

57. One-click semantic enrichment of scientific data

Researchers, publishers, and funding agencies increasingly recognize the importance of publishing the original research data along with traditional journal articles. However, the threshold for publishing data in a way that enables reproducibility and reuse, is still too high for an unsophisticated scientist. To solve this problem, we have developed a web-based dashboard which incorporates a number of techniques for enrichment of research data with appropriate metadata, such as links to relevant external resources and identifiers. It also helps the user to upload the results to a popular research data repository where the data can be discovered, verified, and ultimately reused by other researchers.



ICT science question

How can scientists publish their research data in such a way that their colleagues can easily reproduce or reuse them?

We approach this problem by automating metadata discovery and enabling the user to publish results as Linked Data following standards and best practices. For the convenience of the user we hide the details of the underlying semantic web technologies.

Application

We apply the technology to support researchers who are not necessarily data management experts, but produce and publish original scientific data. A number of related efforts aiming at lowering the barriers for data publication and improving the quality of published data are underway. However, these efforts are platform-dependent and do not comply with emerging standards for



Marat Charlaganov
m.charlaganov@vu.nl
www.data2semantics.org

COMMIT/ project
Data2Semantics From Data to Semantics for Scientific Data Publishers

data publishing. Our approach relies on the latest semantic web standards for data publication and enables easy automatic processing of published results.

Alternative Application

The exploration of texts is broadly needed. Rather than rely on researchers for enriching metadata and publishing Linked Data, repositories might use our technology in their user-facing applications, and for batch processing of existing deposits. We are collaborating with Mendeley (Elsevier's free reference manager and academic social network) and with Figshare (an online digital repository where researchers can preserve and share their research outputs) in order to bring parts of our technology to their products.

Nice to know

Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is often called data about data or information about information.



The value of scientific data is determined by reuse in academia and industry. Therefore every effort should be made to make all research data available and discoverable.



There is only one-way to ensure interoperability with future applications and services: follow the standards!



In the world where high quality scientific data is easy to find and verify innovation accelerates tremendously.



Data publication must provide enough context to be interpreted and allow citation. It's not that different from a journal publication, after all.