Linked data applications of RDA in the Netherlands

RKD – Netherlands Institute for Art History
& Podiumkunst.net

Thank you very much for the invitation to explain two Dutch linked data applications of RDA. I would have loved to join you personally in Helsinki, but unfortunately that is not possible.

Let me first briefly introduce myself. I am Lian Wintermans, chair of the Dutch RDA Committee and I work as an information specialist linked open data at the RKD and in that capacity I am involved in both RKD Research and Van Gogh Worldwide, two linked data platforms managed by the RKD. I am also an independent music and information specialist and trainer. In this capacity, I made a leading contribution to the drafting of the RDA application profile for the Performing Arts Network in the Netherlands, Podiumkunst.net. Previously, I worked at the National Library of the Netherlands for almost a decade and was also involved in EURIG then, as some of you may remember. It is wonderful to be among you again!

National Digital Heritage Strategy & Reference Architecture

https://netwerkdigitaalerfgoed.nl/

The two applications I am going to talk about have some similarities, but otherwise differ in many ways. Let me first say a few words about the shared principles and then I will explain each of the two applications individually.

The main similarity is that both applications implement the Dutch National Digital Heritage Strategy, published by the Digital Heritage Network and the Ministry of Education, Culture and Science. Both applications follow the Digital Heritage Reference Architecture and are set up according to linked open data principles. Another similarity is that both use official RDA.

I will first talk about the RDA application at Podiumkunst.net and then say something about the one at the RKD.

Podiumkunst.net

Vision

Enable the performing arts sector in the Netherlands to share their digital collections and stimulate reuse

In January 2021 the Podiumkunst.net consortium was founded with a 4-year grant of the Dutch Ministry of Culture. It consists of eight leading partners in the music, theatre and dance domain. Two of them – the Institute of Sound and Vision (Beeld en Geluid) and Muziekweb – merged in 2022.

The consortium's vision is to enable the Dutch performing arts sector to share their digital collections and stimulate reuse. To achieve this four main activities were defined: support the network and help it grow, develop shared services, increase expertise, and stimulate creative reuse.

The consortium's activities and projects are broad and diverse, and there is much interesting to say about them. In this presentation, I limit myself to the application of RDA and the work of the application profile working group.

In 2022, Podiumkunst.net chose RDA as its data model for linking performing arts collections and in the next year, the application profile working group started. This working group consists of people with different backgrounds, namely linked open data, information architecture, RDA, cataloguing, music collections (both notated and performed music), and theatre collections. Of course, there is no one person who has knowledge in all these areas, but we are lucky to have a number of people in our midst who combine multiple areas of knowledge.

Performing arts collections

podium
kunst
.net

4

One of the interesting things we had to take into account when creating the application profile was the great diversity of the collections to be connected, not only in terms of content, but also in terms of size, data granularity and degree of accessibility.

This obviously included the collections of the consortium partners and other cultural heritage institutions, often with a long tradition of description and access. But we also had to take into account archival collections of music, theatre and dance companies, often hardly or not at all described.

Or, put another way, in connecting the collections, we are not only dealing with transformations from RDA implementation scenario C to RDA implementation scenario A, but also with new data that is yet to be recorded, native RDA-LOD so to say, and that opens up new possibilities for connecting the collections and encourages to commit ourselves to an application that makes the most of the linked open data opportunities in RDA.
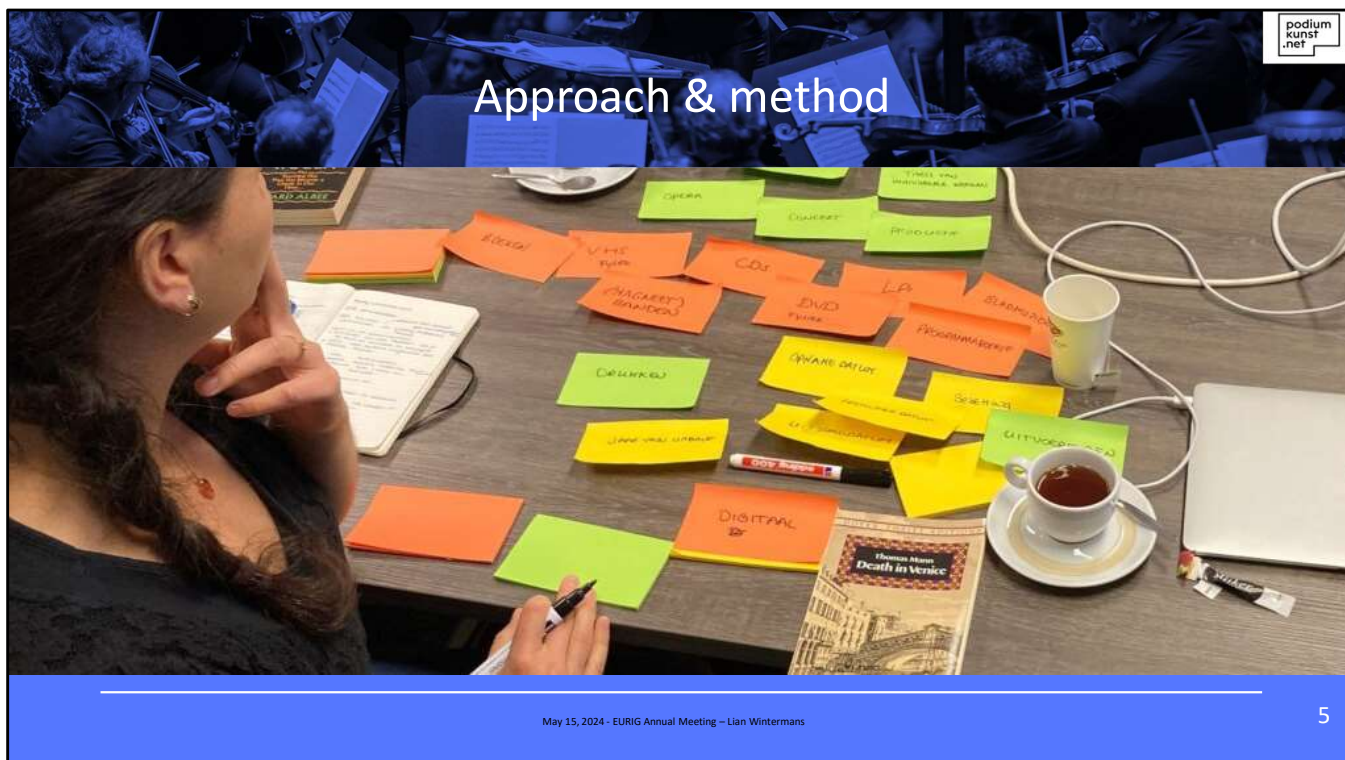
The working group began by confirming the use of RDA implementation scenario A and by formulating a number of principles in the application of that scenario, such as using the IRI as the preferred recording method, and avoiding shortcut elements, inverse elements and soft deprecated elements. And although all entities in RDA are relevant to our application, the working group decided to start with the resource entities, that is Work, Expression, Manifestation, and Item. A separate working group had been formed for the selection of vocabularies, therefore the AP working group limited itself to the RDA vocabularies and those that was already decided upon.

We also discussed the functionalities of the application profile itself. What data should it contain, in what contexts should it be used and by whom. It was quickly clear to us that we wanted to use the application profile for different functions (including validation and data entry), and that it had to be usable by both humans and machines. We also decided to use the csv files in the RDA Registry, and to include explicitly which data elements are not allowed in our application.

The Podiumkunst Application Profile

The first step we took was to reduce the number of elements to be used. We did this mainly by looking at the semantic relationships between the elements in RDA and excluding certain groups of elements. For example, we decided to always record relationships between resources and agents in the domain of the resource entities and to use only the elements with Agent as range, that is to say, we excluded the narrower elements that have an Agent subclass as range. When recording agents, the specific subclass will be used to indicate whether the Agent is a Person, Corporate body or Family.

In theory, this approach could be taken much further, as all classes in RDA are subclasses of the RDA Entity class, but so far we have limited this to the Agent subclasses and to the subject relationship.

We also discussed specific topics, such as representative expressions and access points. We concluded that representative expressions are relevant for our users and that we will record them as Work properties instead of marking specific Expressions as "representative". We excluded access point elements from our application, arguing that in a linked data application they are not needed for the identification of works, expressions, or agents, and that strings like access points can be generated automatically according to a string encoding scheme.

The next step that helped us get a grip on the huge amount of data elements in RDA was to apply levels of description in our application profile. In doing so, we followed the RDA classification into minimum description, coherent description, and effective description. The effective description is further divided into a 'general' section and into specific sections for specific types of Works, Expressions etcetera. We consider these levels of description as 'layers' in our application profile. You could also say that we created a set of interrelated application profiles that can be stacked on top of each other.

Finally, we worked out some examples in RDF. This step provided relevant input for the effective description and helped us evaluate the established principles and the decisions taken earlier.

Next steps

- Elaborate the different layers of the effective description
- Formulate entity boundaries
- Work out how collections with less detailed descriptions can be included in the graph
- Add relevant examples, templates and guidelines
- Add application profiles for the Agent class (and/or its subclasses)

By definition, an application profile is a living document. In the coming period, we want to continue working on the following issues:
- elaborate the different layers of the effective description;
- formulate entity boundaries;
- work out how collections with less detailed descriptions can be included in the graph;

This is about adding second (or third) best options for cases where the data cannot meet the recommendations in the effective description layer of the application profile. There are roughly three methods to deal with this: use a broader element, use a different registration method, or use a different (type of) element, e.g. a note or a manifestation statement instead of a relationship element. We would like to give users guidance in this, also to keep the Podiumkunst.net knowledge graph as usable and manageable as possible.

And we would like to:
- (continue to) add relevant examples, templates and related explanations;
- add application profiles for the Agent class (and/or its subclasses).

We will do this by means of use cases and practical examples from partners in the network.

More information

- The application profile: https://podiumkunstnet.github.io/ApplicationProfileRDA/

- Podiumkunst linked data: https://podiumkunst.triply.cc/

- Podiumkunst knowledge base:
https://www.podiumkunst.net/kennisbank/standaardisatie/

- Dataset Dutch classical musical works:
https://data.muziekschatten.nl/zoekmodule_NL_klassiek.html

- Medium of performance vocabulary:
https://data.muziekschatten.nl/zoekmodule_uitvoeringsmedia.html

Here you will find more information about the application profile and some other links that might be of interest. Now that there are so many online translation tools, I hope it is not a problem that the texts are almost entirely in Dutch. Please, feel free to share and reuse!

Note in case anyone asks for more information on the Medium of Performance vocabulary: The hierarchical structure is derived from that of the Dewey Decimal Classification, and the terms are an amalgamation of the IFLA/UNIMARC list, several existing Dutch vocabularies and – last, but not least - the MIMO (Musical Instrument Museums Online) thesaurus, which represents all musical instruments present in collections around the world and contains about 2,500 instrument names. Our Medium of Performance thesaurus is multilingual, with terms in Dutch, English, German, French, Italian and Spanish, among others. It is available as linked open data.

About the RKD

The RKD is the knowledge centre for the visual arts of the Low Countries and provides worldwide access to knowledge, research and information to museums, academic community and general public

RKD, the Netherlands Institute for Art History is one of the world's leading documentation and research institutes. As a knowledge institute, we work closely with museums, heritage institutions, universities and private institutions. In addition to a research function, the RKD has a collection and a public function. Our staff contribute to research, publications and exhibitions. We manage unique archives, documentation and photographic materials and the largest art history library on Western art from the late Middle Ages to the present, with a focus on the art of the Low Countries in an international context.

Collections

The RKD houses over 2.5 kilometers of archival materials in the field of the visual arts in the Netherlands. In total, it comprises more than 900 archives of artists, artists' associations, art historians and other art-related individuals and companies.

The RKD has more than five million digital and analogue images of works of art, making it one of the largest collections of visual documentation in the world. It also houses collections relating to portrait iconography and topography, and its archives also contain large quantities of historical photographs.

The RKD houses the most important library relating to the visual arts of the Low Countries in an international context, from the late Middle Ages to the present. The library contains more than 500,000 volumes, ranging from general art-historical literature, art dictionaries, monographs, artist's books, and periodicals to early and rare editions. About a third of the collection consists of exhibition catalogues and another third of auction catalogues.

RKD databases
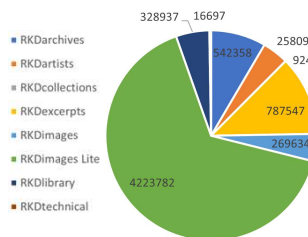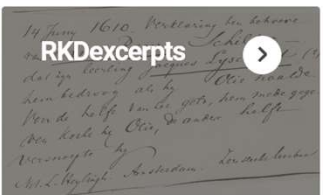
Pie chart legend:
- RKDarchives
- RKDartists
- RKDcollections
- RKDexcerpts
- RKDimages
- RKDimages Lite
- RKDlibrary
- RKDtechnical

Chart values: 328937, 16697, 258095, 924, 542358, 787547, 269634, 4223782

Tiles: RKDartists · RKDcollections · RKDexcerpts · RKDimages · RKDimages Lite · RKDlibrary · RKDtechnical · Visual search — Search the databases using an image

The RKD collections are accessible through eight interrelated databases. The databases are managed in the Adlib system (which will be replaced by its successor Axiell Collections later this year). One of the advantages of this system is that the various databases can be easily linked to one another. One of the drawbacks, at least for the library database, is that the data structure does not align with common record formats, such as the MARC standard. Within the total amount of data, the number of library records is rather modest, as you can see here.

Besides these eight databases, the system contains other databases (thesauri) for (among others) concepts, places, exhibitions and auctions, which also play an important role in connecting the data.

RKD Research
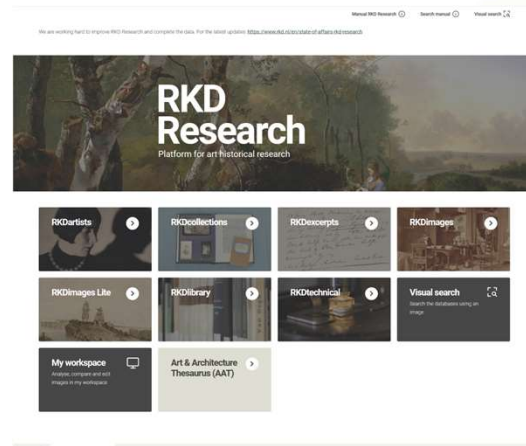
The RKD databases are accessible to the public via the RKD Research platform. This new platform, based entirely on linked open data, was launched last December. So, the reason for this RDA application was a very practical one: the previous platform, RKD Explore, had to be replaced before the end of last year. The same source data had to be included in the application profile and the same functionalities had to be supported.

Note, if necessary, that linked open data was chosen for several reasons, including to better serve users by leaving room for not yet known user queries.

Besides RDA (for the library database), the RKD uses other ontologies for modelling the data, namely Records in Contexts for the archival collections, and Linked Art for all other data, except concepts, which are modelled according to SKOS.

If necessary, note that general decisions regarding the linked open data also had an impact on the RDA application. For example, the RKD uses rdf:list to retain the original order of data. This is applied to all repeatable source data fields, including those in the library database.

RKDlibrary & RDA

Here you see an example of (left) part of the input screen of a library record in Adlib, (center) the XML output from the web API for that same record, and (right) part of the resulting RDA-RDF data for that record.

Not surprisingly, one of the trickiest things about modelling the data was the identification of works, expressions, manifestations and items.

We have assumed that each record in the library database represents a unique manifestation. And while this assumption is largely correct, there are also cases where a record represents multiple manifestations (e.g. a physical resource and its digital version) or where, on the contrary, multiple records represent the same manifestation (see e.g. https://rkd.nl/library/400002112 and https://rkd.nl/library/400002113).

It was practically impossible to identify expressions and works in the library records. Therefore, we decided not to identify expressions and use the shortcut element "work manifested" (rdam:P30135) to relate the manifestation directly to a work. We assumed that each record in the library database represents a unique work. And while we know that this assumption is not always true, we assume it is true for the majority of records.

We also had difficulties in identifying items. The holdings records in our system represent either a proper item or a part (that is, an individual volume or issue) of a multi-part resource. We tried different solutions, but in the end we had to accept that all holdings records are related to the manifestation using the same property.

We encountered another difficulty in the source field "author". It was clear that the values in this field did not always match the author relationship as defined in RDA. In some records, a role field is also used, but because this data is recorded as an unstructured description (transcribed from the resource), it was not possible to convert the data to the matching relationship elements in RDA. We therefore used the broader element "creator".

Next steps

- Improve identification of Works, Expressions, Manifestations, Items

- Examine impact on RKD cataloguing practices, update cataloguing manual, familiarize cataloguers with data model

- Some data cleanup

- Add values from selected RDA value lists

- Decide what to do with square brackets, ISBD punctuation etc. in data values

In the coming period, we will work on the following issues:
- find out whether the identification of Works and Manifestations (and preferably Expressions as well) can be improved;
- examine the impact of the RDA linked data application on the RKD cataloguing manual and practices, and familiarize cataloguers with the data modelling in RKD Research;
- do some data cleanup;
- find out whether values from certain specific RDA value lists (such as Content Type, Media Type, Carrier Type, Extension Plan and Mode of Issuance) can be included;
- decide what to do with data values that include square brackets and ISBD punctuation.

More information

- RKD Research: https://research.rkd.nl/
- RKD Knowledge Graph: https://rkd.triply.cc/
- Questions? wintermans@rkd.nl

The RKD's application profile is not publicly available. It currently takes the form of specifications for the data transformation, supplemented by some worked-out examples to verify the transformation. However, the RKD Research platform and the underlying RKD Knowledge Graph are accessible to everyone.

Concluding remarks

- Every application its own profile!

- Modelling of specialist collections can provide the RDA community with interesting feedback

- Great need for reliable datasets, especially for Works and Expressions

- Need for linked data examples using RDA and for a place to discuss data modelling issues

Thank you for your attention!

I would like to end with some concluding remarks:

- indeed: every application needs its own profile;
- these two applications relate to specialist collections that sometimes (have to) push the edges of RDA and thus can provide the standard with interesting feedback;
- there is a great need for reliable datasets, especially for works and expressions, and I wonder what role national libraries and bibliographic agencies see for themselves in this;
- there is also a need for linked data examples using RDA and for a place to discuss data modelling issues.

Thank you for your attention!

RKD – Nederlands Instituut voor
Kunstgeschiedenis

Prins Willem-Alexanderhof 5
2595 BE Den Haag
Nederland

Contact
T: 070 3339777
E: info@rkd.nl
W: www.rkd.nl

in 📷 f