

FGDC & HFA to BIBFRAME Conversion

- [HFA](#)
 - [Completed Work](#)
 - [Analysis/Modeling](#)
 - [Linked Data Creation](#)
 - [Tool Exploration / Requirements Definition](#)
 - [Collaboration](#)
 - [Community Engagement](#)
 - [Software development](#)
- [FGDC](#)
 - [Completed Work](#)
 - [Analysis/Modeling](#)
 - [Data conversion](#)
- [BIBFRAME Conversion](#)
 - [Code](#)

This work was performed as part of the project to pilot linked data conversion, publication, and visualization of Harvard Geospatial Library metadata and Harvard Film Archive metadata.

HFA

The HFA (Harvard Film Archive) project created linked data descriptions for a set of moving image materials by women directors--work that has previously been underexposed and in many cases is unique to the HFA. The overall HFA project is described at [Harvard Film Archive](#)

Completed Work

Analysis/Modeling

[See Harvard LD4L Labs wiki for documents](#)

Linked Data Creation

HFA subjects & genre mapping to LCSH, Getty AAT, and FAST URIs

Converted a full snapshot of the Harvard Film Archive metadata to the target Moving Image linked data ontology (https://github.com/HLITS/LD4L_Film_Ontology)

Tool Exploration / Requirements Definition

Vitrolib custom form for annotations
Vitrolib lookup specs for ISNI

Collaboration

Discussions with Library of Congress BIBFRAME pilot participants
Pattern documents for LD4P/LD4L Labs BIBFRAME extension group
LD4P/LD4L ontology extension meeting

Community Engagement

Interviews with Harvard Film Archive and Northeast Historic Film staff

Software development

HFA data originated in FilemakerPro format. A java program was created using FilemakerPro database drivers to extract data from two relevant database tables. This data was output to XML format in a large single file for processing by the BIBFRAME converter described below.

FGDC

Harvard created native Linked Data descriptions for a selection of library cartographic resources including printed maps, atlases, digital geospatial datasets, and other cartographic information resources. Together with LD4L-Labs partners, Harvard specifically converted a set of Harvard Geospatial Library metadata records into linked data descriptions. The overall Harvard Cartographic Materials project is described at [Harvard Cartographic Materials](#)

Completed Work

Analysis/Modeling

- Completed field mappings for geospatial metadata conversion: FGDC to bibliotek-o + Geospatial and Cartographic Resources Ontology (GCRO) extensions

Data conversion

- Converted a set of 8,800 [Harvard Geospatial Library](#) and 5,100 [Stanford Earthworks](#) geospatial metadata records to the target geospatial LOD ontology
- Reconciled agent names, topic keywords, places of publication, and place keywords to linked data entities in the converted data sets
- Data loaded into Harvard geospatial data instance of VitroLib

FGDC metadata from Harvard Geospatial Library originated in XML format so no converting of this format was necessary prior to processing by the BIBFRAME converter described below.

BIBFRAME Conversion

The [bib2lod project](#) (see also [MARC -> BIBFRAME Converter Framework](#)) was used as base code for converting both the HFA and FGDC XML data. An extension of this base code was made for each of these input formats. Custom code for each project was necessary due to significant difference between the datapoints available for each format. The XML input for each of these formats was converted to RDF output. This RDF output was imported into a [VitroLib web application](#), one for each format type. During the development process extensive test cases were written for each format type and vetted a domain expert.

Code

<https://github.com/ld4l-labs/fgdc2lod>

<https://github.com/ld4l-labs/hfa2lod>