# Dynamic Data

## What is the issue?

Some dynamic data is generated by sensors which produce data streams that may be temporarily incomplete (owing to latencies or temporary interruptions of the transmission lines between the field sensors and the data acquisition centres) and that may consequently fill up over time (automatically or after manual intervention). Dynamic data can also be generated by massive crowd sourcing where, for example, experimental collections of data can be filled up at random moments. The nature of dynamic data makes it difficult to handle for various reasons:

1. establishing valid policies that guide early replication for data preservation and access optimization is not trivial,
2. identifying versions of such data – thus making it possible to check their integrity – and referencing the versions is also a challenging task, and
3. performance issues are extremely important since all these activities must be performed fast enough to keep up with the incoming data stream. There is no doubt that both applications areas (namely data from sensors and crowdsourcing) are growing in their relevance for science, and that appropriate infrastructure support (by initiatives such as EUDAT) is vital to handle these challenges.

## What objectives have been set?

The EUDAT Dynamic Data Working Group (WG) has been tasked with considering common services and policies for which EUDAT might play a role of benefit to the research communities using and maintaining DDS. This broad mandate could be divided into two problem areas:

1. Assuming a server can accommodate requests for them, what sort of persistent identifiers are needed to support the citation goals, and
2. How can a data centre operator (a server) support these requests today and in the future?

## What are the achievements to-date?

**Dynamic Data at the 3rd EUDAT conference**: The session was composed of three presentations with an introdcution from Daan Broeder on the Dynamic Data activities to date: the WG meetings in Barcelona and Rome. Also the EUDAT 2020 proposal foresees a special Dynamic Data WG and in EUDAT2020 workflow and directive aspects of Dynamic Data play a prominent role.

- Is Dynamic Data a Special case? – Daan Broeder,  MPI for Psycholinguistics

*With an overview of the Dynamic Data (DD) terminology and the use-cases considered until now in EUDAT are sensor data from earth sciences and crowd sourcing, corpus collection and surveys from the Humanities. The*

EUDAT receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 654065.

*challenge with DD for EUDAT is especially in the replication scenario where it is difficult to update already replicated DD. The use of deltas can also enable efficient processing if linear transformations are required. Referring to DD using Handle type of PIDs can be achieved using Handle System templates.*

- Handling Dynamic Data from Sensors – Peter Danecek INGV

*Included a presentation of the DD that result from the use of remote sensors for data acquisition in the seismic monitoring network and how thee is a need for real-time processing and early availability of data to the scientists. There is a need to solve accountability, traceability, reproducibility and citability for two related problems: (near-) real time processing and data delivery from a query system.*

- Dynamic Data in the Humanities – Marc Kemps-Snijders, Meertens Institute

*Marc Kemps-Snijders presented the concept of Dynamic Data in the Humanities as it occurs in the Nederlab project that collects digital Dutch written sources from several providers, in total 37M documents and 12000 M words, in a single Virtual Research Environment (VRE).*

The main session outcome focused on two points:

- that non-converging dynamic data should definitely be a use case
- The use of deltas and PID part identifiers are important mechanisms for efficiency.

**Dynamic Data Session at EUDAT 2nd Conference**: During the service building process and roll out of the Safe Replication (B2SAFE) service dynamic data has been a challenging subject. It is difficult to keep consistency between data objects, which are eligible to change and are replicated in a distributed environment. This use case is prominent within the seismology community (EPOS) dealing with sensor-generated data in earthquake sensitive areas across Europe and data streams that are generated by mobile devices at unpredictable times and in unpredictable order (CLARIN). Dynamic data is a broad subject, not only from sensor-generated data, but is seen within communities who have to deal with many unstructured and independent non-scientists (e.g. citizen scientists or crowdsourcing). The dedicated session on Dynamic Data at the 2nd EUDAT conference (Oct 2013 Rome) presented both the outcome of the Dynamic Data working group discussions and the EPOS and CLARIN community dynamic data use cases.

The Working Group had it's first meeting in Barcelona, September 2013 and the following recommendations for EUDAT emerged:

- EUDAT should consider a consultancy service to provide guidance on paths for different user communities to follow depending on the individual use case scenarios – differentiated by data rate, required granularity and level of accountability, and total data volume.
- EUDAT should not try to develop a single “ad hoc” solution rather it should suggest conventions/standards for fragment identifiers and how to represent time stamps.

## Who is involved?

Co-Chairs

- Daan Broeder, MPIPL, Netherlands - CLARIN
- Robert Huber, University Bremen, Germany - EMSO
- Andreas Rauber, University Austria
- Alberto Michelini, INGV Italy - EPOS

EUDAT receives funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 654065.

Members

- Peter Wittenburg, RZG, Germany - CLARIN
- Reinhard Budich, MPIMET, Germany - ENES
- Tobias Blanke, Kings College London, United Kingdom - DARIAH
- Peter Evans, German Research Council for Geosciences, Germany - EPOS
- Mark Kemps-Snijders, Meertens Institute/Royal Netherlands Academy - CLARIN
- Luca Trani, ORPHEUS/KNMI - EPOS

## Useful Links / Documents

- Full details of the Working group discussions are included in the Dynamic Data Barcelona Workshop Report (pdf)
- Dynamic Data session at the 2nd EUDAT Conference report (pdf)
- Read Dymanic Data at the 3rd Conference report (pdf) for complete detail

Read more

EUDAT receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 654065.