Academy of Finland: FIRI Roadmap 2024

EPOS-FI Data Management Policy

This EPOS-FI Data Management Policy sets out the guiding principles for managing and curating data in the EPOS-FI Research Infrastructure (RI). The EPOS-FI RI is the Finnish national node of the European Plate Observing System (EPOS) research infrastructure. The EPOS-FI consortium is a joint community of universities and research institutes with a task of maintaining geophysical observatories and laboratories in Finland.

1. General description of administrative data of RI

National research infrastructures (NRIs) consist of physical measurement instrumentation, associated data centers, and their personnel. Administrative data within the national EPOS-FI RI includes administrative reporting and planning related documents and personal data of the EPOS-FI RI member organizations' personnel, users and stakeholders. Personal data processed within the EPOS RI includes a person's name, professional title, organization name/type, primary (and secondary) domain of work/expertise, e-mail address, city/country of residence, and address. These are collected, for example, when applying for access to the RI and/or registering for events such as conferences, seminars, or workshops. The personal data collected by the EPOS-FI RI members when hosting activities for the RI are protected and minimized under the host organizations' data security measures and procedures, in compliance of the European General Data Protection Regulation (GDPR) and the Finnish national law. All organizational data are handled according to home organizations' data and privacy policies.

2. General description of research data managed within RI

The EPOS-FI research data include various types of geophysical data sets produced and managed by EPOS-FI consortium NRIs (Tables 1 and 2). Data are shared as data products. The data can be grouped by theme (research domain/subject), type (campaign or continuous) and level (degree of processing). NRIs are using international data formats. In addition, some data are stored in open-access databases in international data centers and will be stored in joint data storage facilities that are planned for EPOS-FI. EPOS-FI aims to establish central data services at the CSC – IT Center for Science. These are currently planned to be mainly based on CSC's Fairdata services (https://research.csc.fi/-/fairdata-services) and, where feasible, services from the EUDAT service portfolio and possible new services that CSC is developing can be included. The Finnish Ministry of Education and Culture offers preservation free of charge for data that is regarded, after separate negotiations, to be valuable on a national level via the Digital Preservation Service (DPS). Importantly, however, this service does not currently include direct access to the data. Thus, data pipelines from the planned intermediate or long-term storage to CSC's Digital Preservation Service would need to be constructed for data that are accepted to be preserved. The EPOS-FI data and data sources are described on the website of the RI: https://www.helsinki.fi/en/infrastructures/FIN-EPOS/data The data are and will be available online through the metadata repositories of EPOS and its Thematic Core Services (TCS),

relevant national data portals, and other relevant national services. The recorded data from NRI's are stored and archived on data servers hosted by the

Table 1. Data types and most common data formats

Data types	Most common file formats
continuous time series measurement data	Seismic: miniSEED, SEG-Y
	GNSS: Rinex-format
	Superconducting gravimeter: IGETS defined
	Absolute gravimeter: AGrav defined Ascii
	Quantum gravimeter: tbd
	Geomagnetic: IAGA ASCII
campaign data	Seismic: SEG-Y
processed event data	Seismic: SEISAN Nordic-format, QuakeML, CSS3.0, ASCII text, Microsoft Excel
laboratory measurements	N/A
images	.tiff, .jpg
instrument and processing configuration files	N/A
descriptive documents	Text
Physical or analogue data sources (produced with seismograms or magnetograms)	Paper or microfilm documents
Metadata about each data type	Compliant with EPOS DCAT ontology extension

3. Ethical and legal compliance of personal or sensitive research data

EPOS-FI research data do not include personal sensitive data (e.g., data subject to GDPR) by nature and do not have ethical problems. However, there may be data that are sensitive based on national/international security regulations such as Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) seismic data, which are protected by means of restricted physical/digital access and firewalls. Servers hosting these data are located in locked premises with authorized access only. More information about CTBTO can be found at https://www.ctbto.org/our-work/international-data-centre.

4. Agreements on research data rights

The NRIs are data suppliers to the international EPOS ERIC that implements the EPOS Data Policy and are thus required to manage and share data following the FAIR principles (findable, accessible, interoperable, and reusable). The EPOS-FI partners own and operate geophysical and geodetic research infrastructures and data in Finland. Each data originator or NRI may have data-type-specific procedures for processing, consolidating, and discarding parts of the raw data. As a result, pre-existing and raw data that may not be part of an EPOS data product, and thus access is granted on a case-by-case basis by the operators of the NRIs. The NRIs follow their host institute data policies but they have adopted or are compatible with the EPOS Data Policy for datasets included in EPOS-FI. (https://www.helsinki.fi/en/infrastructures/FIN-EPOS/FIN-EPOS-data-policies)

EPOS-FI user rights follow the EPOS Data Policy documentation. EPOS-FI NRI's Data, Data products, and Software and Services (DDSS) may be open (freely available or accessible to

Table 2. Different infrastructures within the EPOS-FI framework, amounts of raw and processed data produced in a year, and national and international data centers to which the infrastructure is providing data.

Organization	Equipment/network	Raw data		Processed data		Data centers	
		continuous	Gbytes/a	parameters	Gbytes/a	National rep	International repository
Institute of Seismology,	OBF,HEI,HKI,OTA local seismic networ	online	2800	N/A	N/A	ISUH	N/A
UH (ISUH)	FINES, PS17 seismic array	online	1000	N/A	N/A	ISUH	CTBTO
	FNSN permanent seismic network	online	1000	N/A	N/A	ISUH	GEOFON, ORFEUS
	FNSN permanent seismic network	N/A	N/A	earthquake	N/A	ISUH	EMSC, ISCUH, EFEHR, AHEAD
	FNSN permanent seismic network	N/A	N/A	earthquake	N/A	ISUH	Nordic database
	FNSN seismic research networks	partly online	7000	earthquake, SEG-Y sections	N/A	ISUH	GEOFON, ORFEUS
	DSS, refraction and refraction data	N/A	N/A	SEG-Y sections	40	ISUH	N/A
	DSS, FIRE reflection data	N/A	N/A	SEG-Y sections	30	CSC-IDA, PA	N/A
Department of Geosciences, SEG laboratory, UH	Instruments to measure magnetic properties of solid material	N/A	3	paleomagnetic	3	UH	MagIC Database, PALEOMAGIA database
	Instruments to measure physical properties of solid material	N/A	2	magnetic	2	IDA (CSC)	N/A
National Land Survey (FGI)	Superconducting gravimeter	online	20	N/A	10	FGI	IGETS Data Base
	Absolute gravimeter	online	5	N/A	0.01	FGI	NKG Absolute Gravity Data Base
	GNSS (FinnRef)	online	310	N/A	N/A	FGI	EUREF Permanent GNSS Network
	VLBI	online	8000	N/A	N/A	N/A	International VLBI Service for Geodesy and Astrometry
	VLBI2010 system	online	365000	N/A	N/A	N/A	International VLBI Service for Geodesy and Astrometry
	SLR	online	1	N/A	N/A	N/A	International Laser Ranging Service
Sodankylä Geophysical Observatory (SGO), UOULU	FN seismic network	online	10	N/A	N/A	SGO	GEOFON, ORFEUS
	Magnetometers	online	10	N/A	< 1	SGO	INTERMAGNET, IMAGE
	Finnish pulsation magnetometer chain	online	370	N/A	N/A	SGO	N/A
	SMA historical geomagnetic data from 36 Gough-Reitzel magnetometers	N/A	N/A	Numerical values	N/A	SGO	N/A
	Substorm number from ground magnetic data	online	< 1	N/A	< 1	SGO	N/A
	Rock property laboratory	N/A	N/A	Numerical values	<1	OMS	N/A
	local seimometer network	online	N/A	Induced seismicity episode data		OMS / AH na	TCS AH
Kerttu Saalasti Institute (KSI), UOULU	EMMA Muon Monitor	N/A	N/A	Numerical values	1 TB	KSI	N/A
FMI	Magnetometers	online	5	Numerical values	5	IMAGE	IMAGE
VTT MIKES	Finnish national time-scale UTC(MIKE)	online	1	Numerical values	<1	MIKES ftp-se	BIPM website/ftp-server
GTK	Magnetometers to measure remanent	N/A	N/A	Numerical values	N/A	GTK Hakku	N/A
	Susceptibility bridges	N/A	N/A	Numerical values	N/A	GTK Hakku	N/A
	Devices for density measurements	N/A	N/A	Numerical values	N/A	GTK Hakku	N/A
	Rock magnetic equipment	N/A	N/A	Numerical values	N/A	GTK	N/A
	Seismic p-wave velocity device	N/A	N/A	Numerical values	N/A	GTK	N/A
	Electrical resistivity equipment	N/A	N/A	Numerical values	N/A	GTK	N/A
	Instruments for geothermal studies	N/A	N/A	Numerical values	N/A	GTK	N/A
	Gamma-ray spectrometer	N/A	N/A	Numerical values	N/A	GTK	N/A
FLEX-EPOS	Mobile geophysical instruments	N/A	N/A	Geophysical data in field-specific data formats	N/A	UH, SGO, Aalto	N/A

users), restricted (available under certain conditions determined by the institution), embargoed (only available after a certain time following data collection/generation) or fee-based data. In principle, all the data products available through EPOS TCS are open access, but some data may be available only upon request or may involve an embargo ranging from 1–36 months. Reasonable restrictions (embargoed, user registration, user authorization) may be implemented for specific data sets if divulging them could jeopardize a potential

industrial/commercial application, violate the rules on personal data protection, or they are confidential for security reasons. The DPS for Research Data (Fig. 1) is dedicated for data sets that have great current and future value to an organization or at the national level. This service guarantees the preservation of digital assets from several decades to centuries. To use DPS for Research Data, a formal contract with CSC as well as a preservation agreement for a given data set(s) with the Ministry of Education and Culture are needed.

5. Documentation and metadata

To align with the FAIR data principles, the planned EPOS-FI metadata services (Fig. 2) are especially important. This includes metadata creation and management, catalogues, and persistent identifiers. Metadata can be harvested and presented both on a national level and via the EPOS TCS. For some of the data products metadata production can be automated to a large degree, but other data need substantial curation and even manual documentation. Not all parts of the EPOS-FI NRIs have metadata services, but these will be created based on examples from existing services.

Metadata must fulfill relevant national, EU, and EPOS requirements. Metadata standards accepted by the EPOS scientific community will be used (e.g., the EPOS Data Policy). NRIs have various practices and tools for documentation and producing metadata. When possible, NRIs automatically collect metadata from sensors and/or devices. Quality control of the raw data and metadata is performed by the NRI/supplier. Service Providers of the domain-specific EPOS TCS are also responsible for checking the quality parameters of the metadata descriptions. The EPOS framework will ensure a continuous process of review and assessment to verify that provision of EPOS DDSS is operating as envisioned, continually developing, and preventing and/or resolving problems.

EPOS TCSs are responsible for final inspection of the quality parameters of the metadata descriptions that provide information for discovery, contextualization and action, and provenance and traceability. Each TCS has a user feedback group. TCSs report the usage of the data to EPOS ERICs Service Coordination Committee.

The EPOS-FI DDSS will be published using persistent identifiers (e.g., PIDs, such as Digital Object Identifier (DOI) or Uniform Resource Name (URN)) minted by CSC's related services.In addition to this, citations to scientific papers describing and publishing NRIs data can be used. The users of the EPOS-FI data are supplied with appropriate references and citation information (article or PID) when data are downloaded. The data creators, processors, and curators must be ensured credit, which is supported by promoting and enabling use of ORCIDs and other identifiers as well as securing data lineage tracking and easy data citation.

6. Access control, backup, storage and disposal of administrative and research data

The data collected under the NRIs are stored by their central facilities. Initial storage takes place on servers at the NRIs, where the data are kept as long as economically feasible. Each

EPOS-FI partner organization has backup systems that duplicate or triplicate the national data and mirror it to other secured locations. Some NRIs also use data storing resources available from the CSC. EPOS-FI, however, does not currently have common data services and backups. The data storage services planned within the EPOS-FI are the Common intermediate storage and Long-term access data storage. The Common intermediate storage aims to handle data produced by measurement campaigns. These data need to be available for consortium members and other collaborators, and readily available to computational resources. Thus, a common storage/analysis solution is needed. The Long-term access data storage is needed for data that have been quality checked, processed, and used in publications (i.e., data products). These data are findable through persistent identifiers from various metadata repositories as described in Section 4. DPS is available to the NRIs with formal contracts. The EPOS-FI and CSC, as part of the consortium, will support developing data pipelines from the NRIs to the CSC's DPS once the contracts are in place. Regarding digital preservation, the Finnish Ministry of Education and Culture can offer preservation free of charge for select data that is regarded to be valuable on a national level.

7. Opening research data and/or metadata

Metadata and DDSS descriptions are always free and available at any time through the EPOS data portal and national metadata repositories, even for restricted and embargoed data. For access to restricted data, fees may be charged but they must not exceed the actual cost of making the DDSS available. The longest embargo time allowed is 3 years, and thereafter, data will become either open or restricted.

Different data have their own open publishing platforms according to theme, type, and level. Actual quality-controlled data streams are published by NRIs and international networks when data products are published in bulletin, yearbook, World Data Centre for Geomagnetism (WDC) data catalogue formats or in scientific publications. The research results are openly available and published in peer-reviewed journals.

All metadata, data, data products and software delivered to EPOS must have a data license attached to them. EPOS recommends using Creative Commons licenses CC BY 4.0 and CC BY NC 4.0. Some NRIs also use their own licenses. Data without a license will automatically use the CC-BY 4.0 license.

To better open the NRIs data resources for national use, EPOS-FI aims to establish central data services, including streamlining metadata generation and sharing to relevant portals and the storage for EPOS-FI data (described earlier) in dedicated data storage hardware (planned) or at the CSC in some cases. These services will likely be based on the CSC's service portfolio (including Fairdata services and Allas object storage) as well as EUDAT services where feasible.