Title:	RObot MANipulation of Deformables through dynamIC actions
Acronym:	**ROMANDIC**
Type of Action:	HORIZON Coordination and Support Actions
Grant Agreement No.:	101159522
Starting Date:	01-10-2024
Ending Date:	30-09-2027

ROMANDIC

Robot Manipulation of Deformables through Dynamic Actions

Deliverable Number:	D6.1
Deliverable Title:	Data Management Plan (DMP)[M6]
Type:	Report
Dissemination Level:	Public
Authors:	Cristian Barrué, Guillem Alenyà
Contributing Partners:	CSIC, JSI, KIT

Estimated Date of Delivery to the EC:	31-03-2025
Actual Date of Delivery to the EC:	31-03-2025

Table of Contents

Executive Summary

The ROMANDIC project (RObot MANipulation of Deformables through dynamIC actions) addresses robot manipulation of deformable objects, a rapidly advancing research area leveraging new robot hardware, innovative AI-driven control and perception methodologies, and expanding robotics applications beyond traditional industrial settings.

This Data Management Plan (DMP) has been structured according to FAIR principles—ensuring that all data generated within ROMANDIC are Findable, Accessible, Interoperable, and Reusable:

- **Findable**: All research outputs, including datasets, software, and documentation produced by ROMANDIC, will receive unique persistent identifiers (e.g., DOIs) and descriptive metadata to facilitate easy discovery by both the research community and stakeholders.
- **Accessible**: Project datasets will be openly accessible by default, shared through recognized repositories such as Zenodo or institutional data repositories, ensuring long-term availability. Exceptions, if any, due to intellectual property constraints or third-party agreements, will be clearly documented.
- **Interoperable**: Data and metadata will initially adhere to widely accepted standards and formats within the robotics and AI communities (e.g., ROS-compatible formats, JSON, CSV, XML) to maximize interoperability across various systems, platforms, and applications. These standards might evolve as the project progresses.
- **Reusable**: Clear licensing conditions (initially planned as Creative Commons licenses such as CC-BY or CC0) will accompany all shared outputs, specifying how datasets and tools can be reused by researchers and industry practitioners. Comprehensive documentation, including methodological descriptions, will support easy reuse and reproducibility.

ROMANDIC does not foresee the generation or processing of personal data, thus simplifying ethical considerations and data privacy compliance. The CSIC and collaborating institutions will manage and oversee implementation of this DMP, regularly reviewing and updating procedures throughout the project's lifespan to guarantee ongoing adherence to FAIR principles and best practices in data stewardship.

1. Data Summary

This section provides an overview of the datasets anticipated to be collected or generated within the ROMANDIC project. As the project is currently in its initial phases, some partners are still determining the exact nature, scale, and formats of data that will be produced during the course of their activities.

The following tables summarize planned data descriptors, including dataset types, formats, estimated volumes, and responsible partners. These tables represent an initial snapshot of anticipated data resources. The information provided here will evolve and be refined as the project progresses. A comprehensive update will be provided in the final version of the Data Management Plan, capturing all data actually generated or collected during ROMANDIC.

Table 1 CSIC Dataset Description

Data Purpose	The data is intended to track poses of robots and textiles (cloth) manipulated by humans or robots (teleoperated, automated, or simulated). It supports the development of manipulation primitives, perception methods training, and validation of cloth state estimation metrics.
Data Origin	Initially planned data collection includes RGB-D videos, motion capture systems in lab-controlled environments without identifiable individuals, and simulation environments comprising tables, cloth, and robot arms.
Types of data	1. Observational data: Recorded body, arm, and hand trajectories (position, velocity, acceleration). 2. Experimental data: RGB-D images capturing manipulation of textiles. Cloth trajectories and deformation states (depth, markers, mesh). 3. Simulation data: Cloth trajectories and deformation states (depth, markers, mesh). 4. Derived data: High-level textual descriptions of manipulations, cloth states, and action plans.
Data formats	Initially proposed formats are MP4 (video), ROS bag files, CSV, SQL, JSON, and point clouds. No personal data will be included.
Expected size of data	Estimated not to exceed 100 GB.
Re-use of existing data	Existing datasets will be reused as extensively as possible.
Data utility	The primary audience includes robotics researchers, particularly those working in cloth manipulation and machine learning communities.

Table 2 JSI Dataset Description

Data Purpose	- Robot learning for project execution: The data will be used to train and refine robotic learning algorithms, helping robots to learn and improve their ability to dynamically manipulate textiles. - Project analysis and evaluation: The data will be analyzed to assess the performance of robotic systems, identify areas for improvement, and evaluate the effectiveness of implemented control strategies.
Data Origin	- Existing databases on relevant topics/fields: We will use publicly available datasets, such as cloth manipulation databases and robot motion datasets, which are aligned with the project's objectives. -Self-collection: Data will also be self-collected through robot experiments, where robots perform various tasks on textiles, manipulating different cloth types.
Types of data	- Real robot data: Robot joint positions, joint torque values, and robot control data (torques, impedances), end effector positions and orientations, endeffector forces - Camera data: Images, depth information, and video data will be collected by cameras (e.g., RGB, stereo, depth cameras) to capture the environment, robot movements, and textile manipulation actions. - Robot learning data: This includes datasets for learning-based methods such as textile manipulation databases, models learned through reinforcement learning or imitation learning, and data generated from robot simulations. - State planning data: The state graph (which tracks possible states of the robot and textiles) and the action execution graph (which represents sequences of robot actions and transitions) will be used for robot planning and decision-making.
Data formats	- Real robot data: JSON, CSV. - Camera data: JPEG, PNG, CSV, MP4. - Robot learning data: - Model data: ONNX, TensorFlow SavedModel, or PyTorch models for storing trained models. - Simulation result data type: HDF5, CSV, or JSON, depending on the format of the simulation software. - State planning data: - State graph data type: GraphML, DOT, or JSON. - Action execution graph data type: GraphML, JSON, or a custom format, depending on how the action sequences and transitions are represented.
Expected size of data	The expected size of data generated over the course of the project will be in the range of a few hundred gigabytes, depending on the frequency of data collection, the number of experiments conducted, and the type of data (images, simulation results, etc.).
Re-use of existing data	Existing databases on relevant topics/fields: Existing datasets like annotated cloth manipulation datasets (e.g., from the ICRA or other robot learning challenges), and robot motion data will be reused. This will allow for initial model training to reduce the time required for real-world training.
Data utility	The collected data will be used to train and evaluate robot learning algorithms for better dynamic manipulation of deformable objects like textiles. Generated data will be curated and evaluated for publication along with research results.

Table 3 KTI Dataset Description

Data Purpose	We will record human motion data for learning actions for dynamic manipulation of deformable objects based on human demonstrations. This data is needed for addressing the research questions described in WP4 regarding learning, planning and execution of manipulation strategies.
Data Origin	Data collection will be conducted at the KIT human motion capture studio using a comprehensive multimodal sensor setup consisting of 1) marker-based VICON motion capture system, 2) Cyber Glove III data gloves, 3) 9-DoF inertial measurement units (IMUs), 4) RGB-D cameras and 5) egocentric RGB camera.
Types of data	The data includes whole-body human motion trajectories, object motion, contact forces, and environmental information.
Data formats	The raw sensor data will be captured in their respective native formats. All data streams will subsequently be unified and standardized using the Master Motor Map (MMM) framework to ensure cross-modal compatibility.
Expected size of data	Expected volume is a few TB. All collected data will be systematically processed, annotated, and made available to the broader research community via the KIT Whole-Body Human Motion Database https://motion-database.humanoids.kit.edu
Re-use of existing data	To bootstrap our research, we will initially leverage existing bimanual manipulation datasets from the KIT Whole-Body Human Motion Database. These baseline datasets will provide comparative benchmarks and initial training data before our specialized deformable object manipulation datasets are fully developed.
Data utility	The collected data will serve multiple purposes: 1. Training machine learning models for dynamic deformable object manipulation, 2. Benchmarking and evaluating the developed action and skill learning methods, 3. Enabling comparative studies across different manipulation strategies, and 4. Advancing the broader research field through open access to standardized datasets. The data will be structured to maximize utility for both our research consortium and the wider robotics and AI research community.

2. FAIR Data

2.1. Making data findable, including provisions for metadata

While there is currently no specific 'Minimal Metadata' standard defined for ROMANDIC, the team plans to use essential metadata fields (such as date, experimental conditions, robot

type, sensor model) for future interpretability and discoverability. Each dataset will receive a persistent identifier (DOI).

We will use an electronic lab notebooks to make sure that there is good provenance of the data analysis.

We will be keeping the relationships between data clear in the file names. All the metadata in the file names also will be available in the proper metadata.

2.2. Making data accessible

The project's datasets will be openly accessible, working with the philosophy *as open as possible,* initially through Zenodo or institutional repositories. Data not restricted legally will be publicly available at the project's end or earlier, alongside clear documentation on data access procedures. Potential data embargoes will be transparently communicated.

Metadata will be openly available including instructions how to get access to the data. This metadata will available in a form that can be harvested and indexed (managed by the used repository / repositories).

We have a consortium agreement that arranges Intellectual Property.

2.3. Making data interoperable

ROMANDIC initially plans to adopt standard vocabularies and data formats recognized by the robotics community (e.g., ROS standards, IEEE standards, controlled ontologies). The selected standards may be adjusted, always ensuring maximum interoperability and alignment with FAIR principles.

2.4. Increase data re-use

As stated already in Section 2.2, all of our data can become completely open over time. We do not plan to be archiving data (using so-called *cold storage*) for long term preservation already during the project.

Datasets will initially be shared under Creative Commons licenses (such as CC-BY or CC0) to facilitate maximum reuse potential. The project will provide comprehensive documentation, clearly stating conditions for reuse.

Automated workflows and electronic lab notebooks will document data processing and analysis steps, ensuring transparency, reproducibility, and validation.

3. Other research outputs

We use Data Stewardship Wizard for planning our data management and creating this DMP. The management and planning of other research outputs will be done separately and included as appendix to this DMP. Still, we benefit from data stewardship guidance (e.g. FAIR

principles, openness, or security) and it is reflected in our plans with respect to other research outputs.

4. Allocation of resources

FAIR is a central part of our data management; it is considered at every decision in our data management plan. We use the FAIR data process ourselves to make our use of the data as efficient as possible.
None of the used repositories charge for their services.
Cristian Barrué is responsible for implementing the DMP, and ensuring it is reviewed and revised. To execute the DMP, no additional specialist expertise is required.
We do not require any hardware or software in addition to what is usually available in the institutes.

5. Data security

Project members will not carry data with them (e.g. on laptops, USB sticks, or other external media). All data centres where project data is stored carry sufficient certifications. All project web services are addressed via secure HTTP (https://...). Project members have been instructed about both generic and specific risks to the project.

The possible impact to the project or organization if information is lost is small. The possible impact to the project or organization if information is leaked is small. The possible impact to the project or organization if information is vandalized is small.

We are not using or storing any personal information during the execution of the project.

Compliance with institutional and EU guidelines on data security will be explicitly maintained.

6. Ethics

ROMANDIC does not involve collecting personal data, simplifying ethical considerations. All data collection complies with relevant ethical and legal regulations.

This Data Management Plan will be regularly reviewed and updated throughout the project's lifespan to accommodate changes while maintaining alignment with FAIR principles and good data stewardship practices.