

FAIR Data Sustain

Data Steward: FAIR multi-modal big data for EU Mission CANCER & Data Sustainability

Programm / Ausschreibung	Implementing EU Missions, IMPA 2024, Data Stewards 2025	Status	laufend
Projektstart	01.02.2026	Projektende	31.07.2028
Zeitraum	2026 - 2028	Projektlaufzeit	30 Monate
Keywords	sustainability; FAIR big data; multi-modal data; multi-omics; cancer research		

Projektbeschreibung

Das Institut für Molekulare Biotechnologie (IMBA), ein führendes europäisches biomedizinisches Forschungsinstitut im Vienna BioCenter (VBC), nutzt fortschrittliche Bildgebung (z.B. smRNA-FISH, 3D-Zellkernbildgebung, Lebendzellmikroskopie) und diverse Sequenzierverfahren (z.B. RNAseq, SPRITE, scRNA-seq, räumliche Transkriptomik). Diese Technologien erzeugen riesige, komplexe, multimodale Datensätze, entscheidend für das Verständnis von Krankheiten wie Krebs. Dem IMBA fehlt derzeit ein spezialisierter Data Steward für den gesamten Lifecycle dieser Datentypen. Diese Lücke führt zu heterogenem Datenmanagement, birgt Risiken der Datenunterauslastung, ineffizienter Forschungsworkflows, erschwerter multi-omics Datenintegration und beeinträchtigt die Reproduzierbarkeit. Unsere Motivation ist der dringende Bedarf, eine expertenbasierte Data Stewardship zu etablieren, um diese wertvollen Datenressourcen gemäß FAIR-Prinzipien zu verwalten, ihren wissenschaftlichen Wert zu maximieren, Kollaborationen zu fördern und datengesteuerte biomedizinische Spitzenforschung zu ermöglichen.

In diesem 30-monatigen Projekt wird Rolle des Data Steward mit der Gesamtverantwortung für das "FAIR Data Sustain" Projekt am IMBA eingeführt. Diese Fachkraft entwickelt und setzt einen umfassenden Datenmanagementplan für die multimodalen Daten von IMBA und den VBC-Partnerinstituten um. Die Ziele des Projekts sind: 1) Entwicklung maßgeschneiderter Datenmanagementpläne (DMPs) für die Workflows der Bildgebung und Sequenzierung, die mit den institutionellen Strategien und Förderrichtlinien abgestimmt sind.

2) Einführung standardisierter Metadatenschemata (z.B. Erweiterung von OMERO für Bildgebung; Anpassung von GA4GH-Elementen für krebsrelevante genomische/phänotypische Daten) und robuster Standard Operating Procedures (SOPs) für eine konsistente Datenannotation, -erfassung und Qualitätskontrolle (QC).

3) Durch die Nutzung der VBC-Computerinfrastruktur (einschließlich des CLIP HPC-Clusters) wird eine effiziente Datenspeicherung und strukturierte Archivierung ermöglicht, sowie sichere Datenzugriffsprotokolle definiert.

4) Es werden spezifische RDM-Schulungen angeboten, um die institutionellen Kapazitäten im Bereich Forschungsdatenmanagement aufzubauen.

Die Innovation liegt in der Schaffung einer spezialisierten, proaktiven Stewardship-Rolle, die modernste Datengenerierung mit Best-Practice-Datenmanagement verbindet, spezifisch für anspruchsvolle Bildgebungs- und Sequenzierdaten in einem Spitzenforschungsumfeld. Dies unterstützt direkt die datenintensive Krebsforschung und dient als Betriebsmodell für eine breitere RDM-Implementierung am VBC.

Die erfolgreiche Projektumsetzung wird Schlüsselergebnisse liefern, die Entdeckungen im Rahmen der EU-Mission CANCER: 1. Erstellung hochwertiger, gut annotierter, FAIRer Bildgebungs- und Sequenzierdatensätze, die für Forschung auffindbar, zugänglich, interoperabel und wiederverwendbar sind. 2. Etablierung optimierter, dokumentierter Datenworkflows zur Reduzierung des Datenmanagementaufwands. 3. Signifikante Verbesserung der Kapazität für komplexe Datenintegration – wichtig für systembiologische Ansätze in der Krebsforschung und datenintensiven Feldern. 4. Aufbau einer Kerngruppe von Forschenden, geschult in modernen Datenmanagementpraktiken und nachhaltiger Data Stewardship. Der etablierte RDM-Rahmen dient als skalierbares Modell für andere Datentypen und informiert breitere institutionelle RDM-Strategien.

Abstract

The Institute of Molecular Biotechnology (IMBA), a leading European biomedical research institute within the Vienna BioCenter (VBC), leverages advanced imaging (e.g., single-molecule RNA-FISH, 3D nuclear imaging, live-cell microscopy) and diverse sequencing applications (e.g., RNAseq, SPRITE, single-cell RNA-seq, spatial transcriptomics). These technologies generate vast, complex, and multi-modal datasets crucial for understanding complex diseases like cancer. Currently, IMBA lacks a specialized data steward to manage the lifecycle of these data types. This gap leads to heterogeneous data management practices, risking data underutilization, inefficient research workflows, challenged multi-omic data integration, and impeding reproducibility. Our motivation is the very urgent need to establish expert data stewardship and ensure these valuable data assets are managed according to FAIR principles, thereby maximizing their scientific value, fostering collaboration, and enabling cutting-edge, data-driven biomedical discovery.

This 30-month project aims to establish and embed a Data Steward for the FAIR Data Sustain project at IMBA. This expert will develop and implement a comprehensive data management plan specifically for the multi-modal data generated across IMBA and VBC partner institutes. Project goals: 1) Develop tailored Data Management Plans (DMPs) for core imaging and sequencing workflows, ensuring alignment with institutional strategies and funder mandates. 2) Establish standardized metadata schemas (e.g., enhancing OMERO for imaging; adapting GA4GH elements for cancer-relevant genomic/phenotypic data) and robust Standard Operating Procedures for consistent data annotation, collection, and quality control (QC). 3) Facilitate efficient data storage, structured archiving using VBC's computational infrastructure (incl. CLIP HPC cluster), and defining secure data access protocols. 4) Provide targeted RDM training and support to build institutional capacity.

The innovation of this project lies in creating a specialized, proactive stewardship role that links cutting-edge data generation and best-practice data management, specifically for high-demand imaging and sequencing data within a top-tier research environment. This will directly support data-intensive cancer research and serve as an operational model for broader RDM implementation across the VBC.

The successful implementation of this project will yield several key results that will accelerate discoveries relevant to the EU Mission CANCER and other strategic research areas at IMBA: 1. Create quality, well-annotated, FAIR imaging and sequencing datasets that are findable, accessible, interoperable, and reusable for current and future research. 2. Establish streamlined, documented data workflows to reduce the data management burden on researchers, improving operational efficiency and contribute to research reproducibility and reliability. 3. Significantly improve the capacity for complex data integration - important for advancing systems biology approaches in cancer research and other data-intensive fields. 4. The project will

build a core group of research staff proficient in state-of-the-art data management practices, sustainable data stewardship. The established RDM framework for imaging and sequencing will serve as a valuable, scalable model for other data types and inform the ongoing development of broader institutional and campus-wide RDM strategies.

Projektpartner

- IMBA - Institut für Molekulare Biotechnologie GmbH