

EOSC-Pillar

Coordination and Harmonisation of National & Thematic Initiatives to support EOSC

Bilan EOSC-Pillar/Data Terra-Odatis - Use case #2 « Environnement & Système Terre »

Marine Vernet (Ifremer/Data Terra), Christelle Pierkot
(CNRS/Data Terra),
Joël Sudre (CNRS/Data Terra), Gilbert Maudire (Ifremer),
Valérie Harscoat (Ifremer)



EOSC-Pillar has received funding from the European Union's Horizon 2020 research and innovation Programme under Grant Agreement No. 857650.

This material by the EOSC-Pillar is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

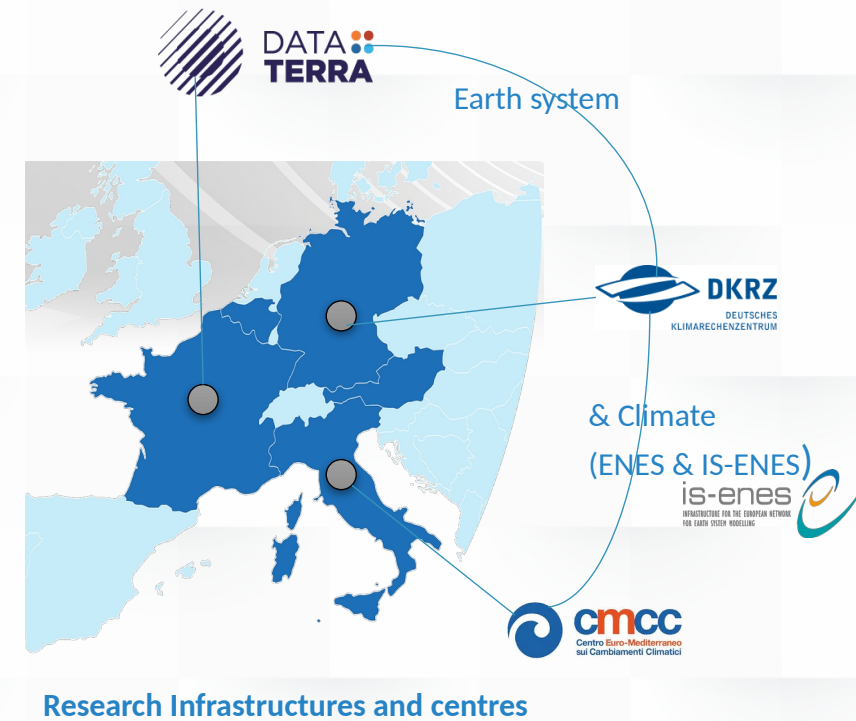


UC#2 – Environnement & Système Terre

Use Cases and Community-driven Pilots

Use cases

<p>Defining procedures and services to enforce data provenance for thematic communities and beyond Due to data exploration complexity, provenance management is a key component in order to guarantee...</p> <p>view</p>	<p>Agile FAIR Data for Environment and Earth System Communities Earth Environment Sciences & Geosciences require a large panel and volume of data from satellite...</p> <p>view</p>	<p>Integration of data repositories into EOSC based on communities approaches The agriculture, food and environment research community faces many challenges common to all: Gas...</p> <p>view</p>	<p>Software source code preservation, reference and access Leveraging the experience of Software...</p> <p>view</p>
<p>view</p>	<p>view</p>	<p>view</p>	<p>view</p>
<p>FAIR principles in data life-cycles for Humanities This task aims to identify and develop use cases based on Social Sciences and Humanities (SSH) co...</p> <p>view</p>	<p>Exploring reference data through existing computing services for the bioinformatics community Galaxy is a widely adopted workflow management system for bioinformatics, aiming to make computat...</p> <p>view</p>	<p>Suitable data formats for seismological big data provisioning via web services Seismological web services have been designed some years ago with particular types of user and da...</p> <p>view</p>	<p>Virtual definition of data sets according to RDA recommendations At GEOIDN data centre is very difficult to offer big pre-assembled datasets to be downloaded, due...</p> <p>view</p>



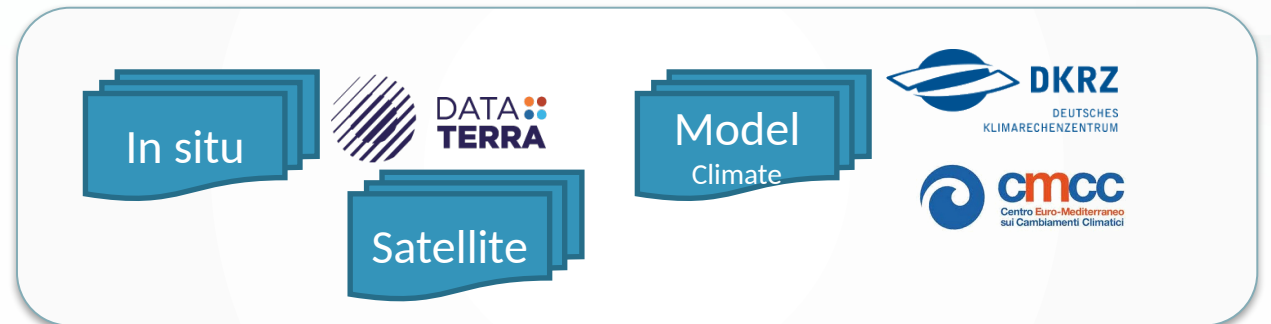
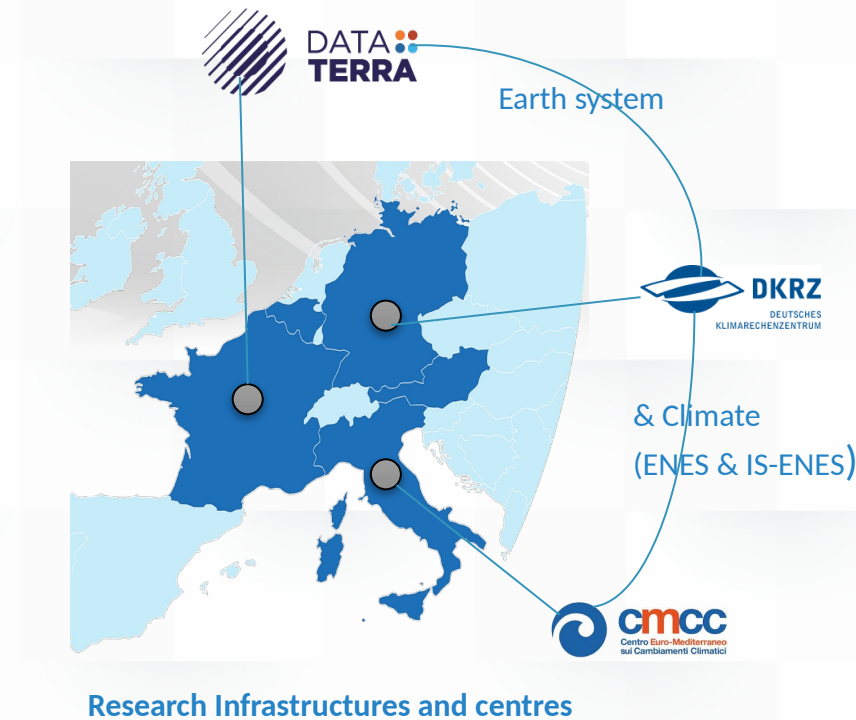
UC6.2 - Partenaires

UC#2 – Contexte

Sciences des données pour les communautés de l'environnement et de la Terre

- * 2 types d'utilisateurs: scientifiques et analystes de données
- * **Inter-domaine**, indépendamment des thèmes scientifiques ou organisations
 - * Intégration des observations satellite et in-situ
 - * Etudes multi ou interdisciplinaires « Sciences de l'environnement et de la Terre »
- * **Large panel de sources**
 - * satellite, in situ et climate model
- * **Volume important**
 - * Plusieurs centaines de pétaoctets pour les données satellitaires et sorties de modèle
 - * Plus d'un million d'observations in-situ (très nombreuses)
- * **Diversité des types**
- * **Avec des métadonnées communes spécifiques**
 - * Géo-localisation (latitude, longitude, high/depth)
 - * Temps

➔ **Entrepôts de données (& services) multiples et distribués**



UC#2 - Objectifs

F_{indable} A_{ccessible} I_{nteroperable} R_{eusable}

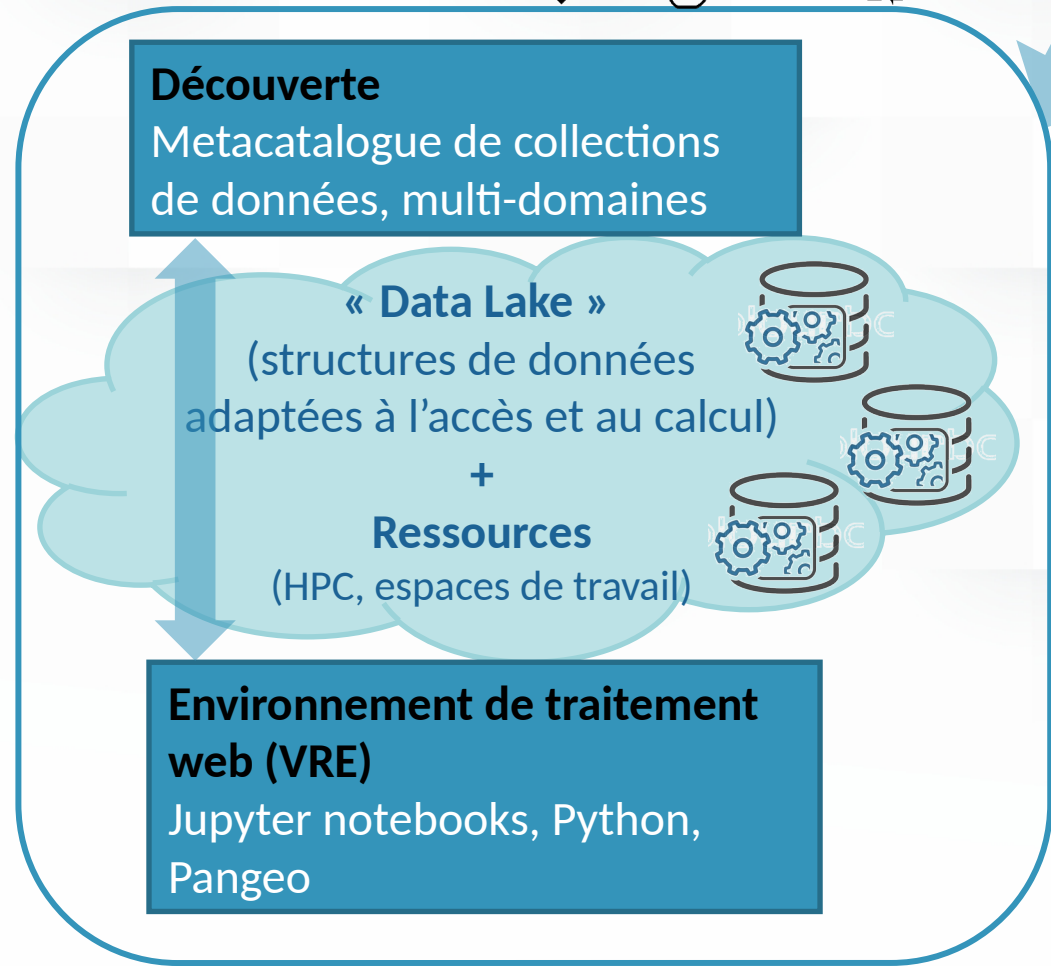
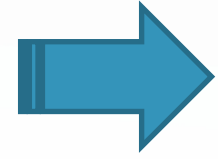
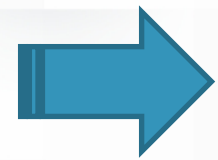
Entrepôts

Scientifique
ou analyste de données

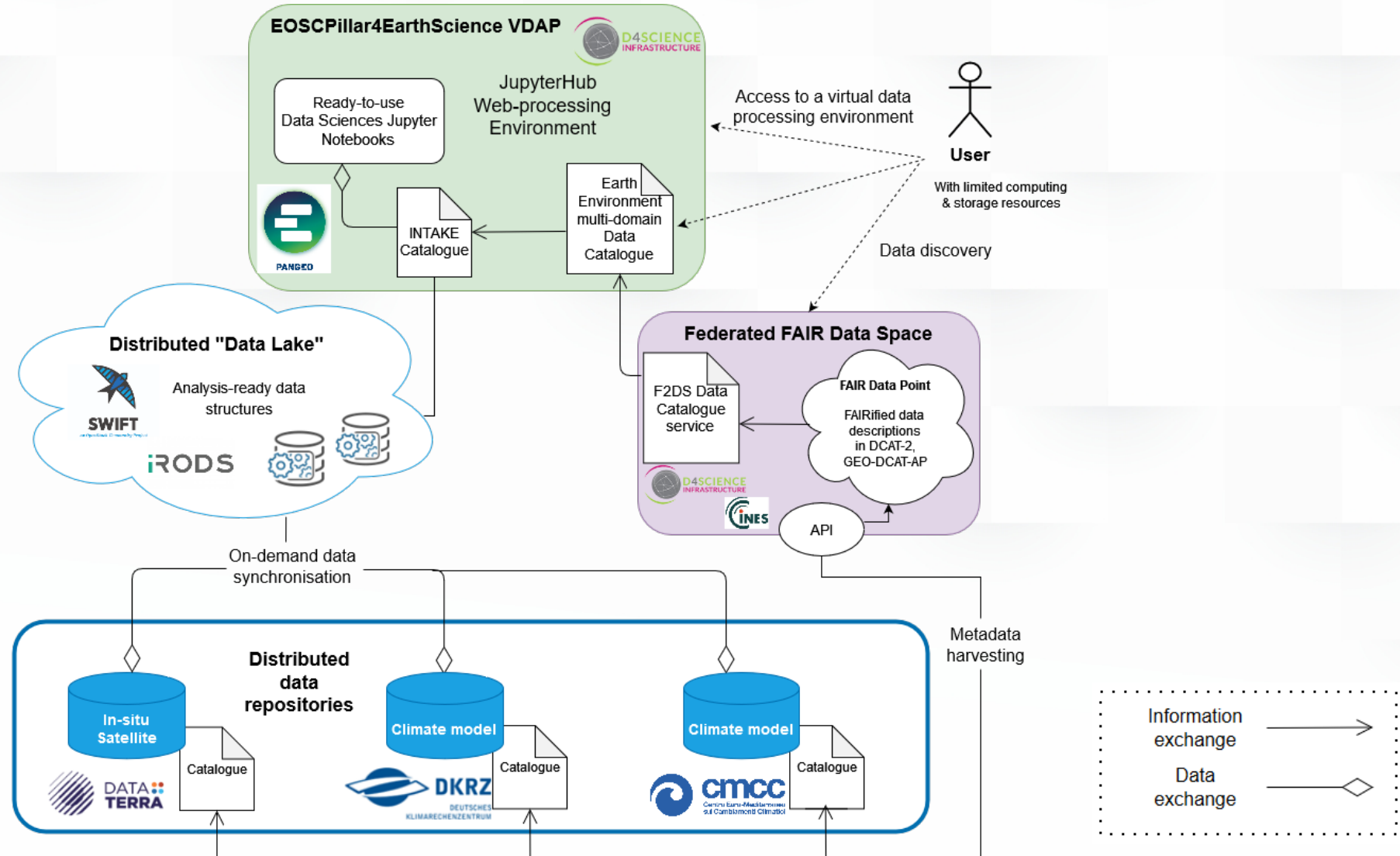


- Traitement Big Data
- Télécharger les résultats pas les données
- Accès avancés aux données

- 1 **Trouver & télécharger** données disponibles & d'intérêt
- 2 **Préparer** rendre les données (inter)opérables
- 3 **Traiter/analyser** sur moyens locaux **distants**



UC#2 - implémentation technique



UC#2 – Bilan

Processing & Infra:

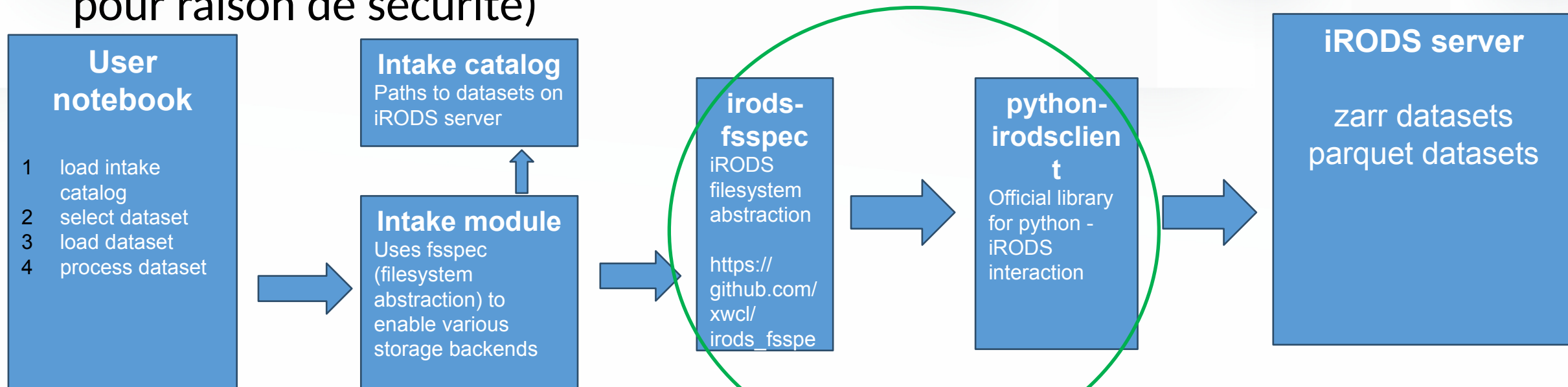
- Ecosystème Python et PANGEO plébiscités par l'ensemble des partenaires
- Formats prêts à l'analyse type Zarr et Parquet
- Intake pour masquer la complexité de l'accès aux données
- Service VRE D4Science améliorable avec infrastructure élargie plus adaptée au calcul parallèle
- Protocoles sécuritaires inhérents aux centres de calcul posent des problèmes d'accès

UC#2 - Bilan

Accès aux données :

iRODS:

- Sub-setting à la volée (mais duplication)
- lecture des données à la demande (mais accès anonyme nécessaire pour raison de sécurité)



UC#2 - Bilan

Découverte des données:

- GEODCAT-AP plus adapté aux spécificités des données géo-référencées
- Extension favorisant l'interdisciplinarité
- Connexion entre le catalogue de découverte et le niveau données à améliorer

UC#2 - Bilan

Conclusion :

- Concept validé par la communauté, remontée des besoins de la communauté à l'EOSC (GEO-DCAT, Infra, etc.)
- Bac à sable permettant d'explorer des solutions et monter en expertise
- Avancées intéressantes qui continueront à être développées via d'autres projets (services FAIR EASE par exemple)
- Elargissement à d'autres pratiques
- Rencontre avec d'autres communautés (Agro, biodiv, SHS, climat), favorisant des partenariats et montage de consortium

EOSC-Pillar

Coordination and Harmonisation of National & Thematic Initiatives to support EOSC

Thank you!

Demonstrateur : EOSCPillar4EarthScience VRE <https://eosc-pillar.d4science.org/group/eoscpillar4earthscience>

Documentation : EOSCPillar4EarthScience QuickStart guide
<https://doi.org/10.5281/zenodo.7128695>

Poster : <https://doi.org/10.5281/zenodo.7194818>

Get in touch with us!



www.eosc-pillar.eu



@[EoscPillar](https://twitter.com/EoscPillar)



[/company/eosc-pillar](https://www.linkedin.com/company/eosc-pillar)



EOSC-Pillar has received funding from the European Union's Horizon 2020 research and innovation Programme under Grant Agreement No. 857650. This material by the EOSC-Pillar Consortium is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

