

PREFIX: ACCESSION Compact Identifier Resolution: An EBI/CDL Collaboration

Tim Clark, Massachusetts General Hospital and Harvard Medical School; Ian Fore, National Institutes of Health; Niall Beard, University of Manchester; Jeffrey Grethe, University of California, San Diego; Greg Janeé, California Digital Library; Rafael Jimenez, ELIXIR; Nick Juty, European Bioinformatics Institute; John Kunze, California Digital Library; Julie McMurry, University of California, Santa Barbara; Sarala Wimalaratne, European Bioinformatics Institute

Many science policy studies insist that robust archiving and citation of primary research data should be a prerequisite for publication of any claims which rely upon such data. If these recommendations were to be implemented – along with parallel actions for research resources and software – we would have a far stronger assurance that published results will be reliable, and may be reused.

However, proper data citation requires that citation references are machine resolvable. That is, any reference should be appended with a globally unique, web-resolvable identifier. DOIs fit the bill here, but are used by a small minority of the over 500 specialized biomedical data repositories. The most common approach is to assign locally-unique accession numbers. How do we implement data and resource citation at scale without a massive re-engineering project across all relevant biomedical repositories?

How can locally assigned identifiers become globally resolvable?

We describe here a project to enable PREFIX: ACCESSION based identifier resolution using a common registry of resource namespace prefixes and provider codes. This common registry and resolution approaches have been agreed by two major identifier and metadata providers in Europe and the U.S.A., the Identifiers.org system of the European Bioinformatics Institute (EBI), and the Name-To-Thing resolver at the California Digital Library (CDL).

This is a subproject of the bioCADDIE Data Citation Implementation Pilot, organized as a FORCE11.org community group. We plan for both the EBI and CDL resolver systems to be capable of handling these coordinated compact identifiers by the end of 2016.