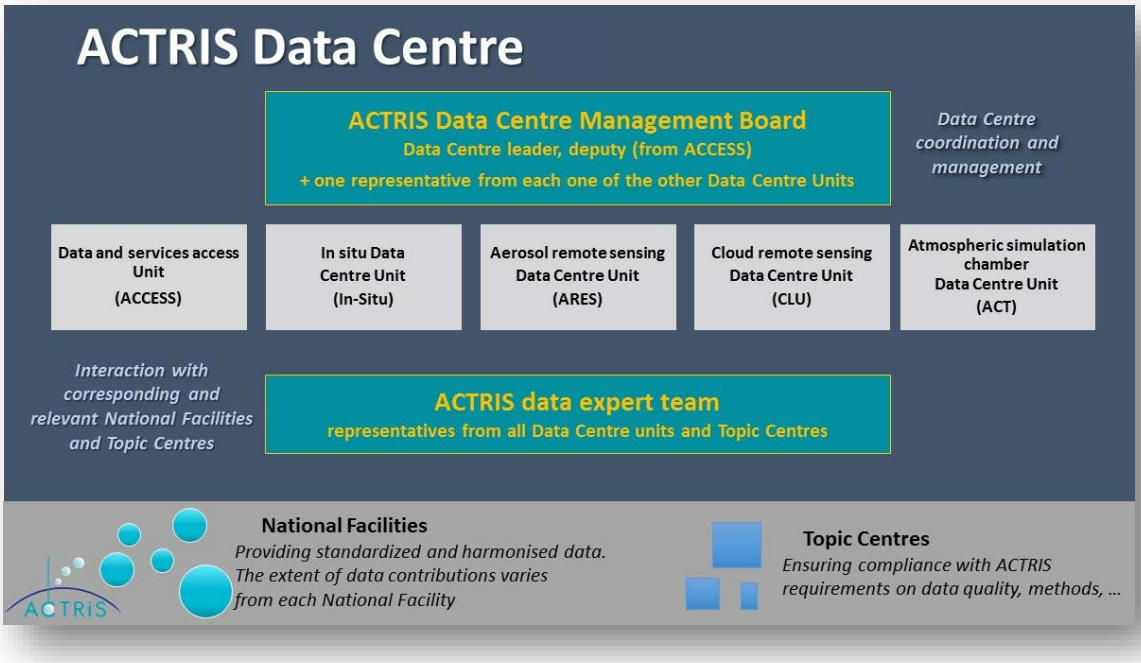
Rationale

The ACTRIS Data Centre offers now an integrative framework in which the EUROCHAMP Data Centre is one of the service units. The inclusion of the EUROCHAMP-2020 infrastructures as the fifth pillar in this system is an important achievement as a strategic political action, and is currently already being effectively implemented. This integrative DC framework is fully described in the Concept document on ACTRIS Central Facilities structure and services prepared in the framework of ACTRIS-PPP by managers of Data Centre units, including EUROCHAMP (D5.1). Figure 1 below describes the overall framework for ACTRIS-DC



Towards interoperability

The task of the ACTRIS DC is to compile, archive and provide access to fully documented and traceable ACTRIS measurement data and data products, including digital tools for visualisation, data analysis and research. ACTRIS measurement data are generated from a wide variety of methodologies (+110 variables from more than 45 instruments or combination of instruments) covering: ground-based in situ and remote sensing; online and offline sampling; observational- (long term stationary facilities) and exploratory platforms (transportable or mobile systems adopting ACTRIS methodologies); and atmospheric simulation chambers.

This range of applied methodologies and specific needs requires a highly advanced, well-organised and structured Data Centre. As the EUROCHAMP-2020 data centre is intended to become a pillar of the future ACTRIS-ERIC data centre, developing formal interaction and improved interoperability between ACTRIS and EUROCHAMP data centres has become a central issue on which the consortium is working since the early stage of the activity. Recent smart-solutions developed within the ENVRI^{plus} reference model are considered as tools to reach this objective. In particular, ACTRIS-DC will seek for common and interoperable solutions for the specific five phases of data management, namely Data Acquisition, Data Curation, Data Publishing, Data Processing and Data Use.

Effective implementation of EUROCHAMP-DC into ACTRIS-DC will demand specific efforts in RP2 and in the larger context of the newly funded ENVRI-FAIR project, considering the different approaches between the original ACTRIS-DC products (mainly time-series of field observations) and the EUROCHAMP products (experiment results of specific

experiments on kinetics and mechanisms). Implementation of interoperable solutions will be strongly connected to the FAIR (Findable, Interoperable, Interoperable and Re-usable) approach in ACTRIS-DC, which is of one the ENVRI-FAIR's objectives.

ENVRI-FAIR develops specific work targeting atmospheric subdomain ESFRIs (thus including ACTRIS-DC and its EUROCHAMP-DC unit) to advance towards FAIRness by implementation of technological solutions on RI level to improve access to data and services. A series of measures increasing FAIRness of the atmospheric subdomain RIs will be developed and tested to demonstrate atmospheric subdomain FAIRness by developing sophisticated QA-QC workflows and web-services with a focus on the I(nteroperability) and R(eusability) in FAIR.

FAIR Maturity Matrix

As a first step in developing FAIRness, a maturity analysis was performed within the DC. The FAIR maturity has been self-evaluated using the matrix developed for the preparation of the ENVRI-FAIR project where the maturity level of different functionalities is assessed.

The matrix is presented below and for each of the 40 items, a single value was attributed corresponding to data section and service section. The Matrix was compiled independently by each of the ACTRIS-DC unit leader: EBAS, CLOUDNET, EARLINET and EUROCHAMP

Table 1a: FAIR criteria for data section

1	Data Centre	Data centre in operation
2	Data Centre	Standardized data format according to international conventions provided
3	Data Centre	Centralised and transparent data quality assessment and control (QA/QC, calibrations, uncertainties) in place
4	Data Centre	System of persistent idenifiers (PID) established, e.g., DOI of data sets
5	Data Access	Restricted access to data, on request only
6	Data Access	Open access to data implemented
7	Data Access	Authorization, authentication, identification, and cybersecurity tools implemented
8	Data Access	Transaprent and open-access documentation of metadata implemented
9	Data Access (FORCE11)	Data and metadata include qualified references to other (meta)data
10	Data Access	Data stored at external trusted long-term repository as succession strategy for preservation
11	Data (FORCE11)	Data are registered or indexed in a searchable resource
12	Data (FORCE11)	Data use a formal, accessible, shared, and broadly applicable language for knowledge representation
13	Data (FORCE11)	Data use vocabularies that follow FAIR principles
14	Data (FORCE11)	Data are released with a clear and accessible data usage license
15	Data (FORCE11)	Data meet domain-relevant community standards
16	Metadata (FORCE11)	Data are described with rich metadata
17	Metadata (FORCE11)	Metadata are registered or indexed in a searchable resource
18	Metadata (FORCE11)	Metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation
19	Metadata (FORCE11)	Metadata use vocabularies that follow FAIR principles
20	Metadata (FORCE11)	Metadata are accessible, even when the data are no longer available
21	Metadata (FORCE11)	Metadata are released with a clear and accessible data usage license
22	Metadata (FORCE11)	Metadata meet domain-relevant community standards

0 = not applicable, 1= =planned, but not started, 2= planned, partially started, 3=fully implemented

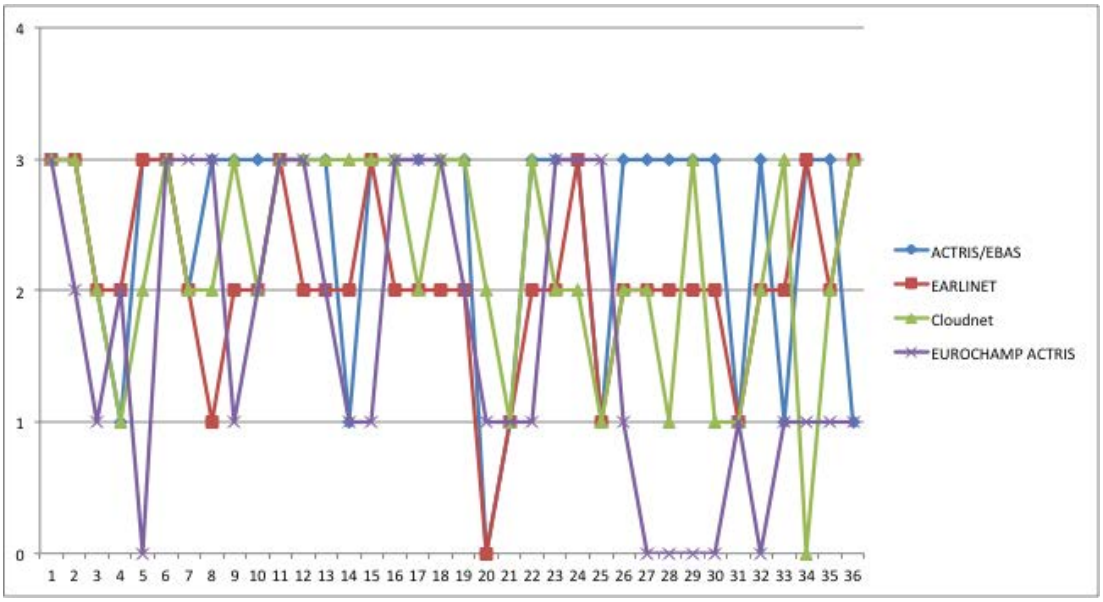
Table 2a: FAIR criteria for service section

23	User Support	Interactive data catalogue is available
24	User Support	User data seacrh is supported by metadata
25	User Support	User data search is supported by semantic instead of purely keyword-based search
26	User support	Traceability of data use is implemented
27	User support	Annotation service for user feeback support is implemented
28	User support	User feedback strategy is implemented
29	Data Provision	Provision of data in near to real real-time to users (e.g. Copernicus, agencies) establisehd
30	Data Provision	On-demand provision of taylored data and services established
31	Data Provision	Data subscription is possible
32	Provenance	Evolution of data is logged
33	Provenance	Metadata model for provenance is establshed
34	Metadata-related Services	Prototypes of added value products of data are available for open access
35	Metadata-related Services	Production of added value products of data is implemented for open access
36	Data Processing Services	The RI operate workflow-executing tools
37	Data Processing Services	External resources, such as e-infrastructures or cloud resources are used
38	Virtual Research Environment	The RI operates a virtual research environment
39	User Community	How many users are registered at present (< 100; 100 - 500; > 500)
40	User Community	How many search queries are placed per month (give a range of values)

0 = not applicable, 1= =planned, but not started, 2= planned, partially started, 3=fully implemented

Maturity assessment

Figure 2 below shows the maturity matrix for the 4 ACTRIS-DC units, as provided by unit leaders. The indicators are:
0 = not applicable, 1= planned, but not started, 2= planned, partially started, 3=fully implemented



A first assessment is that, overall, ACTRIS-DC is amongst the most mature ENVRI-DCs and that none of its specific units lag behind the others considering FAIRness. While EUROCHAMP DC appears to be quite mature as far as data management is concerned, development of service-level must be pursued, which is often the case in relatively new, not-yet operational RIs. Secondly, ACTRIS is among the DCs within ENVRI with the largest variety in methodologies and data, and consists of 5 units, see Figure 1. For ACTRIS, internal harmonisation and implementation of common consistent solutions will be of high priority, also taking the special needs of EUROCHAMP DC into account; the only DC unit dealing with data from exploratory platforms, and non-ambient observations.

It is obvious that being based on self-assessment, the maturity matrix will have to be re-evaluated against proper absolute reference to understand where priority developments are needed in the EUROCHAMP-DC. This will be the starting activity in ENVRI-FAIR. However, a clear sign of mutual integration, EUROCHAMP DC is now included in the design of ACTRIS-DC and will be part of the ACTRIS development towards FAIRness in the ENVRI-FAIR projects (scheduled start 1 January 2019). The use of common terms, semantics, nomenclature and formal translations will be central and early activity in the work, both within ENVRI-FAIR and EUROCHAMP RP2.

Feedback from ENVRI expert (Zhao Zhiming, U. Amsterdam)

Based on the FAIR Matrix (for the ACTRIS-DC and EUROCHAMP-DC sites), and the FORCE 11, metadata and catalogue are important starting points for integration, namely data curation as we defined in the ENVRI RM. I think the following actions can be highlighted:

1) Metadata level harmonization between ACTRIS-DC (different topic databases) and the EuroChamp-DC. The EuroChamp-DC has different assets than ACTRIS: experiments (simulations), configurations and data results. Metadata (including both schemas and control vocabulary) are important to enable all FAIR compliant operations.

2) Semantic search are not well addressed in both catalogues. So far, metadata fields guided search (filtering) is provided by both catalogues. Flexible keyword or semantic search can enhance the findability of the digital assets.

3) Catalogue level integration. I can see almost all data products from ACTRIS can be searched via geo-network like interface, with metadata guidance user can find specific data products. The data products in EuroChamp-DC are mainly created by simulated environments or experiments. They are not geo-location aware. A common metadata catalogue can be an important issue for future community, to enable the findability and accessibility.

4) PID issues. In the self-assessment, neither ACTRIS-DC nor EuroChamp-DC provides real PID solutions. It might be an issue for future interoperability and reusability, in particular when the data products are cited by certain experimental results.

5) Quality control procedures are not well established in EuroChamp-DC. It will be important to keep them compliant so that digital assets in the integrated catalogue can be all at certain consistent level of quality.

6) Provenance is not planned for the EuroChamp-DC. Such information is curtail to assure the reproducibility.

7) Data processing and e-Infrastructure usage, in particular for user on-demand workflows, e.g., for creating customized data products, or for model validation, are under development in Euro-Champ. Those services are important for future data subscriptions, VRE, and user defined workflows of the infrastructure, to enable data (re)use.

Conclusions and future work

ACTRIS, including EUROCHAMP, are engaged in ENVRI-FAIR the strategic project that will ensure a clear and coherent catalogue of services proposed by ENVRI as part of its contribution to European Open Science Cloud. It is expected that developments towards higher maturity data and service levels will be addressed within this project. This will strongly contribute to ensure data from measurements performed for ACTRIS at Exploratory Platforms are made available to users through the ACTRIS Data Centre (DC). Targeted activities will be performed to achieve this goal.
However, substantial work remains to be performed in connexion to the EUROCHAMP community to further develop access to data through the ACTRIS DC architecture, and interoperability with other ACTRIS-DC units:

- Ensuring EUROCHAMP-DC is easily accessed through the ACTRIS-DC portal
- Ensuring data and access policy documents are well-suited to EUROCHAMP and provide feedback to build a sustainable DC in ACTRIS.
- Developing the adequate data quality assessment in EUROCHAMP-DC including instrument status as provided by Topical Centres.