Plan Overview

A Data Management Plan created using DMPTuuli

Title: Maritime Critical Infrastructure Research Platform

Creator: Jerker Björkqvist

Principal Investigator: Touko Herranen, Tomi Westerlund, Jarkko Paavola, Jerker Björkqvist

Data Manager: Vesa Rautio

Project Administrator: Vesa Rautio

Affiliation: Åbo Akademi University

Funder: The Research Council of Finland (former The Academy of Finland)

Template: Data Management Policy (RIFs)

Project abstract:

Finland, strategically positioned on the edge of the European Union, heavily relies on its maritime infrastructure. With over 90% of trade transported by sea and energy and information flowing through subsea pipelines and cables, maritime resources play a critical role. However, both environmental concerns and the geopolitical situation present challenges. Efficient utilisation of digitalisation offers possibilities to address these global challenges. However, research infrastructures are not available for the scientific community.

The novel MarCI infrastructure provides research and research result verification opportunities for the development of next-generation resilient shipping and waterborne activities, where automation and electrification will change how maritime activities are designed and operated. These technologies will environmental sustainability, maritime security, and offer new business opportunities while addressing societal challenges for seafarers. These aspects will provide resiliency in our society.

The Mar-CI infrastructure offers a comprehensive ecosystem for advancing maritime technologies. By providing data sets, a simulation platform, test vessels, port and fairway infrastructures in the picturesque Finnish archipelago, MarCI creates a unified environment. Researchers and technology providers can develop, validate, and demonstrate their ideas, algorithms, and products, ensuring their applicability.

MarCI unites four Higher Educational Institutions in the Turku region, leveraging their collective expertise. These institutions specialise in shipping, AI, communications, drones, IoT, sensor technology, cybersecurity, maritime law, and automation. Together, they build and develop a joint research infrastructure for advancing autonomous and electrified shipping.

On a national scale, MarCI collaborates with research institutions and companies, facilitating the scale-up of these technologies for environmental sustainability and security. This collaborative effort enables fostering innovation, development towards commercialisation, and addressing societal challenges for seafarers.

ID: 24829

Start date: 01-01-2025

Last modified: 15-05-2024

Maritime Critical Infrastructure Research Platform

Data management policy

1. General description of the administrative data of the RI

Administrative data includes project plans, agreements, meeting agendas, meeting minutes, budgets, economical reporting, contact information to researchers and collaborators.
The infrastructure itself does only store personal data needed for maintaining operations of the infrastructure: Names of researchers, emails, business phone numbers, roles in the organization. For collaborators and advisory board, the same information as for the researchers are stored.

· Turku University of Applied Sciences is the provider of data platform for test vessels and ROC. The platform is designed to serve as data storage for their user network. Members of the user network (later referred to as Users) can be different type of organizations including research institutions, educational institutions, and companies.
· Contract is used to define roles and responsibilities between User and Turku University of Applied Sciences. This contract specifies that in the case of administrative personal data, User has a role of data controller and Turku University of Applied Sciences has the role of data processor. Depending on the situation, User can also refer to Turku University of Applied Sciences.
· Administrative personal data (e.g. user names, contact information) processed by the platform includes:
o Name
o Username
o Email address
o Organization
· Link to privacy policy of research, development and innovation activities at Turku University of Applied Sciences:
o https://www.tuas.fi/media-en/filer_public/4e/e9/4ee9e993-8bbc-4d97-8951-8df03ea96ef4/privacy_notice_rdi_2023.pdf

2. General description of research data managed within RI

· On platform the main part of the research data consists of multimodal sensor data stored in blob storage. E.g., AIS (Automatic Identification System) data consists of numerical data about a vessel's location, velocity, heading and a timestamp for the data as well as text providing the vessel's reported destination.
· When downloaded from the platform the file format depends on the type of data. Text and numerical values are typically returned in csv files while video streams are returned as mkv files. The files are compressed into zip archives before uploading to the user.
· The main source for the data is the data platform users who can either upload data automatically by connecting their sensor sources to one of the data interfaces or by sending individual data objects directly to the platform.
· Public open data interfaces are also used, such as Digitraffic's AIS data interface. Datasets can be formed from individual data objects. The user interface for this is currently in the design stage. The user can query the platform for data objects and give the returned result a description as a dataset
The data produced with the research infrastructure includes:

1. Data from sensors onboard test vessels: Images, videos, AIS-data, lidar-data, radar-data, recorded VHF communication, telemetry data, location data, communication systems KPI-data (signal strength, base station information)
2. Syntetic data created using simulations, that includes similar data the data from sensor from actual vessels
3. System monitoring the infrastructure facilities, telemetry data

The data stored in a database is in standard file formats, according to the data they represent.
Estimate of the size of the data produced: About 100 TB of data in year.

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

· The platform does process personal research data. Contract is used to define roles and responsibilities between User and Turku University of Applied Sciences. This contract specifies that in the case of personal research data, Users have a role of data controller and Turku University of Applied Sciences has the role of data processor for test vessel and the ROC. Depending on the situation, User can also refer to Turku University of Applied Sciences.
· Link to privacy policy of research, development and innovation activities at Turku University of Applied Sciences:
o https://www.tuas.fi/media-en/filer_public/4e/e9/4ee9e993-8bbc-4d97-8951-8df03ea96ef4/privacy_notice_rdi_2023.pdf
· The platform does not process D sensitive research data.
Some of the actual data from test vessels captures information that can contain personal data. This might be in images, videos, AIS-data or recorder VHF communication. The evaluation of content of personal data must be evaluated after each test session. Such data must be marked and stored separately.

4. Agreements on research data rights

· The data platform has agreements about data usage rights. These agreements provide data platform users to have usage rights to all the data on the platform. We recommend granting wider usage rights, e.g. open data, to the data objects. In the metadata of the data objects and datasets, users must provide information about the usage rights.
The RI publish its data as open data, with access right to organizations and researchers, which apply for usage rights. Access rights are granted, unless special reason for not granting exists.
The RI research organizations owns the data they produce, but grant free license for data usage.

5. Documentation and metadata

· Guidelines are integrated into the tool the platform provides for uploading data onto the data platform. Metadata is required for all data objects and datasets.
· In the use case of automated sensor data collection, the data platform has integrity token calculation and both data format and content validation mechanisms implemented. No automated validation can be performed on manual uploads, manual validation is possible.
· The metadata used on the the platform conforms to Dublin Core metadata standard.
· The platform doesn't currently provide PID.
In the initial phase of the RI project, tools and guidelines for data handling are built, as this is a part of the research infrastructure. The data needed for successfully providing the services of the RI is identified and proper tools and guidelines for utilization of the data is formed.

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

- Turku University of Applied Sciences is responsible for controlling the technical access to the data in platform for test vessel and ROC.
- Access to the data is secured by authenticating individual users requesting data from the data platform. The user gets a temporary access token by logging in with his credentials.
- Users are unable to connect directly to the databases that store the data. Only query interfaces are reachable from outside the Turku University of Applied Sciences ICT laboratory network.
- Physical access to the servers in the data storages is restricted.
- Estimate roughly how much research data is produced or received in the RI over a given period of time (e.g. daily, monthly, yearly). Use GB, TB, or other units.
- As a baseline, the platform receives 300 MB of data per day from open data sources. Through sensor sources of particular Users, the platform can receive up to 9GB's of data per minute in ROSbag format.
- The platform offers data storage services for its user network. Data will be stored in the platform as long as the User has contractual relationship with ARPA platform.
- The platform does not provide Users backup or long-term preservation options.

The administrative data needed for administration of the RI is stored in Microsoft Teams, utilizing the access control, backup, storage and disposal functionalities that system provides.
The actual research data is stored in data storage system that is determined during the initial phase and is appropriate for achiving the the control, backup and storage capabilities needed. Data storage will be in the EU.

7. Opening research data and/or metadata

All data research data is provided as open data, with access rights to anyone applying for it, unless special reason exist for denying usage rights.

- Users are recommended to license their data with CC0 or CC-BY license before they store it in the platform.