D1.3: Data Management (Update, RP1)

Effective data management is central to TRIAD's mission of fostering transparency, reproducibility, and long-term usability of research outputs. Since the project's launch, the Data Management Plan (DMP) has been implemented as a dynamic framework that aligns with FAIR principles and adapts to evolving project needs, technological advances, and regulatory requirements.

1. Electronic Documentation and Lab Practices

TRIAD partners have adopted electronic lab notebooks (ELNs) across all sites to ensure structured documentation of experimental workflows. The ELNs are being used in day-to-day practice, with common templates for experiment metadata, data provenance, and version control.

- **Enhanced collaboration**: Joint protocols for telomere RNA sequencing and live-cell imaging are already documented and shared through ELNs, enabling researchers in Lisbon, Drevníky, and Heraklion to annotate and cross-validate procedures in real time.
- **Compliance and audit readiness**: TRIAD has introduced audit trails within ELNs, ensuring compliance with EU data governance requirements and journal reproducibility standards.
- **Security and backup**: Data are encrypted and stored on institutional servers with automatic nightly backups, with local policies harmonised across partners.

2. Data Deposition and Repository Use

During RP1, TRIAD has already deposited or prepared datasets in recognized repositories, demonstrating that the DMP is operational and not only aspirational.

- **Genomics and transcriptomics**: Pilot long-read sequencing datasets of telomeric repeat-containing RNAs (TERRA) are being formatted for deposition in **ArrayExpress**, with standardised metadata templates in place.
- **Proteomics and metabolomics**: Initial proteomics results generated from telomere-associated protein complexes are being curated for submission to **ProteomeXchange**.
- **CRISPR/Cas9 screens and molecular tools**: Plasmids engineered for TERRA knockdown have been submitted to **Addgene**, ensuring early dissemination and reproducibility. Discussions are ongoing for registering pooled CRISPR screen results in **ORCS**.
- **Cell lines**: New telomere-reporter cell lines generated within TRIAD are undergoing authentication and will be deposited at **DSMZ** in RP2.

This early use of repositories underscores TRIAD's commitment to embedding open science practices directly into the research workflow, rather than as an end-of-project obligation.

3. Adaptability of the Data Management Plan

TRIAD's DMP has been refined during RP1 to incorporate lessons learned and new developments:

- **Integration of AI-assisted analysis**: With the adoption of AlphaFold2 in telomere protein structure analysis, the DMP was updated to include procedures for documenting model training data, reproducibility criteria, and file formats.

- **Standardisation across partners**: Initial heterogeneity in metadata annotation was identified during cross-lab dataset comparison. A harmonised metadata standard—based on ELIXIR recommendations—has now been implemented.
- **Licensing and reuse conditions**: A Creative Commons Attribution (CC-BY) license has been adopted as the default, ensuring clarity of reuse for external researchers.

4. Evolving Management Strategy

The DMP is positioned as a living document, with biannual reviews incorporated into TRIAD's governance structure. These reviews ensure responsiveness to:

- **New data modalities** (e.g., RHINO-based live imaging, EV-mediated delivery data), which require specific annotation and storage strategies.
- **Policy changes** at EU and national levels, including GDPR compliance for potential patient-derived cell data in RP2.
- **Consortium dynamics**, such as onboarding new early-stage researchers, who receive training on ELN use, data deposition workflows, and FAIR compliance.

5. Training and Capacity Building

Data management has been embedded in TRIAD's training programme. At the Lisbon and Drevníky retreats, dedicated sessions on FAIR principles, repository submission, and research integrity (EMBO workshop) ensured that >50 researchers gained hands-on skills in responsible data stewardship. This ensures continuity and sustainability beyond the lifetime of the project.

Concluding remarks: TRIAD's DMP is fully operational and has already demonstrated measurable progress in data documentation, deposition, and training. By embedding electronic documentation into routine practice, initiating repository submissions during RP1, and refining the plan to accommodate new technologies and metadata standards, TRIAD ensures that its outputs are not only scientifically impactful but also reusable, reproducible, and aligned with European Open Science policies. The living nature of the DMP guarantees that it will remain responsive and fit-for-purpose throughout the project lifecycle.