IFLA Section Bibliography

Webinar: Linked Open Bibliographies: Exposing, Linking, Reusing

27 November 2024, 3:00 – 5:00 pm CET

Q&A Report

This document presents questions submitted by participants of the webinar *"Linked Open Bibliographies: Exposing, Linking, Reusing"*, that were not addressed during the live Q&A session. The responses are provided by our esteemed speakers and are identified by the acronyms of their respective libraries.

The IFLA Bibliography Section extends its sincere gratitude to all speakers for generously dedicating their time and expertise to answer these questions. Their contributions are invaluable to the library community.

About the Webinar

As National Libraries and Bibliographic Agencies across the globe continue to evolve, the transformation of National Bibliographies and Authority data into linked open data has become a critical step in modernizing information access, discovery, and integration. The webinar provided a forum for six of the National Libraries that took the lead in this direction to share their experiences, insights, and showcase advancements in working with linked open data.

Presentations

- *The British National Bibliography Experience in the Share Family Linked Open Data Environment*. **Thurstan Young**, British Library and **Anna Lionetti**, Casalini Libri/Share Family.
- *FAIR Authority Data, a First Step Towards a Linked Open Belgian Bibliography*. **Sven Lieber**, Royal Library of Belgium (KBR).
- *Linked Open Japanese National Bibliography: The New NDL Search and Web NDL Authorities*. **Kazue Murakami**, National Diet Library (NDL).
- *Renewing the German National Library's Linked Open Data Ecosystem: Challenges and Opportunities*. **Tracy Arndt**, German National Library.
- *Non-MARC-ish BIBFRAME*. **Andreas Andersson**, National Library of Sweden.
- *Envisioning the Future of Finnish Library Data with BIBFRAME and RDA*. **Matias Frosterus**, National Library of Finland.

Speakers' Biographies

Thurstan Young is the British Library's Collection Metadata Standards Manager. He is a member of the Share Family Advisory Council and of the Sapientia Entity Identification Working Group (SEI WG). He also currently serves as the British Library's representative to the European RDA Interest Group (EURIG) and as the U.K. Representative to the MARC Advisory Committee (MAC).

Anna Lionetti works for Casalini Libri as R&D Assistant and facilitator for research and development projects focused on Linked Data, supporting the project management and acting as a liaison with stakeholders. In her previous experience, she has taken part in international projects for the development of new services for the publishing industry based on metadata standards.

Sven Lieber works as data manager at The Royal Library of Belgium (KBR), where he supports the improvement of data quality, mainly by using computational methods and Linked Data. Sven is currently involved in the research infrastructure project MetaBelgica and the research project BELTRANS. For both projects, Sven helps to make data available according to the FAIR principles. He studied Computer Science at the University of Freiburg in Germany and pursued a PhD related to Knowledge Graphs at the University of Ghent in Belgium, where he was also involved in several interdisciplinary research projects at the IDLab research group.

Kazue Murakami is in charge of the Japanese National Bibliography. Since 2019 she has been responsible for ensuring that it and other bibliographies from the National Diet Library are available free of charge to anyone, whether for commercial or non-commercial use. She has been a member of the Japan Library Association's Committee on Cataloging since 2018, and was involved in the release of the 2018 Edition of the *Nippon (Japan) Cataloging Rules*. She is a Standing Committee member of the Bibliography Section since 2021.

Tracy Arndt is responsible for the Linked Data Service of the German Nation Library since 2021. Her responsibilities encompass e.g. data modeling, software and ontology development, documentation and requirement management. She holds a M.Sc. in Library Informatics from Wildau Institute of Technology. She is speaker of the working group Competence Center for Interoperable Metadata (DINI AG KIM).

Andreas Andersson is a Metadata Specialist and Systems Librarian at the National Library of Sweden. He is team leader of the Format Team, a cross-functional team maintaining and developing KBV – Sweden's application vocabulary of BIBFRAME.

Matias Frosterus is an information systems manager at the National Library of Finland. His main interests lie in interoperability of metadata from many different angles. He has been working with linked data in library context for over ten years and his doctoral thesis dealt with utilizing linked data techniques for legacy data. For the past few years, he has been seriously getting to grips with BIBFRAME.

Acronyms List

BL: British Library (incorporating Share Family's response where appropriate/ applicable)
KBR: Royal Library of Belgium
NDL: National Diet Library (Japan)
DNB: German National Library
NLS: National Library of Sweden
NLF: National Library of Finland

1. *Is it only about linking authorities (mostly people, institutions, and subjects), or are you also connecting records through other fields in your catalog or to external resources that would be truly helpful to users? Additionally, are you already testing a version of BIBFRAME?*

   BL: At the moment in the Share Family environment, it is mainly about entities such as persons, organizations and subjects but the initiative keeps developing so there is no reason why in future, places, which enable the production of map visualizations, are not included. The underlying data model used by the Share Family initiative is an extension of the BIBFRAME model, i.e. the BIBFRAME ontology is used as much as possible and only when there is no adequate BIBFRAME property / class, a Share-VDE property / class is created.

   KBR: Since we use RDA terms (Resource Description and Access terms) in our bibliographic records (e.g. type of binding), we also refer to the external resources of RDA. We are not already testing BIBFRAME.

   NDL: The NDL creates links based on identifiers not only in authority data but also in bibliographic data. The NDL Search has links not only to Web NDL Authorities but also to external databases. Links are made automatically based on identifiers in bibliographic data. For example, if the LCCN is recorded in the bibliographic data, the NDL Search can provide a link to the catalog data on the Library of Congress website. Currently, external links are available except for the Japanese National Bibliography, but we aim to include them in Japanese National Bibliographic data in the future.

   DNB: We provide a BIBFRAME prototype for our bibliographic records. The connections in our data are manifold. Links to subject headings, people, themes, locations as well as links to classifications e.g. from publishers.

   NLF: We have a robust linked data ecosystem for managing authorities, such as subjects, agents, places, and cataloging phrases, delivered through the national thesaurus and ontology service, Finto. Additionally, we are preparing for a transition to BIBFRAME and have completed version 1.0 of our data model, which is based on BIBFRAME. However, full implementation will require additional time.

2. *From the perspective of the different presenters, what is the most appropriate platform for storing and disseminating bibliographic Linked Open Data (LOD): Git, Wikidata, or another option?*

BL: We are not in a position to answer technical questions of this nature, but Share Family may be able to offer a view on the matter. Depending on the use case, different solutions could be adopted that can even be put in relation with other linked data platforms / storage systems through interoperability processes. The Share Family is currently working on an installation of Neptune triple store that will host BIBFRAME/RDF data of participating institutions. Neptune is a product of AWS - Amazon Web Services, and is included in the suite of components that the Share Family has adopted to offer a data consumption method compatible with other infrastructure components that the Share Family technology is already using (ie. AWS cloud infrastructure).

KBR: Of course this depends on the use case. Personally, I would not recommend Git. This was created for version control and popular platforms implementing the Git protocol such as GitHub are not meant to store data, but code or configuration (even though I also have seen people using it to store data). Wikidata could be an option, but since it is a crowd-sourcing platform you in a sense lose authority over the data. In general you could either set up your own workflows and infrastructure to disseminate and publish Linked Open Data or you kind of source it out to trusted parties such as Share-VDE. I would say that the former requires more maintenance and skills in-house but gives more flexibility, whereas the latter has potential to already provide linking and matching as well as a platform for discoverability.

NDL: My answer may be a slight departure from the purpose of your question; the NDL is currently investigating the possibility of adding various identifiers to our authority data and is focusing on Wikidata as an important source of information.

DNB: This depends on the use case. There is no "platform" for storing LOD. It is a federated and decentralized approach.

NLF: Very much depends on the use case and the nature of the data. For authority type data we use Finto (based on the open source Skosmos software) but we have links to Wikidata for many of our vocabularies. We have examined Wikibase quite a bit and it is certainly promising for certain types of use cases. Finally, for bibliographic data intended for metadata repository use we are looking into various options currently with FOLIO looking like the most promising candidate when it comes to open-source solutions.

3. *It could represent revolutionary progress in information systems worldwide, but could it also be misused for global surveillance or control of individuals?*

BL: Issues of data protection may be relevant to linked open data. By ensuring that personal data is processed in a transparent manner which corresponds with a regulatory framework such as GDPR (General Data Protection Regulation), it can lower the barrier to publishing and re-using linked open data.

KBR: This highly depends on the data. My following answer assumes that your question refers to authority data of national libraries. Even though internally we might hold personal information such as author contact information, the publicly available authority data does not. Minimal author identification, e.g. via ISNI identifier, is needed, also in the interest of the author. For example, for royalties or simply for *correct* data processing. One could see it like the "required technical cookies only" from websites nowadays. Authority data is already available in most online catalogues or on platforms such as Wikidata anyway. Even though this is a practical fact, this of course does not mean that as national libraries we are free from legal responsibility. We usually operate under national law that allows us to process and publish certain data. Personally, I don't see any surveillance or control of individuals-aspect if privacy regulations such as the GDPR are followed. On the contrary, high quality authority data from authoritative sources can even be helpful to counter misinformation.

NDL: As you may be concerned, the possibility of misuse cannot be denied when various metadata is easily accessible in the form of Linked Open Data. However, I believe that national bibliographic agencies have a particular obligation to make their national bibliographies accessible to all, serving as a reflection of their country's publishing activities.

NLF: We feel that high-quality and neutral (in the sense that there are no commercial or political motives) library data is an important building block for safer data in general. We cannot stop the progress, but we can make sure that the data used for it is as high-quality and free from bias as possible.

4. *Regarding the overlap between national bibliographies, there would be some, but it would not constitute a large proportion. For example, bibliographies that include publications by national authors published in other countries, or publications in their national language published worldwide, or publications about the country or its nationals. These publications should also appear in the national bibliography of the country where they were published. However, this is likely to be less than 10%, and most publications will appear in at most two national bibliographies. This line of thinking aligns with the concept of Universal Bibliographic Control (UBC).*

   BL: Different countries have different definitions of what to include in their national bibliographies. The British National Bibliography (BNB) includes publications published and /or distributed in the United Kingdom and the Republic of Ireland. It does not include publications by national authors published in other countries or publications in English published worldwide or publications about the UK and the Republic of Ireland. When there is more than the BNB in the Share Family National Bibliographies tenant, it is hoped that this aggregation of metadata will help produce a more extensive view of knowledge about, for example, particular authors. By this means, we will have a resource that is greater than the sum of its parts.

   KBR: I agree, the overlap in terms of bibliographies probably is small. But there certainly is an overlap due to things such as multiple nationalities of authors, different policies to what constitutes a national bibliography, or publication forms such as anthologies that combine the works of several authors.

   NDL: The Japanese National Bibliography (JNB) also includes foreign-published Japanese-language materials, which make up around 5% of its total content. In any case, I agree that the perspective aligns with the concept of the UBC.

5. *We truly need multilingualism. Bilingual and multilingual national libraries should collaborate on this effort, including countries like Finland, Belgium, Canada, Switzerland, India (sic!), and others.*

BL: Multilingual (and multiscript) linked open data is already available in the context of bibliographic description. The RDA Registry contains linked data and Semantic Web representations of the entities, elements, terminologies and value vocabularies approved by the RDA Steering Committee (RSC). Current and pending translations from English include Arabic, Catalan, Chinese, Danish, Dutch, Estonian, Finnish, French, German, Greek, Hebrew, Hungarian, Italian, Latvian, Norwegian, Portuguese, Slovak, Spanish, Swedish, Ukrainian and Vietnamese. RDA in Many Metadata Formats (RIMMF) offers a practical application of the RDA Registry; using this tool, one can create RDA complaint data, then display its element labels and controlled vocabulary terms in languages which are supported by the Registry.

KBR: I agree that multilingualism is needed and helps in collaboration, e.g. by easier identifying the overlap of publications in different national libraries. Out of context, however, I wonder what you mean with multilingualism. We apply it more and more for our bibliographic descriptions. And this in fact is where Linked Open Data helps us, because we get a single identifier with different possible multilingual labels.

NLF: Multilingualism is a very interesting challenge and certainly something that can be tackled with linked data. Moving from the term level to the concept level certainly helps though it is not a silver bullet as such. There are culture-specific differences to concepts and though they can be represented in each language they might not be something that is used by the end-users. For the General Finnish Ontology YSO we decided to use Finnish and Swedish as the main languages and English and Northern Sami as translational languages. When there are clashes between Finnish and Swedish, we try to rely on the culture of Finland to make the choice but this isn't always trivial.

6. *Which of the tools you are using are available as open source?*

BL: While the tools and components of the Share Family LOD Platform technology are not open source, the linked data output (ie. "new-born" linked data objects) produced by the processing of original library records are openly available, in accordance with the policies of Share Family participating institutions.

KBR: Our main Library Management System (LMS) that we get from a software supplier is, as far as I know, not open source. Web interfaces of other institutions that we use probably aren't open source either. However, tools we develop ourselves are usually available under different open-source licenses: https://github.com/kbrbe/.

NDL: The Web NDL Authorities has a unique interface developed through outsourcing; therefore, its system cannot be made available as open source software.

DNB: Metafacture: https://github.com/metafacture;
Raptor: https://librdf.org/raptor/rapper.html;
Tarql: https://tarql.github.io/

NLF: Skosmos is an open-source SKOS vocabulary publication and browsing tool: https://skosmos.org/
Annif is a tool for automatic subject indexing and classification: https://annif.org/.

7. *What systems or software are used to manage metadata for bibliographic records? Does the library use any tools for automated metadata generation or enrichment?*

KBR: We use our own Library Management System Syracuse from the company Archimed in France. We can perform bulk operations via the UI, or export data in CSV, enrich the records manually, via OpenRefine or via Python, and re-import the CSV.

NDL: The NDL uses an in-house developed system to manage bibliographic and authority data. The metadata format in the system is MARCXML. We don't use tools for automatic metadata creation, but we are beginning to discuss whether AI could assist in the creation of bibliographic data in the future.

DNB: We use CBS from OCLC to manage our data. We use tools like annif: https://annif.org/ to create more metadata. More information about our AI Project can be found here:
https://www.dnb.de/EN/Professionell/ProjekteKooperationen/Projekte/KI/ki_node.html.

NLF: We currently use Aleph for our union catalogue but will be switching to a linked data capable system in the future.

8. *What are the steps in building linked data for a National Bibliography and is it necessary to create PIDs such as urn:nbn for URI first?*

BL: Publishing a national bibliography is, in many ways, no different to publishing any other dataset. The W3C Best Practices for Publishing Linked Data list the main steps required. These include minting and using HTTP URIs. They can be HTTP URN:NBN-based URIs as implemented by the National Library of Finland.

KBR: I would say it is necessary to have a good PID strategy first, because Linked Data relies on identifiers. And the last thing you want is keeping track of who took copies of your data when such that in case of switching your identifier system you can keep all the links working. If URNs are a good solution, is another question. There are several possibilities, each with their pros and cons. However, there is a lot of momentum regarding PIDs due to (inter)national research infrastructure initiatives. For example: https://zenodo.org/records/14092489, https://zenodo.org/records/7258286, or https://doi.org/10.5281/zenodo.7330527. Regarding the first part of the question I refer to my presentation: https://doi.org/10.5281/zenodo.14217617.

NDL: In order to build linked data, not just national bibliographies, it is first necessary to design a URI as the foundation for representing the resource. A URI allows a resource to be associated with various external resources. The Web NDL Authorities and the NDL Search have URIs for authority and bibliographic data based on bibliographic IDs and authority IDs, rather than urn:nbn.
For more information about URIs for the NDL resources, please refer: Uniform Resource Identifiers (URI) for National Diet Library resources
https://www.ndl.go.jp/en/dlib/standards/lod/uri.html.

NLF: The previous answers covered the subject well, but just to add something, NLF maintains a directory of open source URN tools:
https://www.kiwi.fi/display/URN/URN%3ANBN+Code+Directory.

9. *How do you treat Creative Works in your data? There are a few different bibliographic models, is there one that is superior?*

BL: IFLA's Library Reference Model (LRM) offers an entity-relationship framework with which to structure and describe bibliographic data. RDA is a practical implementation of the LRM and follows the basic structure of the LRM's entities and relationships. In RDA a resource entity may be a work, expression, manifestation or item. For published books, serials, maps and scores our data has historically tended to conflate these entities at the local level of the library catalogue. However, the Share Family initiative offers us the opportunity to express our cataloguing output as linked open data in an LRM/RDA and Bibframe context going forward. From an LRM/RDA perspective, BIBFRAME conflates the concept of a work and expression into a single entity: i.e. a BIBFRAME work. Under these circumstances, BIBFRAME may be considered to offer a less user-friendly means of retrieving bibliographic resources and a more inefficient means describing them than RDA does; practical consequences could include more scrolling through flatter sets of search results and less recording of data by exception. On the other hand, the BIBFRAME model corresponds more closely to its predecessor, MARC 21, than the LRM/RDA does. Bibframe consists of three entities rather than four for the purpose of resource description; a Bibframe work, instance and item broadly equate to the MARC 21 Authority, Bibliographic and Holdings formats. Practical benefits of this may include a simplified mapping and transformation process from MARC 21 to Bibframe and vice versa. But whether data is structured using the LRM/RDA or Bibframe entity model, it is important that it can coexist effectively for the mutual benefit of the bibliographic community as a whole; the Share Family initiative offers a means by which the two could be made interoperable in future.

KBR: We are internally still discussing which model suits our needs best.

NDL: The NDL creates authority data for classical works and some modern and contemporary works written in foreign languages. This authority data is created based on the FRBR model. At the NDL, we are still in the process of investigating the Works in the BIBFRAME model and have not yet decided whether they will be transferred to the BIBFRAME model.

NLF: We have been testing MARC work authority records and have been preparing for the clustering of works when the conversion to BIBFRAME happens. In our testing so far, the most promising results have been through the SVDE clustering tool. We follow the RDA WEMI model in our BIBFRAME-based data model, so we need both Works and Expressions, eventually.

10. *Could you please specify some best practices for creating serials metadata and distributing it in the Linked Open Data (LOD) format?*

BL: In terms of serials the ISSN data model offers one approach: https://www.issn.org/understanding-the-issn/assignment-rules/issn-linked-data-application-profile/.

KBR: Unfortunately, I don't have much experience with serials metadata myself.

NDL: Although we are not sure whether it is the best practice, the NDL uses the same format to create both books and serials. The bibliographic data is created using the MARC 21 format and distributed as Linked Open Data (LOD) with our own vocabulary (DC-NDL) based on the Dublin Core.

NLF: We implemented the RDA aggregate model for our BIBFRAME based data model so will be going with that for serials.

11. *What metadata standards are currently used for bibliographic records in the National Library (e.g., MARC, Dublin Core)? Are there any plans to adopt or transition to new metadata standards in the future?*

BL: The standards the British Library currently follows are listed in the table below:

Current standards

The table below provides an overview of the principal metadata standards followed by the Library for collection description.

	Published books, serials, maps, scores	Unpublished archives and manuscripts	Sound / video recordings
Description	RDA: Resource Description & Access Descriptive Cataloguing of Rare Materials International Standard Bibliographic Description	General International Standard Archival Description (ISAD(G))	Internal
Name authority	Library of Congress/ Name Authority Co-operative (LC/NACO)	NACO applied selectively	Internal
Subject classification	Dewey Decimal Classification		
Subject indexing	Library of Congress Subject Headings (LCSH) FAST (Faceted application of subject terms)	LCSH/FAST applied selectively	Internal
Encoding	MARC 21	Encoded Archival Description Text Encoding Initiative: Manuscript Description	Internal

We are migrating to Alma, an Ex Libris product. MARC 21 is being used for the time being.

KBR: We have MARC-based records that contain RDA terms.

NDL: At the NDL, the bibliographic and authority data are created using MARC 21 format. They are also distributed as LOD using our own vocabulary (DC-NDL) based on the Dublin Core. We have currently started discussing whether we could implement BIBFRAME.

DNB: We use different external vocabularies to describe our bibliographic data. You can see the namespaces in our records: https://d-nb.info/1286333938/about/lds.

NLF: We currently use MARC and RDA terms (collected into our Metadata Thesaurus (https://finto.fi/mts/en/) and will be adopting Official RDA in early 2025. We are preparing for a move to a BIBFRAME based data model.

12. How is metadata created or enhanced for digitized materials?

KBR: I'm not sure about the details, because digitization is handled by another department. But I know that some basic metadata fields are included in the raw .tiff images, such as the related bibliographic identifier.

NDL: The NDL creates metadata for digitised material based on the metadata of the original material, adding digitisation-specific metadata such as digitised manufacturer, date of manufacture, PID, and other relevant information.

13. Are there specific challenges in creating metadata for older or fragile materials?

KBR: I would say no specific challenges that I am aware of. It rather is that each type of document has their own specialties that might need to be catalogued and that are not always already part of standards such as RDA.

NDL: In my personal opinion, when we create metadata for older or fragile materials, it is necessary to ensure access to digitised content. In this case, access should not be limited to the digitised content provided by the library itself; it should also be considered to link to digital resources in other institutions. It is essential to preserve the original materials as much as possible while ensuring that users can easily access digital content.