Managing Data Packages

- Overview
- Add a data package
- Update a data package
- Delete a data package
- Available Data Packages

Overview

Data packages are sets of data represented using VIVO RDF in one of the supported VIVO RDF file formats – Turtle (.ttl), Triples (.nt), Notation3 (.n3) or RDF-XML (.rdf) Data packages are typically produced as semi-static – they can be loaded into VIVO and updated as needed. Data packages typically deliver statements about entities outside the management of the particular VIVO.

VIVO manages data packages by creating a new graph for each package, containing the asserted triples for the data package and named with the name of the data package file. The VIVO inferencer creates inferred triples for the data packages and stored the inferred triples in the inference graph. When changes are made to the data package file, the VIVO inferencer must be run to bring the inference graph up to date with the changes made in the asserted graph for the data package.

An example of a data package would be the Grid data, representing the research organizations of the world. This data set, maintained by Digital Science, contains more than 65,000 university, research institutes, funding agencies and other organizations involved in research across the world. The data set contains the official name and alternate names as well as abbreviations of names of the each organization, its geographic location, its type, date of founding, parent, child and affiliated organizations, as well as persistent identifiers for the organization. The data is available as a data package for VIVO at https://github.com/openvivo/grid-rdf

Add a data package

To add a data package to VIVO,

1. Place a copy of the data package file in `vivo/home/rdf/abox/filegraph`
2. Restart Tomcat. VIVO will add a new graph to the triple store containing the asserted triples in the data package file. See Graph Reference for additional detail. The VIVO inferencer will infer additional triples regarding the data package and add those triples to the vitro-kb-info graph. Again see Graph Reference. Note: the inferencer may take quite awhile to complete. Adding a package with tens of thousands of new entities, each with dozens of attributes may take hours to reinference.

Update a data package

To update a data package:

1. Place a copy of the updated data package file in `vivo/home/rdf/abox/filegraph`
2. Restart Tomcat. VIVO will compare the contents of the triple store with the contents of the data package file, and update the triples in the associate graph as needed. VIVO will then reinference the triple store. Note: the inferencer may take quite awhile to complete.

Delete a data package

To delete a data package:

1. Remove the data package file from `vivo/home/rdf/abox/filegraph`
2. Restart Tomcat. VIVO will detect that the file is no longer present and remove the associated graph. The inferencer will be run and triples in the inference graph associated with the deleted data package file will be removed from the inference graph.

Available Data Packages

Data packages are available at the following locations:

1. Research organizations of the world. From http://gid.ac Available as CC-0 data. https://github.com/openvivo/grid-rdf
2. Journals of the world. Compiled from CrossRef and NIH PubMed. More than 40K journals, each with title and ISSN. https://github.com/OpenVIVO/OpenVIVOjournals
3. Dates. Dates with simple URI, known URI. Avoid creating multiple date entities for the same date. Link all references to a date to a single date entity. https://github.com/OpenVIVO/date-rdf
4. Cities of the United States. Data for all cities in the United States with population 100K or more. Includes lat/long. https://github.com/mconlon17/vivo-add-cities