

EarthLinks – Easy Access to EOD for Social Scientists

Dennis Abel¹, Stefan Jünger¹

¹GESIS – Leibniz Institute for the Social Sciences, Cologne

June 21, 2024

Abstract

For many researchers in the social sciences, Earth observation data (EOD) still represents a black box. "EarthLinks" aims to close the gap and create an automated interface to EOD and complementary resources for social science research. The project's core is **creating an open-source tool to link time- and space-sensitive social science datasets with data from Earth observation programs**. The pilot's main aim is to advance the automatization of these data integration processes between social science data and EOD based on an open-source, user-friendly tool that does not require users' programming skills. Detailed documentation and a beginner-friendly tutorial will complement the tool to showcase the capability of our project. The social science community is the main target group of our tool. At the same time, Earth system science researchers may profit from having easy access to integrated social science studies, which comprise information about changing individual perceptions of the environment, climate, and sustainability. "EarthLinks" opens new avenues for inter- and transdisciplinary research made difficult because of technical, disciplinary, and organizational barriers. In sum, "EarthLinks" will advance interdisciplinary research data management workflows based on FAIR and Open Science principles.

I. Introduction

A growing interest in economics and the social sciences in earth observation data has led to a broad thematic spectrum of publications in recent years. They range from studying environmental attitudes and behavior, economic development, the role of climate change in violent conflicts and causes of flight, and electoral behavior. However, social science researchers also face many obstacles in applying and using these data, resulting from 1) a lack of technical expertise, 2) a lack of knowledge of data sources and how to access them, 3) unfamiliarity with complex data formats, such as high-resolution, longitudinal raster datacubes, and 4) lack of expertise in integrating the data into existing social science datasets. Despite the increased interest in the data, for many researchers in the social sciences, Earth observation data (EOD) represents a black box after all.

"EarthLinks" aims to close the gap and create an automated interface to open data from Earth observation programs and complementary resources for social science research in a comprehensive R Shiny App. The project's core is creating a user-friendly open-source tool to link time- and space-sensitive social science datasets with data from Earth observation programs. "EarthLinks" advances

the automatization of the data integration processes between social science data and EOD by piloting a tool that does not require users' programming skills.

The project extends the georeferencing services at GESIS. In the team "Survey Data Augmentation" (SDA), we have been active since 2022 in exploring EOD for social sciences and developing workflows for data integration with GESIS survey data, which comply with FAIR and Open Science principles. A first basic R package for retrieving, transforming, and linking data has been conceptualized and is currently being developed for internal usage. We also strengthened interdisciplinary networks by organizing an international workshop on EOD in the social sciences in 2023 and 2024 (for a few impressions, visit <https://youtu.be/iW4onh38UmA>). A key finding is that this work should be further intensified and developed, and our tool concept should be piloted and streamlined for a larger audience, as described in this proposal. Our team knows how to develop and implement open-source tool software¹, and the pilot phase will translate the existing R workflows into a comprehensive R Shiny App.

The project collaborates closely with other EOD stakeholders to institutionalize long-term interdisciplinary expertise and generate high-impact use cases. We work in close cooperation with partners in the field through third-party funded infrastructure projects like the DFG project "**Social-Spatial Research Data Infrastructure**" (SoRa) and its follow-up project SoRa+², as well as the recently started Horizon Europe project "**Integrated Research Infrastructure Services for Climate Change risks (IRISCC)**"³. We are also active members of the working group on data integration and management of the **Leibniz research network "Earth & Societies" (E&S)**⁴ and the newly founded **Leibniz-Lab "Systemic Sustainability"**⁵. Thus, we are well positioned at the intersection of social sciences and Earth system sciences to test, further develop and promote the pilot tool through these channels.

II. Pilot description

"EarthLinks" offers direct and user-friendly access to openly available EOD products in a **comprehensive Shiny App in R**, which allows easy linking to social science data such as surveys or digital behavioral data. R Shiny is an extension of the R programming language often used in the social sciences to create interactive dashboards, data management, and analysis tools. Embedding the tool in a Shiny App enables its use through a simple user interface, suitable for users with little programming experience or a different software focus (e.g., STATA, SPSS). Detailed documentation and a beginner-friendly tutorial will complement the tool.

The unique feature of the tool should be the possibility of carrying out both geographically and temporally high-resolution queries of data from Earth observation data sources like the Copernicus

¹ Lieth, J., Abel, D., & Jünger, S. (2024). GRETA Analytics: Communicating the spatial dimension of energy citizenship using a collaborative Shiny app. In: SoftwareX, *forthcoming*.

² <https://www.gesis.org/en/research/external-funding-projects/details/project/206/ausbau-der-sozial-raumwissenschaftlichen-forschungsdateninfrastruktur>

³ <https://cordis.europa.eu/project/id/101131261>

⁴ <https://www.leibniz-gemeinschaft.de/en/research/leibniz-research-networks/earth-societies>

⁵ <https://www.leibniz-gemeinschaft.de/en/research/leibniz-labs/systemic-sustainability>

data services, which at the same time function efficiently on simple workstations albeit large amounts of data. Our tested workflow development has identified five major levers: indicator type, indicator intensity, focal time period, baseline time period, and spatial buffer. Flexibility on these five attributes should be maximized for users. The tool also offers the functionality to automatically derive spatio-temporal links with other georeferenced data (e.g., surveys, digital behavioral data). The technical implementation and coordination of data integration from other work packages will follow digital project work steps (initial planning, scoping, development, testing, and deployment).⁶

Users should benefit from a core set of variables integrated into the interface for social research. Preparatory work is currently being carried out to select appropriate indicators from the corpus of indicators offered by data providers and conceptualize the data integration logic. Furthermore, in exchange with users and other stakeholders, we compile data products from EOD that are particularly relevant for social scientists. Examples include data on local air quality and pollutants, extreme weather events, or land use changes. The primary data sources will be the Copernicus Monitoring Services on Climate Change, Atmosphere, and Land.

III. Relevance for the NFDI4Earth

Expanding its services to the social science community is part of the objectives of NFDI4Earth. "EarthLinks" provides an opportunity to do this in a low-threshold way. At the same time, although addressing social scientists, we see a strong potential for transferability to other domains and engaging with a user community beyond social sciences. Particularly, Earth system science researchers may profit from having easy access to social science studies, which comprise information about changing individual perceptions of the environment, climate, and sustainability. These data open new avenues for interdisciplinary research that are difficult because of technical and organizational barriers for non-social scientists.

Moreover, R is used in many disciplines, and the open-source nature of the tool will allow researchers from all disciplines to use it. In order to support the uptake, our tool will be complemented by comprehensive, open-source tutorials and training materials, which will be hosted online and include real-world sample data. The inclusion of the tool into the NFDI4Earth Living Handbook and Knowledge Hub will further disseminate the project among the NFDI4Earth community. A publication outlining the methodology behind the tool will support the documentation.

GESIS strongly supports the principles of FAIR data and Open Science. The selection of R as the programming language for the tool supports open-source infrastructure development and shareability, as well as quality control via online repositories. Publishing the R scripts for data management and analysis ensures the reproducibility of all research steps. Regarding data sources, the Copernicus data services were explicitly designed as an open data source, and the linked GESIS survey data are available through secure access facilities. GESIS offers a Secure Data Center, which allows GDPR-compliant access to sensitive, georeferenced, and integrated survey data.

⁶ Fay, C., Guyader, V., Rochette, S., & Girard, C. (2022). Engineering production-grade shiny apps (First edition). CRC Press, Taylor & Francis Group.

IV. Deliverables

We are planning with a project duration of 9 months and propose the following deliverables for this pilot:

Publication of open-access R Shiny App programming code in a public repository (M9) and hosting online (M9): Complying with FAIR research data management workflows, we will publish the entire R code for the Shiny App in an open repository. The tool will be furthermore hosted online from M9 onwards.

Submission of a software article to a scientific journal (M9): Our goal for this deliverable is to write a method report on the development and deployment of the tool. We aim to submit this article to a data management or data science journal.

Final report (M9). The final report includes a description of the entire pilot roadmap, an evaluation of the project development, and an outlook for future avenues after the finalization of the project.

V. Work Plan & Requested funding

The main task within "EarthLinks" is programming the R Shiny App. This work requires a position for a social or environmental data scientist who will work closely with the SDA team at GESIS. Therefore, we would like to apply for a full-time position for 9 months (12 PM, TVL E13). Due to the open-science approach and usage of open-source software, we do not need any funding for material resources or software. Server capacities for hosting the tool also exist in our team. Please see the Gantt chart below for the entire pilot project work plan.

Table 1: Gantt chart for the project work plan. Yellow cells indicate milestones and correspond to deliverables as outlined in the section above.

		Module 1: Pilot									
		Apr M1	My M2	Jn M3	Jl M4	Ag M5	Sp M6	Oc M7	Nv M8	Dc M9	
AP	Module	Status									
	AP0 Project management	Not started									
	Scheduling	Not started									
	Documentation	Not started									
	AP1 Tool development: Shiny App	Not started									
	Tool concept and connection to existing R workflows	Not started									
	User Interface design	Not started									
	Setting up back end of the tool	Not started									
	Deployment checks and user evaluation (internal)	Not started									
	Finalizing version 1 and storage in open repository	Not started									
	AP2 Academic contribution	Not started									
	Data management article	Not started									
	AP3 Dissemination	Not started									
	Hosting of the online tool	Not started									