

## FAIR2earth

A FAIR Data Infrastructure for Earth System Sciences

<b>Programm / Ausschreibung</b>	F&E Infrastruktur, F&E Infrastruktur, F&E Infrastruktur 3. Ausschreibung	<b>Status</b>	laufend
<b>Projektstart</b>	01.01.2021	<b>Projektende</b>	31.12.2025
<b>Zeitraum</b>	2021 - 2025	<b>Projektlaufzeit</b>	60 Monate
<b>Keywords</b>	Earth System; global; monitoring; IT Infrastructure; climate change		

### Projektbeschreibung

Mit einer Weltbevölkerung von nun fast 8 Milliarden Menschen hat die Menschheit ihre Spuren bereits in allen Teilen unseres Planeten hinterlassen. Der Klimawandel ist nur eine der unbeabsichtigten Auswirkungen, die Verschlechterung des Ökosystems und der Verlust von Arten sind weitere zutiefst besorgniserregende Entwicklungen. Ein Verständnis dieser Veränderungen erfordert eine ganzheitliche Sicht auf das System Erde, in der Daten und Fachwissen aus einem breiten Spektrum wissenschaftlicher Disziplinen zusammengeführt werden müssen. Leider sind der Datenaustausch und die interdisziplinäre Zusammenarbeit zur Weiterentwicklung der Geo-Wissenschaften immer noch mit vielen praktischen Problemen verbunden. Eines davon ist die Tatsache, dass viele wertvolle Datensätze nicht leicht auffindbar, zugänglich, interoperabel und wiederverwendbar sind (Findable, Accessible, Interoperable, and Reusable (FAIR)). Darüber hinaus mangelt es weiterhin an Rechenressourcen, die es ermöglichen, große Datenmengen mithilfe von maschinellem Lernen und anderen aktuellen Modellierungstechniken zu kombinieren und gemeinsam zu analysieren. Ziel des vorgeschlagenen FAIR2earth-Projekts ist es daher, die vorhandenen Infrastrukturfähigkeiten des Erdbeobachtungsdatenzentrums (EODC) zu erweitern, um Erdbeobachtungs- und andere geowissenschaftliche Datensätze in der Größenordnung von mehreren Petabyte zu speichern und die Abbindung und Kooperation mit der European Open Science Cloud (EOSC), den Copernicus Data and Information Access Services (DIAS) und anderen europäischen Infrastrukturen zu stärken. FAIR2earth wird die Nutzung des EODC für jeden interessierten Benutzer auf diskriminierungsfreie Weise (gleiche Bedingungen gelten für alle Benutzer) gemäß den FAIR-Grundsätzen erheblich vereinfachen.

Die FAIR-Prinzipien sind von zentraler Bedeutung für die Open Science-Bewegung, die darauf abzielt, wissenschaftliche, öffentliche und private Daten, Codes und Veröffentlichungen für alle zugänglich zu machen und die wissenschaftlichen Standards (Reproduzierbarkeit usw.) zu erhöhen und so den gesellschaftlichen Nutzen zu verbessern. Um der Intention von FAIR Data and Open Sciences gerecht zu werden, brauchen Wissenschaftler, die ihre Datensätze kontinuierlich verbessern und sie offen zu teilen wollen, Infrastrukturen, die den gesamten Datenlebenszyklus ab der Datenerfassung abdecken (Qualitätssicherung, Beschreibung der Metadaten, langfristige Aufbewahrung, wissenschaftliche Nutzung, etc). Das FAIR2earth-Projekt soll die vorhandene EODC-Infrastruktur (Netzwerk, Speicher, Rechenleistung) signifikant erweitern, um alle Phasen des FAIR-Datenlebenszyklus effektiver zu unterstützen. Viele der einzelnen Komponenten existieren bereits, sind

jedoch entweder nicht leistungsfähig genug noch ausreichend integriert, um großen Anzahl an Nutzer die einfache Verwirklichung von FAIR-Daten und Open Science-Prinzipien zu ermöglichen.

## **Abstract**

With the global population approaching 8 billion people, humans have started to leave their imprints on all parts of our planet. Climate change is just one of the unintended impacts of human activities, ecosystem degradation and the loss of species are other deeply worrying developments. An understanding of these changes requires a holistic view of the Earth System, bringing together the data and expertise from a broad range of scientific disciplines. Unfortunately, data sharing and interdisciplinary collaboration to advance Earth System Science is still met by many practical problems, one of which is the fact that many valuable data sets are not easily Findable, Accessible, Interoperable, and Reusable (FAIR). Furthermore, there is a lack of computing resources allowing to combine and jointly analyse large data sets using machine learning and other advanced modelling techniques. Therefore, the aim of the proposed FAIR2earth project is to scale up the existing infrastructure capabilities of the Earth Observation Data Centre (EODC) to be able to store and process multiple Petabyte-scale Earth observation and geoscientific data sets on site, and improve its connectivity with the European Open Science Cloud (EOSC), the Copernicus Data and Information Access Services (DIAS), and other European infrastructures. In this way, accessing and exploiting EODC's Petabyte-scale data repository will be made much more easy for any interested user in a non-discriminatory manner (same conditions apply for all users), complying to the FAIR principles.

The FAIR principles are central to the Open Science movement which, on a practical level, aims to make scientific, public and private data, code and publications accessible to all, but with the more ambitious goals to elevate scientific standards (reproducibility etc.) and enhance societal benefits. Living up to the true spirit of FAIR Data and Open Sciences requires that scientists, who are committed to keep on improving their data sets and share them in an open spirit, can work on infrastructures that serve the complete data life cycle, from data collection, quality assurance, metadata description, long-term preservation, scientific exploitation, regular re-processing, to publishing. Recognising this need, the FAIR2earth project will scale up the existing EODC infrastructure to more effectively deal with all stages of the FAIR data life cycle, from a powerful network to collect, share and re-distribute data, dedicated processing clusters, multi-tiered storage systems for on-line data usage, backup, and public repositories, analysis platforms for scientific analysis, and high-performance computing for large-scale analysis and processing efforts. Many of these individual components already exist, but are either not powerful enough or are not sufficiently integrated to let a large and diverse Earth System Science community work towards realising FAIR Data and Open Science principles.

## **Projektkoordinator**

- EODC Earth Observation Data Centre for Water Resources Monitoring GmbH

## **Projektpartner**

- Technische Universität Wien