Fennica RDF data model

Introduction

This document specifies the RDF data model used for the Finnish national bibliography Fennica Linked Data set, which consists of approximately 40 million RDF triples generated from 1 million MARC bibliographic records and auxiliary sources. The data model is heavily based on Schema.org, including the bibliographic extensions. The OCLC WorldCat Linked Data model has been used as a reference whenever possible. The separation between Works and Instances is modelled according to BIBFRAME 2.0.

This document specifies the available entity types, their relationships and properties.

Publishing Fennica as Linked Data is a work in progress. For more detailed information about current issues, see the open issues on the bib-rdf-pipeline GitHub project that implements the conversion of data from MARC records and auxiliary sources into the published RDF.

Accessing the data

The data set is currently available as:

- a SPARQL endpoint on http://data.nationallibrary.fi/bib/sparql with the following named graphs:
  - <http://data.nationallibrary.fi/bib/fennica> - Fennica bibliographic data
  - <http://www.yso.fi/onto/yso/> - YSO concepts
  - <http://www.yso.fi/onto/yso-paikat/> - YSO places
- downloadable data dumps at http://data.nationallibrary.fi/download/ in HDT and gzipped N-Triples formats
- Linked Data access (HTML, RDF/XML, JSON-LD, Turtle, N-Triples) via URI dereferencing
- **TODO** Linked Data Fragments access

URI patterns and stability

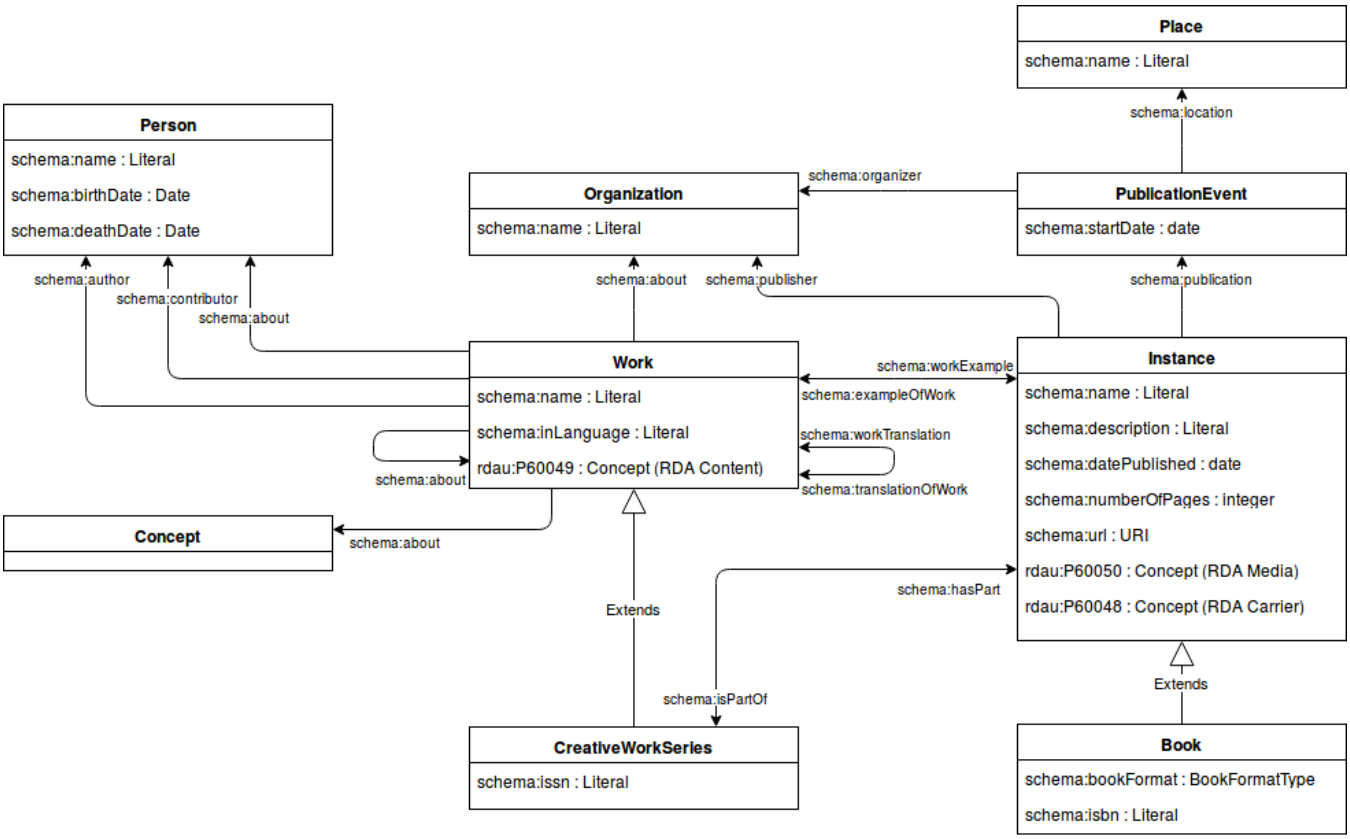
The data set currently uses URIs of the form `http://urn.fi/URN:NBN:fi:bib:me:`*`Tnnnnnnnnxx`* where

- *T* is a single capital letter representing the entity type (see below)
- *nnnnnnnn* is the numeric identifier of the MARC record where the entity originated
- *xx* is a two-digit sequence number ensuring uniqueness of entities of the same type from the same record

The URI patterns are in draft status and may still change. Some entities are currently represented only as blank nodes in the RDF graph, but may later be given URIs.

Entity types

Overview



This diagram shows the main entity types and their relationships as a UML class diagram.

Work

URI pattern: **`http://urn.fi/URN:NBN:fi:bib:me:W`***nnnnnnnnnxx*

Number of entities: approx. 930 000 Works, of which 450 000 are Series and 270 000 are Periodicals.

The Work entity type represents an abstract creative work, very similar to the BIBFRAME 2.0 notion of Work. Derived works such as translations are modelled as separate Work entities. In FRBR terms, this Work entity is a combination of a FRBR Work and Expression.

Field name	RDF property	Expected Value / Range	Definition	Cardinality	Data Quality Notes
Type	rdf:type	Class	Type of resource. Always both schema:CreativeWork and bf:Work. May also have the more specific types schema:CreativeWorkSeries and schema:Periodical (see below).	2..*	
Title	schema:name	Literal	Title of work	1..*	
Subject	schema:about	skos:Concept, Work, Person, Organization or Literal	Subject matter of the work	0..*	YSO concepts are used whenever possible. Literal values are used in cases where no entity was found matching the label.
Has instance	schema:workExample	Instance	Example/instance/realization/derivation of the concept of this work. eg. The paperback edition, first edition, or eBook.	0..*	
Language	schema:inLanguage	Literal (language code)	Language of the work, expressed as a language code following BCP 47 rules (i.e. ISO 639-1 or 639-3 code)	0..*	Needs cleanup. There are a few bad values such as numeric values
Author	schema:author	Person or Organization	The main author of this work	0..1	
Contributor	schema:contributor	Person or Organization	A secondary contributor to the work	0..*	
Content type	rdau:P60049	skos:Concept from RDA Content Type vocabulary		0..1	Should generally be available for most Works, but in practice, missing for some of them.

Is part of (series)	schema: isPartOf	Series	The series which this work is a part of.	0..*	
Is translation of	schema: translationOfWork	Work	The work that this work has been translated from. Inverse of "Has translation"	0..*	
Has translation	schema: workTranslation	Work	A work that is a translation of the content of this work. Inverse of "Is translation of"	0..*	

Periodical

The Periodical entity type is a sub-type of Work and represents a publication series.

Field name	RDF property	Expected Value / Range	Definition	Cardinality	Data Quality Notes
Type	rdf:type	Class	Always schema:Periodical	1	
Has part	schema: hasPart	Work	A work that is included in this series	1..*	
ISSN	schema:issn	Literal (ISSN code)	The International Standard Serial Number (ISSN) that identifies this periodical	0..1	

Instance

URI pattern: **`http://urn.fi/URN:NBN:fi:bib:me:I`***nnnnnnnnnxx*

Number of entities: approx. 1.1 million

The Instance entity type represents a specific edition (e.g. a hardcover book or a specific DVD release of a film) of a Work. It is similar to the BIBFRAME 2.0 notion of Instance. In FRBR terms, it is similar to a FRBR Manifestation.

Field name	RDF property	Expected Value / Range	Definition	Cardinality	Data Quality Notes
Type	rdf:type	Class	Type of resource. Always both schema:CreativeWork and bf: Instance. May have an additional, more specific type (see below).	2..*	The more specific types such as schema: Book are for the most part not implemented yet.
Title	schema: name	Literal	Title of instance	1..*	
Description	schema: description	Literal	A textual description of the instance	0..*	This field is used to represent many kinds of notes extracted from the bibliographic record. Some of these would probably deserve their own fields or a more structural way of expressing the information.
Is instance of work	schema: exampleOfWork	Work	A work that this work is an example/instance/realization /derivation of.	1	
Date published	schema: datePublished	Literal (date value)	Date of first publication	1	Needs cleanup. May contain brackets or other expressions indicating uncertainty
Publication	schema: publication	PublicationEvent	A publication event of the instance	0..1	
Publisher	schema: publisher	Organization	The publisher of the instance	0..1	
Media type	rdau: P60050	skos: Concept from RDA Media type vocabulary	Relates a resource to a categorization reflecting a general type of intermediation device required to view, play, run, etc., the content of a resource.	0..1	Should generally be available for all Instances, but in practice, missing from some.
Carrier type	rdau: P60048	skos: Concept from RDA Carrier type vocabulary	Relates a resource to a categorization reflecting a format of a storage medium and housing of a carrier in combination with a type of intermediation device required to view, play, run, etc., the content of a resource.	0..1	Should generally be available for all Instances, but in practice, missing from some.

Number of pages	schema: numberOfPages	Literal (integer)	The number of pages in the book	0..1	Needs cleanup. Often the values are structured page counts (including Roman numerals), not plain integers.
URL	schema:url	URI	The URL where the electronic version is available	0..1	

Book

The Instance sub-type Book represents a book edition, e.g. hardcover, paperback or electronic book.

Field name	RDF property	Expected Value / Range	Definition	Cardinality	Data Quality Notes
Type	rdf:type	Class	Always schema:Book	1	
Book format	schema: bookFormat	schema: BookFormatType	The value schema:EBook is used for electronic books. Other values are currently not used.	0..1	
ISBN	schema:isbn	Literal (ISBN code)	The ISBN code of the book	0..*	

Person

URI pattern: **`http://urn.fi/URN:NBN:fi:bib:me:P`***nnnnnnnnnxx*

Number of entities: approx. 1.3 million. Note that there is a lot of duplication within these entities. **TODO**: reconcile the Person entities with the person authority file.

The Person entity type represents a human being (e.g. author, contributor or subject of a work). A person may be a pseudonym or fictitious.

Field name	RDF property	Expected Value / Range	Definition	Cardinality	Data Quality Notes
Type	rdf:type	Class	Always schema:Person	1	
Name	schema:name	Literal	The name of the person	1	
Birth date	schema: birthDate	Literal	The year of birth of the person	0..1	Available only for deceased persons due to privacy concerns.
Death date	schema: deathDate	Literal	The year of death of the person	0..1	Available only for deceased persons.

Organization

URI pattern: **`http://urn.fi/URN:NBN:fi:bib:me:O`***nnnnnnnnnxx* **`or`** `blank node` **`or`** `CN identifier (TBD)`

The Organization entity type represents an organization (e.g. publisher of a work).

Field name	RDF property	Expected Value / Range	Definition	Cardinality	Data Quality Notes
Type	rdf:type	Class	Always schema:Organization	1	
Name	schema:name	Literal	The name of the organization	1	

Place

URI pattern: blank nodes only

Number of entities: approx. 1.0 million. Note that there is a lot of duplication within these entities. **TODO**: reconcile the Place entities with YSO places.

The Place entity type represents a physical place (e.g. a country or a city).

Field name	RDF property	Expected Value / Range	Definition	Cardinality	Data Quality Notes
Type	rdf:type	Class	Always schema:Place	1	
Name	schema:name	Literal	The name of the place	1	Needs cleanup. May contain abbreviations, inflected forms etc.

PublicationEvent

URI pattern: blank nodes only

Number of entities: approx. 980 000

The PublicationEvent entity type represents the event when an instance of a work was published.

Field name	RDF property	Expected Value / Range	Definition	Cardinality	Data Quality Notes
Type	rdf:type	Class	Always schema:PublicationEvent	1	
Organizer	schema:organizer	Organization	The publisher	1	
Location	schema:location	Place	The place of publication		
Date	schema:startDate	Literal (date string)	The date of publication		