INF
Data management

Z03/ INF

FLUXES
TRANSPORT
ARCTIC AMPLIFICATION
$(AC)^3$
CLOUDS
SURFACE
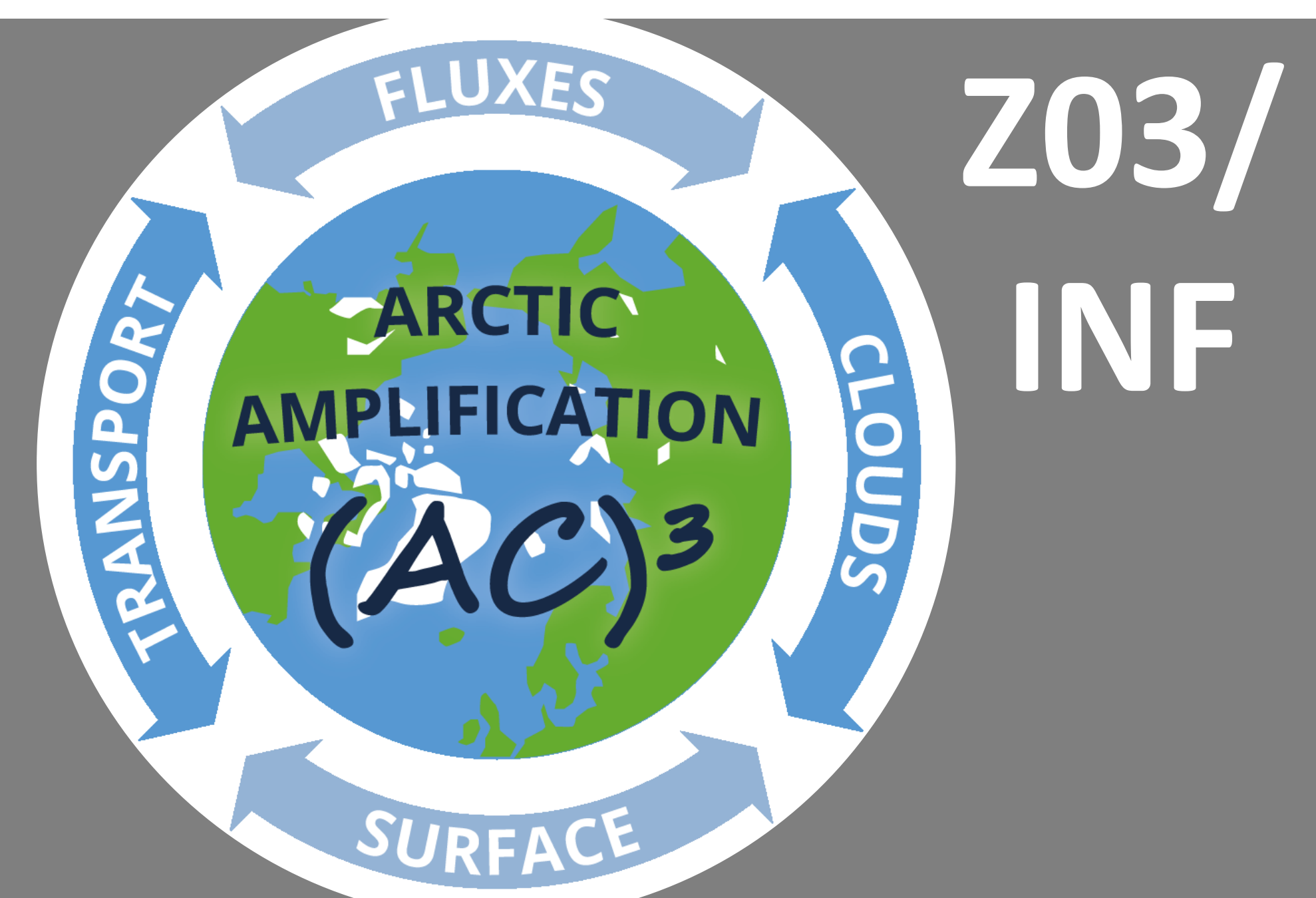
Matthias Buschmann
Justus Notholt

1. Summary

Th INF project will facilitate $(AC)^3$'s open access data publication strategy, following the FAIR (Findable, Accessible, Interoperable, Re-usable) principles and the DFG guidelines.

To achieve this we will provide:

- Training of users
- Additional education in the framework of the IRTG
- User support during dataset preparation
- Upload and integration into PANGAEA
- Long term storage and public availability of the data
- Data integration

2. Achievements phase I

Number of $(AC)^3$ data publications per month

Cumulative sum of $(AC)^3$ data publications

Jan 2016 Jul 2016 Jan 2017 Jul 2017 Jan 2018 Jul 2018 Jan 2019 Jul 2019

ıds to PANGAEA within phase I of $(AC)^3$

3. INF Workflow phase II

MOSAiC

Halo-$(AC)^3$

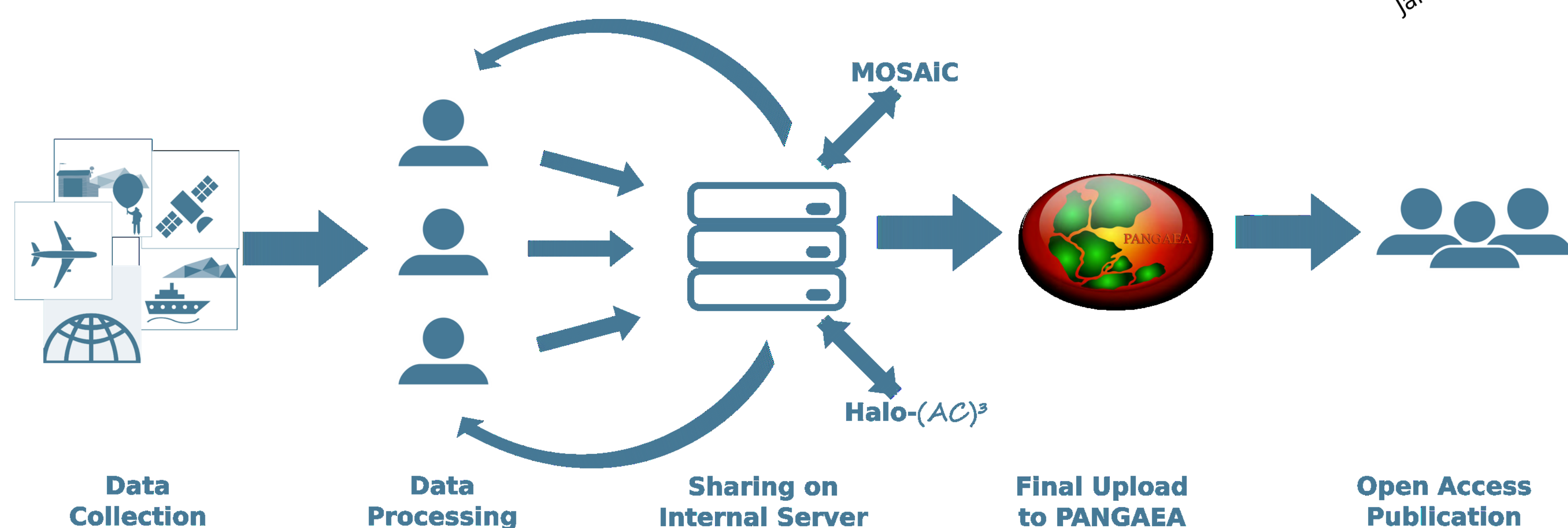| Data Collection | Data Processing | Sharing on Internal Server | Final Upload to PANGAEA | Open Access Publication |

Fig. 2: Data publication workflow in phase II of $(AC)^3$

- Data collection/generation by the project partners during measurement
- campaigns, permanent stations, satellites and model simulations
  - Training of project partners and PhD students
  - Special focus on early career scientists (IRTG) education
- Data processing by the project partners
  - Personalized, continuous support on metadata selection, formatting and
  - data set preparation
- Upload of preliminary datasets to internal server
  - Enabling convenient data sharing within $(AC)^3$ and partner projects
  - (MOSAiC and Halo-$(AC)^3$)
- Open access publication of data sets via PANGAEA (@AWI)
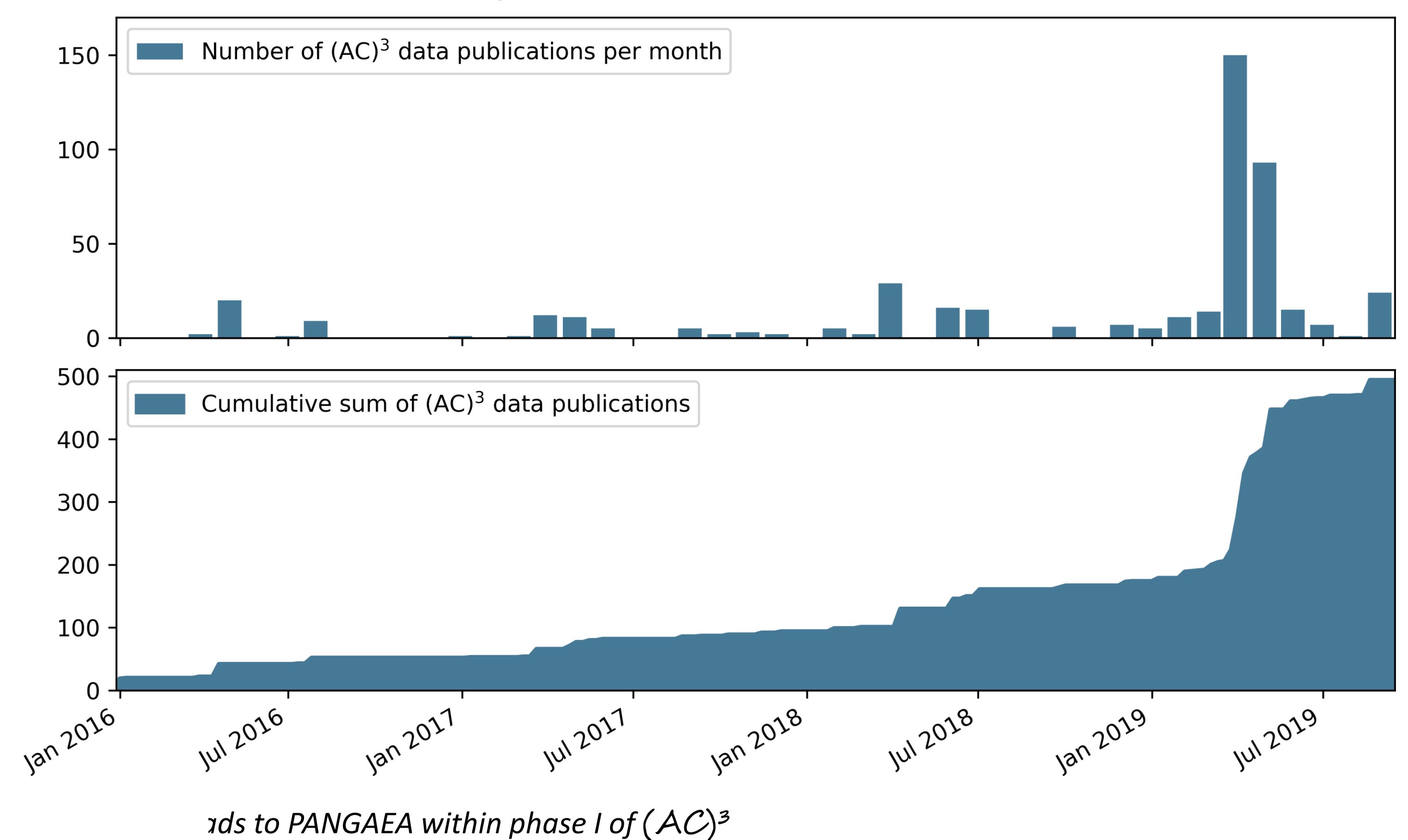
Challenges

- Consolidate different formats and data structures and prepare for
- stand-alone usage
- Model result publication in part incompatible with standard data publication approach
- Ensure centralized early data access for $(AC)^3$ partners
- Advocating the integration of data publications into the established
- publication workflow (e.g. ESSD – AMT - ACP)
- Improve time distribution of final publication date
- Transferring best practices to data management facilities at partner
- institutes

Live access to the data published in phase I

f published data sets under

http://www.ac3-tr.de/data

Scripts for convenient access to our data uploaded to PANGAEA have been created and made publicly available at:

https://github.com/ac3-tr

Research data management infrastructure at the partner institutes

- Different levels of maturity of RDM infrastructure at
- partner institutes
- Hosting integrated at library, computing centre or
- stand-alone
- Collaboration and knowledge transfer appreciated

4. Role within $(AC)^3$ & perspectives

- Collaboration with all projects for data upload and publication
- Training courses in open science and data publishing for the IRTG
- Collaboration with data management facilities of the partner institutes

Perspectives

The learned lessons and developed best practises will be communicated to the partner institutes and the wider research data community (e.g. through the NFDI4Earth)