

1. Summary

The INF project provides services to all project members to facilitate a smooth and efficient data publication process in accordance with our data policy, the guidelines of good scientific practice and the FAIR (Findable, Accessible, Interoperable, Reusable) principles throughout the whole data life cycle. In addition we offer education and guidance for the members and especially the PhD candidates of the integrated research training group (IRTG) with respect to open data and research data management.

2. Achievements phase II

Data publications

The INF project assists all project members with their data publications. As a metric to track data publication we track the number of registered DOIs of datasets published from the project.

- 1768 DOIs are registered.
- More than 28.000 views and over 4000 downloads in PANGAEA
- 9 data papers (3 in phase I and 6 in phase II)

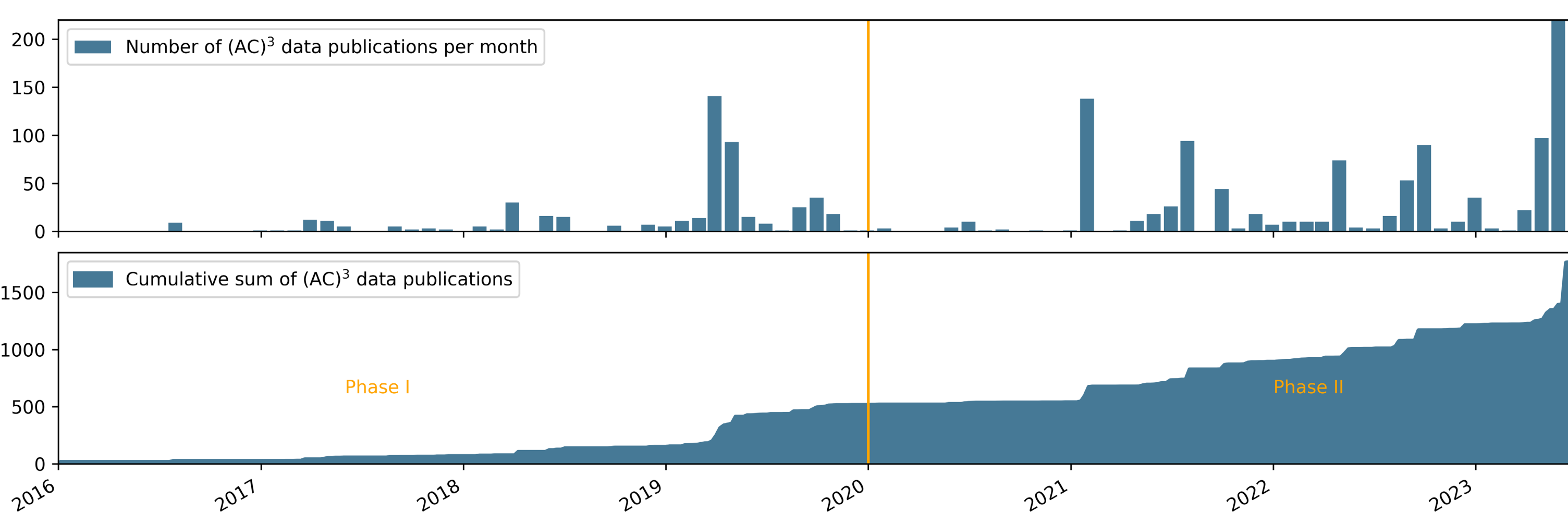


Fig. 1: Data publications from the project so far.



Fig. 2: Links to our data portal in PANGAEA and our published data policy

Infrastructure advances

In phase II we introduced an internal data sharing server, based on a self-hosted Nextcloud instance, which has seen great adoption in the community. It enabled integration of unpublished datasets during measurement campaigns and workshops. A chat server has been set up to foster communication between members



Fig. 3: Links to the Nextcloud and Mattermost instances and our collection of Jupyter Notebooks

Education and outreach

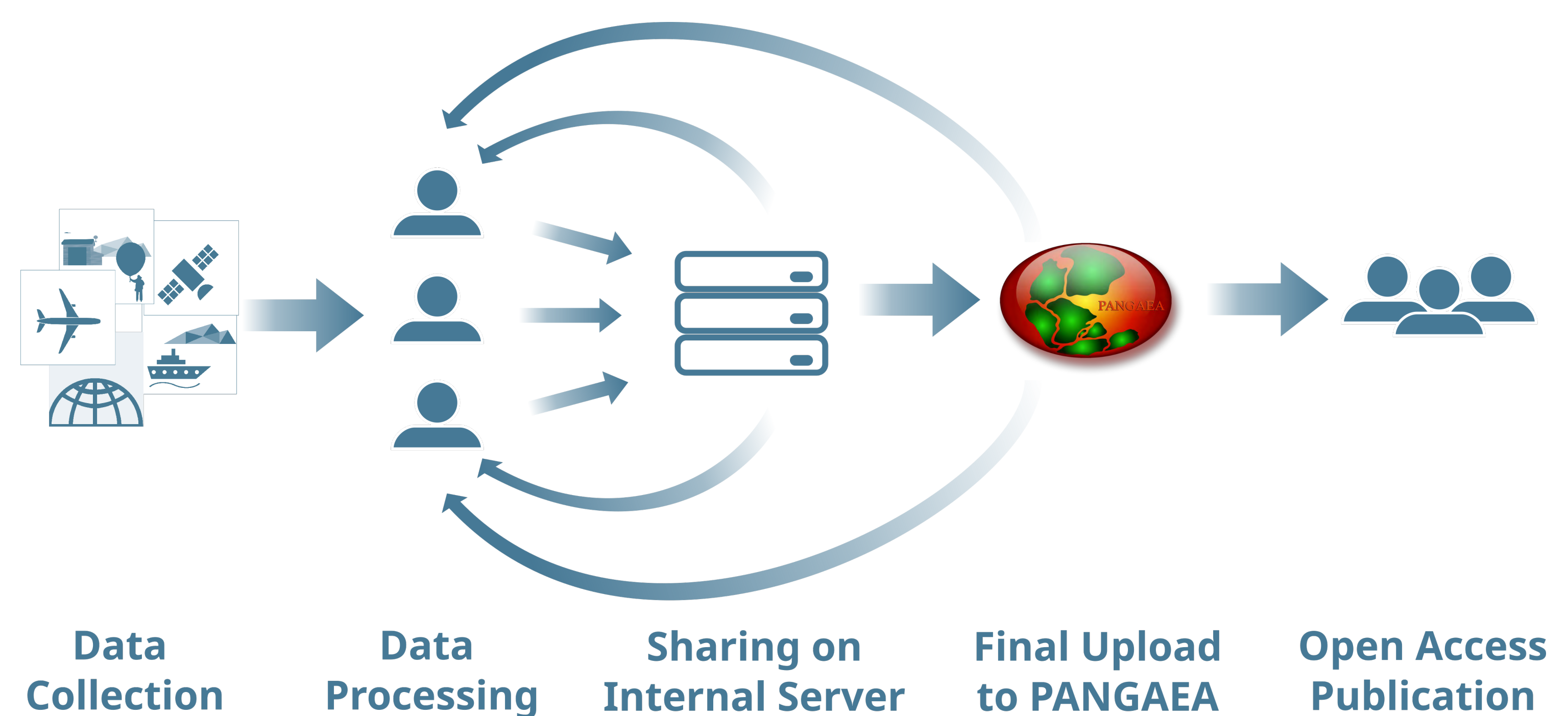
- In close collaboration with the IRTG, we promote Open Science and Open Data among the early career scientists.
- We assist all members by offering individual consultation on RDM from data collection to publication.
- Jupyter Notebooks highlighting several data sets are made available to the public for easier access to our data.

4. Legacy & Major expected results

Project Legacy

- A large collection of high quality datasets of essential climate data published
- Our experience in the successful management of a research data from a large research center will be used in future projects
- Our domain specific best practices will be adopted by the partner RDM projects

Research data workflow



3. Research plan phase III

WP1 Data Publications and Data Stewardship

- Facilitating an efficient research data workflow for all members from data collection to publication.
- PANGAEA as main repository, other options can be chosen in consultation with the INF project
- Increased focus on data from model results
- Monitoring scientific paper drafts to comply with our data policy

WP2 Data Management Training

- Offering training in RDM for all members in collaboration with the partner institutes
- Special focus on courses for IRTG members
- Providing access to partner university programs are open to all members (e.g. DataTrain Bremen, RDWG Leipzig and C³RDM Cologne)
- Workshops and hackathons for IRTG members

WP3 Infrastructure

- Hosting a data exchange server for yet unpublished data
- Expanding user management to allow access to quick look datasets for students and workshop participants
- Hosting a chat server to foster communication between different partners
- Web portal to make the large number of published data sets more easily findable

WP4 Communication of Results

- Communicating best practices to RDM projects at partner institutes
- Interfacing with NFDI4Earth
- Integrating our experience on domain-specific RDM approaches to partners

WP5 End-of-project transition

- Publication and archival of all scientific data is covered by WP1
- Aim to archive the large collection of additional information aggregated in the 12 year project lifetime
- Including e.g. lectures, newsletters, project webpage

Major expected results within phase III

- Increased number of published, high quality data sets
- All project members, especially early career researchers, will adopt concepts of RDM in their future research approaches