Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data

In many disciplines, data are highly decentralized across thousands of online databases (repositories, registries, and knowledgebases). Wringing value from such databases depends on the discipline of data science and on the humble bricks and mortar that make integration possible; identifiers are a core component of this integration infrastructure. Drawing on our experience and on work by other groups, we outline 10 lessons we have learned about the identifier qualities and best practices that facilitate large-scale data integration. Specifically, we propose actions that identifier practitioners (database providers) should take in the design, provision and reuse of identifiers. We also outline the important considerations for those referencing identifiers in various circumstances, including by authors and data generators. While the importance and relevance of each lesson will vary by context, there is a need for increased awareness about how to avoid and manage common identifier problems, especially those related to persistence and web-accessibility/resolvability. We focus strongly on web-based identifiers in the life sciences; however, the principles are broadly relevant to other disciplines.

General information
State: Published
Organisations: Department of Bio and Health Informatics, High Performance Computing, University of Manchester, Oregon Health and Science University, European Bioinformatics Institute, Wellcome Trust Genome Campus, University of California at Berkeley, Maastricht University, University of Oxford, Helmholtz Zentrum Muenchen German Research Center for Environmental Health, University of California at San Diego, Babraham Institute, European Molecular Biology Laboratory, California Digital Library, Daresbury Laboratory, University of Groningen, Heidelberg Institute for Theoretical Studies, Bern University of Applied Sciences, University of Stellenbosch, Lawrence Berkeley National Laboratory, Leiden University
Number of pages: 18
Publication date: 2017
Peer-reviewed: Yes

Publication information
Journal: P L o S Biology
Volume: 15
Issue number: 6
Article number: e2001414
ISSN (Print): 1544-9173
Ratings:
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Scopus rating (2017): CiteScore 6.79 SJR 4.941 SNIP 1.996
Web of Science (2017): Impact factor 9.163
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 6.01 SJR 5.06 SNIP 1.896
Web of Science (2016): Impact factor 9.797
BFI (2015): BFI-level 2
Scopus rating (2015): CiteScore 6.12 SJR 5.996 SNIP 2.025
Web of Science (2015): Impact factor 8.668
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): CiteScore 7 SJR 6.814 SNIP 2.32
Web of Science (2014): Impact factor 9.343
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): CiteScore 8.47 SJR 8.223 SNIP 2.619