

BOMOS: Linked data 3.0.0

Logius Guide

Definitive version June 07, 2022

**This version:**

<https://gitdocumentatie.logius.nl/publicatie/bomos/linkeddata/en/3.0.0/>

Latest published version:

<https://gitdocumentatie.logius.nl/publicatie/bomos/linkeddata/en/>

Latest editor's draft:

<https://logius-standaarden.github.io/BOMOS-LinkedData/>

Editors:

Gül Işık ([Logius](#))

Edwin Wisse ([Logius](#))

Author:

Wouter van den Berg ([TNO](#))

Participate:

[GitHub Logius-standaarden/BOMOS-LinkedData](#)

[File an issue](#)

[Commit history](#)

[Pull requests](#)

This document is also available in these non-normative format: [pdf](#)



This document is licensed under
[Creative Commons Attribution 4.0 International Public License](#)

Abstract

BOMOS (Beheer- en OntwikkelModel voor Open Standaarden) is een hulpmiddel van en voor de standaardisatiewereld. Dit deel bevat een aanvullende module over linked data. Het is een verkenning van de rol die linked data speelt en kan gaan spelen in moderne semantische standaarden.

Het beheer van dit document ligt bij TNO (de oorspronkelijke auteur). Geïnteresseerden en experts worden van harte uitgenodigd om feedback te delen, het document aan te vullen of anderzijds te verbeteren.

Status of This Document

This is the definitive version of this document. Edits resulting from consultations have been applied.

Table of Contents

Abstract

Status of This Document

Documentbeheer

Colofon

1. Introduction

- 1.1 Purpose
- 1.2 BOMOS structure
- 1.3 Reading guide

2. LINKED DATA & ONTOLOGIES

- 2.1 Emergence of Linked Data
- 2.2 Public sector
- 2.3 Private sector
- 2.4 The FAIR movement
- 2.5 Why Linked Data?
- 2.6 Impact on standardisation and management
 - 2.6.1 Reuse
- 2.7 Training
- 2.8 Selecting a compromise
- 2.9 Further developments
- 2.10 User analysis
- 2.11 Tools
 - 2.11.1 NeoNeoLogism
 - 2.11.2 PoolParty
 - 2.11.3 Protégé
 - 2.11.4 Semantic Treehouse
 - 2.11.5 TopBraid Composer
 - 2.11.6 VocBench
 - 2.11.7 Matey

§ Documentbeheer

Datum	Versie	Auteur	Opmerkingen
13/11/2020	1.0	Wouter van den Berg	Oplevering module
01/06/2022	2.0	Wouter van den Berg	Aanvullende modules in eigen documenten ondergebracht

§ Colofon

Logius Servicecentrum:
Postbus 96810
2509 JE Den Haag
tel. 0900 555 4555 (10 ct p/m)
email servicecentrum@logius.nl

§ 1. Introduction

The BOMOS Substantiation discusses the components that make up BOMOS. In addition to operational, tactical and strategic layers, it also deals with implementation support and communication. The BOMOS documentation is divided into various sections (see [1.2 BOMOS structure](#)).

§ 1.1 Purpose

The purpose of this publication is to assist organisations in compiling and improving the management of standards. This publication provides answers among others to the following questions:

- How can we successfully (further) develop and manage the standard in our organisation?
- How can be structure development and management in such a way that it results in an open standard?
- How can we improve the adoption rate of our standard among users?

These specific questions were the original reason for drawing up the Management and Development Model for Open Standards (BOMOS) with its best practice guidelines for an open structure for management. Since that time, BOMOS has been used in practice, and users have expressed the need to share more knowledge and experience with the management of standards. Other issues such as improving interoperability based on standards, transparency and the manageability of standards have been added. Finally BOMOS is now used as the common language in the world of standardisation.

§ 1.2 BOMOS structure

BOMOS consists of:

- BOMOS Part 1: [The Foundation](#)
- BOMOS Part 2: [The Substantiation](#)
- BOMOS Supplementary modules: Linked Data (this document) en [Stelsels](#)

The heart of BOMOS is the [Foundation](#). This consists of a basic description of the Management and Development Model and a further elaboration based on literature and experiences gained in practice. In essence, the Management and Development Model is an activity diagram which also offers a definition of the roles relevant in the process of managing and developing standards.

The *Substantiation* (this document) offers further insight in particular by sharing best practices from the world of standardisation.

Together Part 1 and Part 2 form the basis for BOMOS.

On top of this basic structure, the community has produced a number of BOMOS expansions which can be useful in deploying BOMOS in concrete situations, some of which may involve a slightly different context. We refer to these as the BOMOS Supplementary Modules or a Body of Knowledge, which will remain dynamic over time.

When we talk about BOMOS, what we are actually referring to is the basis as described in [Part 1](#) and Part 2. Although the supplementary modules are clearly linked to BOMOS, they have their own governance, which can result in their being given their own name, their own target group, their own management system, etc. The [BOMOS management process](#) also describes the requirements that are imposed before something can be added as a BOMOS supplementary module.

The first two supplementary modules are: - [Linked Data & Ontologieën](#): the specific use of Linked Data for semantic standards. - Structure for [BOMOS for the management of trust framework](#): the use of BOMOS in the specific situation governing trust frameworks.

§ 1.3 Reading guide

If from your policy making or administrative role you are only interested in the primary level, the foundation (part 1) will offer sufficient background and context. If however you are personally active in standardisation communities, you can seamlessly continue with reading part 2: The substantiation with best practices, which includes more background and practical tips for standardisation.

If you actually intend to make use of BOMOS, it is advisable that you also study the supplementary modules. These contain examples and tools that could prove useful for implementing open standards. The supplementary modules also contain variants on BOMOS. These implementation profiles make BOMOS suitable for use with more than just semantic standards.

§ 2. LINKED DATA & ONTOLOGIES

Linked Data has been an emerging trend for some time. The standards and technologies covered by Linked Data -- such as RDF, SPARQL en OWL -- are becoming increasingly widely used in practice. It started in the academic world but over the past few years has expanded into both the public and the private sector.

Linked Data results in changes to the semantic standardisation discipline described by BOMOS. This chapter explains those changes. It is not merely a question of using new technologies and tools. Linked Data means a different approach to semantic interoperability: the emphasis is shifted from standardisation of messages (or ‘transactions’) to the standardisation of the underlying domain description.

This chapter does assume a basic knowledge of the functioning of Linked Data. We start by discussing the differences between Linked Data and conventional semantic standards in terms of both management and further development and in application by users. Finally, we provide an overview of the digital tools that can be useful in the management and (further) development specifically of ontologies.

However, we will start by explaining the meaning of the word ‘ontology’. There is a clear distinction between what is designated in the world of Linked Data as ‘vocabularies’ and ‘ontologies’. [According to W3C](#) there is a clear trend of using the word ‘ontology’ for more complex and formal collections of terms, while ‘vocabulary’ is used for simpler variants or in situations in which less strict formalism is employed. In this chapter, the term ontology is used to refer to the entire spectrum.

Another fundamental element at the start of this chapter is the following distinction between the two types of semantic standards (see also figure 1):

1. Conventional semantic standards. These standards solve the interoperability problem by means of common agreements on the content structure of messages or transactions.
2. Semantic standards based on Linked Data. These standards solve the interoperability problem by recording a common language for a description of the elements or events in a domain (in an ontology). Thanks to this common language, the community is no longer dependent on fixed message structures for uniform interpretation of the shared data. (The exchanging of standardised messages does however remain possible if there is a demand for this practice within the community.)

§ 2.1 Emergence of Linked Data

In the past few years, the Linked Data approach has become increasingly common in both the public and private sector. This chapter describes initiatives in the public sector and private sector and one initiative that occupies the ground between the two: the FAIR movement.

§ 2.2 Public sector

The cultural sector has always been a pioneer within the public sector. The Europeana project (launched in 2008) was ground-breaking: Today, Europeana Collections offers access to more than 50 million digitised items - books, music, art and more - with advanced search and filter functions to simplify the search process. As in other sectors, here too, it started with the development of the models in which the metadata of collections could be described. Other cultural institutions have expressed different models in Linked Data, for example the Royal Netherlands Library with the Dutch Thesaurus of Authors' names and the Cultural Heritage Agency among others with its Archaeological Key Register.

In the field of building and infrastructure, so-called Object Type Libraries (OTLs) are a well-known phenomenon. An OTL is a type of semantic library within an organisation. ProRail and Rijkswaterstaat are well-known organisations that employ such libraries. There there is COINS, another important Linked Data based standard, intended for the exchanging of BIM models. Organisations in the geographical information sector were also early adopters, with Geonovum as the driving force behind the Platform Linked Data Nederland. The Land Registry (Kadaster) soon published the first models of the Key Registers for Addresses and Buildings, followed by other datasets. Following the publication of the models, the key registers themselves were published as Linked Data, at 'production' level, in other words with persistent URIs, both updated and

monitored. Moreover, on labs.kadaster.nl as a source of inspiration, a large number of examples are given of the possibilities offered by Linked Data. A vital follow-up step is that the NEN3610 (the semantic ‘parent model’ for all Dutch geoinformation models) has also now been made available as Linked Data.

These are just some of the many initiatives in the public sector, for example in agriculture and the organisations cooperating in the immigration process. Equally, the educational sector has taken numerous steps towards Linked Data based on ROSA, in which once again the semantic models are the starting point. There is not sufficient space in this chapter to list them all.

§ 2.3 Private sector

Linked Data developments are taking place at every level in the private sector: at the level of sector and chain processes, in individual companies and in specific business processes.

There are sectors and communities in the private sector in which the role of Linked Data is constantly growing, but for the time being mainly in the form of a reconnaissance phase. The private sector above all appears curious about the huge volume of work being carried out in the academic world in the field of semantic technologies and what this could mean in real terms for addressing their problems.

There is one clear example in the temporary employment sector. Parties from that sector have joined the research organisation TNO in establishing the multiyear Flexible IT (FIT) with Ontologies project. Using this project, they are testing the benefits of a switch from exchanging information via the SETU messaging standards to more flexible exchange using the SETU ontology.

Elsewhere there are individual companies working with Linked Data. For example the company NXP Semiconductors, which won prizes in Dutch and European Linked Data competitions in 2015, with their Enterprise Datahub. This system enables data on their numerous products (more than 20,000) from different databases to be integrated. Wherever possible, NXP uses existing standard ontologies such as the Dublin Core, SKOS, FOAF and Schema.org. Because a number of the terms and concepts needed for their sector were not available, an NXP vocabulary was developed, and added to the already mentioned set. As a consequence, NXP Semiconductors is able to sufficiently enrich its data description so that external users can more easily identify the meaning of the terms.

Finally, Linked Data technologies are becoming increasingly widely used in business processes. One well-known example is online marketing, more specifically search engine optimisation or SEO. Using SEO, businesses strive to ensure that their webpage scores as high as possible in the search results of a search engine. Today, this is often achieved using JSON-LD. JSON-LD, a method for transferring Linked Data via the well-known JSON format, is used by search engines to gather information about websites in order to determine the desired search results. Google, owner of the

world's most widely used search engine, recommends use of JSON-LD. The emergence of JSON-LD, then, is another example of Linked Data creating a connection with widely used data formats.

§ 2.4 The FAIR movement

The volume of data used by scientists and other data analysts is constantly growing. Hence also the growth in demand for more machine support in data processing and analysis. However, the data infrastructure required for these services simply did not exist. The FAIR Data Principles are a response to this situation. FAIR stands for a set of guidelines for making data more **F**indable, **A**ccessible, **I**nteroperable and **R**eusable. The focus of FAIR is on expanding the capacity of machines to automatically find and use data.

The FAIR principles were introduced in an academic article in 2016. The FAIR movement massively expanded shortly thereafter. Leaders of the G20, for example, in the same year, expressed their support for FAIR, and in 2017, Germany, the Netherlands and France together launched the GO FAIR initiative.

The 'I' in FAIR underlines the importance of automated data interoperability. The aim is to achieve and secure interoperability, even if data are produced in a very different format in different languages and in different search domains. This makes the link between FAIR and Linked Data very solid. As reflected by the three FAIR principles that come under the 'I' of 'Interoperability*':

1. (meta)data use a formal, accessible, shared and widely applicable language for knowledge representation.
2. (meta)data use vocabularies that comply with FAIR principles.
3. (meta)data contain qualified references to other (meta)data.

To implement the Interoperability principles of FAIR, use is often made of ontologies (for point 1) and other Linked Data standards (for points 2 and 3).

Links:

- [The FAIR Guiding Principles for scientific data management and stewardship](#), Nature, 2016, *het* academic article about FAIR
- [FAIR Principles](#), GO FAIR Initiative, the FAIR principles

§ 2.5 Why Linked Data?

Linked Data makes it possible to share and open up datasets in a controlled manner without the dataset having to leave the source. The Europeana Collections is an excellent example. Not only the data within a dataset of this kind are linked, but the datasets are also linked together. Using a query language like SPARQL, it then becomes possible to consult the entire set of datasets. This is the core of the vision of the Semantic Web. The more Linked Data, the more powerful the queries that become possible.

In particular government authorities are increasingly using Linked Data to make data available to the public, preferably in a manner in which interested parties can easily automatically request and process the data. This is subject to the well-known five star model:

- 1 star:** the data is available via internet, no matter what the format.
- 2 stars:** the available data has a machine-readable structure (no scan for example).
- 3 stars:** the data is available in a non-patented format (such as .csv and not MS Excel).
- 4 stars:** Linked Data published using W3C standards (RDF and SPARQL).
- 5 stars:** all the above and the dataset is linked to other Linked Open Data.

Linked Data is also used by digitalising organisations, to further integrate their data and systems, as with the example of Rijkswaterstaat and ProRail with their OTLS. The scope of efforts of this kind is often initially internal, before being extended to include major suppliers and other fixed partner organisations. This extension generates additional benefits from the step to Linked Data, with relatively little effort.

Linked Data also makes it possible for machines to automatically read in the meaning (semantics) of data and in this way to create new links, links which would normally speaking only be created by people. Linked Data is an essential component of new generation Artificial Intelligence (AI) techniques, in which enormous volumes of data are combined with formalised domain descriptions to build ever smarter machines.

Finally, Linked Data is the core of a new approach to expanding interoperability between parties in a sector or value chain. The semantic standard recorded for this purpose then contains one more more ontologies. Just as with other types of semantic standards, the aim is to make it as simple as possible for users to establish mutual interfaces that satisfy all their wishes. The Flexible IT with Ontologies project for the temporary employment sector is an example of a user community that is investigating the step to information exchange via a standardised ontology.

The use of Linked Data is subject to the well-known network effect. The following applies to each of the above applications: the more widespread the use of Linked Data becomes, the greater the benefits for users.

Go to the website of Platform Linked Data Nederland (www.pldn.nl) for more information about how Linked Data can be used.

§ 2.6 Impact on standardisation and management

Originally, BOMOS made no real distinction between the types of semantic standards. It is applicable to both the conventional approach and the Linked Data approach. However, there is certainly a clear difference between the two in terms of the precise structuring of the activities. It was always quite a challenge to implement the BOMOS model because each context requires a custom approach. With the Linked Data approach to semantic standards, new challenges have also emerged.

Below we list a number of the challenges and lessons learned from the FIT with Ontologies innovation programme within the temporary employment sector. The order roughly runs from development phase to management phase.

§ 2.6.1 Reuse

Ontologies rarely emerge in a ‘green field’ scenario. Specifically digital value chains in which far-reaching exchange agreements have already been reached are quicker to make the switch to a solution based on Linked Data. The documentation of these standards often contains a great deal of information valuable for the new solution, such as detailed descriptions for chain processes and business rules. It is important that that knowledge and information be reused as far as possible. Reuse not only ensures time saving in the development process for the new solution but also greater confidence in the end result and hence better adoption by the user community.

§ 2.7 Training

The switch to standardised exchange Linked Data almost always means that a different technology stack is used (for example with OWL, RDF and/or SPARQL). This in turn calls for different knowledge and skills, both for those involved in the development and management of the standard and among its users.

The management organisation must have sufficient expertise in these technologies on hand, to be able to implement their management tasks, such as estimating the impact of proposals for change on the standard or answering technical questions. These are obvious caveats. However, also on the part of the user organisations, there will probably be a need for further learning. Experience has

shown that the majority of software developers have little or no experience with these technologies. To safeguard active participation by users in the development process, it is essential that sufficient training be made available and that users be given the time to acquire knowledge of the technology. The management organisation can meet these needs in many different ways, for example by writing guides, providing demonstrations and even organising a Linked Data hackathon.

Training is also essential if the user community was already used to using standards for information exchange before making the switch to Linked Data. In that situation, the users and even the managers are used to thinking in terms of the old technology stack and the methods used for data exchange. In those cases, the switch to the Linked Data approach is often treated as a change in syntax, while in reality it also needs a new approach to overall thinking. This applies both to the development phase and the actual application of the semantic standards. Without this, during the development phase, you run the risk that the status quo will have too much of an influence on how the ontology is modelled, resulting in impoverishment of the model's semantics. If the old thought processes are applied, the result is cumbersome approaches to requesting or exchanging data, which in turn leads to impoverishment of the semantics of the data. For example, the managers of the Key Register of Addresses and Buildings (BAG) saw that many users also unintentionally requested the entire history of the data, with their data request.

§ 2.8 Selecting a compromise

It should be remembered that the choices available are not only making the complete switch to Linked Data, or leaving everything as it was. Conceivably there are gradations or intermediate steps.

One technology that plays a vital role in the switch is RML (RDF Mapping Language), developed by a team at the University of Ghent. Using RML, it is possible to transform semi-structured data like JSON, XML and CSV into Linked Data. You do this by first writing down the RML rules that must be applied to your data. You then feed these RML rules plus the original data into a so-called RMLMapper, which converts them into Linked Data triples.

The use of RML could be the best alternative balance for many communities with standards, as compared with the complete switch to Linked Data or doing nothing. In principle, it is sufficient to use RML mappings to create an intermediate layer that makes the translation between the relevant internal systems of the user and the common ontology. Thanks to those mappings, using RESTful APIs, it is still possible to exchange Linked Data, while retaining the existing systems.

EXAMPLE 1

This strategy was also employed in the FIT with Ontologies project. One key aim of that project was to safeguard the backwards compatibility with message exchange according to the conventional SETU standards. In other words, using the new technology, it still had to be possible to receive and process SETU messages. This led the project group to work on the FIT engine: a software component that applies RML to SETU messages to convert them to Linked Data. Using the FIT engine, the parties that had made the switch to exchanging Linked Data were able to reap all the benefits, without having to redefine their existing links with their customers.

§ 2.9 Further developments

Imagine, the first version of the new semantic standard based on Linked Data is complete. The user community is hard at work on the implementation phase. From this point onwards, what characterises the management and further development?

To start with, the managed material, the new standard ontology, is more modular in nature than you are used to as the manager of a more conventional standard. Any further development is aimed more at compiling existing ontologies and other building blocks than was the case with the conventional approach. Take for example the Geo standards. In this application, new ontologies have added relatively little. Numerous existing vocabularies have been assembled. The common agreements discuss how those items should be compiled. Using the conventional approach to semantic standards, this was less the case, for example using XML Schema, but it is more relevant with ontologies and for larger blocks.

Nevertheless, part of the domain description is modelled in its own ontology. That own ontology must be further developed in order to move with the times. The challenge is to ensure that this continues to be done as carefully and precisely as possible. The advanced applications of ontologies, such as complex queries or automatic reasoning, are more sensitive to errors. More than with other semantic standards, therefore, precision and coherence of the ontology are of critical importance.

Not every request for change will result in changes in the common ontology. It is sometimes better to develop an extension, which is then used by the part of the user community that has most to benefit from its use. Ontologies and other Linked Data standards were designed precisely for that purpose. For you as manager, the issue is to make these extensions explicit and accessible to the users. Before carrying out your own checks and analyses on those expansions.

§ 2.10 User analysis

In the FIT with Ontologies project, a web application is used to click together message definitions from the SETU ontology. These specific message definitions may look different, but all comply with the ontology and can be uniformly translated to facts from that ontology. Storing these different message definitions centrally means that detailed user analysis is possible, comparable to the code coverage concept from software engineering. This term refers to identifying which parts (e.g. concepts, relationships) of an ontology are actually used in which messages, and by whom, together with the related trends. Findings from code coverage tests provide important clues for further development and can support discussions about the standard, with real user data.

§ 2.11 Tools

Below we list a number of tools that could be valuable for managers of common ontologies. Pool Party, Protégé and TopBraid Composer demonstrate considerable overlap; it is possible to model ontologies using all three. Semantic Treehouse is not intended for modelling but for the management and further development process, and is the only one of the four tools with a direct link to BOMOS. Matey (and the related RML and YARRRML languages) is focused on the transformation of traditional data formats such as JSON, XML and CVS to Linked Data that complies with a specific ontology.

Tool	Developer	Description
NeoNeoLogism	Netage B.V.	An RDF Schema vocabulary editor and publication environment to be used with the web browser.
Pool Party	Semantic Web Company GmbH	Knowledge management, data analytics & content organization
Protégé	Stanford University School of Medicine	Ontology editor (free and open source)
Semantic Treehouse	TNO	Publication and management environment for common ontologies and other data models
TopBraid Composer	TopQuadrant	Integrated Development Environment (IDE) for Semantic Web, including ontology editor
VocBench	EU Publicatiebureau	Open source publication and management environment with multilingual vocabularies

Tool	Developer	Description
Matey	University of Ghent	Matey is a browser-based editor for writing YARRRML rules.

§ 2.11.1 NeoNeoLogism

NeoNeoLogism is a simple tool for writing and publishing vocabularies in RDF(S), and is nothing like TopBraid, Protégé, VocBench or PoolParty. It is intended to be simple but sufficiently expressive to support subclasses and sub attributes and different forms of relationships.

§ 2.11.2 PoolParty

The PoolParty Semantic Suite offers various functions to enable organisations to generate more value from their own data. The product commercialises Semantic Web technologies. The first release of the PoolParty Semantic Software was placed on the market in 2009. Since that time, the project has evolved from a taxonomy management tool into a multifunctioning semantic software platform that enables businesses to use ontologies to integrate structured and unstructured data.

§ 2.11.3 Protégé

Protégé is a free, open source platform that offers a series of tools for building domain models and applications with ontologies. There is both a web version and a desktop version. WebProtégé is an ontology development environment for the web for creating, uploading, adjusting and sharing ontologies. Users can view and process the ontologies, in combination. Protégé Desktop also contains additional functionality such as ‘semantic reasoners’ and visualisation tools.

§ 2.11.4 Semantic Treehouse

Semantic Treehouse is the community platform that supports the management of ontologies and other types of data models. What makes Semantic Treehouse unique is that it combines the publication, maintenance and management of your data models in a single place. Semantic Treehouse is based on more than 10 years’ experience with the development, maintenance and sharing of data standards according to BOMOS. Semantic Treehouse falls into the category

Software as a Service, which offers customers the opportunity to structure their own visual identity for the user interface, to provide a recognisable user experience.

§ 2.11.5 TopBraid Composer

TopBraid Composer is a desktop tool for creating ontology models and integrating data sources. It supports visual processing of ontologies and can be linked to database backends such as Oracle and different popular triplestores. The expanded edition is an IDE that adds the possibility of rapidly developing semantic web applications.

§ 2.11.6 VocBench

VocBench 3 was financed by the European Commission's ISA2 programme. It is a free platform that enables the joint processing and management of code lists, taxonomies, thesauri and ontologies. In addition to community management (management of multiple users, role definition, access control), VocBench 3 offers a mechanism that enables users to link their model to Jira, a type of issue and bug tracking platform for software.

§ 2.11.7 Matey

The idea behind Matey is to make RML more accessible.

The RDF Mapping Language (RML) is a generic mapping language for expressing rules that allocate data in heterogeneous structures and serialisations (e.g. JSON, XML, CSV) to the RDF data model. See <https://rml.io/docs/rml/introduction/> for more information about RML.

YARRRML is a human readable, text-based reproduction for declarative rules for the generation of Linked Data. It is a subset of YAML, a commonly used data serialisation language designed to be human friendly. YARRRML can be used for reproducing RML rules. See <https://rml.io/yarrml/> for more information about YARRRML.

