



Facilitate the reuse of data from public authorities in research

Hannes Thiemann, Ivonne Anders, and Martin Schupfner

DKRZ, Data Management, Hamburg, Germany (anders@dkrz.de)

Governmental organizations collect and manage diverse data types at various levels to fulfill their official duties. This includes geographic, environmental, meteorological, population, health, traffic, transport, financial, and economic data. Traditionally, access to such data was restricted, but over the past decade, there has been a global shift towards more open data policies, influenced in part by directives like GeoIDG, the PSI directive, and INSPIRE. In Germany, federal states and their offices have also embraced open data policies, with some data being made publicly available (Open Government Data) through portals like Destatis or GDI-DE. This data serves multiple purposes, such as identifying locations, analyzing environmental trends, traffic planning, health service planning, and more. Public authorities' data is increasingly utilized for scientific investigations, yet the full potential remains untapped, particularly for large datasets. Despite the high quality of governmental data, further alignment with FAIR principles (Findable, Accessible, Interoperable, and Reusable) is necessary to enhance its efficiency for reuse in research. Privacy regulations and legal frameworks may impose limitations, necessitating data anonymization or adherence to modern data standards. Nevertheless, governmental data remains a valuable resource contributing significantly to expanding knowledge across scientific disciplines.

In a pilot project funded by the NFDI4Earth, in a collaboration, the German national meteorological service (DWD) and the German Climate Computing Centre (DKRZ) aimed to facilitate access to data from public authorities, increase data visibility, as well as the number of users from different disciplines, and make these data available in standardised and FAIR formats for easy use in research but also for other public applications. As an example, the COSMO-REA6 reanalysis dataset from DWD (Kaspar et al. 2020) was selected, crucial for climate modeling, analyses, and energy applications in Europe. The standardization process involved mapping public authority standards to domain-specific standards in climate research, requiring close collaboration between DWD and DKRZ. After detailed curation and quality checks, the dataset was made accessible through the ESGF infrastructure and long-term archived in the WDCC, addressing licensing and authorship considerations.

The project's insights and lessons learned were incorporated into a blueprint, providing guidance on making data from other authorities accessible and usable for both research and the public. Overall, the entire process can be divided into 5 sub-steps: (1) determination and classification of the need, (2) survey of the feasibility, (3) implementation, (4) feedback and follow-up, (5) dissemination. This blueprint outlines generalizable steps and aspects applicable across domains and collaborators, offering a framework for optimizing the use of governmental data in diverse

fields.

References:

Kaspar, F., et al., 2020: Regional atmospheric reanalysis activities at Deutscher Wetterdienst: review of evaluation results and application examples with a focus on renewable energy, *Adv. Sci. Res.*, 17, 115–128, <https://doi.org/10.5194/asr-17-115-2020>, 2020.