

SMEAR II DATA MANAGEMENT PLAN

INTRODUCTION

SMEAR (Station for Measuring Ecosystem - Atmosphere Relations) stations are research stations of the University of Helsinki (UH) and the University of Eastern Finland (UEF). SMEAR II station in Hyytiälä is run by the Institute for Atmospheric and Earth System Research (INAR) and the Faculty of Agriculture and Forestry at UH. This Data Management Plan describes the general curation practices of data collected at SMEAR II.

GENERAL DESCRIPTION OF DATA

Types and file formats of data

SMEAR stations produce in situ and laboratory observations on the Earth system processes, for instance, concentrations, compositions and fluxes of greenhouse gases, trace gases and aerosols, terrestrial and aquatic ecosystem processes, and meteorological parameters.

Raw data from continuous field observations and laboratory experiments are largely recorded in widely used file formats, most frequently as tables of whitespace or comma separated plain text (.csv). A few instruments that produce very large amounts of data record the data in binary data formats.

Control of data consistency and quality

Relevant measurement, data processing and documentation guidance, e.g. WMO, ICOS, ACTRIS and eLTER protocols and instructions are followed to produce data that are of high quality and consistent with data collected within other networks of Earth system research. The quality of the observational data is continuously monitored by employing extensive automated production of preliminary data, visualizations, diagnostics and log files that are checked by human experts. The data curation personnel of INAR help local and visiting scientists to follow the guidelines and embedding their data into centralized data processing and curation as far as possible. The original raw data and data products are stored in separate locations so that no original data are accidentally modified during use.

DOCUMENTATION AND METADATA

All types of data include documentation, such as the context of the data, field sampling design and measurement setups, calibrations and maintenance of instruments, descriptions of the data files, and information on the processing of raw data into end-user data products. These are often compiled as free-form “readme” files in widely-used formats; plain text, .pdf, MS Word documents, MS Excel or OpenDocument spreadsheets.

Continuous documentation of field operations, such as instrument configuration, calibration, maintenance, and problem notifications, is mostly done using electronic logbook software (elog) that runs on SMEAR II central data server. The underlying file format is plain text, and there is also a possibility to export in commonly used file formats (.csv or .xml). Additional documentation files can be saved as logbook attachments or in the same location as the data on the data server.

Processing software or scripts and the software for reading proprietary data formats are also saved and may be opened with data if the data processing is complicated and the original licence allows redistribution. Additional living documents with online tools such as wiki (<https://wiki.helsinki.fi/display/SMEAR/The+SMEAR+Wikispace>) can be provided to support accumulating observational data.

Formal administrative metadata that describes the ownership and the terms of use, as well as additional context metadata, such as geolocation and keywords/themes, are created during the opening of data using the tools provided by data publishing services.

STORAGE AND BACKUP OF DATA

Raw environmental observation data and documentation collected at SMEAR II are first stored on measurement computers and data loggers in the field and copied automatically to SMEAR II data server. Measurement computers not in SMEAR II network can upload data to INAR's sftp server or UH cloud service (Datacloud). Manually collected data and documents are saved on SMEAR II data server or in Datacloud.

All field and laboratory observation data, documentation, and processing scripts are eventually stored on data servers that are under central administration of the IT centre of UH. This is done automatically for continuous measurements. Security and backups are taken care by the IT centre, access rights controlled by INAR data curation personnel. The original field data and documents, excluding very large datasets, are kept on SMEAR II data server as long as it is economically feasible.

UH secure service (Umpio) is employed in the unlikely case that sensitive data must be stored.

OPENING, PUBLISHING AND ARCHIVING THE DATA

End-user data from continuous measurements are inserted into SMEAR database and distributed at the AVAA portal (<https://smear.avaa.csc.fi/>) hosted by CSC – IT Centre for Science. In many cases this is done automatically near real time. New versions of final quality-checked SMEAR data are exported from the database and published with persistent identifier (PID) annually in Fairdata Etsin research data catalogue (<https://etsin.fairdata.fi/datasets/SmartSMEAR>).

Campaign data and documents are initially shared on request and opened in trustworthy repositories (e.g. Etsin, Zenodo or B2SHARE), latest during publication of scientific papers on the data or after embargo of no more than 2–3 years.

INAR aims to keep most SMEAR II data available for as long as possible. This is ensured by using UH and CSC storage services and trustworthy data repositories that are expected to persist far beyond typical project life times. Also European research infrastructures (ACTRIS, ICOS) provide safe long-

term storage for some observation data. The large accumulation rate of raw data may result in need to delete most space-consuming raw data at some point however.

Selected datasets (such as long time series of key measurements) will be proposed for long-term preservation (decades) periodically, tentatively every 5 years. The preservation follows the national data preservation specifications and employs CSC's Digital preservation service (<https://www.digitalpreservation.fi>).

ETHICAL AND LEGAL COMPLIANCE

Data ownership and the terms of use are in most cases determined by the data policy of UH (<http://hdl.handle.net/10138/338778>). The original contributors of the data transfer the rights to use and distribute the data to the University of Helsinki. The data are published under licence that restricts further use as little as possible while enforcing data users to follow the principle of fair scientific use. The recommended licences are Creative Commons Attribution (CC-BY) for data and MIT for software (models and data processing codes). Different terms of data use can be applied if the collected data contain sensitive information or there are financial interests (for instance, data collected during the development project of a commercial device).

Visitors from other institutes than UH doing research at SMEAR II are expected to follow the SMEAR II data policy and share with their host researchers and UH all data they have collected at SMEAR II during the visit. The visitors naturally retain the authorship of the data and the licence requires the authors be acknowledged when the data are used.

DATA MANAGEMENT RESPONSIBILITIES AND RESOURCES

SMEAR II staff and INAR data curation staff coordinate and guide the data collection, storage, processing and visualization, and the opening and long-term preservation of the data to ensure fluent data flows, reliable data and open access to the data. The data curation costs are largely covered via the Academy of Finland research infrastructure programme and other long-term funding.