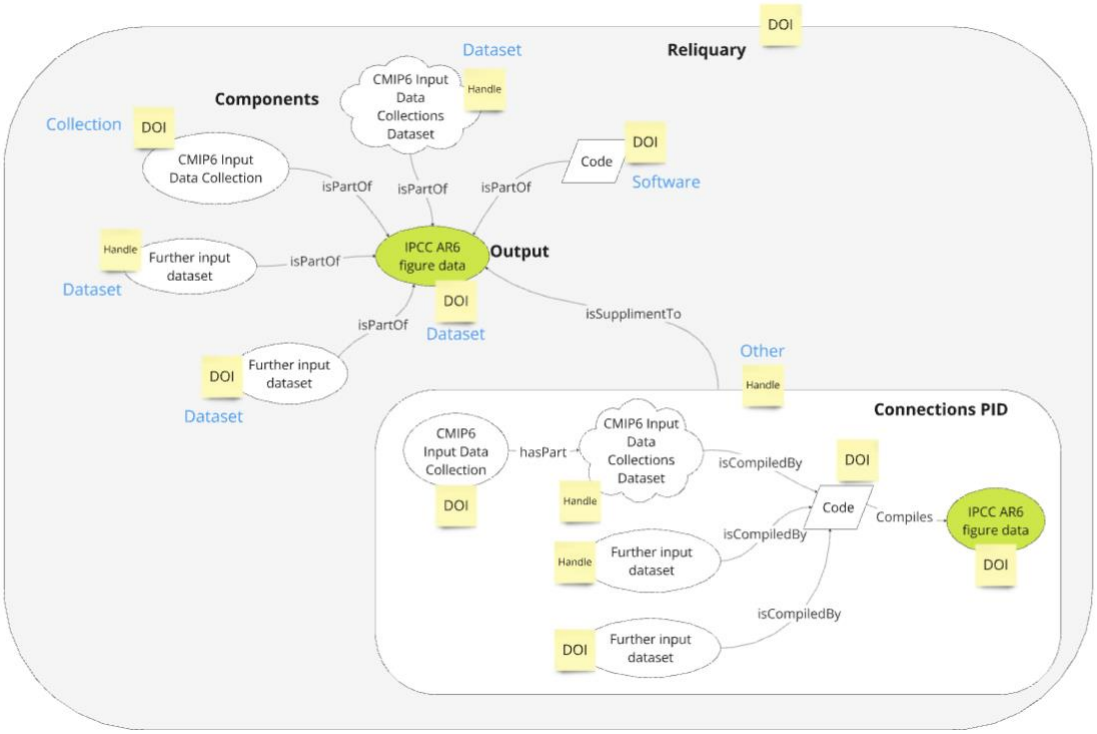

Figure 3 IPCC Use Case

3. *Physical Samples Use Case Study*

The ESIP Physical Samples Curation Cluster outlined community and technical needs to facilitate sample discovery, integration, and credit https://doi.org/10.31223/X5ST2K (currently in

review). Some key conclusions are directly applicable to the needs for Complex Citations for samples, including the following needs informed by specific use cases outlined in the manuscript:

a. General Requirements for Physical Samples:
   Tracking the use and provenance of physical samples requires the use of standard metadata and promotion of sample PIDs. Additionally, a community change towards citing sample-based datasets including numerous individual samples is needed. The Complex Citation Object can meet these general requirements.
b. Data Repository tasks:
   Data repositories need to implement and provide guidance for an aggregation of datasets and the provision of an associated Complex Citation Object for the dataset. The individual sample PIDs need to be related with appropriate relationship types. The Complex Citation Objects need to be integrated in data discovery and export of citation information e.g., in portals and APIs, and generally machine actionable. Advanced analysis tools of data usage and data tracking should be provided.
c. Physical sample repository tasks:
   Physical sample repositories, laboratories, data repositories, doi registration agencies, and indexers should encourage, and develop tools that incentivize and support use of sample PIDs and usage tracking. This is particularly important for showing the value and the reuse of Physical Samples on the long-term to funders and scientists but also to support tracking and analyzing the use of Physical samples, partly as subsamples, in multiple interdisciplinary studies and Complex Citation Objects.

Gap analysis on requirements of pilots

Following the design and build of the Complex Citation Object pilots, a gap analysis was undertaken against each of the requirements in **Table 4**. The pilots successfully cover many of the requirements, particularly as the pilots use DOIs as the basis for their Complex Citation Object. By defining the DOI as a type of Complex Citation Object, we still need to work out how to give credit to the underpinning objects. This will need work with indexers and underpinning infrastructure to ensure the Linked Digital Objects can be given direct credit. Additional gaps exist around the creation and standardization of provenance as well as expectation on versioning.

Table 4 Gap analysis of selected pilots against requirements. Green: indicates that the Complex Citation **will fill the gap**. Amber: indicates that the Complex Citation **will partially fill the gap**.

Requirement	IPCC Gap	BODC-NERC Gap
R1: Each Complex Citation Object must capture a sufficient level of granularity to ensure the credit and provenance to the specific cited material is included.	*Green* Related identifiers with appropriate markup are captured within the Complex Citation Object	*Green* Example provides flattened related identifiers direct to the citable element and encourages relations across the connection nodes once flattened.

Requirement	IPCC Gap	BODC-NERC Gap
R2: Each Complex Citation Object must enable the mechanism of automated attribution and credit of each of the individual objects that are referenced	*Amber* The use case can allow the capture of information through relatedWorks from the objects, but the richer knowledge graph within the Complex Citation Object needs application by indexers and harvesters down the line.	*Amber* The use case can allow the capture of information through relatedWorks from the objects, but the richer knowledge graph within the Complex Citation Object needs application by indexers and harvesters down the line.
R3: Each Complex Citation Object must be a clearly demarcated object that is not a primary output and should not accrue credit itself.	*Amber* DOIs are created for Zenodo records. Creation uses DataCite's metadata standard, which does not allow to classify a resourceTypeGeneral="ComplexCitationObject" meaning the object cannot be identified by indexers.	*Amber* Complex Citation Objects are created with DOIs with connections contained within the metadata schema. This is built on DataCite DOIs with objects underneath containing appropriate relationships.
R4: Each Complex Citation Object must be identifiable, referenceable, remain stable and resolvable through use of a suitable PID mechanism. As such they may not be deleted. The reference to the linked digital object content within a Complex Citation Object cannot be changed; any reference update to the content would require a new version to be issued (i.e a new Complex Citation Object).	*Green* Using DataCite's metadata standard DOI infrastructure the creation of the Complex Citation Object is identifiable, referenceable and resolvable.	*Green* Using metadata standard and DOI (DataCite) infrastructure the creation of the Complex Citation Object is identifiable, referenceable and resolvable.
R5: Each Complex Citation Object should employ a suitable versioning mechanism.	*Green* Zenodo support versioning for DOIs	*Amber* PIDs can have versioning support the application of which will be on the provider of the service. Versioning of DOIs and how they are versioned is dependent on the provider.

Requirement	IPCC Gap	BODC-NERC Gap
R6: Each Complex Citation Object should adhere to a standardized structure, agnostic of any particular provider, utilizing linked data defined elements to ensure interoperability and machine actionability. It should not have any depth greater than 2 layers of connections.	*Green* Structure uses related identifiers but these should be able to be understood and mapped across providers akin to the credit infrastructure	*Green* Structure uses related identifiers but these should be able to be understood and mapped across providers akin to the credit infrastructure
R7: Each Complex Citation Object should primarily support the use of Persistent Identifiers (PIDs) for listed objects within a Complex Citation where possible but permit non PID-items to be provided via a stable resolvable reference, e.g. URLs.	*Green* The Complex Citation is based around DOIs and Handles for listed objects and can link to other PIDs using related identifiers. Zenodo implementation supports a range of PID types and also use of URLs for linked items Standard (DataCite) relationTypes used. Embedded Complex Citation Object provides additional graph detail for specific relationships between linked items.	*Green* The Complex Citation is based around Handles - a specific PID and can link to other PIDs using related identifiers.
R8: Each Complex Citation Object should be sufficiently flexible to allow easy adoption by providers of Complex Citation Objects for a wide range of use-cases.	*Green* Using existing infrastructure that would support multiple domains and workflow actors. Adjustments may require cost compensation.	*Green* Using existing technology that would support multiple domains and workflow actors. It may require cost to set up such infrastructure from scratch.
R9: Each Complex Citation Object should be provided, such that its metadata and content are always as open and accessible as possible/reasonable.	*Green* Follows DataCite's DOI principles, which does allow restricted objects to occur; to be added resourceTypeGeneral="ComplexCitationObject" should not allow restricted objects.	*Green* Follows DataCite's DOI principles, which does allow restricted objects to occur; to be added resourceTypeGeneral="ComplexCitationObject" should not allow restricted objects.

Requirement	IPCC Gap	BODC-NERC Gap
R10: Each Complex Citation Object should have sufficient provenance metadata for its creation.	*Amber* Using existing infrastructure that captured Complex Citation Object metadata for provenance purposes and allowed credit; added machine-actionable JSON-LD file providing further provenance information, which requires standardization.	*Amber* Allows for metadata used in existing infrastructure that supports PID creation. Provenance of the underpinning metadata will be limited to the relationships that underpin the Complex Citation Objects, which may be limited by the generator, or the creator, of the Complex Citation Object.

The gap analysis helps to showcase how generators of Complex Citation Objects and DOIs can adhere to the requirements but will need to support and be supported by other elements to ensure the Complex Citation Object is able to support all the requirements. **Table 5** showcases the workflow actors against the process with comments and recommendations on how the actor should approach the process attributed to their role. The actor's workflow also details the expected flow from creation, support, dissemination and consumption of a Complex Citation Object. This highlights the depth of communities and the importance of community engagement and use of Complex Citation Objects for wider adoption across multiple scientific domains.

Table 5 Workflow actors against the process with comments and recommendations.

Workflow Actor	Process	Comment	Recommendations
Complex Citation Object Generator	A means to create a Complex Citation Object	Should be agnostic to existing PID/DOI providers and allow users to build a Complex Citation Object	Any tool will be provider agnostic
			Credit is not given for creation. A recommendation for the creation of a custom object type for Complex Citations
		Indexers need to be able to unzip Complex Citation Objects	Complex Citation Object can be unzipped by indexers
		Complex Citation Objects should be FAIR	Complex Citation Object is encoded in a manner which is both human interactable and machine actionable
Complex Citation Object Host (i.e. Repository)	Ongoing management of Complex Citation, preservation, version	A recognized place where a Complex Citation may be managed and maintained for	Hosts commit to maintaining the Complex Citation Object

Workflow Actor	Process	Comment	Recommendations
	control, reference support	the longevity of citation purposes.	Hosts will not edit contents of a Complex Citation Object
			Hosts can update metadata around the Complex Citation Object
			Hosts should be able to supersede Complex Citation Objects
Complex Citation Object Host	Registration of Complex Citations objects	Provides the mechanism and infrastructure to register Complex Citations for the long-term	Providing globally indexed objects supporting FAIR data
Author team for journal article of report / Researcher	Placing the Complex Citation in the paper, or other scholarly document	To support wider citation researchers should provide a Complex Citation should be in-line with journal requirements	Researchers are responsible for providing objects with the Complex Citation Object
			Researchers are responsible for ensuring appropriate relationships exist for their objects
Journal / Publisher - to include their third-party providers	Where the Complex Citation is in the Reference Section, described in the Availability Section, and supports in-text citation, included in the Crossref files.	A lot of journals already encourage use of DOIs to cite data, Complex Citations will be an extension to this requirement. Third party providers i.e. copy editors	Publishers are responsible for ensuring/determining Complex Citation Objects are provided where individual DOIs are not appropriate
		Use of third-party editors	Third-party providers should accept Complex Citation Objects as valid digital objects
Scholarly Indexers	Unzip package	Indexers need to be able to unzip a Complex Citation Object to build the graph to support citation	Complex Citation Objects should be accessible by indexers
			Complex Citation Objects should be machine readable
Consumers incl. Harvesters	Recognizes and integrates Complex Citation Objects into their schema	Complex Citation elements, Linked Digital Objects, are brought into Schema for wider digital ecosystem i.e.	Complex Citation Objects should be consumed as a new persistent identifier construct.

Workflow Actor	Process	Comment	Recommendations
		Crossref Event Data, DataCite Commons etc. They can be used by other services i.e. a person's ORCID	
	Supports metrics around data and data use	Repositories can consume information to inform and update data granules i.e. Citation numbers may be updated	Complex Citation Objects should be consumed to inform on data granules (provide attribution)

A Complex Citation Object collates information for the purpose of enabling attribution and credit, supporting traceability, and providing transparency for data use.

A Complex Citation Object should be assigned a DOI[31] to support persistence of the object, and to facilitate access for indexers, harvesters and other consumers. The DOI metadata should allow identification of Complex Citation Objects through a new resource type such as "ComplexCitationObject". This identification of a new resource type will allow interpretation of the Complex Citation to support credit to authors of the Linked Digital Objects and not to the authors of the Complex Citation Object itself. The Linked Digital Objects need to be primary citations in order to facilitate the implementation of the knowledge graph. Although not mandatory, PIDs are strongly encouraged[32] to support the full capability of knowledge graphs, and the automation of information, although URLs or other identifiers that are well maintained may be used as an alternative[33].

In summary (Figure 4):

Complex Citation Objects needs:

1. DOI Persistent Identifier
2. Creator(s) and originator(s) of the Complex Citation Object
3. Funder and Grant information for the research supported by the Complex Citation Object
4. Linked Digital Objects

Linked Digital Object needs:

1. Persistent Identifier, URL
2. Originator / Creator / Contributor
3. Repository or host location of digital object

[31] The DOI as the recommended persistent identifier is stated here because it represents the current status of technology for publishers. Nearly all publishers support DOIs as a way to create semantic links when listed in the Reference Section of a paper.

[32] PIDs for linked data objects would be used to support data granules, authors, funders etc

[33] Using https://archive.org/ for URLs in order to avoid unresolvable URLs as recommended by RAID are an option (https://metadata.raid.org/en/latest/core/relatedObjects.html).

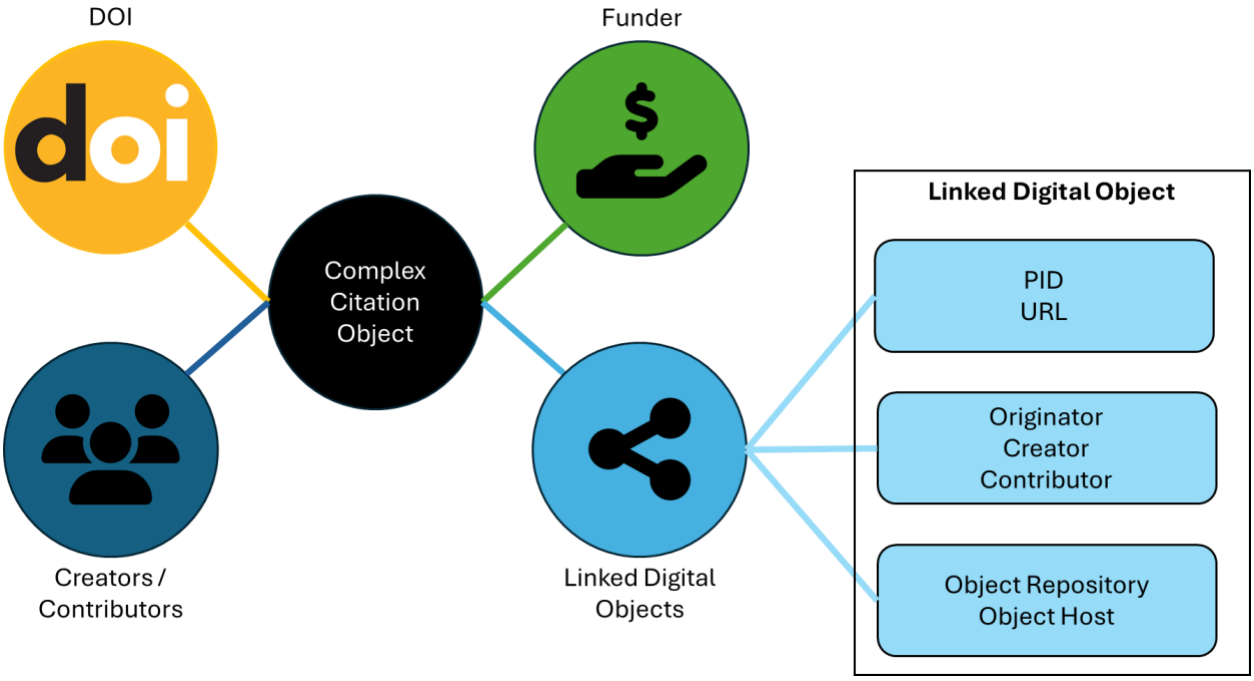

Figure 4: Complex Citation Object and Linked Digital Object with Required Elements

Requirement from Indexer Community:
It is important that a Complex Citation Object NOT be included as a Linked Digital Object that is also a Complex Citation i.e. a complex citation object should not include a complex citation object as this would make the graph difficult to implement.

Discussion on the Challenges Yet to be Addressed

The Complex Citation Object seeks to support a solution for citation although challenges remain in citing data that are outside these recommendations. Table 6 shows how the Complex Citation has addressed the requirements initially identified in Table 2 (page 10).
Citation for research objects still needs to be better supported by journals and publishers beyond papers to support the flow of credit. This includes credit being provided for contributors such as funders, repositories, engineers, lab technicians etc. Understanding what credit is and how credit is enabled for these contributors still needs to be defined. Likewise, distinguishing between the use of Linked Digital Objects for provenance or respecting the alternative citation requirements of the Linked Digital Object providers remains a challenge. [34]Workflow actors will need to define credit to consolidate how Complex Citation Objects can be used.
Independent scrutiny of (meta)data used in publications and a process of peer review are not widely adopted and are not a consideration for Complex Citation Objects but something which may support this process.

[34] Julien Colomb provided an example during community review of software being a citable object, still included in the CCO for provenance, but should have its alternate resource listed in the citation.cff file as the required citation route.

Governance of the metadata schema will need to be considered to support machine actionable and interoperable Complex Citation Objects across domains and workflow actors. Guidelines from key workflow actors are likely to support this work going forward to ensure multi domain adoption and use.

Through the engagement with the different communities Complex Citations also encountered new requirements such as reducing the knowledge graph to meet indexer requirements to prevent the graph becoming too large. This also highlights the use of workflow actors to ensure engagement across domains, and throughout the Complex Citation described in Figure 1.

Table 6 Comparison of Complex Citation use case against the requirements of Table 2

	Use Case: Complex Citation	
Requirements	Meets Requirement	Comment
Requirement 1 (citing subsets of larger datasets)	Y	Allows large listings of linked digital objects from subsets of large datasets
Requirement 2 (Underpinning data for graphs, chapters)	Y	Allows graphs to identify data used for the graph
Enable credit	Y	Through the provision of authors in the linked data objects allows the attribution of credit to the underpinning data but needs further work
Citing larger group of datasets	Y	Allows large listings of linked digital objects to collate information akin to one larger dataset
Compression of knowledge graph levels	Y	Provides nodes that allow the compression of the knowledge graph to be unzipped by indexers
User base beyond academia	Y	Complex Citation Objects are to be provider agnostic so as to allow non academics to create their own Complex Citation Objects and could be used for gray or white papers
Traceability (historical connections between data, objects and citation outputs)	Y	By using DOI services this allows historic traceability and transparency of (meta)data used even when the data are no longer available - to include relations to other digital objects

	Use Case: Complex Citation	
Requirements	Meets Requirement	Comment
Knowledge graph to be limited to 2 levels	Y	Indexers require a restriction to the knowledge graph to ensure it is usable, manageable and sustainable.

Discussion of Risks

1. Do Nothing

Currently there is not an optimal solution for citing large numbers of digital research objects in a manuscript such that all the requirements for enabling credit, transparency, and traceability are met.

The increase in research complexity using more and more digital objects to address multidisciplinary, multinational challenges with transparency and integrity requires a flexible, global approach.

There is no current workaround that provides the necessary, machine-actionable resolution to the problem.

Without addressing this concern, the current drawbacks will continue:

- Scientists do not publish research objects, as they will not get credit and therefore don't have any incentive to create them
- Research objects are not regarded as valuable scientific product
- Transparency of research results is not provided
- Trust in research results decreases as they are not traceable nor reproducible
- Reuse of research objects is limited especially for machine-readiness, hampering AI developments and multidisciplinary use
- Research object discoverability is reduced through missing or not-machine-accessible relationships and provenance information are not machine-accessible
- Non-standardized and non-machine-actionable workarounds are in use to comply with author guidelines from publishers or funding agency requirements
- It is difficult for funders to understand how research was used -- due to the complexity of the research requiring the data aggregation

2. Not fully supported / implemented / maintained Complex Citation approach

A partially or poorly implemented Complex Citation approach is associated with several risks related to different workflow actors (Table 7; see responsibilities of workflow actors in Table 5).

Table 7 Risks for Complex Citation implementation

Risk	Workflow Actor	Impact	Comment
Risk 1: Complex Citation Object is not maintained	Complex Citation Host	Acceptance of Complex Citation Objects deteriorates over time	A Complex Citation is part of the scientific record as an index of digital objects used in research. It is expected that the object is persistent.
Risk 2: Complex	Registry Provider	Misinterpretation	The Complex Citation

Risk	Workflow Actor	Impact	Comment
Citation Object is not accurately described in the metadata		leads to wrong attribution assignment	Object, represented in a new DOI type ensures that the relationship with the Linked Digital Objects correctly identifies what is to be cited.
Risk 3: Dataset incorrectly has the object type of “Complex Citation”.	Researcher / Complex Citation Host	The wrong attribution assignment worsens the current situation	Thorough check of resource type Complex Citation required before DOI registration to ensure a Complex Citation Object resource type is correct.
Risk 4: Complex Citation Object not unzipped	Indexer	Misinterpretation leads to wrong or no attribution assignment	Indexers have a pivotal role ensuring attribution for digital objects used in research.
Risk 5: Complex Citation Object accidentally includes a Linked Digital Object that is also a Complex Citation	Complex Citation generator	Indexers requested that the Complex Citation not include a Complex Citation Object as a Linked Digital Object.	Too many levels in the knowledge graph would hinder use, speed and access to the citation ecosystem
Risk 6: Complex Citation Object not recognized	Harvester	Credit, traceability and transparency will not be provided	Harvesters play a pivotal role ensuring the dissemination of the information to provide credit, traceability and transparency
Risk 7: Creation of Complex Citation Object too cumbersome	Researcher / Complex Citation Host	Complex Citation approach is not taken up	Making creation and adoption of Complex Citation Objects is important for the adoption of the Complex Citation Object
Risk 8: Complex Citation Object not accepted in	Journal Publisher	No improvement of current situation	Journals play a vital role in supporting the use and adoption of

Risk	Workflow Actor	Impact	Comment
publications			Complex Citation Objects
Risk 9: Complex Citation Object makes it more difficult for the reader to locate the data or digital object used for the research.	Journal Publisher	It will become more difficult to locate a digital object with the additional layer of connection from using a Complex Citation Object. This would be especially problematic if the use cases for the Complex Citation were expanded to include all papers, even those requiring only a few digital object citations.	Journals may find it simpler to provide author guidance that is the same for all authors. It is difficult to know what the adoption process will be at the moment. However, for the researchers that need a Complex Citation, we feel this risk can be mitigated.
Risk 10: Complex Citation Object not provided in unzipped consumable version by Indexers	Indexer / Consumers	Uptake by consumers is hampered and leads to a high number of misinterpretations	Indexers play a vital role in supporting the use and adoption of Complex Citation Objects

Next Steps

In Review of this Recommendation:
Sharing this work with the RDA Complex Citation community will be the first step in obtaining feedback on the requirements and recommendations of Complex Citation Objects. This will showcase the evolution of the Complex Citation Object idea with active demonstrators and provide a mechanism to identify if community requirements have been addressed, identify any further concerns from the community, and gain feedback on knowledge gaps or assumptions made. This will then allow the Complex Citation writing group to publish the work to get global peer review to consolidate the work for a resolution going forward. Conferences, outside of RDA, to be targeted for researchers and repositories include International Data Week (IDW), SciDataCon/World Data Service (WDS)/ Committee on Data of the International Science Council (CODATA), Coalition for Publishing Data in the Earth and Space Sciences (COPDESS), EGU and AGU meetings. Meetings for journals include: COPDESS, International Association of Scientific, Technical & Medical Publishers (STM), Society for Scholarly Publishing (SSP), and Council of Science Editors (CSE).

Following Publication of this Recommendation:

This work forms the basis around feedback into “credit” and how the attribution of credit needs to function. Using the outcomes and progress from the initial pilots, this working group will transition into “maintenance”, further expanding the use cases across the disciplines, refine the implementation approach of a Complex Citation, heavily promote adoption and implementation across the workflow actors, and bring awareness to the broad research ecosystem as to the value of a Complex Citation, and encourage capacity building to support Complex Citation Objects.

Each workflow actor requires an implementation guide that defines the detailed integration needed for their role. This working group can support this effort in partnership with each workflow actor. To date we have engaged representatives from each of the workflow actors in preparation for this step. Expanding the number of participants will be paramount to reaching a level of implementation where a Complex Citation is easy for a researcher to use.

Further, the concept of “enabling credit” for data and other digital objects is intertwined with this work. Partnering with other RDA entities to bring this discussion to the level of practice and policy is desirable.

Additional Workflow Elements Where Others Can Contribute

The research ecosystem is broad with systems for participation that are entrenched in culture. Providing accurate attribution has always been a mainstay for scholarly literature. Our work in Complex Citation supports the importance of identifying and tracking the digital objects necessary to provide attribution and provenance of the research findings.

- The need for digital object creation / citation to be valued in scholarly impact and productivity metrics. Groups working on that include: CoARA, MakeDataCount; Tenure/Promotion Evaluation Changes (AAU/APLU)
- The ability for the physical sample repositories to know when a sample is cited. Groups working on this include: iSamples, ESIP Sample Cluster
- The ability for all PIDs to be recognized in the scholarly workflow. Publishers have implemented the Digital Object Identifier (DOI) PID, and most don’t support other types of PIDs for automated attribution. Groups working on this include: DataCite, MakeDataCount, RAID community, IGSN community, PIDFest

Appendix

Use Cases and alignment with the exemplar use cases

The exemplar use cases and the role of Complex Citation are described in more detail together with further gathered use cases and their alignment with these exemplar use cases.

Detailed IPCC Use Case

General Background:
Every five to seven years, the Intergovernmental Panel on Climate Change (IPCC) convenes the climate science community to assess the latest knowledge on climate change relevant to policymakers. This generally takes the form of Assessment Reports (AR) covering the scientific basis of climate change, its impacts and future risks, and options for adaptation and mitigation. With each cycle, these reports have grown in scope, length, number of referenced papers, and underpinning datasets. During the sixth assessment cycle, a large-scale collective effort went into archiving digital products assessed and generated through the IPCC process. The main objectives driving this initiative are making IPCC's work more transparent, improving the reproducibility and reusability of the assessment outcomes, better utilization of the services of the IPCC Data Distribution Centre (DDC), and, more generally, compliance with best practices in open science (Pirani et al., 2022). In particular, a new focus was put on archiving data and scripts underpinning key figures and tables from the assessment reports transparently by references between data, scripts and reports and basic provenance documentation - ie. to ensure implementation of Open Science and FAIR principles. This is a process that will continue to be built upon to support future Assessment Reports (TG-Data, 2023; Stockhause et al., 2024).

Role of Complex Citation:
Complex Citation is an integral part of the IPCC FAIR approach, supporting the traceability and reproducibility of key statements of the Assessment Reports, especially their figures (Stockhause, 2023). The aim is the inclusion of Complex Citation references in figure captions to connect figures with the figure creation process - both input data and associated processing code - to enable traceability and credit assignment for data (both derived and source), and code input data providers. However, ensuring both these aspects is challenging, primarily due to the differing levels of granularity in play: citations for credit assignment exist on larger dataset collections, whilst traceability is established at a lower level by links to specific subsets of datasets across these collections. This situation, therefore, cannot be supported by existing citation/linked-data techniques for journals and indexers, and needs to utilize Complex Citation Objects. Furthermore, the Complex Citation approach needs to be one that can be embedded into the IPCC FAIR implementation into the IPCC Assessment process across chapters and Working Groups, e.g. by author training and provenance-enabling tools for authors/utilization of provenance records. ComplexCitation publication must be part of the AR editorial process and be coordinated with the DDC Partners responsible for long-term data preservation and the AR7 publisher.

References:

Anna Pirani, Andrés Alegria, Alaa Al Khourdajie, Wawan Gunawan, José Manuel Gutiérrez, Kirstin Holsman, David Huard, Martin Juckes, Michio Kawamiya, Nana Klutse, Volker Krey, Robin Matthews, Adam Milward, Charlotte Pascoe, Gerard van der Shrier, Alessandro Spinuso, Martina Stockhause, & Xiaoshi Xing. (2022). The implementation of FAIR data principles in the IPCC AR6 assessment process. Zenodo. https://doi.org/10.5281/zenodo.6504469

Intergovernmental Panel on Climate Change. (2023). TG-Data Recommendations for AR7 (1.0). Zenodo. https://doi.org/10.5281/zenodo.10059282

Stockhause, M. (2023, February 28). RDA WG Complex Citation: Use case IPCC. Zenodo. https://doi.org/10.5281/zenodo.7684261

Stockhause M, Huard D, Al Khourdajie A, Gutiérrez JM, Kawamiya M, Klutse NAB, et al. (2024) Implementing FAIR data principles in the IPCC seventh assessment cycle: Lessons learned and future prospects. PLOS Clim 3(12): e0000533. https://doi.org/10.1371/journal.pclm.0000533

Initial Collection of Use Cases (RDA Complex Citation WG: Use Cases)

No	Name	Type/ Alignment	Domain	Short Description	Reference/Link	Complex Citation Draft
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	https://docs.google.com/presentation/u/1/d/1D8etrhvZIJmoHiqg-LM6CNwQ3p0AY9U8	https://docs.google.com/spreadsheets/d/10v58s_ZeMlffXy_pxjleK08LEhbVhYOetqh80BzGT_8
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	Copy of AGU community of practice - uses cases - BODC	Reliquary-BodcSeriesData
3	Intergovernmental 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	https://docs.google.com/presentation/d/1yYjtZAyNdi296hbgPzdkYJ71ITJ92n6morl9KgUehGw	https://docs.google.com/spreadsheets/d/1jZgdAHoSiFn3dxvYFPN_SXK4j35pjR2-wwLIJSJORC8

No	Name	Type/ Alignment	Domain	Short Description	Reference/Link	Complex Citation Draft
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/	https://www.clarin.eu/content/virtual-collections	example: https://collections.clarin.eu/details/1000?2&backPage=0 https://doi.org/10.34733/vc-1000
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?	https://journals.ametsoc.org/view/journals/mwre/148/9/mwrD190404.xml?tab_body=abstract-display	
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	https://docs.google.com/document/d/1KGVJa7QbVOOjwYXeyS8jCr32nMYBoIvhkO4fjnbj89Q/edit?usp=sharing	
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.	https://doi.org/10.20387/bonares-fm2j-c233 chapter: Citation of data from the BonaRes Repository - special cases; Dataset as part of a data collection	

No	Name	Type/ Alignment	Domain	Short Description	Reference/Link	Complex Citation Draft
8	NASA Astrophysics Data System	Linking data and software to the scholarly literature and capturing their citations	Astrophysics, Heliophysics, Planetary Science, Earth Science	The ADS 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.	https://ui.adsabs.harvard.edu/abs/2022BAAS...54b.022A/abstract	
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 to attribute uses to funders in coming months.	https://www.gbif.org/citation-guidelines https://www.gbif.org/data-use	example: https://doi.org/10.15468/dl.5n4h8v , cited in https://doi.org/10.1007/s10841-023-00488-6

No	Name	Type/ Alignment	Domain	Short Description	Reference/Link	Complex Citation Draft
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	https://doi.org/10.7283/T57H1GGM, https://doi.org/10.7283/T5B27SN9, https://doi.org/10.7283/T5V98697,	

No	Name	Type/ Alignment	Domain	Short Description	Reference/Link	Complex Citation Draft
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.	Borton, et al. (2022) GROWdb US River Systems - Samples. [Data set]. DOE Systems Biology Knowledgebase. https://doi.org/10.25982/109073.30/1895615. OSTI record (auto-sent to DataCite): https://www.osti.gov/dataexplorer/biblio/dataset/1895615	
13	SeaDataNet/ Emodnet	Citation of subsets from up to millions of files	Marine (but could be generic)	Marine observation data are organised in files, millions of files. In aggregators like SeaDataNet, EMODnet and others these files come together in a cloud. The newly developed BEACON software allows 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?	https://beacon.maris.nl/	
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.	https://docs.google.com/presentation/d/1QsR1NTjcD8pSPnZI7kuSyxnWnjLSFfZGWdB1j5sUWVI/edit#slide=id.g2def643904c_0_180	Example published datasets that include numerous samples with associated PIDs (IGSNs), ideally documented in dataset metadata as Related Identifiers with relation type “References” and/or “HasPart” https://doi.org/10.26022/IEDA/112300 https://doi.org/10.5880/ICDP.5059.001

No	Name	Type/ Alignment	Domain	Short Description	Reference/Link	Complex Citation Draft
						https://data.ess-dive.lbl.gov/view/doi:10.15485/1603775 https://doi.org/10.1594/PANGAEA.917685
15	Research Resource Identifier (RRID)	Physical Sample Persistent Identifier	Biological	The RRID represents antibodies, model organisms and software projects) in the biomedical literature to improve transparency of research methods.	https://www.rrids.org/	
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 PIDs - Matt Mayernik	Citation of instrument used in the research	All domains		Instrument PID RCN (Matt Mayernik)	
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).	https://docs.google.com/presentation/d/1g0IDGDIT8qEKDfgKq7KQlZg1ot36MYGQ/edit?usp=sharing&ouid=107571526871591327843&rtpof=true&sd=true	

Complex Citation Development Activity

When we first started this work, we did not want to confuse the Complex Citation problem with existing vocabulary terms that are similar. We humorously selected the term 'reliquary', a box of precious things, that was not intended to be permanent. This was replaced by 'Complex Citation Object'. Historical project text uses the older term.

During 2021 we held three large working sessions as well as smaller group development efforts

April 2021

Develop a common agreement on the use case (and variations) as well as hear from those whom it affects. Materials and link to recording: Agarwal, Deborah, Coward, Caroline, Stall, Shelley, & Erdmann, Christopher. (2021, April). Data Citation Community of Practice – 8 April 2021 Workshop. Zenodo. http://doi.org/10.5281/zenodo.4673622

June 2021

Presentations from different repository use Cases:

- RO-Crate, Carole Goble
- BioStudies, Ugis Sarkans
- GBIF, Daniel Noesgaard
- Pangaea, Uwe Schindler

Infrastructure Elements:

- DOI Collection, Martin Fenner, DataCite
- Make Data Count, Martin Fenner, DataCite
- Scholix / OpenAire, Paolo Manghi

Workshop materials and link to recording: Agarwal, Deborah, Goble, Carole, Soiland-Reyes, Stian, Sarkans, Ugis, Noesgaard, Daniel, Schindler, Uwe, Fenner, Martin, Manghi, Paolo, Stall, Shelley, Coward, Caroline, Erdmann, Chris, 2021. Data Citation Community of Practice – 8 June 2021 Workshop. https://doi.org/10.5281/zenodo.4916734

October 2021

We leaned into the term "reliquary" as a temporary word for a collection/package of other DOIs, PIDS, or links.

"Reliquary" Use Cases:

- British Oceanographic Data Centre, Justin Buck and James Ayliffe
- German Climate Computing Center/IPCC DCC, Martina Stockhause
- Ameriflux, Deb Agarwal

Workshop materials and link to recording: Stall, Shelley, Buck, Justin, Ayliffe, James, Stockhause, Martina, Agarwal, Deb, Coward, Caroline, & Erdmann, Chris. (2021, October 29). Data Citation Community of Practice – 29 October 2021 Workshop. Zenodo. https://doi.org/10.5281/zenodo.5641236

We are hopeful to bring awareness of this effort to the broad RDA community and move forward using the RDA structure.

May 2022
Reviewed initial attempts to create reliquaries for pilot use cases:

- Ameriflux, Deb Agarwal
- German Climate Computing Center/IPCC DCC, Martina Stockhause
- British Oceanographic Data Centre, Justin Buck and James Ayliffe

Stall, Shelley, Agarwal, Deb, Buck, Justin, Ayliffe, James, Stockhause, Martina, & Coward, Caroline. (2022, May 3). Data Citation Community of Practice - 3 May 2022 Workshop. Zenodo. https://doi.org/10.5281/zenodo.7455108

Community Outreach

December 2021 - AGU Fall Meeting 2021
April 2022 - EGU 2022 General Assembly
December 2022 - AGU Fall Meeting 2022

At the end of 2022 the effort needed to expand internationally and be applicable to all/most domains. It was decided to establish an RDA Working Group.

RDA Plenary Activities (Link to WG activity details)

February 2023: Complex Citation working group endorsed by RDA. Link to WG page.
March 2023: RDA P20, Gothenburg, Sweden, session Complex Citations in the Earth and Space Sciences: Formulating Requirements for a Demonstration Prototype. Reviewed the use cases and invited participation.
October 2023 - RDA P21 / IDW23, Salzburg, Austria: Complex Citations: Next steps towards a Demonstration Prototype
May 2024 – RDA VP22: Complex Citations: Working towards recommendations
November 2024: RDA P23, San Jose Costa Rica, Presenting the draft recommendations.

Community Outreach

April 2023: European Geoscience Union (EGU) General Assembly Town Hall, Complex Citations: Current Work to Ensure Proper Credit for 100+-cited Data and Software Objects
May 2023: PV2023 Poster, Focusing on Scalable Citations to Improve Data Usability and FAIRness (Where PV means Preservation and Value)
December 2023: AGU23 Town Hall: Complex Citations Are Needed in Wide-Open Science: Ensuring Transparency, Reproducibility, and Credit for All Supporting Research Contributions
December 2024: AGU24 Town Hall: TH15N Complex Citations: Ensuring Transparency, Reproducibility, and Credit for All Supporting Research Contributions