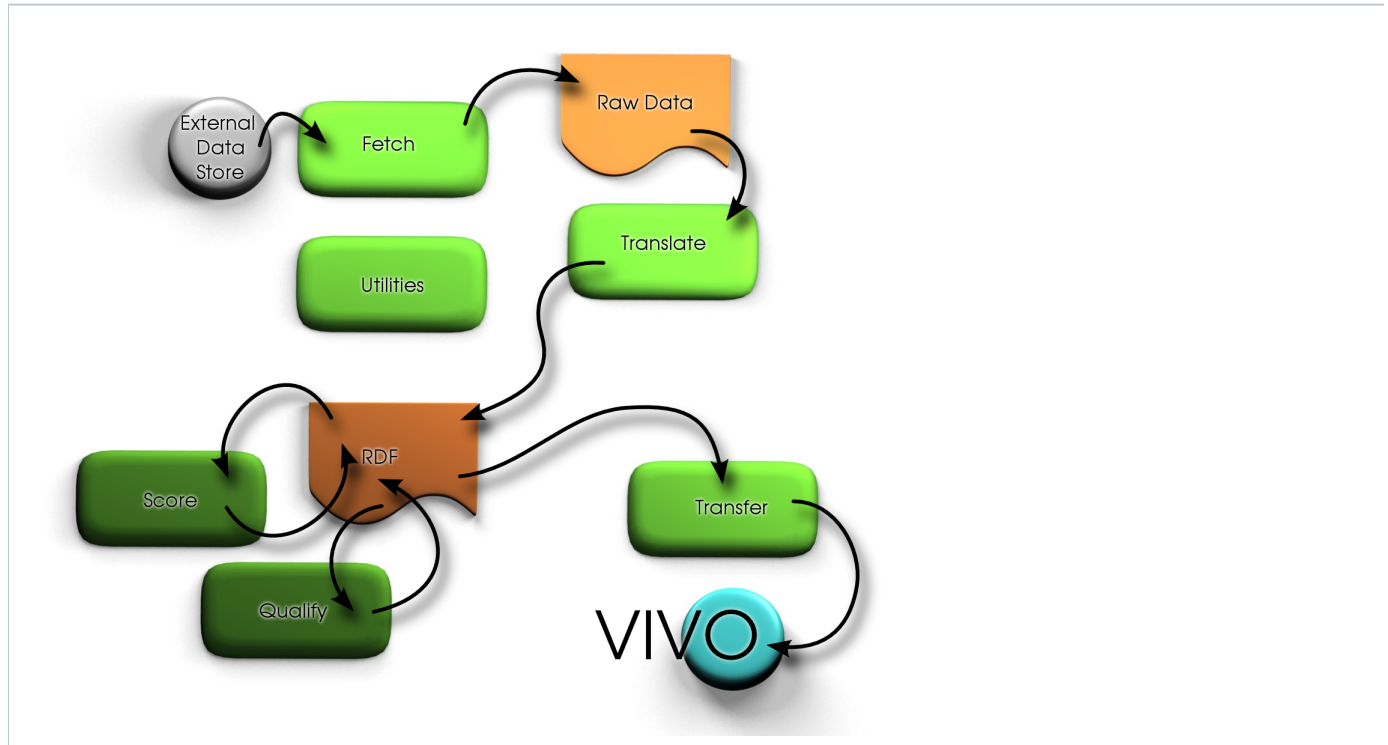


Fetch

Overview

The first step of a typical harvest is the get you data from your target source. We call this the Fetch. For example, let us suppose we have a VIVO installation containing researchers at our university, and we want to harvest from [Pubmed](#) information on publications written by researchers at our university. In this case we would use Harvester's [PubmedFetch](#) tool to send a query off to Pubmed, which will return the results of that query to us in its own XML format. The harvesters [Fetch](#) package ([org.vivoweb.harvester.fetch](#)) contains various methods for retrieving data from external data sources, including as CSV files, as JSON, through JDBC calls, or via OAI Harvest. Fetch is the first stage in a harvest. It accesses the external source and orders the information in a simple RDF/XML file. The configuration task file determines properties of the fetch (ex. for the [JDBCFetch](#) it includes queries, table names, where clauses, fields, delimiters).

Process Diagram



1. External Data Source - This is the foreign source
2. Fetch - Retrieves data from foreign source
3. Raw Data - A simple database or as simple XML
4. Translate - Turns the raw data into Ontological RDF
5. RDF - RDF models which can be dumped into RDF/XML.
6. [Score](#) - First find similarities and rate them, second determine and apply matches based on a threshold of difference.
7. [Qualify](#) - Changes any unmatched data
8. [Transfer](#) (Update) - move into a vivo model (through an update process if possible.)
9. Vivo - Final model in RDF visible from the webapp. |

Tools

- [OAIFetch](#): Tool for fetching from [OAI](#) repositories
- NIH Fetches
 - [PubmedFetch](#)
 - [NLMJournalFetch](#)
- RDB Fetches
 - [JDBCFetch](#)
 - [D2RMapFetch](#)