

Dashboard



Preview

Publish



Plans

B / ↻ ▾ ...

Posts

All Posts

Add New

Categories

Tags

Copy a Post

Media

Links

Pages

Comments

Feedback

Contact Form

Polls

Appearance

Plugins

Users

Tools

Settings

Collapse menu

Simple standard for linking mappings

A pain point for anyone working in bioinformatics is mapping . Many databases have overlapping content, or provide different identifiers for the same entities, such as genes. Typically every database mints identifiers in its own namespace. This means that the same gene may have multiple different identifiers in different databases (see for example the issues with [SARS-CoV-2 protein identifiers](#)). Anyone doing an analysis that combines data from different databases must do some kind of normalization or mapping.

Automatically mapping is fraught with problems. Simply finding the correct mappings can be a challenge, and for any given pair of databases there can be multiple different mappings from different providers. The provider can be one of the source databases, or a 3rd party provider such as [BridgeDb](#). The ambiguity of a mapping may not be clear: does a mapping denote a many-to-many relationship, or at least a 1:1 relationship? This is particularly challenging when trying to [build a knowledge graph](#) from multiple sources, where we want to merge information about the same entity.

These are a big deal for ontologies too. There is an [entire field of research on alignment/matching](#). In theory ontologies should be able to make the meaning of mappings explicit, [yet somehow we have messed this up](#), leading to multiple alternate ways to say the same thing (OWL logical axioms, SKOS, and classic loose bioinformatics 'dbxrefs').

In the [Open Bio Ontologies](#) project we attempted to avoid the mapping