

Changing theory into practice: playing the Metadata Game

*Jorien Weterings MA
Advisor Recordkeeping at the National Archives of the Netherlands*

Abstract

This paper describes how the international standard on metadata for records ISO 23081 has recently been incorporated in the Dutch legislation on archives and gives an introduction to initiatives that have been pursued by the central government to develop metadata schemas and implement application profiles. The National Archives of the Netherlands has developed an e-Depot (Trusted Digital Repository). For the transfer of electronic records to the e-Depot, the records and the metadata linked to them are wrapped as a Submission Information Package (SIP) by the use of an XML-schema called MeDuSa. This XML-schema defines the structure of the SIP as well as a set of metadata elements, and is developed simultaneously and compliant with the metadata schema for the Dutch government. To make colleagues, archivists, records managers, information architects and others familiar with the most important concepts concerning metadata, we developed the Metadata Game.

Contents

- Abstract 1
- 1. Introduction..... 2
 - 1.1. Purpose of this paper 2
 - 1.2. Metadata, why again? 2
- 2. Dutch context: legislation, standards and application profiles 3
 - 2.1. Metadata in the Ministerial Regulation on Archival Records 3
 - 2.2. Applying the ISO 23081 standard in the Netherlands..... 4
 - 2.2.1. A metadata schema for the Dutch government..... 4
 - 2.2.2. A generic application profile for the central government..... 5
 - 2.2.3. Specific ‘in-house’ application profiles..... 5
 - 2.2.4. Current state of affairs 6
- 3. Metadata at the National Archives 7
 - 3.1. MeDuSa: an XML-schema for the e-Depot..... 7
- 4. Playing the Metadata Game..... 8

1. Introduction

The digital environment of the 21st century requires proper use of metadata to ensure long term authenticity, accessibility and sustainability of electronic records. International organizations, such as the International Standardization Organization (ISO) have developed standards and guidelines to support agencies with understanding metadata concepts and implementing metadata schemas for records management.

In accordance with the international standard ISO 23081 (Information and documentation – managing metadata for records), the Dutch central government developed an application profile that identifies metadata elements and provides guidelines for the use of metadata in records systems. The National Archives of the Netherlands translated these national guidelines into an XML-schema that is used for the ingest of electronic records in the e-Depot (Trusted Digital Repository) of the National Archives.

1.1. Purpose of this paper

This paper gives an example of the Dutch practice, by first outlining the Dutch legal context in respect to metadata and the development of application profiles according to ISO 23081. It will then give a glimpse into the application of metadata in the e-Depot of the National Archives of the Netherlands. The final chapter describes the Metadata Game that we developed to explain the general concepts of metadata and the necessity of metadata schemas and mappings.

1.2. Metadata, why again?

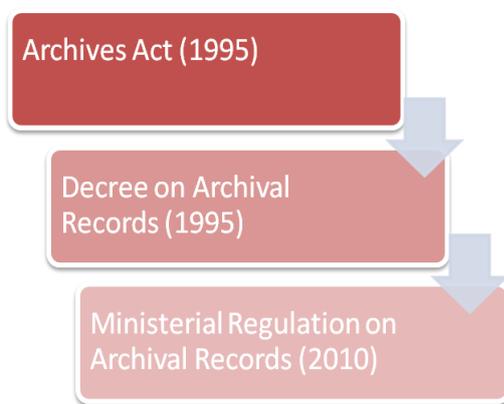
A commonly heard answer to the question: ‘What are metadata?’ is: ‘Data about data’. In a records management context, this definition is rather thin. The ISO-standards define metadata as: ‘data describing the context, content and structure of records and their management through time (ISO 15489-1:2001, 3.12). As such, metadata are ‘structured or semi-structured information that enables the creation, registration, classification, access, preservation and disposition of records through time and within and across domains’ (ISO 23081-1:2006, 4).

Metadata are not only essential to ensure authenticity, integrity, usability and reliability of records, they’re equally indispensable for interoperability. This means that records can move easily between information systems and organizations. In order to support this exchange of records between systems, arrangements have to be made about the way in which this exchange takes place, in what ‘language’ systems communicate (like XML) and how metadata and records should be interpreted. When it comes to metadata, these arrangements are made in metadata schemas. Arrangements about definitions and semantics of metadata values can be made in so-called encoding schemes. These are ‘authoritative sources, including pre-defined lists, classifications, controlled vocabularies or taxonomies. Using encoding schemes that are formally documented aids in ensuring the quality and consistency of metadata values’ (ISO 23081-2:2007, 10.3.3).

Exchange of records and metadata between systems can occur within an organization as well as between organizations, for instance when electronic records and their custody are transferred to the National Archives.

2. Dutch context: legislation, standards and application profiles

First of all, I would like to give a brief introduction to Dutch legislation on records management, with a focus on metadata for records. Government organizations in the Netherlands have to fulfill a number of legal obligations regarding the creation and management of their records. The *Wet Openbaarheid van Bestuur* (Freedom of Information Act), the *Wet Bescherming Persoonsgegevens* (Personal Data Protection Act) and the 1995 *Archiefwet* (Archives Act) all include provisions about how the government should deal with the information it produces and receives.¹ The Archives Act applies to all 'government bodies'. These are defined as 'any body forming part of a legal entity established under public law and any other entity or body entrusted with public authority'. In the Netherlands, government comprises the central government and three other tiers: provinces, municipalities and water board authorities. Central government operates at a national level, the other tiers at regional or local level.²



A number of articles contained in the Archives Act are elaborated in its principal executive instrument, the 1995 *Archiefbesluit* (Decree on Archival Records). More detailed aspects concerning the long term sustainability and accessibility of records are elaborated in the Ministerial Regulation on Archival Records that has been published in 2010: the so-called *Archiefregeling*. This regulation concerns both analog and electronic records that are to be preserved permanently.

2.1. Metadata in the Ministerial Regulation on Archival Records

The above mentioned Ministerial Regulation is where the rules concerning the use of metadata for records are delineated. Chapter three contains several articles regarding metadata. Basically these come down to the following principles:

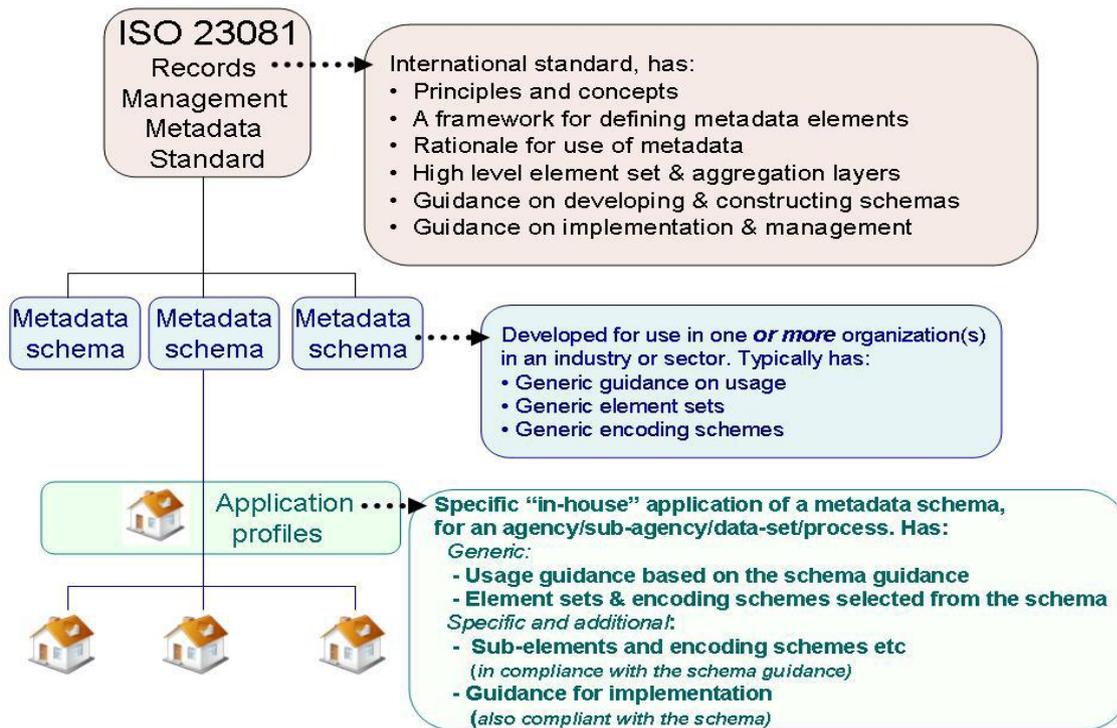
- A government body should develop a metadata schema in accordance with the international standard for metadata for records ISO 23081.
- In order to attest the authenticity and creation context of records, metadata should provide for essential information about content, structure and type of the records, their provenance (when, by whom and within what processes are the records received or created and captured?), their relation with other records, the recordkeeping processes that are performed on the records and the software used to manage these records.
- When it comes to electronic records, metadata should provide for information about the initial and current technical specifications of the records as well as the hard- and software that has been used to create and manage them, in order to support reproduction at all times.
- Records should be made accessible and interpretable within a reasonable time-frame, for example through the use of metadata linked with the records.
- Records metadata should be persistently linked with electronic records.

¹For an English version of the Dutch Archives Act: http://en.nationaalarchief.nl/sites/default/files/docs/wetten-regels/Dutch_Public_Records_Act_1995.pdf.

²For an overview of the organization of the Dutch government, see: <http://www.government.nl> and <http://overheid.nl/english/aboutgov/government>.

2.2. Applying the ISO 23081 standard in the Netherlands

As described in the previous paragraph the Dutch ministerial regulation stipulates that a government body should design a metadata schema in accordance with ISO 23081. The standard defines a metadata schema as a ‘logical plan showing the relationships between metadata elements, normally through establishing rules for the use and management of metadata specifically as regards to the semantics, the syntax and the optionality (obligation level) of values’ (ISO 23081-1:2006, 3.3). The ISO committee that developed ISO 23081 published a very useful document on how to ‘build’ a metadata schema³. It explains the relationship between ISO 23081, metadata schemas and application profiles with the diagram below:



2.2.1. A metadata schema for the Dutch government

In 2009 the *Richtlijn metagegevens overheidsinformatie* (Guideline on metadata for government information) for the Dutch government was developed and published⁴ as a part of the Dutch *e-Government Reference Architecture*.⁵ According to the ISO 23081 definition this guideline is in fact a metadata schema, with generic guidance, developed for use by organizations in every tier of the Dutch government sector.

The guideline describes:

³ K.P. O’Kane, *Creating a metadata schema – where to start?* (2008). The document can be downloaded at <http://isotc.iso.org/livelink/livelink?func=ll&objId=8800147&objAction=browse&sort=name> (in English).

⁴ The description of elements and entities of this guideline can be downloaded at: <http://kennislab.files.wordpress.com/2010/01/richtlijn-metagegevens-overheid-schema-entiteiten-en-elementen-versie-2-5.pdf>, and the commentary on the guideline can be downloaded at: http://www.nationaalarchief.nl/sites/default/files/docs/richtlijn_metagegevens_overheid_toelichting_0_0.pdf (both in Dutch only).

⁵ <http://www.e-overheid.nl/onderwerpen/e-overheid/architectuur> (in Dutch only).

- what entities can be distinguished. An entity can be any concrete or abstract thing that exists, did exist, or might exist, including associations among these things (ISO 23081-2:2009). The guideline distinguishes the following entities: record, agent, (business) process, mandate and relationship.
- what metadata elements can be distinguished, including semantic definitions;
- the obligation level of metadata elements for each of the entities; obligation levels can be: mandatory, mandatory if applicable, comply or explain, recommended and optional.
- whether, and if so what sub-elements are required to further specify a metadata element;
- what aggregation levels can occur within an entity (e.g. 'archive', 'series', 'case file', and 'record' are aggregation levels of the entity 'record'; 'legislation', 'policies' and 'business rules' are aggregation levels of the entity 'mandate').

2.2.2. A generic application profile for the central government

The guideline is a mere framework that will not be applied as such, but needs to be translated into a dedicated application profile. An application profile 'delineates the use of metadata elements declared in an element set. While an element set establishes concepts, as expressed via metadata elements, and focuses on the semantics or meanings of those elements, an application profile goes further and adds business rules and guidelines on the use of the elements. It identifies element obligations and constraints, and provides comments and examples to assist in the understanding of the elements. Application profiles may include elements integrated from one or more element sets thus allowing a given application to meet its functional requirements.'⁶

In 2009 a commonly accepted generic application profile for the central government has been developed. This application profile describes a consensus based minimum set of metadata elements and definitions including which ones are mandatory and which conditional or optional. It is part of the so called *Baseline Informatiehuishouding Rijksoverheid* (Baseline for Information Management of the Central Government), a government-wide set of standards and guidelines for information management that promotes the accessibility and reliability of government information.

2.2.3. Specific 'in-house' application profiles

Each ministry has to translate the generic application profile for the central government into more specific application profiles for each organizational unit, a working area, a function or a chain of functions, processes or organizations. The application profile for the central government defines the following steps that could be undertaken to develop an organization-specific application profile:

1. The ministry determines what metadata elements are mandatory, in addition to the elements that are already mandatory in the generic application profile. The ministry decides whether elements that are mandatory if applicable in the generic application profile should be mandatory in their own profile and chooses what elements of the recommended, optional or 'comply or explain' levels should be applied.
2. The ministry can add sub- and sub-sub-elements to the basic element set of the generic application profile. However, it is not allowed to add basic elements or to add sub-elements that don't fit within the original definition of the basic element as this would interfere with the aim for interoperability.

⁶ O'Kane (2008), p. 2.

3. Elements, sub-elements and sub-sub-elements can be refined by a description of the syntax or technical choices.
4. The ministry looks after unambiguous semantic interpretation of (sub)elements by using encoding schemes (controlled list of all the acceptable values in natural language and/or as a syntax-encoded text string designed for machine processing; e.g. controlled vocabularies).
5. The ministry looks after unambiguous semantic interpretation of (sub)elements by determining specific sources for specific data.
6. The ministry can choose to use other element names than those in the generic application profile (e.g. 'agent' instead of 'actor'). In that case however, a mapping must be made to describe to which element in the generic application profile the element name refers. This supports the interoperability.
7. The ministry decides how many entities are applied. The Dutch guideline on metadata for government information lists the five entities of Actor, Record, Business, Mandate and Relationship. The generic application profile is based upon a so-called one-entity model that relates all metadata to the entity 'Record'.
8. For some elements, the ministry may refer to existing registration systems that store metadata. A persistent and reliable association of the metadata to the record, however, remains a stringent requirement. The organization thus needs to meet the following conditions:
 - a. the organization describes exactly and clearly how the metadata will be related to the records;
 - b. the relationship between records and metadata (within or between systems) has to be persistent and has to remain persistent as well when systems are replaced or when data is migrated from one system to another;
 - c. both records and metadata meet records management requirements.

2.2.4. Current state of affairs

As explained in paragraph 2.1 the Ministerial Regulation on Archival Records requires that every government body should design a metadata schema in accordance with ISO 23081. This requirement is met as soon as an organization has developed its own application profile. The generic application profile for the central government was approved by the board of CIO's at the end of 2009. In the meantime, several specific application profile initiatives have started, for instance by *Rijkswaterstaat* - the department of Waterways and Public Works of the Dutch Ministry of Infrastructure and the Environment.

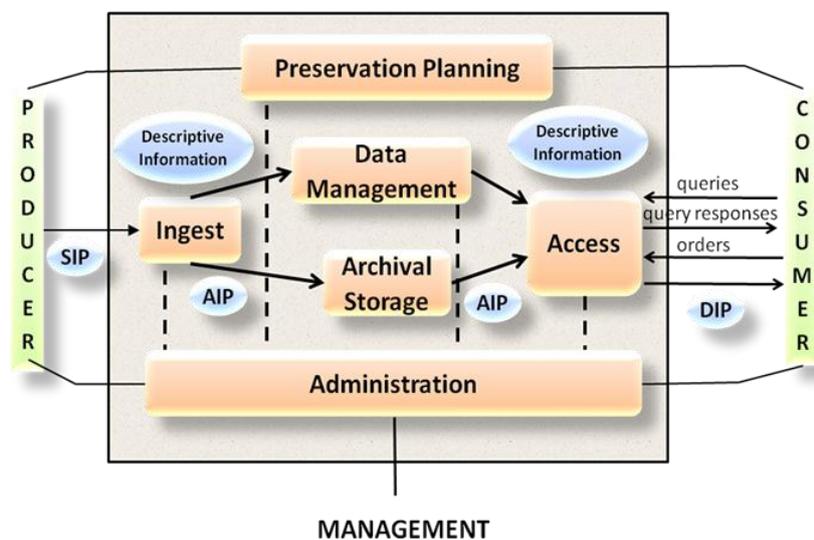
Apart from the central government there are other initiatives to be mentioned, for instance the *Werkgroep Voorbereiding Implementatie e-Depot (WVI)* - a collective working group of the eleven Regional Archives in the Netherlands and the National Archives. The WVI develops business rules, describes processes, defines functions and user roles and designs a generic architecture that can be used for the implementation of a Trusted Digital Repository for the Regional Archives. One of the deliverables is a generic application profile for the provinces, municipalities and water board authorities that are affiliated with the Regional Archives. This application profile will be heavily based

on the application profile for the central government. Furthermore the *Kwaliteitsinstituut Nederlandse Gemeenten* (KING: quality institute for Dutch municipalities) will investigate the possibilities of re-using the application profile for the central government for a generic application profile for the municipalities as well.

Unfortunately, the subject does not seem to have a high priority on the records and information management agendas. By participating in working groups, expert meetings and through lobbying the National Archives tries to ‘spread the word’, as one of the parties involved. The Metadata Game – as I will describe in chapter 4 – may be a useful instrument to address the issue and explain the main concepts.

3. Metadata at the National Archives

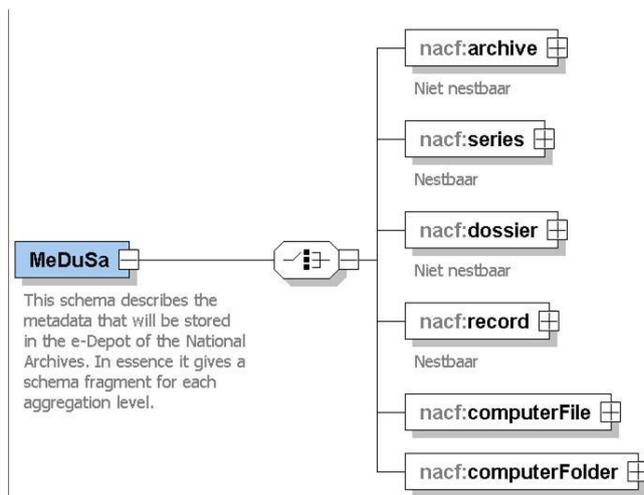
The information architecture of the National Archives e-Depot is based upon the concepts of the international standard ISO 14721 *Open Archival Information System*. The image below outlines the process of acquiring/ingesting, managing and representing records within this architecture. Records transferred to the e-Depot are ingested as a so-called Submission Information Package (SIP). This SIP contains the records and the metadata linked to those records. To create a SIP, we use an XML-schema that defines the structure of the SIP as well as a set of metadata elements that are linked to the records on different aggregation levels. This XML-schema is called MeDuSa⁷.



3.1. MeDuSa: an XML-schema for the e-Depot

MeDuSa was developed simultaneously with the *Guideline on metadata for government information* and the *Generic application profile for the central government*, and thus contains all metadata elements that occur in the other two. The *structure* of the XML-schema allows to link metadata on different aggregation levels:

⁷ The schema and its documentation can be downloaded at <http://www.nationaalarchief.nl/onderwerpen/overbrenging-vervreemding/het-overbrengen-digitale-overheidsarchieven/e-depot/MeDuSa> (Dutch only).



This schema describes the metadata that will be stored in the e-Depot of the National Archives. In essence it gives a schema fragment for each aggregation level.

- Archive
- Series
- Case file
- Record
- Computer file
- (Computer folder)

The level 'Computer File' is not an aggregation level in archival terms, but since a record can consist of one or more computer files (for instance ten TIFF-files that together form one digitized letter that is a record), it is necessary to relate

metadata on that entity level as well. This would mainly be technical metadata about for example application software for rendering, file format or integrity of the computer file. 'Computer Folder' is a separate category as well, designated to retain the original folder structure of the original data. This occurs for complex objects that in their original form already consist of a folder structure, like websites.

The image above is a view of the XML-schema; each of the aggregation levels is expandable and contains the metadata elements. MeDuSa consists of a set of generic metadata elements that can be used for archives of any record creator, public or private. As said, these elements are the same as those described in the guideline and application profile. This supports interoperability with organizations that apply the Dutch guideline. However, only a few of these elements are mandatory in MeDuSa since the National Archives acquires not only the public archives of the government, but also private archives that are not bound to archival legislation or application profiles. Some names of the metadata elements in MeDuSa differ from the names of the elements in the guideline and application profile, their meaning however remains the same.

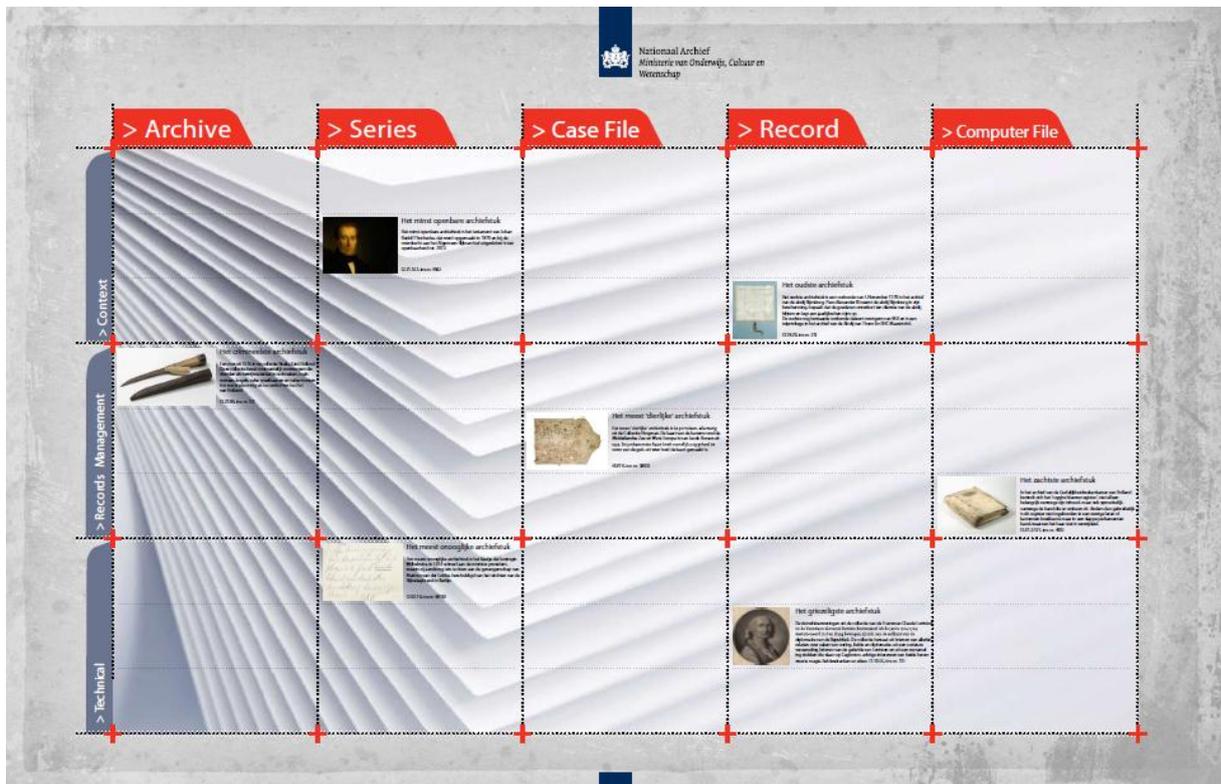
Apart from the generic set of metadata elements that is applicable for any organization, MeDuSa also has a container for 'Agency Specific' metadata. This container can enclose any other schema, which allows validation to other structures than MeDuSa.

4. Playing the Metadata Game

From time to time we heard jokes like: "I've just returned from the doctor to get a treatment for my metadata" and "This metadata-ache has been bothering me for ages now!". There was a frequent buzz that all this metadata stuff was way too theoretical. Apparently – and understandably – people find it hard to understand why metadata schemas are crucial in an environment where exchange of information between systems and organizations becomes more and more common (and necessary). Besides, it's not always obvious how to apply metadata schemas. One year ago my colleague Gijsbert Kruithof⁸ and I decided to add some fun to all this theory and technique and make it more tangible. Thus we developed a game to explain the metadata basics. We simply call it the Metadata Game.

⁸ Gijsbert Kruithof is information architect at the department of Information, Infrastructure and Innovation of the National Archives.

The purpose of the Metadata Game is to explain the most important concepts concerning metadata. It shows the difference between a metadata schema (a set of arrangements) and an XML-schema (an instrument), and it demonstrates the importance of standardization. We also want to show how a mapping of the metadata from the source system of a records creator to the target system – the e-Depot of the National Archives – can be made. By playing the game, we aim at educating colleagues that have to give advice on creating metadata schemas or support records creators with the transfer of their electronic records to the e-Depot.



The game consists of an A0-size magnetic game board divided in five columns and three rows (see image above) and three sets of ten magnetic cards. The columns indicate the aggregation levels archive, series, case file and record, as well as a level for computer file (as explained in 3.1). The rows indicate three different types of metadata⁹:

- metadata about the (business) *context* : when, by whom and within what processes are the records received or created and captured, their relationship with other records and processes etc.;
- metadata about the *records management* processes that are performed on a record or aggregation of records (e.g. migration, destruction, change of access rules);
- *technical* metadata that provides for information about the technical specifications of the records as well as the hard- and software that's been used to create and manage them (e.g. file format, integrity algorithm, creation application).

⁹ This trichotomy is based upon Hans Hofman, 'Een uitdijend heeal? Context van archiefbescheiden', in: *Context. Interpretatiekaders in de archivalie* (2000), pp. 55-65.

The 'playing cards' contain a selection of ten metadata elements that are part of the following three different metadata sources:

- the application profile for the central government;
- the XML-schema for the e-Depot (MeDuSa);
- an example metadata export .

The example metadata export derives from a set of metadata elements that was transferred with electronic records of the Ministry of the Interior to the e-Depot of the National Archives. The 'playing cards' of this set only contain the element tags, not the values. The course of the game follows three steps:

1. First make a 'mapping' by making combined sets of elements of the three metadata sources that have the same meaning. This results in ten sets of mapped metadata elements.
2. The next step is to place these sets in the column on the highest aggregation level on which they can occur (even when they can occur on a lower level as well).
3. Finally - or simultaneously with the previous step - place the sets in the row that correctly indicates the type of metadata they concern.



Some of the results of these steps are unambiguous, but we purposely left some of the outcomes open for discussion as well. These discussions reveal the importance of determining and documenting the *meaning* of metadata elements, in order to translate metadata elements from one system to metadata elements in another system and thus enable exchange of information. They also show that it depends on an organization's specific implementation of a metadata schema on what aggregation level metadata elements are linked to the records (what rules for the inheritance of metadata values are defined?).

Until now we've played the game with groups of colleagues at the National Archives and with colleagues from the Regional Archives – with a maximum of five small groups of four to five participants at a time. Beforehand we couldn't quite estimate the effect of the game, but one of the best results was undoubtedly that, while playing the game, the participants within one small group started to explain the most important concepts to each other. The overall feedback we received from the 'gamers' so far is very positive.¹⁰ Most of them were even convinced that metadata can indeed be fun! The game does not immediately train them in making a metadata schema or a metadata mapping, but by focusing on the basics it takes away uncertainty and makes the gamers familiar with existing standards and the relatively new requirements on metadata required by the Ministerial Regulation.

We intend to play the game with records managers, information architects, archivists etc. at the ministries and other organizations as well in the near future.

¹⁰ See for instance Christian van der Ven's blog: <http://www.digitalearchivaris.nl/2012/06/playing-metadatamappinggame.html> (in Dutch only).