

Using compact identifiers in project reports

Egon Willighagen 

Published March 29, 2026

Citation

Willighagen, E. (2026). Using compact identifiers in project reports. In *chem-bla-ics*. chem-bla-ics. <https://doi.org/10.59350/re9j2-hk972>

Keywords

Identifier, Semweb, Cito

Abstract

This document describes how you can improve the FAIR-ness of your project report by using compact identifiers. Of course, it can be applied to any other document too, and has been used in, for example, journal articles and online documentation already.

Copyright

Copyright © Egon Willighagen 2026. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

chem-bla-ics

This document describes how you can improve the FAIR-ness of your project report by using compact identifiers. Of course, it can be applied to any other document too, and has been used in, for example, journal articles and online documentation already.

Compact identifiers find a balance between compactness in writing and being a persistent, unique, and global identifier. It “is a string constructed by concatenating a namespace prefix, a separating colon, and a locally unique identifier (LUI)” (doi:10.1038/sdata.2018.29). For example, for proteins it can represent the PDB structure [2gc4](#) as *pdb:2gc4*. There is a clear similarity with the SciCrunch [Research Resource Identifiers](#) (RRIDs) as used by several journals, like [eLife](#) (doi:10.1007/s12021-015-9284-3).

When the prefixes are defined by community standards, then a compact identifier can be resolved. There currently are multiple providers of prefix files (doi:10.1038/sdata.2018.29), including Identifiers.org (doi:10.1093/bioinformatics/btaa864) and Bioregistry (doi:10.1038/s41597-022-01807-3). The Bioregistry has an overview of more than twenty registries of prefixes and their metadata (doi:10.1038/s41597-022-01807-3). The metadata commonly includes information on the URL pattern for each identifier. Often this is more than one pattern, as there may be several databases with information for the same identifier.

It is the URL pattern in the database that allows services to *resolve* the compact identifier into a link to a database. The above registries correspond to three existing *resolvers* that will take a compact identifier as part of a resolver URL and redirect to the database with the record matching that identifier:

- Name-to-Thing (N2T): <https://n2t.net/>
- Identifiers.org: <https://identifiers.org/>
- The Bioregistry: <https://bioregistry.io/>

Each of these URLs can be extended with a compact identifier. For example, a taxon record from the NCBI databases or the PDB entry mentioned earlier:

- <https://bioregistry.io/pdb:2gc4>
- <https://identifiers.org/col:6MB3T> (**col** is the prefix for the Catalogue of Life)

Why use in reports?

Using persistent identifiers is generally accepted as a good practice that benefits science and has been part of the ideas of FAIR data (doi:10.1038/sdata.2016.18) and of Open Science. Compact identifiers make it easy to be precise in reports about what things the reports talk about: they are relatively short but very precise at the same time. also, that has the benefit that they are much easier to reuse than labels of things and concepts that intrinsically have a certain level of uncertainty; a database entry has commonly a very specific meaning.

Examples uses

The use of compact identifiers can be used in two ways. The simplest is to just put the compact identifier as plain text in the document, possibly in parentheses (with the compact identifier highlighted here in bold):

- *This report is only about the experimental data of the human (**NCBITaxon:9606**) cell lines.*

Or:

- *We found that BRCA1 (**ensembl:ENSG00000012048**) played an important role.*

Alternatively, you can add a hyperlink with one of the resolvers, for example, Identifiers.org:

- *We found that BRCA1 ([ensembl:ENSG00000012048](https://identifiers.org/ensembl:ENSG00000012048)) played an important role.*

Compact identifiers for material identifiers

The European Registry of Materials proposes to use the compact identifier for their ERM identifiers (doi:[10.1186/s13321-022-00614-7](https://doi.org/10.1186/s13321-022-00614-7)):

- *For example, the NanoSolveIT project registered a material with the ERM000000001 identifier. The full Uniform Resource Identifier (URI) for this compound is <https://nanocommons.github.io/identifiers/registry#ERM000000001> which is too long to be used in documentation. The corresponding compact identifier **erm:ERM000000001** is easy to use in written material, analogous to the use of Protein Data Bank (PDB) identifiers for proteins in journals.*

Compact identifiers for citation intent annotations

The compact identifier has also been used as the method to include citation intentions in journal articles (doi:[10.1186/s13321-020-00448-1](https://doi.org/10.1186/s13321-020-00448-1), compact identifier here highlighted in bold):

- *We take advantage here of the ability to add notes to full form [...] references in bibliographies. These are referred to as bibnotes. The content of the note will be strictly formatted: it will use the syntax [**cito:usesMethodIn**] and formatted in bold. That is, the bibnote starts with the [character, followed by one of the CiTO types, and ends with the] character. If you wish to provide more than one annotation, you can repeat this syntax, separated by one or more spaces, for example: [**cito:usesMethodIn**] [**cito:citeAsAuthority**].*

Note that in this use, the square brackets and bold typeface are used to make them easier to be recognized. Also, note that this document uses this approach to indicate the intention of why the cited articles are cited.

Conclusion

This document described what the compact identifier is, how it helps linking to online databases, and how they can be used in written reports as plain text, optionally hyperlinked with one of the compact identifier resolvers.

Acknowledgments

I thank [github:tabbassidaloi](#), [github:cthozt](#), and [github:larsgw](#) for their comment on [this GitHub repo](#).