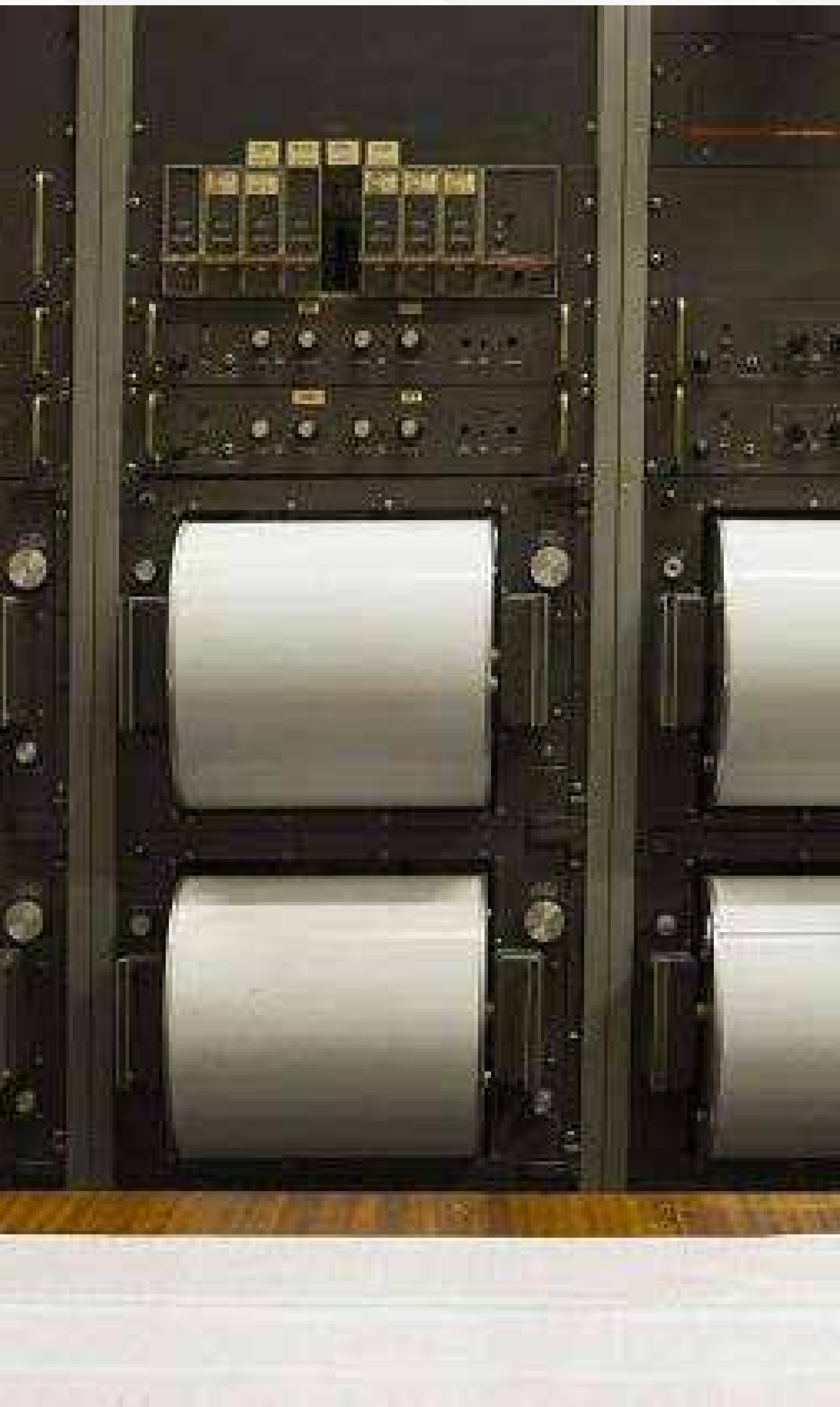


Introduction

At GEOFON data centre is very difficult to offer big pre-assembled datasets to be downloaded, due to the resources needed for their storage. In this context, the idea of using a Data Collections System (DCS) in order to define and save this type of dataset is very appealing, because we can define collections containing only “pointers” (e.g. PIDs, URLs) to the files which are included. This implies almost no extra storage, as only the pointers are saved. Therefore, we implemented this DCS based on an extended version of RDA WG specification on research data collection and make it generic enough and ready to be adopted by different communities within EOSC. Currently, 6000+ collections and 1.5+ million members were defined in the internal service.



Challenges addressed

This use case aims at making the Data Collections System, designed within the activities of RDA, generic enough and ready to be adopted by different communities within EOSC.

The system aims at:

- ★ Building aggregations or data collections of different data objects as an essential element in the Research Data Management practice.
- ★ Describing the research data collections in a standardized way to make them actionable by automated processes in order to be able to cope with the increasing amounts and volumes of data.

Benefits through EOSC-Pillar

This use-case relies on – and demonstrates the benefits of – EOSC-Pillar services as follows:

- ★ Build a generic, multi-disciplinary Data Collections System based on the experience and requirements of 13 communities within the RDA Working Group,
- ★ Improve the current system by revisiting the RDA specifications based on the feedback we received from our partners in EOSC-Pillar.
- ★ Deploy the service in production and foster its usage within communities taking part in the project.
- ★ The system is available via containers for communities who would like to run them on their own.

Highlights

Here is what we have achieved:

- ★ MS28 Code for the improved Data Collection System.
- ★ We revised the requirements from the seismological community for such a system.
- ★ After revising the requirements, we modified the system in order to make it completely generic and ready to be tested by other communities.
- ★ One of the best aspects, regarding the resources needed to put the service in production, is that the members of a collection are identified by PIDs (DOIs, ePIC), what means that almost no space is needed to define them. Within the context of this project, we added the capability to identify resources by URL, making the system independent from a Handle server in case that some resources (or the collection itself) needs to be identified.
- ★ The latest stable version of the system has been recently released and simultaneously is being deployed at GEOFON.

Here is what remains to be done:

- Implement extensions and improvements to the RDA Recommendations. For instance, the automatic export to the Federated FAIR Data Space developed.
- Advertise the service in production.

Partners



Contact details

Task leader: GFZ

EMAIL

Javier Quintero: javier@gfz-potsdam.de

COMMUNITIES

Geosciences (Seismology)