Sharing and Analyzing Remote Sensing Observation Data for Linked Science

Tomi Kauppinen[1], Benedikt Gräler[1], and Giovana Mira de Espindola[2]

[1] Institute for Geoinformatics, University of Münster, Germany
tomi.kauppinen|ben.graeler@uni-muenster.de
[2] Earth System Science Center, National Institute for Space Research (INPE), Brazil
giovana@dpi.inpe.br

Abstract. In order to make research settings transparent and reproducible there is a need for publishing both data and methods behind the research. In this paper our contribution is to show how large amounts of remote sensing observation data about the Brazilian Amazon Rainforest has been published as Linked Spatiotemporal Data. Moreover, we show how this data can be further accessed and analyzed using R statistical computing environment by openly available methods. This all is a contribution towards Linked Science, where not just publications, but data, methods, tools, and other scientific assets are interconnected and shared online.

1 Opening and Linking Science

Open Science needs Open Data to maximize the transparency, reproducibility and reuse of scientific efforts. An example of a high demand for data is the research about climate change, for example about the role of deforestation in it.

Deforestation and its related phenomena such as market prices of agricultural products form together a complex system. There is an urgent need to share and publish research data about it, as it would enable other researchers to interconnect their data to the published ones. The benefit is that these explicit interconnections allow for the analysis of all of the resulting linked data in a transdisciplinary manner. Thus the whole complex socio-economic and environmental system could be modelled and not just subsets of it.

Opening up of scientific assets like data and methods behind scientific settings, and interconnecting them is called Linked Science [1]. Tools are major enablers of Linked Science. One crucial aspect is how to access and analyse data, and especially how to get only that part of data which is of interest for a given research question. Linked Data solves the access part, and SPARQL allows to query only a subset of the data. For statistical computing there are tools like R[1], and a separate package[2] for it supports querying Linked Data.

---

[1] http://www.R-project.org/
[2] http://linkedscience.org/tools/sparql-package-for-r/

2 Case for Linked Science: Sharing and Analyzing Observation Data

For analyzing processes and operations of complex systems such as environmental and societal systems there is a need to have 1) well interconnected data about a system and 2) techniques for statistical computing and for other types of reasoning to find new information, and means to 3) explore and visualize this information.

Our contribution is the Linked Brazilian Amazon Rainforest Data[3] which is openly available for anyone to use for non-commercial research. The data can be accessed in a Linked Datafashion via a SPARQL-endpoint, and via dereferenciable URIs. The data consists of 8250 cells—each of size of 25 km * 25 km—capturing the observations of deforestation in the Brazilian Amazon Rainforest and a number of related and relevant socio-economic indicators. This spatiotemporal deforestation data was created using a number of aggregation methods from different sources. The data covers the whole Brazilian Amazon Rainforest.

In summary, our contribution was to show

– how to access Linked Spatiotemporal Data about the complex environmental system of the Brazilian Amazon Rainforest, including deforestation statistics, and a variety of socio-economic and environmental indicators
– how to store the spatiotemporal data in a meaningful way as Linked Data
– how to plot[4] (see below an example) and handle the data within R.
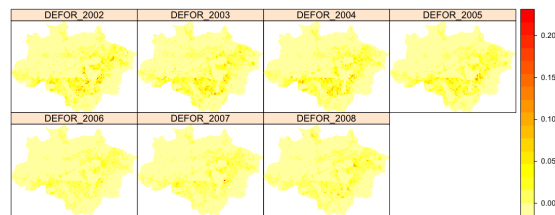
Acknowledgements

References

1. Kauppinen, T., Baglatzi, A., Keßler, C.: Linked Science: Interconnecting Scientific Assets. In Critchlow, T., Kleese-Van Dam, K., eds.: Data Intensive Science. CRC Press, USA (forthcoming 2012)

---

[3] http://linkedscience.org/data/linked-brazilian-amazon-rainforest/

[4] An online tutorial is provided to explore the data from R, see
http://linkedscience.org/tools/sparql-package-for-r/
tutorial-on-sparql-package-for-r/