GFBio Consensus Document for citation pattern of ABCD datasets

[Introduction] [Suggested dataset citation pattern] [Pattern] [Rules applied] [Examples] [References]

Introduction

Data publications are becoming more and more common and accepted in scientific practice. Thus, a proper citation of this publication type should be considered in scientific writing. Martone (2014) provides Data Citation Principles that indicates in more detail why it is relevant to cite data.
In order to support this good practice, the GFBio data centers recommend how to cite ABCD data archives accessed from the GFBio portal. This document represents the version agreed upon by the GFBio Steering Committee as of 2018-06-13.

You used published data in your analysis and want to cite it correctly? Follow the citation guidelines detailed below!

Ready to explore more data? Browse published datasets in the **GFBio Data Portal** ↗

Questions regarding availability and resubality of data? **Contact our Helpdesk**! ↗

Suggested dataset citation pattern

The following formula is the recommendation on how to cite the dataset once it is publically available on the GFBio portal.
This should be accommodated in the ABCD element */DataSets/DataSet/Metadata/IPRStatements/Citations/Citation/Text*

Pattern

```
<Authors> (<Publication_year>). <Title>. [Dataset]. Version: <VersionNr>. Data Publisher: <Data_center_name>. <URI>.
```

The following string might be appended to the citation on the GFBio portal, but should not be included in the ABCD element */DataSets/DataSet/Metadata/IPRStatements/Citations/Citation/Text*:
`Accessed via <GFBio portal link to dataset> at <access date YYYY-MM-DD>.`
<access date YYYY-MM-DD>. is the date on which the end user found and/or downloaded the dataset from the GFBio portal. *Caution*: This MUST NOT identify the version of a dataset at a certain point in time, because if I access the same dataset tomorrow, I need to get the same data as of today! It just attributes to the state of the portal link where the dataset was accessed.

Rules applied

- **<Authors>** (mandatory) can be one to many authors or an institution / organization / working group
  - if there are more that three authors the list of names COULD be shortened by "et al." in the ABCD concept */DataSets/DataSet/Metadata/IPRStatements/Citations/Citation/Text*. But in this case, it is highly recommended to add a full citation (with all author names written out) in the concept */DataSets/DataSet/Metadata/IPRStatements/Citations/Citation/Details.*
  - the authors can be real people, but also legal entities (e.g. consortia, working groups etc.)
  - if author(s) is/are real people
    - the family name MUST be the first
    - the given name SHOULD be abbreviated with a dot
    - the family name SHOULD be separated from the given name initial by a comma
    - thus, use the format `<family name>, <initial of the given name>` e.g. for *Marianne Mustermensch* use `Mustermensch, M.` and for *Karl-Heinz Mustermensch* use `Mustermensch, K.-H.` and for *Wolfgang Amadeus Mozart* use `Mozart, W. A.`
  - if provided as a list of names use semikolon-separation and before the last author the *&*-symbol : `<family name1>, <given name1>; <family name2>, <given name2> & <family name3>, <given name3>`
  - be aware that the author(s) COULD be the same person that is mentioned in the ABCD element */DataSets/DataSet/ContentContacts/ContentContact*, BUT it not necessarily has to be the same.

- **<Publication_year>** (mandatory) is the (foreseen) year of publication (the date the dataset was made available, the date all quality assurance procedures were completed, and the date the embargo period (if applicable) expired); according to: http://www.dcc.ac.uk/resources/how-guides/cite-datasets#sec:elements
  - the publications year by chance COULD be the same year like in */DataSets/DataSet/Metadata/RevisionData/DateModified*. That's when the last modification and the publication were conducted in the same year.

- if old data (modification date is in the past) is published today it MUST be the current year (provided that the publication will be conducted in the current year).
- if the preparation of the datacenters is done close to the turn of a year, but the the publication is shifted to the next year NEXT year MUST be applied as publication year).
- if the publication is paused due to embargos, the citation recommendation makes no sense until the publication is out. Anyways, by definition the publication year MUST be the FUTURE year in which the dataset will be published.

- **<Title>** (mandatory) is the data set title from the ABCD element /DataSets/DataSet/Metadata/Description/Representation/Title

- **<VersionNr>** (optional) is the version number of the dataset.
  - This is useful for submissions of another snapshot (or next version) of continued datasets (e.g. monitoring data), but also generally recommended.
  - The version number's format is arbitrarily defined by the datacenters.
  - Thus version number COULD be the date or timestamp of the last modification /DataSets/DataSet/Metadata/RevisionData/DateModified, but doesn't have to be.
  - If a version number is provided, the qualifier *Version:* must be used (as shown in the pattern above)

- **<Data_center_name>** (mandatory) is the name of the publishing GFBio datacenter as given in the GFBio portal at https://www.gfbio.org/about/data-centers

- **<URI>** (optional) is a stable URI provided by the data center or the data provider. This could use a centralized service like DOI, PURL etc. or an institutional URL of a dataset landing page as for example provided by an institutional BioCAse Provider Software installation.
  - DOIs are cited according to DataCite as URIs (without "doi:", as a resolver domain with https, without the dx-subdomain) https://doi.org/10.DD/XXXX
  - institutional URIs need to be as stable as possible (e.g. according to best practice) and must **not** be the link to the ABCD-archive in the BioCASe Provider Software
  - the URI MUST NOT be the data center website, but a dataset landing page URL COULD be used.
  - A dataset landing page might be similar to a (project) website, but describes a certain dataset. The usage of a project website SHOULD be used with caution and MAY NOT be confused with the institution website.

- punctuation rules and any missing details should be implemented according to APA Citation Style)

Examples

- Kohlbecker, A. (2016). Coleoptera observations in orchards of South Western Germany. [Dataset]. Data Publisher: Botanic Garden and Botanical Museum Berlin. https://data.bgbm.org/dataset/gfbio/0002/.

- Van der Vos et al. (2018). Ontogeny of Hemidactylus. [Dataset]. Data Publisher: Museum fuer Naturkunde Berlin (MfN) - Leibniz Institute for Research on Evolution and Biodiversity. https://dx.doi.org/10.7479/qs1j-mnjp.

- Melzer, R. & Spelda, J. (2018). The Arthropoda Varia Collection at the Zoologische Staatssammlung München. [Dataset]. Version: 20180705. Data Publisher: Staatliche Naturwissenschaftliche Sammlungen Bayerns – SNSB IT Center, München. http://www.snsb.info/DatabaseClients/ZSMarthrovariacoll/About.html.

- ZFMK Ichthyology Working Group (2018). The Ichthyology collection at the Zoological Research Museum Alexander Koenig. [Dataset]. Version: 2.0. Data Publisher: Zoological Research Museum Koenig - Leibniz Institute for Animal Biodiversity. https://doi.org/10.20363/ZFMK-Coll.Ichthyology-2018-03.

References

1. 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