TA 5 Data Irreversibility *(Leads: M. Kramer, A. Redelbach)*

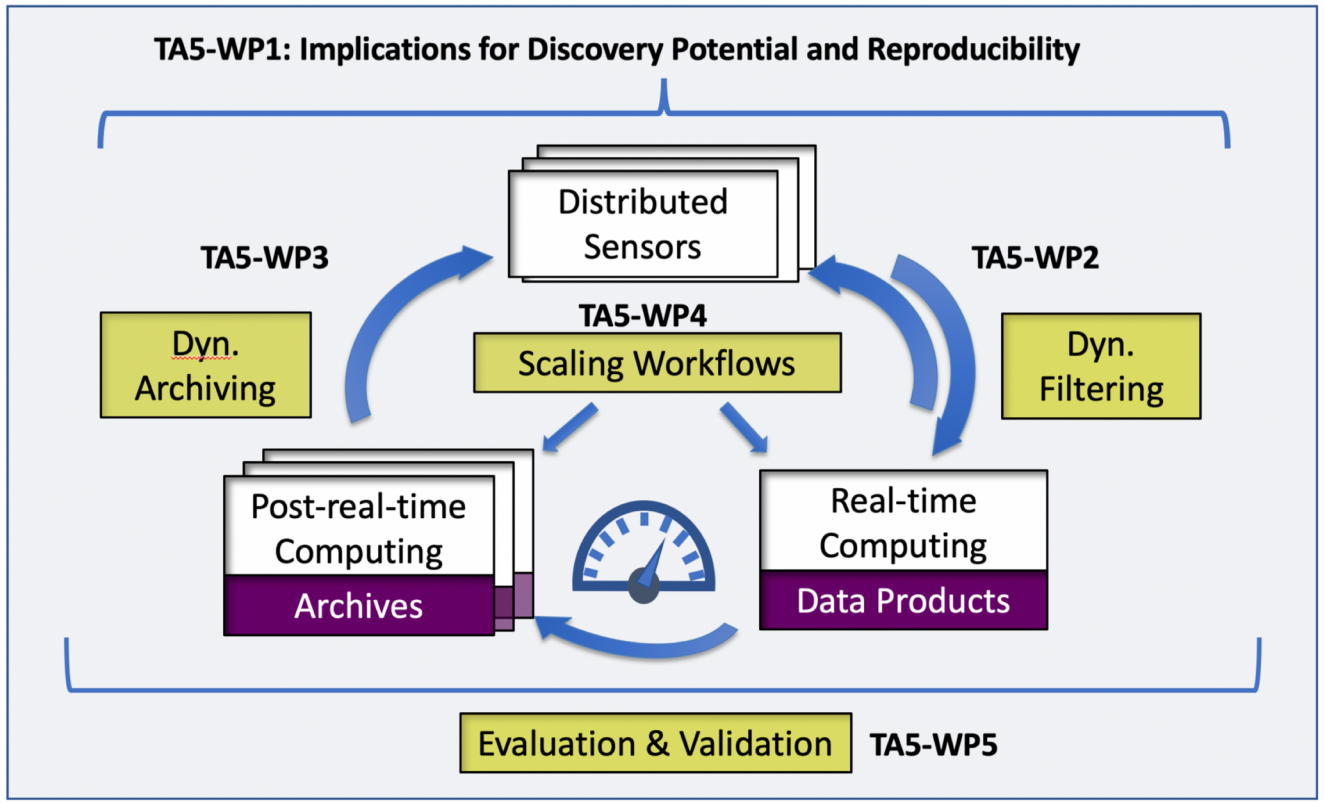
Real-time data reduction required for many PUNCH experiments

Increase of future data rates and also more complex signatures

Irreversible selections and data loss have **implications for discoveries and reproducibility of results**

Concepts and workflows for anomaly-based triggering required

First focus on definition of metadata in order to capture dynamic workflows, enable validations and define interfaces



Dynamic Life Cycle Model

→ **Blueprint for other fields of data-intensive science**

Trigger decisions and metadata

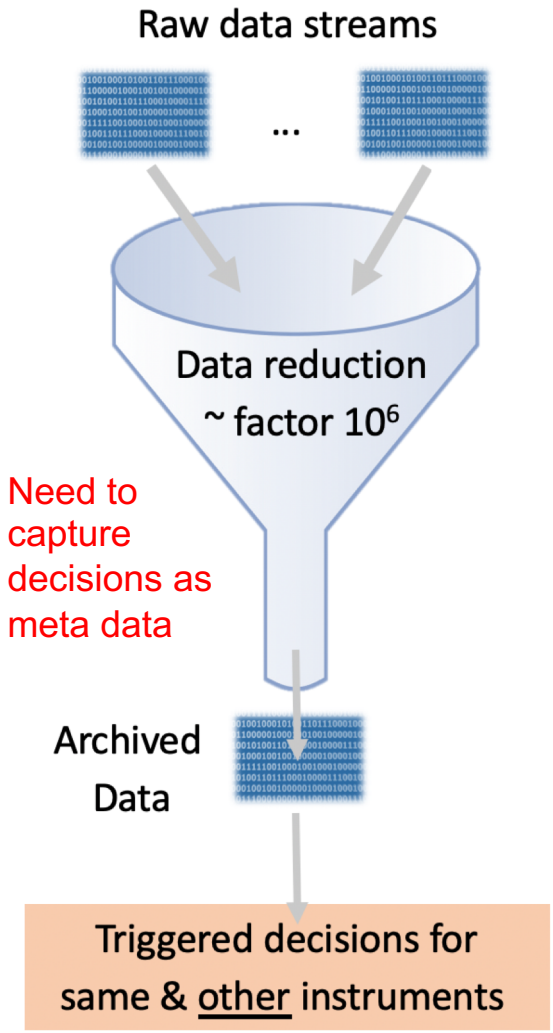
Concepts for **metadata** in all workflows related to **real-time data processing** need **to be established**

Description of underlying decisions has to be included in metadata

High energy physics experiments:
- Trigger menu specifies all rules and conditions → stored in database
- Verification of trigger rules via simulation and e. g. “random” triggers or dedicated support triggers

Dynamic metadata from dynamic archives:
Feedback from offline workflows to sensor control in near real-time

Astrophysics experiments:
- Problem of irreversibility in transient discovery surveys where the original data is not available for reprocessing
- See also related slides (AMPEL, FRB) in last general meeting



→ **Studies needed as requirements analysis for dynamic metadata**

WP2 and WP3: Dynamic filtering & archiving with metadata

Generally not sufficient to have just the discovery signal from transient factories, but one needs **follow-up observations** to determine the nature of the event

Challenge: Follow-up resources are limited and heterogeneous

Some **relevant questions** to be addressed:

- Was an observatory sensitive to the transient's position/time?
- Would an object of type Z even be stored in a public database?
- Can object X be detected as a public alert or in archival data?
- Did object X trigger any follow-up events?
- How sensitive is observatory X to object Y?

Related **technical investigations**:

- How sensitive is pipeline X to object Y?
- Document metadata from Effelsberg useful for triggering

Simulations are essential for validation and understanding
→ also addressed in WP5

→ Starting point:
List of abstract questions for dynamic archive

WP4: Scaling workflows and metadata

Scaling:

- Online (e. g. parallelisation of workflows during data taking)
- Offline (e. g. analysis of data monster)

Drastic **increase of metadata volumes** anticipated

- Constant updates of “quality measures” of archived data
- Flexible data models (e.g. JSON data model)

Properties of relational vs. noSQL data bases

SQL DBs:

- Good for complex queries
- Not designed for fast change
- Typically limited in volume

noSQL DBs:

- Running on distributed data
- Good if queries are simple
- Huge data volumes

Some generic frameworks for metadata:

- UNICORE
- XTENS 2
- MASi (Metadata Management for Applied Sciences)
- Rucio (based on data management of ATLAS@LHC)

→ Studies needed for performance and processing of dynamic metadata
→ Hybrid systems (SQL, NoSQL) promising

TA 5 Basic strategies for metadata and next steps

Extending trigger decisions in high energy physics:
Triggers in hardware/software established since years, extensions for more flexible workflows using e. g. intermediate results, alternative machine learning algorithms or quality control

Extended sets of metadata **towards anomaly-based tiggering**:
Not selected by standards trigger scheme, but anomalous patterns detected

Huge metadata volumes to capture dynamic filtering/archiving workflows and enable validations of underlying decisions and data quality

Cross-experiment data management system Rucio as promising option
→ Dedicated PUNCHLunch with Rucio experts foreseen to discuss critical points

Regular meetings + INDICO for TA & WPs: https://indico.desy.de/category/893/
TA5 meeting time Thursday 9-10 (monthly), next TA5 meeting: April 21, 9-10 am
TA5-WP4 meeting time Monday 15-16 (biweekly)

TA 5 overview of meeting activities and underlying WPs

One TA5 meeting per month established

WP1 (*Leads: D. Schwarz, S. Wagner*):
still to commence

WP2 (*Leads: J. Albrecht, M. Kramer*):
Acquiring of additional real-time data sources where development can be done locally

WP3 (*Leads: J. Nordin, R. Karuppusamy*):
Hand-over meetings between previous WP leader and new one

WP4 (*Leads: H. Heßling, A. Redelbach*):
Biweekly meetings began in March, one student project started

WP5 (*Leads: J. Albrecht, R. Karuppusamy*):
Both previous WP leads needed to be replaced, new leads found

Regular meeting slots + INDICO for TA5 & WPs:
https://indico.desy.de/category/893/
TA5 meeting time Thursday 9-10 (monthly)
TA5-WP4 meeting time Monday 15-16 (biweekly)

Connection to institutes:

- TA5-WP1: FZJ, HTW, MPIfR, UBiel, UHeid
- TA5-WP2: DESY-HH, FIAS, MPIfR, TUDo, TUDd, UHeid, UMainz
- TA5-WP3: DESY-Z, HTW, MPIfR, UBiel
- TA5-WP4: FIAS, FZJ, HTW
- TA5-WP5: MPIfR, TUDo, TUDd, UHeid

TA5 workshop with contributions from all institutes on Feb 21