

CryoET mass manage

Establishing FAIR CryoET Data Stewardship to maximize scientific value, reusability, and EU Mission impact

Programm / Ausschreibung	Implementing EU Missions, IMPA 2024, Data Stewards 2025	Status	laufend
Projektstart	01.10.2025	Projektende	31.03.2028
Zeitraum	2025 - 2028	Projektlaufzeit	30 Monate
Keywords	Cryo-electron tomography; Mission Cancer; Big data management; FAIR data; cross-project integration		

Projektbeschreibung

Ausgangssituation, Probleme oder Motivation:

Die Kryo-Elektronentomographie (CryoET) ist eine Schlüsseltechnologie der Struktur- und Zellbiologie, die es erlaubt, makromolekulare Strukturen im nativen Zustand hochauflösend darzustellen. Dabei entstehen jedoch extrem große und komplexe Datensätze (1-2 TB pro Sitzung), die erhebliche Herausforderungen im Datenmanagement mit sich bringen. Obwohl CryoET am IMBA eingesetzt wird, fehlt derzeit eine strukturierte, FAIR-konforme (Findable, Accessible, Interoperable, Reusable) Datenverwaltung. Dies gefährdet Datenqualität, Wiederverwendbarkeit und die Einhaltung von Anforderungen durch Fördergeber (z. B. Horizon Europe) und kann den Nutzen teurer Infrastruktur einschränken. Das Projekt reagiert auf die dringende Notwendigkeit, professionelles Datenmanagement für CryoET zu etablieren.

Ziele und Innovationsgehalt:

Ziel des 30-monatigen Projekts ist die Einrichtung einer spezialisierten Stelle für CryoET-Datenmanagement. Diese Rolle entwickelt und implementiert ein umfassendes, FAIR-konformes Datenmanagementsystem entlang des gesamten Data lifecycle – von der Planung bis zur Archivierung. Kernpunkte sind:

1. Standardisierte Metadatenmodelle und Datenorganisation (z. B. auf Basis des EMDB-Modells),
2. Unterstützung effizienter Datenablage in öffentliche Repositorien (EMDB, EMPIAR, CryoET Data Portal),
3. Qualitätskontrolle und Provenienzverfolgung,
4. Schulungen für Forschende zu FAIR-Prinzipien und spezifischem Datenhandling. Die Innovation liegt in der Schaffung einer maßgeschneiderten Datenmanagement-Funktion für CryoET und der nachhaltigen Verankerung von Best Practices in der Forschungskultur.

Erwartete Ergebnisse:

Erreicht werden soll ein robustes System zur FAIR-konformen Verwaltung von CryoET-Daten. Erwartete Ergebnisse sind:

1. Hochwertige, öffentlich zugängliche Datensätze,
2. Effizientere Abläufe und geringere Belastung der Forschenden,
3. Verbesserte Datenqualität und Reproduzierbarkeit,

4. Höhere Sichtbarkeit und Nachnutzbarkeit der Daten,
5. Kuratierte Datensätze für KI-/ML-gestützte Analyse,
6. Aufbau einer qualifizierten Kerngruppe für Datenmanagement und eine nachhaltige Datenkultur am IMBA.

Abstract

Initial situation, problems, or motivation:

Cryo-electron tomography (CryoET) is a transformative imaging technique providing unprecedented insights into macromolecular and cellular structures in their native state, critical for advancing structural and cellular biology. However, CryoET inherently generates exceptionally large (1-2 TB per session) and complex datasets, presenting significant data management challenges ("big data" volume, variety, velocity). IMBA utilizes CryoET but currently lacks dedicated expertise to systematically manage this data lifecycle according to FAIR (Findable, Accessible, Interoperable, Reusable) principles. This gap risks inconsistent practices, data underutilization, difficulties meeting funder mandates (e.g., Horizon Europe), and diminished returns on significant infrastructure investments. This project is motivated by the critical need to establish expert data stewardship specifically for CryoET to ensure data quality, long-term value, and compliance.

Goals and innovation content:

This 30-month project aims to establish a dedicated CryoET Data Steward role within the organisation. The primary goal is to develop and implement a comprehensive, FAIR-compliant data management framework covering the entire CryoET data lifecycle, from experimental planning to public archiving. Key objectives include: 1) Developing standardized metadata schemas (based on community standards like the EMDB data model) and data organization conventions. 2) Establishing streamlined workflows and providing support for efficient deposition of CryoET data (raw images, processed volumes, particle stacks, segmentations, models, metadata) into public archives (EMDB, EMPIAR, CryoET Data Portal). 3) Implementing robust data quality control procedures and provenance tracking. 4) Developing and delivering targeted training for researchers on FAIR principles and CryoET-specific data management. The innovation lies in creating a specialized stewardship function tailored to the unique demands of CryoET data, fostering a data-aware culture and integrating best practices into the research environment.

Desired results:

The project will result in a robust, sustainable system for managing CryoET data according to FAIR principles. Expected outcomes include: 1) High-quality, well-annotated, FAIR CryoET datasets deposited in public archives, enhancing their visibility and reusability by the global scientific community. 2) Increased efficiency in CryoET research workflows through standardized processes and reduced data management burden on researchers. 3) Improved data quality, integrity, and potential for reproducibility. 4) Enhanced capacity for finding, accessing, and reusing CryoET data, accelerating internal discoveries and facilitating collaborations. 5) Creation of valuable, curated datasets suitable for developing and training advanced computational tools, including AI/ML algorithms for automated tomogram analysis. 6) A core group of researchers skilled in CryoET data stewardship, contributing to a sustainable data management culture within the organisation.

Projektpartner

- IMBA - Institut für Molekulare Biotechnologie GmbH