

EGU2020-22592

<https://doi.org/10.5194/egusphere-egu2020-22592>

EGU General Assembly 2020

© Author(s) 2020. This work is distributed under the Creative Commons Attribution 4.0 License.



## Federated and intelligent datacubes

**Chris Atherton**<sup>1</sup>, Peter Löwe<sup>2</sup>, and Torsten Heinen<sup>3</sup>

<sup>1</sup>GÉANT Association, Amsterdam, The Netherlands (chris.atherton@geant.org)

<sup>2</sup>DIW Berlin, Berlin, Germany (ploewe@diw.de)

<sup>3</sup>Earth Observation Center, German Remote Sensing Data Center, Oberpfaffenhofen, Weßling, Germany (torsten.heinen@dlr.de)

We face unprecedented environmental challenges as a species, that threaten our existing way of life. We are still learning to understand our planet, although we have a good idea how it works. The speed of research needs to accelerate to provide information to decision makers, to better respond to our societal challenges. To do this we need to move towards leveraging large datasets to speed up research, as proposed by Jim Grey in 'The Fourth Paradigm'. In the world of research infrastructures we need to provide a means for scientists to access vast amounts of research data from multiple data sources in an easy and efficient way. EOSC is addressing this but we are only scratching the surface when it comes to unleashing the full potential of the scientific community. Datacubes have recently emerged as a technology in the Environmental and Earth system domain to store imagery data in a way that makes it easier and quicker for scientists to perform their research. But with the scales of data volumes that are being considered, there are many challenges to curating, hosting, and funding this information in a centralised centre. Our proposal seeks to leverage the existing National Research and Education (NRENs) infrastructures to store national repositories of regional Environmental and Earth system domain data, for this to be shared with scientists in an open, federated but secure way, conforming to FAIR principles. This would provide levels of redundancy, data sovereignty and scalability for hosting global environmental datasets in an exascale world.