Metadata Application Profile for the Wikidata for Digital Preservation Portal

Katherine Thornton

July 6, 2018

ABSTRACT

This report describes a metadata application profile (MAP) for the WikiDP Portal. This MAP lists the set of metadata elements used by the application along with requirements for how to use them.

1 Organization of this MAP

This MAP describes the properties related to software and the properties related to file formats.
    In the two tables of properties the columns represent:

- **Property**- the English label for this property in Wikidata

- **PID**- the Wikidata Property ID

- **Expected Value**- Wikidata data type required for the value of this property

- **Obligation**- whether or not this property is required, and the number of times the property is expected to be used

The definitions for all properties discussed in sections 2 and 4 are provided in section 5.

2 File Formats

Property	PID	Expected Value	Obligation
file extension	P1195	string	0,n
media type	P1163	string	0,n
developer	P178	item	0,n
PRONOM file format identifier	P2748	external id string	0,1
LoCFDD	P3266	external id string	0,1
File Format wiki page ID	P3381	external id string	0,1
based on	P144	item	0,n
file format identification pattern	P4152	string	0,n
described at URL	P973	URL	0,n
described by source	P1343	item	0,n
software version	P348	string	0,1
publication date	P577	date	0,1

3 Software

The Wikidata community added more than 15,000 software items to the knowledge base in the first half of 2018. This is a very encouraging number that reflects the strength of interest in this area.

Property	PID	Data Type	Obligation
developer	P178	item	0,n
operating system	P306	item	0,n
publication date	P577	date	0,n
readable file format	P1072	item	0,n
writable file format	P1073	item	0,n
based on	P144	item	0,n
programming language	P277	string	0,n
described at URL	P973	URL	0,n
described by source	P1343	item	0,n
software version	P348	string	0,1

4 MAP Usage

This MAP is the basis for property checklists feature of WikiDP. We can use this MAP in collaborative data modeling discussions in the Wikidata community. We can use the MAP as a tool to gather feedback from stakeholders. The MAP will be encoded as ShEx, including a shape for software and a shape for file formats. We will then be able to validate entity data from Wikidata to test for conformance to these shapes. We have the option to write additional shapes to extend this MAP in order to support additional use cases.

5 Property Definitions

These are the current descriptions for these properties in Wikidata.

5.1 developer P178

organization or person that developed the item

5.2 operating system P306

operating system (OS) on which a software works or the OS installed on hardware

5.3 publication date P577

date or point in time when a work was first published or released

5.4 software version P348

version(s) of the software, current and past

5.5 readable file format P1072

file format a program can open and read

5.6 writable file format P1073

file format a program can create and/or write to

5.7 based on P144

the work(s) used as the basis for subject item

5.8 file format identification pattern P4152

pattern or string which is used to identify a file as having a particular known format

5.9   described at URL P973

item is described at the following URL

5.10   described by source P1343

dictionary, encyclopaedia, etc. where this item is described

5.11   software version P348

with datatype string that is not an external identifier, for software

5.12   programming language P277

the programming language(s) in which the software is developed

5.13   File Format wiki page ID P3381

identifier on the "Just Solve the File Format Problem"/File Format(s) wiki

5.14   LoCFDD P3266

Library of Congress Format Description Document ID

5.15   PRONOM file format identifier P2748

identifier (PUID) for a file format in the technical registry PRONOM

5.16   media type P1163

IANA-registered identifier for a file type

5.17   file extension P1195

identifier for a file format (e.g. txt for a text file) used as suffix to the file name