

Report: Rådata nå!

Rurik Thomas Greenall, Lene Elisabeth Bertheussen, Ole Husby, Keith Alexander, Roger Aunet & Marit Brodshaug

[@brinxmat](#) radatana means "while crying" in [#marathi](#)

Abhijeet Shinde's Twitter status for 2010-11-18
<http://twitter.com/abhishinde/status/5244169471664129>

Partners:



Funded by:



This work is licensed under the Creative Commons Attribution 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0/> or send a letter to Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.

Executive summary: NTNU University Library and BIBSYS have, in a project funded by the Norwegian archive, library and museum authority, produced a data set containing over one million names. The data set is structured as linked data and is available both for bulk download in addition to search using SPARQL.

Table of Contents

[0.1 Structure](#)

[1.0 Introduction](#)

[1.1 About the project participants](#)

[1.2 Rådata nå?](#)

[1.1.1 Linked data](#)

[1.1.2 NTNU University Library, BIBSYS and linked data](#)

[1.1.3 The raw data: BIBSYS' authority file](#)

[2.0 Operational project](#)

[2.1 Operational phase 1: training together](#)

[2.1.1 Finding training](#)

[2.1.2 Training days](#)

[2.2 Operational phase 2: creating data structures](#)

[2.2.1 RDF vocabularies](#)

[2.2.2 Unique information for the different authority types](#)

[2.2.2.1 Personal name authorities in 100:](#)

[2.2.2.2 Corporation names in 110](#)

[2.2.2.3 Conference names in 111](#)

[2.2.3 Properties that occur in every record](#)

[2.2.4 Explanatory links.](#)

[2.2.5 Example](#)

[2.2.6 Handover](#)

[2.3 Creating a linked data representation from authority file data](#)

[2.3.1 Initial considerations](#)

[2.3.2 Creating the basic data set](#)

[2.3.2.1 Creating a functional platform for working with linked data](#)

[2.3.2.2 Triplestore](#)

[2.3.2.3 JOSEKI](#)

[2.3.2.4 Web application](#)

[2.3.2.4 Linking data with DBpedia, the National library of Germany and VIAF](#)

[2.3.2.5 Dereferenceable URI](#)

[2.3.2.6 RDF Example](#)

[2.3.2.7 Huge number of triples](#)

[3.0 The data](#)

[4.0 Conclusion](#)

[5.0 Glossary](#)

[6.0 Appendices](#)

[6.1 Appendix A](#)

[6.2 Appendix B](#)

[6.3 Appendix C](#)

[6.4 Appendix D](#)

[6.5 Appendix E](#)

7.0 References

0.0 About this document

This is documentation of the work done on project Rådata nå! It covers the background for the project, the methods used in the project and some conclusions and recommendations for future work.

This document is aimed at general readers from within the field of library science, therefore some understanding of concepts related to authority files and cataloguing rules is assumed. Terms in general and specifically those related to linked data are explained in the glossary.

0.1 Structure

Section 1 provides information about the project: why the project was initiated, who participated in the project, its funding and the products that were expected from the project. Section 2 provides an overview of the operational part of the project and how the different parts of this were carried out. Section 3 provides information related to the project's results. Section 4 provides concluding remarks.

1.0 Introduction

In 2009, NTNU University Library and BIBSYS applied to the Norwegian Archive, Library and Museum Authority for funding for a joint, one-year project that aimed to produce a linked data representation of the BIBSYS authority file. Funding was awarded in Q1 2010, and work began later that year, culminating in a completed linked data representation of the authority file in Q1 2011.

1.1 About the project participants

The project was designed to promote strategic and operational collaboration between NTNU University Library and BIBSYS, which are separate institutions (administered within the framework of the Norwegian University of Science and Technology in Trondheim, Norway).

BIBSYS “is a key supplier of products and services for higher educational institutions, other research institutions in Norway, public administrative institutions and the National Library of Norway.” [\[3\]](#) and is administered by the Norwegian Ministry of Education and Research.

NTNU University Library is the library of the Norwegian University of Science and Technology.

NTNU University Library is project owner with Arnt Dragsten (Head of Section, Development and Co-ordinator), and Rurik Thomas Greenall as contact person in the funding application. (See Appendix A). Greenall was also appointed project manager for the operative part of the work carried out by NTNU University Library, with Lene Elisabeth Bertheussen and Ole Husby as project participants.

BIBSYS appointed Marit Brodshaug as project manager and contact person for the operative portion of the project. Further, Roger Aunet, Gregor Gabriel and Mari Ingberg participated in the project for BIBSYS.

1.2 Rådata nå?

“Rådata nå!” is Norwegian for “raw data now”, which is the current mantra of Tim Berners-Lee [\[1\]](#), it signifies a move to provide the vast wealth of raw data that is being created around the world to be open for review and use by the public. Berners-Lee’s talk [\[1\]](#) deals with the philosophy of sharing data for use and re-use and the mechanism for doing it: Linked data.

1.1.1 Linked data

Linked data is a way of publishing data that allows users (whether humans or machines) to navigate “the web of data”. The basic outline for linked data can be found in [\[2\]](#), and can be stated as follows:

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names

3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
4. Include links to other URIs. so that they can discover more things

Berners-Lee [\[2\]](#)

The reasons both BIBSYS and NTNU University Library are interested in linked data include that they have a common interest in using semantic technologies and a belief that linked data is an appropriate way to represent bibliographic metadata. Additionally, it is held that opening data for use in this way is not only a way of making the valuable data created by the libraries in the BIBSYS consortium available for use, but also a necessary step in making library metadata relevant in the modern world.

1.1.2 NTNU University Library, BIBSYS and linked data

At NTNU University Library, the trend towards openness and use of semantic technologies began in 2007 with the creation of the project UBiT2010, while BIBSYS had a number of staff who expressed an interest in semantic technologies.

The perceived benefits of linked data to the running of both institutions are great, in terms of modernization, cost-efficiency and appropriateness of the tools used for specific tasks. The rich nature of semantic metadata combined with the number of developers working with linked data points to this being an area of specific interest to those working with bibliographic data, and one which ought not to be ignored.

1.1.3 The raw data: BIBSYS' authority file

BIBSYS' authority file was chosen as an interesting data set for presentation in the web of data following discussions at LIBRIS during iWorkshop Sthlm! (2009-06-08 – 2009-06-09). It was decided that personal names were a particularly interesting starting point for users, and a particularly interesting point to link to. These factors — as well as the unavailability of this sort of data due to constraints of the systems in use, licensing issues and formats common within library IT — convinced NTNU University Library that the authority file was not only of interest for use in their own projects, but also to the general public. BIBSYS was approached regarding the issue of publishing the data as linked data and the project application (Appendix A) was the outcome.

1.3 Requirements

In the award letter (Appendix B), the Norwegian Archive, Library and Museum Authority stipulated the following requirements:

- The project should be completed in line with the description and plan in the application
- The project should co-ordinate efforts with the project PODE (contact person: Anne-Lena Westrum, Oslo Public Library) as regards contact with LIBRIS
- A complete overview of the project accounts (Appendix C) should be made available

- A progress report due, 2010-10-01, with expected date of completion
- A final report

In order to fulfill the requirements stipulated here, several different measures were taken:

- NTNU University Library and BIBSYS drafted and signed a contract regarding their work (Appendix D); the division of labour was specified in this, and its contents were verified as being acceptable to the Norwegian Archive, Library and Museum Authority (p.c. Arnt Dragsten, NTNU University Library – Anders Olsson, Norwegian Archive, Library and Museum Authority).
- Due to difficulties in finding a point of contact at LIBRIS and issues regarding the non-transferability of methods used by LIBRIS to the Norwegian context (LIBRIS use a MARCXML transformation methodology, whereas BIBSYS use a pure Java workflow), the collaboration and co-ordination between PODE and NTNU University Library was arranged as joint training in linked data.
- The Norwegian University of Science and Technology project management and accounting system was used to manage details related to the funding of the project. The report from the accounting system can be found in Appendix C.
- It was not possible to provide a progress report to the Norwegian Archive, Library and Museum Authority because this organization ceased to exist from 2010-07-01 [4] and it was unclear to whom the report should be addressed. Additionally, efforts were made to publish the project's current status in an open blog (<http://radatana.wordpress.com>), so that the progress report could be reviewed there.
- This document is the final report.

The collaborating organizations expected two things:

- A usable authority file structured as linked open data
- A report outlining the work that has been done

The project has responded to these expectations in the following ways:

- An authority file structured as linked open data has been produced
- This document outlines the work that has been done

2.0 Operational project

NTNU University Library and BIBSYS met in May 2010 to discuss a number of topics [5]. The topics discussed included:

1. the relationship of Rådata nå! to BIBSYS' work on the national authority file
2. details of the collaboration
3. participation in training

(1) was seen to be potentially problematic due to extensive overlap between work on two projects related to personal name authority data. Closer inspection, however, showed that the two projects were of quite different natures: Rådata nå! dealt with linked data representations of personal names, whereas the national authority file deals with controlled authorities. The projects were, in fact, seen as complementary because Rådata nå! data provides a way of discovering controlled authorities.

Under (2), it was agreed that the project should be completed before end-of-year 2010 and that the labour should be divided as follows:

- BIBSYS provides example data for analysis
- NTNU University Library analyses the data and provides a data structure
- BIBSYS structures data according to the provided data structure

(3) relates to training for project participants; training was planned for May 2010, but was not carried out until June 2010 for logistical reasons.

2.1 Operational phase 1: training together

It was deemed necessary to receive training in the proper and effective use of linked data as the project group wanted to know how things are actually done in practice. In the case of linked data, standards relate to the form and format of information, and these were well enough understood by the majority of the group because they were conversant with RDF [6] from previous work; what was not, however, entirely well understood was what should be linked to and how much information should be expressed. This kind of information was needed in order to achieve the project goals as effectively as possible.

2.1.1 Finding training

Various training options were looked at, asking within the linked data community as to who provided training. Given the small number of offerings — several of which were web-based and therefore less interesting to the project because the working groups needed answers to specific questions related to the project — and similarity in price, services from Talis were chosen because the offering was provision of an experienced linked data hacker who also had experience with library applications.

2.1.2 Training days

Talis sent Keith Alexander initially to provide a two-day (2010-06-14–2010-06-15) basic training session in which the group worked together in order to create a common set of terminology which would help communication internally within the project, and additionally with Pode who were invited to join this session. Because the number of participants could not exceed eight, it was decided that the three project participants from NTNU University Library should attend along with two from BIBSYS and two from Pode.

During this first session, the basics of linked data and techniques for finding appropriate design patterns for use with different kinds of data were learnt. All of this was achieved through a mixture of information, examples and practical exercises. In addition, the group received a practical introduction to working with existing data sets using the RDF query language, SPARQL [\[7\]](#).

After the training session was completed, the NTNU University Library team worked together to look again at our data in light of our newly gained knowledge. Ole Husby provided an overview of the authority file format of BIBSYS-MARC in relation to MARC21 [\[8\]](#) and BIBSYS-MARC XML [\[9\]](#) while Greenall & Bertheussen worked on reviewing the current situation as regards personal information representation within library and information science. The concrete subjects looked at included FRAD (Functional Requirements for Authority Data)[\[10,11\]](#), ORCID (Open Researcher & Contributor ID)[\[12\]](#), EAC (Encoded Archival Context)[\[13\]](#), MADS (Metadata Authority Description Schema)[\[14\]](#) and ISAAR-CPF (International Standard Archival Authority Record for Corporate bodies, Persons and Families)[\[15\]](#). The various standards reviewed by Greenall & Bertheussen were deemed to be largely irrelevant for the project because they did not involve a concept of linking to external sources. Additionally, the level of complexity in the existing data limited the extent to which we could represent data outlined in, for example, FRAD. It was also felt that these technologies were related to a type of role better served by a national authority file, rather than a linked data data set.

What did arise from the informal meetings and correspondence during this period was that the authority file should be modelled as “people” and not “authors” or “names”, and that this distinction is important in relation to how data is structured.

The second set of training days (2010-06-25–2010-06-26) were designed to provide a springboard for the NTNU University Library working group. The group worked on creating a common concept of what an RDF representation of data from the authority file would be and how it should be structured. The details of this work are contained in the next section.

2.2 Operational phase 2: creating data structures

This part of the document — essentially a lightly edited variant of [\[16\]](#) — describes the RDF data structure used for the BIBSYS personal name authorities, corporations and conferences

(although only the first of these has actually been modelled, the structures for the latter two have been included for others who might use this information).

The aim of this data structure was to provide usable linked data for users who want a reference to positively identified individuals. One way of doing this is by providing a name in context with publications related to that name.

BIBSYS-MARC is the assumed format for authorities throughout, for details of this see [\[9\]](#), which describes both the format of BIBSYS personal name authorities and the differences between this format and the equivalent MARC21 format, see also [\[8\]](#) for details on BIBSYS-MARC XML.

2.2.1 RDF vocabularies

It was felt to be necessary to establish a vocabulary, here called “radatana” — the ASCII equivalent of “rådata nå”. The reasons for the vocabulary are simple: the data set is specifically related to a set of “pseudo-keys”, some of these are meaningful and they are set up in a specific way (personal names are inverted, the leading articles are dropped in corporate names, etc.). Thus, the following property was created:

```
radatana:catalogueName a rdfs:Property ;
    rdfs:description "The name used in the BIBSYS name
    authority register, structured according to the current
    cataloguing standards at time of modification";
    rdfs:isDefinedBy radatana: .
```

This relates specifically to names in \$a fields 100, 110 and 111.

The remaining vocabularies used for the data structure are:

- BIBO [\[17\]](#)
- DCTERMS [\[18\]](#)
- FOAF [\[19\]](#)
- OWL [\[20\]](#)
- RDFS [\[21\]](#)
- SWC [\[22\]](#)
- SKOS [\[23\]](#)
- WHOIS [\[24\]](#)

2.2.2 Unique information for the different authority types

This section defines the unique data that is to be represented for each of the authority types.

2.2.2.1 Personal name authorities in 100:

```
<bibsys-xidentifier> a foaf:Person
```

Each personal name authority with a name in the 100 field is modelled as a foaf:Person. This is in line with a conceptualization that the names are modelled as “people”.

```
foaf:name "string"
```

This is a representation of the full name of the person, with all of the trimmings, so to speak, i.e. all of the additions that can be found distributed about the authority record.

```
rdfs:label "string"
```

This is a representation of the full name of the person in uninverted form. This is a repetition of the value in foaf:name and is provided to ensure that a generic label is available (i.e. one that does not entail an understanding that the data is modelled as persons); this ensures that the data is as accessible as possible to those who wish to use it.

```
bibo:suffixName "string"
```

This is a representation of the name suffix found in 100 \$b.

2.2.2.2 Corporation names in 110

```
<bibsys-xidentifier> a foaf:Organization
```

Each personal name authority with a name in the 110 field is modelled as a foaf:Organization. This is in line with a conceptualization that the names are modelled as “organizations”.

```
foaf:name "string"
```

This is a representation of the full name of the corporation, with all of the trimmings, so to speak, i.e. all of the additions that can be found distributed about the authority record. In this case, 110 fields \$a + \$b (+ \$w).

```
rdfs:label "string";
```

The same value as foaf:name.

2.2.2.3 Conference names in 111

Conferences are a special case, and an extra “packaging” is included in order for them to make sense. The basic pattern is as follows:

```
<bibsys-xidentifier> a swc:Conference ;  
    rdfs:label "string" ;  
    swc:relatedDocument <y>;
```

```

    bibo:organizer <z> .

    <y> a swc:Proceedings ;
        rdfs:label "Proceedings of string" .

    <z> a foaf:Organization
        foaf:name "Organizers of string" ;
        rdfs:label "Organizers of string".

```

Each conference is an instance of swc:Conference. The name of the conference is identified as an rdfs:label, all related documents are registered as such and the organizing body is identified. Each conference has the documents related to it listed as an swc:relatedDocument. These are in turn registered themselves as triples of the type <x> a swc:Proceedings. The organizer of the conference (typically the conference itself?) can also be registered as a standalone triple. This triple is thereby standardized as a foaf:Organization, and this triple contains the information necessary to represent the full information.

2.2.3 Properties that occur in every record

```

    radatana:catalogueName "string"

```

The established authority name is represented as-is from the authority file; this replicates the functionality expected when using bibliographic databases.

```

    skos:altLabel "string"

```

This represents the unestablished authorities related to the established record and is repeated to represent the name in both in its catalogue and natural forms.

```

    dcterms:identifier "string"

```

The identifier for the authority record (matched by regular expression /x[0-9]{8}/)

```

    dcterms:modified "string"

```

The date of the last update of the authority record.

```

    skos:note "string"

```

Any notes contained in the authority record.

```

    whois:since "date"^^xsd:year

```

Start-date from 100/110/111 \$d.

```
whois:until "date"^^xsd:year
```

Stop-date from 100/110/111 \$d.

```
owl:sameAs <link.to.dbpedia>
```

The authority as represented in dbpedia.org (a linked data representation of content from Wikipedia).

```
owl:sameAs <link.to.libris>
```

The authority as represented in LIBRIS (the National Library of Sweden).

```
owl:sameAs <link.to.NL_Hungary>
```

The authority as represented in National library of Hungary.

```
owl:sameAs <link.to.NL_Germany>
```

The authority as represented in National library of Germany.

```
owl:sameAs <link.to.VIAF>
```

The authority as represented in VIAF (Virtual International Authority File).

```
foaf:page <link.to.BIBSYS_OPAC>
```

The results page that the authority creates in BIBSYS Ask.

```
foaf:page <link.to.bibsys SRU authority search>
```

The link to the SRU MARC21 representation of the authority (see below).

```
foaf:page <link.to.bibsys SRU bibliographic search>
```

The link to the SRU MARC21 bibliographic results set for the authority (see below).

```
foaf:made <links.to.the.bibsys.objektid.RDF.representations>
```

A link to an RDF representation of each and every item registered that has the authority as creator in the bibliographic database (note that this RDF data set doesn't exist as of today.)

2.2.4 Explanatory links.

```
<link.to.bibsys SRU authority> a foaf:Document; dcterms:format
"application/xml+marcxml"; dcterms:description "BIBSYS authority
record for {name}" .
```

This link provides the information contained in the coming national authority file that BIBSYS and the National library of Norway are working on. The link needs to be defined as a standalone triple in the RDF in order to explain what it is and how the resource it represents is formatted.

```
<link.to.bibsys SRU bibliographic search> a foaf:Document;
dcterms:format "application/xml+marcxml"; dcterms:description
"A link to the bibliographic records related to {name} in
BIBSYS" .
```

This link provides the information contained in the BIBSYS bibliographic database. The link needs to be defined as a standalone triple in the RDF in order to explain what it is and how the resource it represents is formatted.

2.2.5 Example

```
<x90052737> a foaf:Person ;
  radatana:catalogueName "Shakespeare, William" ;
  dcterms:modified "19910422" ;
  foaf:name "William Shakespeare" ;
  rdfs:label "William Shakespeare" ;
  skos:altLabel "Uil'jam Sekspir";
  skos:altLabel "William Shakspeare";
  skos:altLabel "William Shakspear";
  skos:altLabel "William Shakspere";
  skos:altLabel "William Shakespear";
  dcterms:identifier "x90052737";
  owl:sameAs <http://dbpedia.org/page/William_Shakespeare> ;
  owl:sameAs <http://libris.kb.se/resource/auth/198702> ;
  owl:sameAs <link.to.NL_Hungary> ;
  owl:sameAs <http://d-nb.info/gnd/118613723/about> ;
  owl:sameAs <http://viaf.org/viaf/96994048/rdf.xml> ;
  foaf:page
  <http://ask.bibsys.no/ask/action/result?cmd=&kilde=biblio&fid=forfa
  tter&term=Shakespeare%2C+William&op=and&fid=bd&term=&bibliografi=in
  gen&arstall=&sortering=sortdate-&treffPrSide=10> ;
  foaf:page <bibsys_authority_william_shakespeare> ;
  foaf:page <bibsys_william_shakespeare> ;
  foaf:made <links.to.the.bibsys.objektid.RDF.representations> .
```


2.2.6 Handover

The completed structure document was handed over to the operational team at BIBSYS 2010-08-18.

2.3 Creating a linked data representation from authority file data

This section is based on information from [\[25\]](#) and describes the work carried out by the operational team at BIBSYS.

2.3.1 Initial considerations

The main considerations for this part of the project are outlined in the requirements for linked data in section 1.1.1, repeated here for convenience:

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
4. Include links to other URIs. so that they can discover more things

Berners-Lee [\[2\]](#)

In order to satisfy (1) and (2), the delivery platform needs to present a HTTP URI that will return some coherent data. To satisfy (3), the data returned needs to be formatted as some variant of RDF and be queryable using a standard query language for RDF, such as SPARQL. The final point (4), is satisfied by enriching the original data by referring to third parties with information about the same topic.

In the next sections, it will be shown how these requirements were satisfied.

2.3.2 Creating the basic data set

Before any work on producing “linked data”, the data had to be structured as RDF. The data was transformed to the structure described above, the mappings can be summarized as follows:

- foaf:person - The type of the resource.
- foaf:name - The established authority name. Inverted name from the authority record. (Marc field 100 \$a)
- rdfs:name - The established authority name. Inverted name from the authority record. (Marc field 100 \$a)
- radatana:catalogueName - The established authority name. Uninverted name as-is from the authority record. (Marc field 100 \$a)
- whois:since - Birth date. (Value selected from before - sign in marc field 100 \$d)
- whois:until - Death date. (Value selected from after - sign in marc field 100 \$d)
- dcterms:modified - The date of the latest update of the authority record.

- dct:identifier - Local identifier of the authority record.
- skos:altLabel - The unestablished authorities name. Inverted name from the authority record. (Marc field 900 \$a)

Additionally, the data was enriched with links to the following data sets (this process is described below):

- owl:sameAs - <link.to.dbpedia> The authority as represented in DBpedia.org.
- owl:sameAs - <link.to.nl_germany> The authority as represented in the National library of Germany.
- owl:sameAs - <link.to.viaf> The authority as represented in VIAF.

Note that the unestablished authority name (marc field 900) is represented as a *sameAs* relation only. Note also that the non-personal-name authority structures were ignored.

2.3.2.1 Creating a functional platform for working with linked data

In order to work with the data set that had been created, it was necessary to implement a platform for linked data development and delivery. This platform comprises the following components:

2.3.2.2 Triplestore

JENA was chosen as the main tool for triple-handling; it is a semantic web framework for Java. With JENA, triples are created and then stored in a triplestore. The JENA framework makes this job easy. See [26] for more details of JENA.

2.3.2.3 JOSEKI

JOSEKI is an HTTP engine that supports the SPARQL Protocol and the SPARQL RDF Query language [7]. See [27] for more details of JOSEKI.

2.3.2.4 Web application

Using the JOSEKI HTTP engine, a web application is instantiated which responds to dereferenceable URI's (fulfilling the requirement that each logical cluster of information is represented by an HTTP URI). For example, requesting <http://data.bibsys.no/data/notrbib/authorityentry/x90061718> returns the resource in RDF/XML.

The web application supports other syntaxes for RDF by adding the format information to the URL:

- <http://data.bibsys.no/data/notrbib/authorityentry/x90061718/rdf-abbrev>
- <http://data.bibsys.no/data/notrbib/authorityentry/x90061718/triple>
- <http://data.bibsys.no/data/notrbib/authorityentry/x90061718/n3> (turtle)

Also embedded in this web application is an interface supporting SPARQL queries:

- http://data.bibsys.no/data/query_authority.html

2.3.2.4 Linking data with DBpedia, the National library of Germany and VIAF

Using matching algorithms, it was possible to link the data to with the authority representation in other authority systems such as DBpedia, the National library of Germany and VIAF. The triplestores were run locally with copies of triples from each system; running SPARQL queries against these triplestores to match representations in the different data sets. Thus:

- owl:sameAs - http://dbpedia.org/resource/Henrik_lbsen
- owl:sameAs - <http://viaf.org/viaf/71378383>
- owl:sameAs - <http://d-nb.info/gnd/118555286>

2.3.2.5 Dereferenceable URI

The authority resource is of type *Person*. The resource is identified by a dereferenceable URI, for instance <http://data.bibsys.no/data/notrbib/authorityentry/x90061718>.

2.3.2.6 RDF Example

This is an example of the RDF representation of *William Shakespeare* in XML/RDF:

```
<rdf:RDF
<rdf:Description rdf:about="http://data.bibsys.no/data/notrbib/authorityentry/x90052737">
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
  <foaf:name>William Shakespeare</foaf:name>
  <dc:identifier>x90052737</dc:identifier>
  <dc:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2001-09-11</dc:modified>
  <rdfs:label>William Shakespeare</rdfs:label>
  <radatana:catalogueName>Shakespeare, William</catalogueName>
  <skos:altLabel>William Shakspere</skos:altLabel>
  <skos:altLabel>William Shakespear</skos:altLabel>
  <skos:altLabel>Uil'jam ?ekspir</skos:altLabel>
  <skos:altLabel>William Shakspeare</skos:altLabel>
  <skos:altLabel>William Shakspear</skos:altLabel>
  <owl:sameAs rdf:resource="http://dbpedia.org/resource/William_Shakespeare"/>
  <owl:sameAs rdf:resource="http://viaf.org/viaf/13092507"/>
  <owl:sameAs rdf:resource="http://d-nb.info/gnd/117743445"/>
</rdf:Description>
</rdf:RDF>
```

Printing the representation *William Shakespeare* as N-triples looks like this:

```
<http://data.bibsys.no/data/notrbib/authorityentry/x90052737>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://xmlns.com/foaf/0.1/Person> .
<http://data.bibsys.no/data/notrbib/authorityentry/x90052737> <http://xmlns.com/foaf/0.1/name>
"William Shakespeare" .
<http://data.bibsys.no/data/notrbib/authorityentry/x90052737>
<http://purl.org/dc/terms/identifier> "x90052737" .
<http://data.bibsys.no/data/notrbib/authorityentry/x90052737>
<http://purl.org/dc/terms/modified> "2001-09-11"^^<http://www.w3.org/2001/XMLSchema#date> .
<http://data.bibsys.no/data/notrbib/authorityentry/x90052737>
<http://www.w3.org/2000/01/rdf-schema#label> "William Shakespeare" .
<http://data.bibsys.no/data/notrbib/authorityentry/x90052737>
```

```
<http://www.bibsys.no/ns#catalogueName> "Shakespeare, William" .
<http://data.bibsys.no/data/notrbib/authorityentry/x90052737>
<http://www.w3.org/2004/02/skos/core#altLabel> "William Shakspere" .
<http://data.bibsys.no/data/notrbib/authorityentry/x90052737>
<http://www.w3.org/2004/02/skos/core#altLabel> "William Shakespear" .
<http://data.bibsys.no/data/notrbib/authorityentry/x90052737>
<http://www.w3.org/2004/02/skos/core#altLabel> "Uil'jam ?ekspir" .
<http://data.bibsys.no/data/notrbib/authorityentry/x90052737>
<http://www.w3.org/2004/02/skos/core#altLabel> "William Shakspeare" .
<http://data.bibsys.no/data/notrbib/authorityentry/x90052737>
<http://www.w3.org/2004/02/skos/core#altLabel> "William Shakspear" .
<http://data.bibsys.no/data/notrbib/authorityentry/x90052737>
<http://www.w3.org/2002/07/owl#sameAs> <http://dbpedia.org/resource/William_Shakespeare> .
<http://data.bibsys.no/data/notrbib/authorityentry/x90052737>
<http://www.w3.org/2002/07/owl#sameAs> <http://viaf.org/viaf/13092507> .
<http://data.bibsys.no/data/notrbib/authorityentry/x90052737>
<http://www.w3.org/2002/07/owl#sameAs> <http://d-nb.info/gnd/117743445> .
```

2.3.2.7 Huge number of triples

The representation of *William Shakespeare* includes 14 triples. For the 1.5 million persons represented in the BIBSYS Authority file we got approximately 13 million triples. From DBpedia we got approximately 350.000 triples, and from National library of Germany we got 40 million triples.

3.0 The data

The data is now hosted at <http://data.bibsys.no/data> and is made available in two forms:

- [SPARQL endpoint](#)
- [Bulk download](#)

The radatana vocabulary is hosted at <http://def.bibsys.no/xmlns/radatana/1.0/index.xml>

The data is licensed under the [ODC-PDDL](#) (Open Data Commons Public Domain Dedication Licence), which entails free use without attribution.

A [CKAN](#) page has been created for the dataset: <http://ckan.net/package/radatana>

4.0 Conclusion

From the point of view of NTNU University Library and BIBSYS, the central aim of the project has been achieved in that the data set is relatively functional and complete, and is available both for download as a large data set and as a queryable resource. The data is made available in a form that conforms to current best practice for publication of linked open data; additionally, the work is groundbreaking in two senses: it is a large data set that opens Norwegian authorities to international audiences and provides a new way of modelling authority data.

Both NTNU University Library and BIBSYS agree that linked data is a technology that is relevant to their work and to libraries in general, and both parties have a commitment to using the technology in further projects. NTNU University Library currently use the Rådata nå! metadata in projects related to presentation of the library's special collections, and BIBSYS will use the technologies described in this document as the core of the national authority file discussed above.

Further, we consider the collaborative work that has taken place, both between the project partners and PODE, to have been central in creating an embryonic linked bibliographic data community that will continue to grow and form the way that linked bibliographic data will evolve in Norway and the wider linked data community.

Concrete actions regarding the data set include ensuring a permanent domain name, maintaining the data description in CKAN and updating the data as new authorities are added to the BIBSYS authority file.

On a more general note, it is recommended that libraries continue to look at the benefits of creating the kind of data sets that are gradually emerging from both within and outside the library community, and using distributed data sets of this kind in their work. Through the concept of distributed data sources, it is possible to nurture the kind of collaboration found within this project and share the load of metadata creation — ensuring that work is not unnecessarily duplicated and therefore that focus can be given to provision of the most appropriate and usable form for metadata.

5.0 Glossary

Authority file	Index of terms maintained to ensure consistency and reduce redundancy in bibliographic systems
BIBO	Bibliographic Ontology, an RDF vocabulary for describing bibliographic resources
BIBSYS	State-run Library Management System supplier based in Trondheim, Norway
BIBSYS-MARC	The dialect of MARC used by BIBSYS
Conference	Standard forms of names of academic conferences
Corporation name	Standard forms of names of institutions, organizations, etc.
DBPedia	Subset of information from Wikipedia presented as Linked Open Data
DCTERMS	Dublin Core Metadata Terms, an RDF vocabulary for describing resources
Dereference	Process of accessing and retrieving data
FOAF	Friend Of A Friend, an RDF vocabulary for describing resources about people
HTTP	HyperText Transfer Protocol, a protocol used to transmit information over the web
LIBRIS	The national library of Sweden
Linked data	A method of presenting data using RDF and hyperlinks
MARC	See MARC21
MARC21	Machine Readable Cataloguing 21 , syntax and encoding rules for representation of machine readable bibliographic records

N-triples	Simple representation format for RDF, made by presenting each RDF statement in full as URIs or named nodes
Norwegian Archive, Library and Museum Authority	Now-defunct authority that provided financial support to development and coordinated various aspects of operational management within the library, archive and museum sector
NTNU University Library	The library of the Norwegian University of Science and Technology, Trondheim, Norway
OWL	Web Ontology Language, an RDF ontology
Personal name	Standard forms of names of people
Pode	Project related to utilization of bibliographic data in new contexts
Rådata nå!	The name of the project described by this document; it means “Raw data now!” in Norwegian
RDF	Resource Description Format, a triple-based description format for metadata
RDF vocabulary	Collection of (related) metadata terms used to express the relationship between a resource and some metadata about the resource
RDF/XML	Representation of RDF in XML
RDFS	Resource Description Format Schema, an RDF vocabulary
SKOS	Simple Knowledge Organization Schema, an RDF vocabulary
SPARQL	SPARQL Protocol and RDF Query Language, a query language for RDF
SWC	Semantic Web Conference, an RDF vocabulary for describing conferences

Talis	Library management system, triplestore and consultancy company
Triple	RDF statement. These have the form subject-predicate-object (the three elements of a triple)
Triplestore	Storage for RDF data
UBiT2010	Three-year, library-modernization project at NTNU University Library (http://ubit.blogs.bibsys.no/ubit2010)
URI	Uniform Resource Identifier, HTTP URIs (URLs — Uniform Resource Locators) are web addresses
VIAF	Virtual International Authority File, a collective of OCLC and several national libraries that present cross-linked online representations of their national authority files in various formats (including HTML and RDF)
Web application	Software that runs on a web server, often presenting a user interface for use in web browsers
WHOIS	Who's who description vocabulary, an RDF vocabulary for describing resources about people
XML	Syntax for encoding machine readable documents containing arbitrary data structures

6.0 Appendices

6.1 Appendix A

[Text of application to Norwegian Archive, Library and Museum Authority](#) [in Norwegian]

6.2 Appendix B

[Award letter from Norwegian Archive, Library and Museum Authority](#) [in Norwegian]

6.3 Appendix C

[Financial report](#) [in Norwegian]

6.4 Appendix D

[Contract between NTNU Library and BIBSYS](#) [in Norwegian]

6.5 Appendix E

[Covering letter](#) for this document as delivered to the National Library of Norway [in Norwegian]

7.0 References

- [1] Berners-Lee, T. The next web of linked, open data. Paper presented at TED2009: the great unveiling; 2009 Feb 4; Long Beach, California, USA; Available from: <http://www.w3.org/2009/Talks/0204-ted-tbl/>
- [2] Berners-Lee T. Linked Data - Design Issues [Internet]. [cited 2011 Jan 11]; Available from: <http://www.w3.org/DesignIssues/LinkedData.html>
- [3] BIBSYS Website - Main page [Internet]. [cited 2011 Jan 11]; Available from: <http://bibsyst.no/english/pages/index.php>
- [4] ABM-utvikling — ABM-utvikling - Statens senter for arkiv, bibliotek og museum [Internet]. [cited 2011 Jan 11]; Available from: <http://www.abm-utvikling.no/om-abm-utvikling>
- [5] Bertheussen, Lene Elisabeth & Greenall, Rurik Thomas. #rdn Minutes 2010-05-06 [unpublished meeting minutes for project Rådata nå!]:13:00; 2010; Trondheim, Norway.
- [6] RDF - Semantic Web Standards [Internet]. [cited 2011 Jan 15]; Available from: <http://www.w3.org/RDF/>
- [7] SPARQL Query Language for RDF [Internet]. [cited 2011 Jan 15]; Available from: <http://www.w3.org/TR/rdf-sparql-query/>
- [8] Husby, O. BIBSYS authority records in MARC21 XML; 2010 [Unpublished internal project documentation]; Trondheim, Norway; Available from: <https://docs.google.com/Doc?docid=0AW3mo82XcjtZGNuM3p0MzhfMTI4cXZyc2RwZHE&hl=en>
- [9] Husby, O. BIBSYS authority records in BIBSYS-MARC XML [Unpublished internal project documentation]; 2010; Trondheim, Norway; Available from: <https://docs.google.com/Doc?docid=0AW3mo82XcjtZGNuM3p0MzhfMjY1ZjV2a3NtZmc&hl=en>
- [10] Patton, G. E. Functional requirements for authority data: a conceptual model : final report, December 2008. München: Saur; 2009.
- [11] Bertheussen, L. E. & Greenall, R. T. FRAD overview [Unpublished internal project documentation]; 2010; Trondheim, Norway.
- [12] The Initiative | ORCID [Internet]. [cited 2011 Jan 23]; Available from: <http://www.orcid.org/>
- [13] Encoded archival context: corporate bodies, persons and families [Internet]. [cited 2011 Jan 23]; Available from: <http://eac.staatsbibliothek-berlin.de/>
- [14] Metadata Authority Description Schema (MADS) - (Library of Congress) [Internet]. [cited 2011 Jan 23]; Available from: <http://www.loc.gov/standards/mads/>
- [15] ICA/CDS | Standards & Guidelines [Internet]. [cited 2011 Jan 23]; Available from: <http://www.icacds.org.uk/eng/standards.htm>
- [16] Bertheussen, L. E., Greenall, R. T. & Husby, O. BIBSYS authority to linked data. [Unpublished internal project documentation]; 2010; Trondheim, Norway; Available from <https://docs.google.com/document/pub?id=1rzxLvyVgf5aDVtiKdZ4qV6HLdIveILF4cH5Ey9cQNK0>
- [17] BIBO: Bibliographic Ontology [Internet]. [cited 2011 Jan 18]; Available from: <http://bibotools.googlecode.com/svn/bibo-ontology/trunk/doc/index.html>
- [18] DCMI Metadata Terms [Internet]. [cited 2011 Jan 18]; Available from:

<http://dublincore.org/documents/dcmi-terms/>

[19] FOAF Vocabulary Specification [Internet]. [cited 2011 Jan 18]; Available from:

<http://xmlns.com/foaf/spec/>

[20] OWL Web Ontology Language Overview [Internet]. [cited 2011 Jan 18]; Available from:

<http://www.w3.org/TR/owl-features/>

[21] RDF Vocabulary Description Language 1.0: RDF Schema [Internet]. [cited 2011 Jan

18]; Available from: <http://www.w3.org/TR/rdf-schema/>

[22] Semantic Web Conference Ontology [Internet]. [cited 2011 Jan 18]; Available from:

http://data.semanticweb.org/ns/swc/swc_2009-05-09.html

[23] SKOS Simple Knowledge Organization System - home page [Internet]. [cited 2011 Jan

18]; Available from: <http://www.w3.org/2004/02/skos/>

[24] Who's who description vocabulary [Internet]. [cited 2011 Jan 18]; Available from:

<http://www.kanzaki.com/ns/whois>

[25] Aunet, Roger. Utkast til Statusrapport: Radatana. [Unpublished internal project documentation]; 2010; Trondheim, Norway.

[26] Jena Semantic Web Framework [Internet]. [cited 2011 Jan 15]; Available from:

<http://jena.sourceforge.net/>

[27] Joseki - A SPARQL Server for Jena [Internet]. [cited 2011 Jan 15]; Available from:

<http://www.joseki.org/>