

# URI Tool

Description	Merge URI in a VIVO from a web interface
Type	Data management
Status	AVAILABLE
Owner	
Language	perl
Team	
Location	<a href="https://github.com/vsposato/uri-tool">https://github.com/vsposato/uri-tool</a>
License	

## URI Tool

The VIVO Cornell team has developed a tool that they use for a variety of data cleaning.

- How many URIs do we have that represent organizations such as Cornell University? Can we merge them into one?
- Similarly for Organizations, People, Journals.

The URI tool is built mostly from perl scripts, and works by making HTTP requests against VIVO, just as a user would. These are invoked by CGI scripts, in an Apache server.

- The tool now includes a simple menu to generate the comparison sets of people, organizations, journals, and grants, using an algorithm that does string distance calculations after removing differences in capitalization or spacing.
- To some extent, the URI tool is being converted to using a SPARQL endpoint instead of HTTP.
- The URI tool is dependent on the VIVO pages and forms, and must be revised to deal with new releases of VIVO.

URI tool v2.0 required considerable modification to adjust for the use of multiple data graphs in newer VIVO releases, but these changes make it more useful for sites like the USDA that have ingested data in different named graphs and only want changes to be allowed in a specific graph.

The URI tool is not yet easily used by other sites while we ferret out remaining assumptions based on the Cornell environment – thanks to Nicholas Rejack of Florida for helping with that process and in uncovering other bugs. Additional effort will be needed to make it more generally usable.

- It was developed for in-house use, and work continues to be oriented around day-to-day problems encountered by the VIVO Cornell team.
- Installation instructions are needed. What do I do with this pile of perl code?
- A version of URI tool is on GitHub, but presently without support.
  - Some code is specific to Cornell's ontology and data sources. How can those pieces be identified and modified by other users