A Generalizable, XSLT Based RDF Ingest Example

Joseph McEnerney, Jon Corson-Rikert, and Chris Westling

Cornell University Library, Albert R. Mann Library Information Technology Services, Ithaca, NY 14850
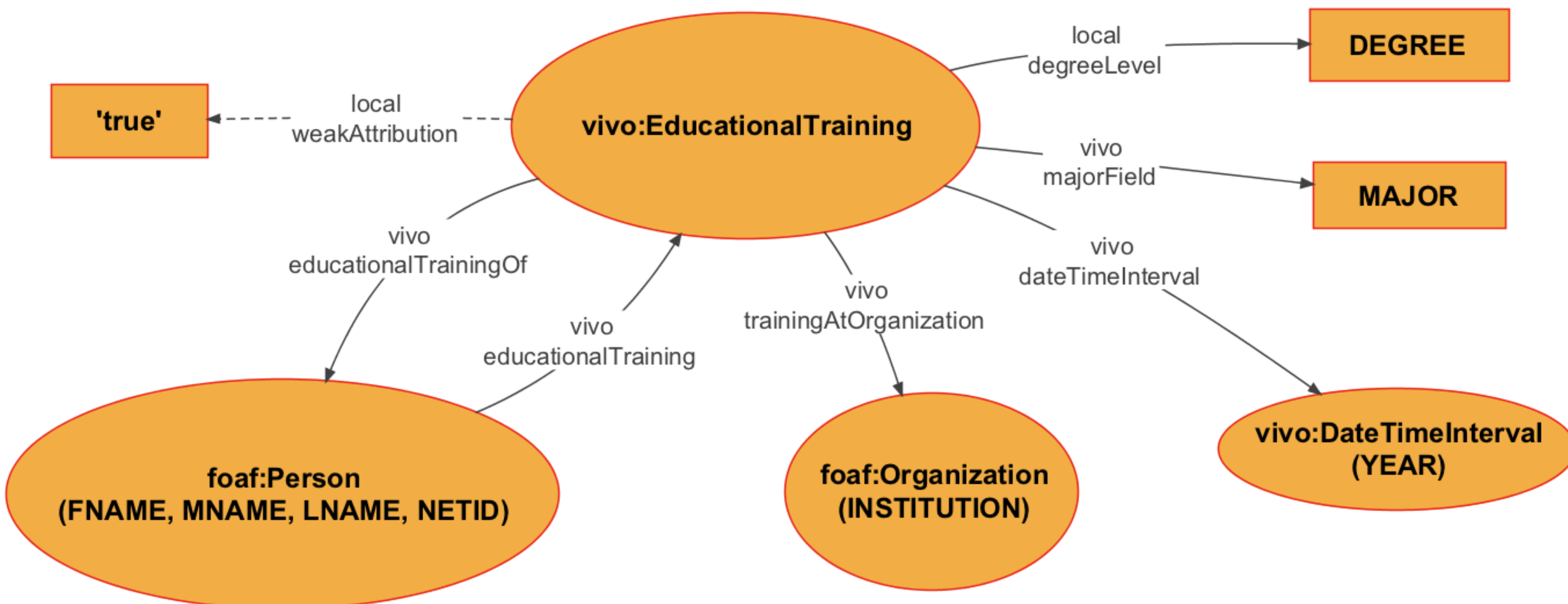
Abstract

The purpose of this example is to illustrate XSLT based techniques that have been used successfully to ingest data from more than a dozen sources into VIVO at Cornell. Instead of a simplified 'toy' example, the source data used will display many of the data quality problems often found in practice. The goal is to transform this source data into RDF that conforms to a specific data model and can be loaded into a VIVO instance. The example is based on educational credentials and the central objects in the RDF data model are instances of the class **vivo:EducationalTraining**. In addition, we want to prevent duplication of Person and Organization RDF. Experience has taught us that this XSLT transform methodology performs well in terms of processing time and is scalable to tens of thousands of source data records.

Tutorial and full download on VIVO wiki at https://wiki.duraspace.org/x/4dwQAg

Target RDF Classes and Properties

Sample input data

Typical data issues

Existing persons

Mapping URIs onto a group of groups

A - Associate
B - Bachelor
M - Master
() - No Netid
Un - nth URI