

Flanders Marine Institute (VLIZ)

Read the LOD



Annual report 2023

Joanna Goley, Marc Portier, Geert Limpens
1-25-2024



Contents

| | |
|--------------------------------|---|
| TERMINOLOGY | 2 |
| Executive Summary | 3 |
| Introduction | 3 |
| Read the LOD | 3 |
| Papers and meetings | 4 |
| Announcements | 4 |
| Statistics/ achievements | 4 |
| Special events..... | 4 |
| Meeting highlights | 5 |
| Feedback | 7 |
| Conclusion..... | 7 |
| Next year (2024)..... | 7 |
| Acknowledgments..... | 7 |
| Appendix | 7 |

TERMINOLOGY

| Terminology/Acronym | Description |
|---------------------|---------------------------|
| VLIZ | Flanders Marine Institute |
| VMDC | VLIZ Marine Data Centre |
| LOD | Linked Open Data |
| LDES | Linked Data Event Streams |

Executive Summary

Read the LOD is a grassroots Linked Open Data (LOD) initiative that organizes reading-and-discussion clubs around relevant contemporary Linked Data and Semantic Web documents. Coordinated by the VLIZ Data Centre and partnered with Redpencil and IDLab, we kicked off the first year from November 2022 - December 2023. This is the first annual report of the Read the LOD initiative.

Introduction

Flanders Marine Institute (VLIZ) seeks to strengthen science-based knowledge about our coasts, seas, and the ocean and share it as widely as possible. That knowledge matters to everyone. Within VLIZ, its Marine Data Centre (VMDC) provides services, technologies, tools, training, and support to scientists, policymakers, the blue economy, and citizens. The Data Centre strives to promote, use, and ensure that Open Science data and data systems follow the FAIR principles. Open access and data-sharing principles have been integral to VLIZ since its launch in 1999.

The Open Science division within the VMDC aims to achieve an elevated level of semantic interoperability to breach the existing domain barriers.

The three main goals of the Open Science division are:

- *Promote Open Science (advocate and pioneer)*
- *Improve FAIRness of data and data systems*
- *Implement technologies that will foster Open Science and Open Data*

To that extent, Linked Open Data (LOD) and the semantic web are applied as the main instruments.

To help achieve the first goal of the Open Science division at VMDC, the "Read the LOD" initiative was launched.

Read the LOD

Read the LOD was first initiated in November of 2022 as part of VMDC's open science mindset. Read the LOD is a grassroots LOD initiative that organizes reading-and-discussion clubs around relevant contemporary Linked Data and Semantic Web documents.

The goals of Read the LOD are:

- *to read papers, articles, journals, ... you always wanted to read but never took up time for*
- *to be able to ask those questions that have been bothering you in a collaborative environment.*
- *to share ideas, strategies, motivational spins, use cases, and components.*
- *to build and accelerate the LOD community, particularly where it intersects with open science.*

A total of 6 meetings took place, one in-person and five online (MS Teams). We held two special summer sessions which drew in up to double the average number of participants.

Papers and meetings

Members are encouraged to submit papers for reading sessions via ReadtheLOD GitHub, email, or chat channel (chat.semantic.works). Potential reading club topics varied about the review of the semantic web, software architecture, knowledge graphs, Linked Data Event Streams (LDES), FAIR signposting, interoperability, Ro-crate specifications, etc. If you would like to see the full list of reading topics, please visit [📄 Preliminary reading list | Read the LOD](#). For the papers discussed in 2022-2023 please see Appendix.

Polls to determine the paper and date become available on our ReadtheLOD chat channel and remain open for two weeks for member voting.

Meetings took place on a predetermined Monday from 14:00 to 17:00 either online (Microsoft Teams) or in person. The one in-person session was held at the VLIZ-library in the InnovOcean Campus in Oostende. Meetings follow the same general agenda:

1. Welcome and practicalities
2. Helping each other to understand the paper
3. Formulating critique or missed opportunities
4. Reflecting if or how the paper met the expectations from the title and voting process
5. Effectively applying the knowledge from the paper in potential use cases
6. Closing off with "Any other business" and agreements on the next gatherings

Announcements

Dissemination of announcements was done via the member's mailing list: news@readthelod.org and/or the Read the LOD chat channel. Inquiries about joining or regarding meetings can be sent to info@readthelod.org.

Announcements, as well as promotional posters (see Appendix Figure A1), were shared within the community via social media (LinkedIn and Basecamp) to spark interest and recruitment.

Statistics/ achievements

From the launch of Read the LOD in November of 2022 until the end of December 2023, there was a total of six meetings during the 13 months (Table 1). We had an average of 6.5 guests joining each session. The most participation was a total of twelve participants during our 4th meeting in July for our special "*Summer-of-Linked-Data*" session.

Special events

We introduced two special summer sessions where we delved into the practical application of LOD in the realm of research data. The two summer sessions each had a special guest the paper's author joining us for the meeting. This also doubled participation from previous sessions.

- RO-CRATE Specification

Annual report Read the LOD 2023

- Paper: [artifacts with Ro-create](#)^[OBI]
- DOI: 10.3233/DS-210053
- Special guest: Stian Soiland-Reyes
- FAIR SIGNPOSTING
 - Paper: [about 15 years of interoperability efforts](#)^[OBI]
 - DOI: 10.1045/november2015-vandesompel
 - Special guest: Herbert Van De Sompel

Read the LOD “*Summer-of-Linked-Research-Data*” announcement on LinkedIn received 855 impressions with thirteen reactions and eight reposts (Figure 1).

Meeting highlights

#1 Nov 2022 - “Is the LOD cloud at risk of becoming a museum for datasets? Looking ahead towards a fully collaborative and sustainable LOD cloud” by Jeremy Debattista et al.

The first meeting kicked off in-person with a good starting paper, it was a nice entry level, allowing members to cover some introductory concepts and have some good discussions. The title having high clickbait, it was felt that the LOD cloud (as the actual corpus under investigation in the paper) does not equate to the whole “web of data” idea. Fun fact that none of us has ever used the LOD cloud while we still are happily pursuing the linked data publication of many datasets contributing to the global knowledge graph.

#2 February 2023 - “Linked Data Notifications: A Resource-Centric Communication Protocol” by Sarven Capadisli et al.

For our members, no one had used Linked Data Notifications before. there was some ambiguity on how you can use it and how it competes with LDES. There was no real critique of the paper but little mention of ActivityPub, it would have been nice to have some more information on that, but in terms of who depends on whom this is not strange at all. We discussed the use case of taking the environmental hazards case as an application exercise.

#3 April 2023 - “Are We Better Off With Just One Ontology on the Web?” by Haller et al.

Applying knowledge from April's session we realise the OSLO traject in Flanders is trying to be a governing strategy for creating ontologies in a way that it tries to ensure some of the needed criteria; quality, trust, shared commitment... Only to end up with yet another ontology to be released in this winner-takes-all arena, where we just have to wait for the (possibly temporary) de facto end-result “popular uptake”.

All in all, it looks like we all have learned to survive in this largely undefined space of “ontology engineering” and are finding some balance of local advancement and super local alignment and collaboration, advocating what we already use in the hope others will commit to the same ontologies.

And all of this is not even narrowed to the realm of ontologies – it relates to a human strategy not to end up “standing alone” - something we recognize in the ancient quote that “nobody ever got fired for choosing blue (IBM)”.

#4 July 2023 - "Packaging re-search artifacts with RO-Crate" by Soiland-Reyes et al.

This meeting was the summer of linked data special sessions, the author asks, "what should profiles be to you?" here is an overlap with ontologies and vocabularies. Any experience RDF Shapes (SHACL, ShEx) and W3C Profiles vocabulary?

How to manage and start with profiles? More about types than domain-specificity. What fits well into that domain? Spatial metadata for instance. With alternative metadata files it is unclear when the RO-Crate profile ends and when is it a separate metadata file. It has a sort of open-endedness. It would be good to have easy markers for consumers on what to expect. It is then less of a constraint and more enabling. But how do you ensure that they are working on the same thing without making another alternative profile? In fact, flexibility adds a cost to end-user if they have all done it differently. This is a general problem for the Semantic Web. The comparison is if you didn't have Semantic Web technology, we at least know we have a common way to represent structures, even if our models differ. This becomes a social aspect, how to gather and agree on how to express metadata in a common way. And do this in a way without forcing everybody to do it all in the same way. This also ties back to how profiles should not be confused with types. It is more like duck-typing, a lot more flexible way to grow and extend and change over time, also more inviting people to have slight alternatives that work together for the common parts, and clearly mark and expose parts that are not.

The idea is also to have a profile registry to be able to make these discoverable. We also must agree that certain cleaning and fixing will never go away. The hope is that it would just get a bit simpler though. The flexibility is great on the creator side – not sure yet it will work out at the end of usage.

#5 August 2023 - "Reminiscing About 15 Years of Interoperability Efforts" by Herbert Van de Sompel.

During the discussion of the Paper *Reminiscing About 15 Years of Interoperability Efforts* we particularly celebrated the "red pill moment" of getting this key understanding about the web that was very eloquently put down on page 4 explaining the "Resource Oriented View" of the Web «that a repository does not exist in the architecture of the World Wide Web, it does not even entail the notion of a web server. The web is all about resources identified by URIs. In this web-centric perspective, a compound object consists of any number of URI-identified resources that exist somewhere on the web. Constituent resources may be hosted by one or more repositories but that is not considered an essential characteristic. This 180-degree shift in perspective became a veritable aha moment for us, a Kool-Aid we drank, something we saw that can never be unseen. »

#6 November 2023 - "Consent Through the Lens of Semantics: State of the Art Survey and Best Practices" by Aneila Kurteva et al.

The question came up of why anyone would want to solve this niche problem with something like the semantic web, that is the problem of consent and personal data. It was mentioned that consent can narrow down better to purpose use, which is not covered by these more general alternatives like access or licensing. We agree GDPR has put consent on the table, so it needs to be dealt with, unlike the United States. There is an essence only a basic yes /no outcome. Different levels of consent would then end up being distinct contexts for which one gives this yes /no. There is a need for more standardization as you can then do an external audit of them following the GDPR rules and having proper consent on records for all. Such standardization, provided they follow the rules, could also allow us to gauge the 'privacy-

protectiveness' of services, which we could then use to decide which ones we prefer (like apps on your phone that require less permissions to provide this service). Even onto the next level we could envision disclosing the policies in a semantic way (outside scope of this paper).

Feedback

A feedback form was sent out to members of the Read the LOD. Only two participants responded and rated high for recommending Read the LOD to a friend. It was liked that participants came prepared even if they did not understand the paper fully. Also, the ability to ask open questions. The feedback from members who responded at the club has positively improved their knowledge of linked data that they wouldn't have otherwise. Feedback from participants found mutual opinions allowed gathering more insight on subject matters.

It was suggested from the form improvement could be made by involving more people from the city sector.

Conclusion

Read the LOD was noticeably more active during the two summer sessions. Inviting the authors of the articles significantly increased participation. It gave a special opportunity for participants to ask questions in an educational environment. We received positive feedback that the book club contributed to the knowledge of linked and open data that would not have been otherwise.

Next year (2024)

For the next year, 2024 having more special guest sessions will increase outreach and members. Continuing to come prepared to the meetings provides the most effective time during the meeting.

From the feedback form, it's suggested to pick themes on more generic, technical linked data topics. It was suggested to incorporate less research-focused topics. Attempts should be made to introduce the book club at regular knowledge-sharing meetings with different organizations to expand participation.

Acknowledgments

The Read the LOD team would like to thank all members and participants of the meetings throughout the year. We would like to thank our partners for getting this grassroots movement started, especially Redpencil and IDLab.

For more information, please visit [Read the LOD](#) or email info@readthelod.org.

Appendix

Table 1: Papers discussed 2022- 2023.

| Month | Title |
|---------------|---|
| November 2022 | “Is the LOD cloud at risk of becoming a museum for datasets? Looking ahead towards a fully collaborative and sustainable LOD cloud” by Jeremy Debattista et al. |
| February 2023 | Linked Data Notifications: A Resource-Centric Communication Protocol by Sarven Capadisli et al. |
| April 2023 | Are We Better Off With Just One Ontology on the Web? by Haller et al. |
| July 2023 | Packaging re-search artifacts with RO-Crate by Soiland-Reyes et al. |
| August 2023 | Reminiscing About 15 Years of Interoperability Efforts by Herbert Van de Sompel. |
| November 2023 | Consent Through the Lens of Semantics: State of the Art Survey and Best Practices by Aneila Kurteva et al. |

Figure 1: Promotional posters for 2023 meetings.

