TA5 Metadata document

Status and planning of Inter-TA-Meeting

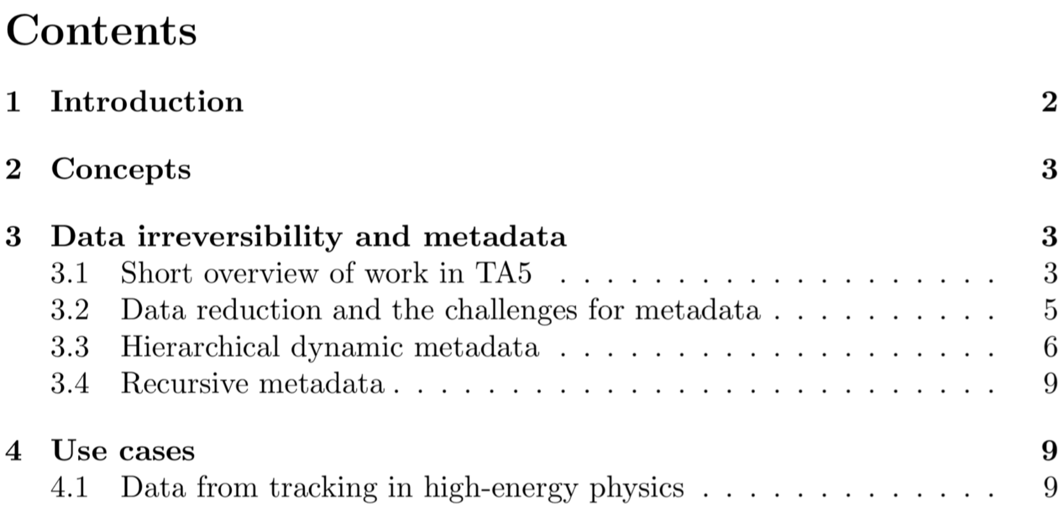
20/04/2023

A. Redelbach, M. Kramer

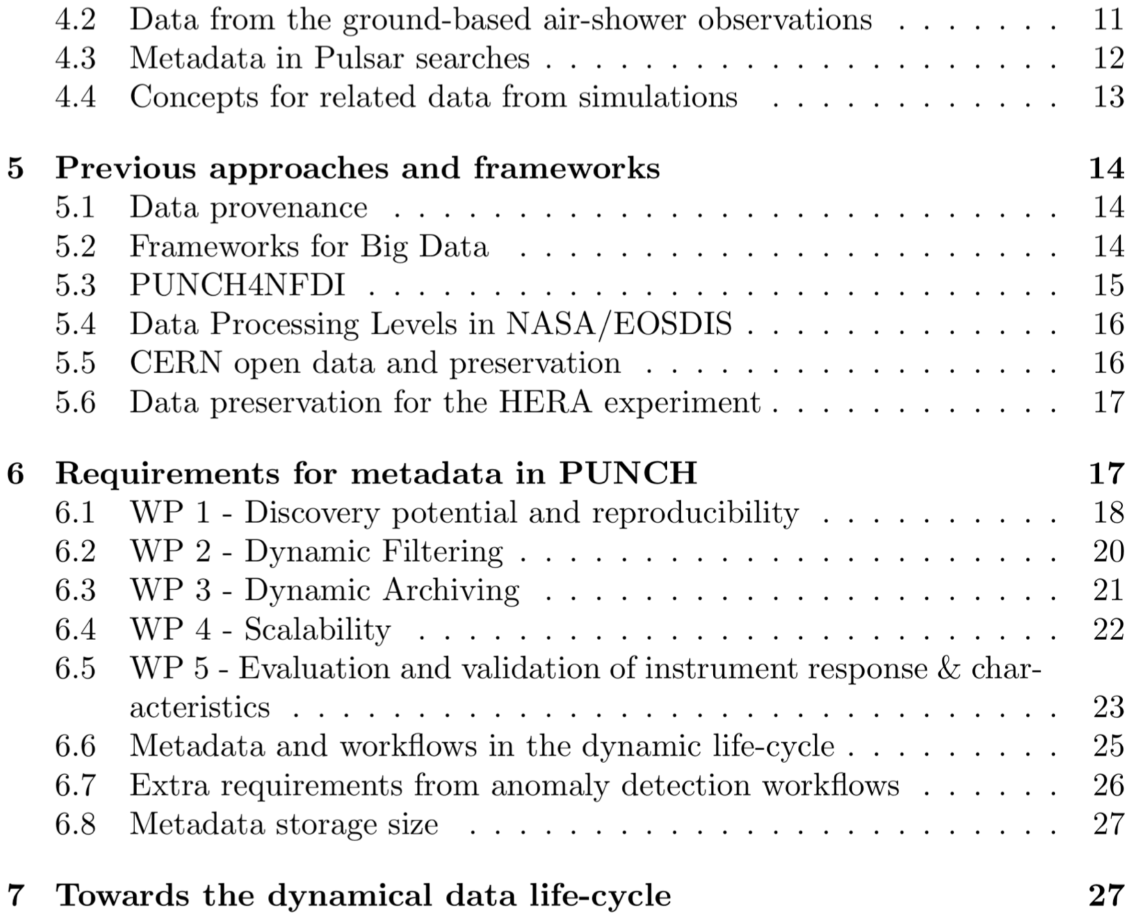
TA5 Metadata document

Status, overview

- Metadata document sent to MB on December 12
- Some modifications recently, see also link to document: https://www.overleaf.com/4394671859tvxrcknqksxr
- Coordination of metadata document with other TAs:
  → Inter-TA meeting scheduled on April 26

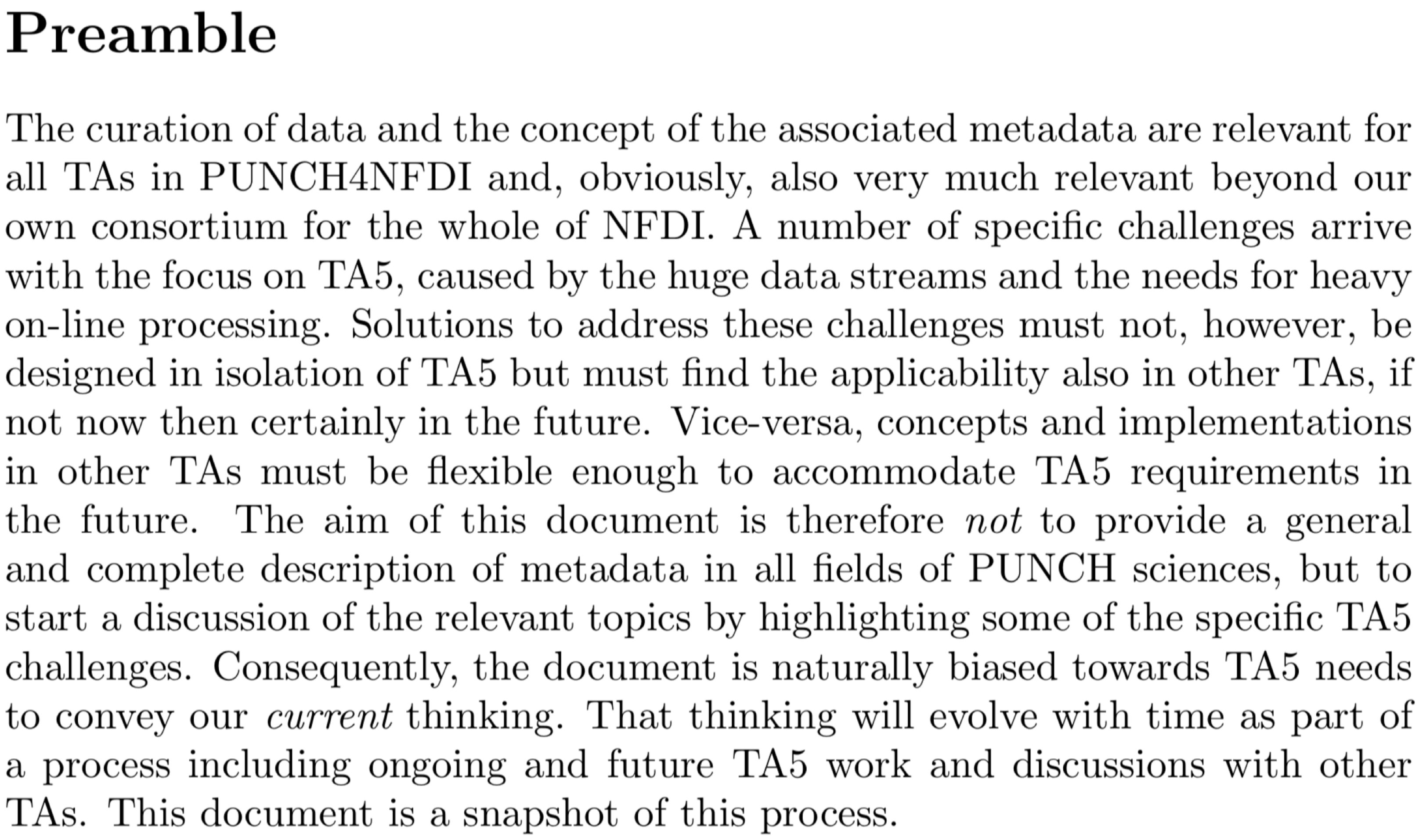


Contents



TA5 Metadata document

Update: Preample



Preamble

The curation of data and the concept of the associated metadata are relevant for all TAs in PUNCH4NFDI and, obviously, also very much relevant beyond our own consortium for the whole of NFDI. A number of specific challenges arrive with the focus on TA5, caused by the huge data streams and the needs for heavy on-line processing. Solutions to address these challenges must not, however, be designed in isolation of TA5 but must find the applicability also in other TAs, if not now then certainly in the future. Vice-versa, concepts and implementations in other TAs must be flexible enough to accommodate TA5 requirements in the future. The aim of this document is therefore *not* to provide a general and complete description of metadata in all fields of PUNCH sciences, but to start a discussion of the relevant topics by highlighting some of the specific TA5 challenges. Consequently, the document is naturally biased towards TA5 needs to convey our *current* thinking. That thinking will evolve with time as part of a process including ongoing and future TA5 work and discussions with other TAs. This document is a snapshot of this process.

Added after discussions in the CollabTools meeting yesterday

→ Sharpening of context, relations to other TAs and scope

TA5 Metadata document

Update: Figures illustrating workflows

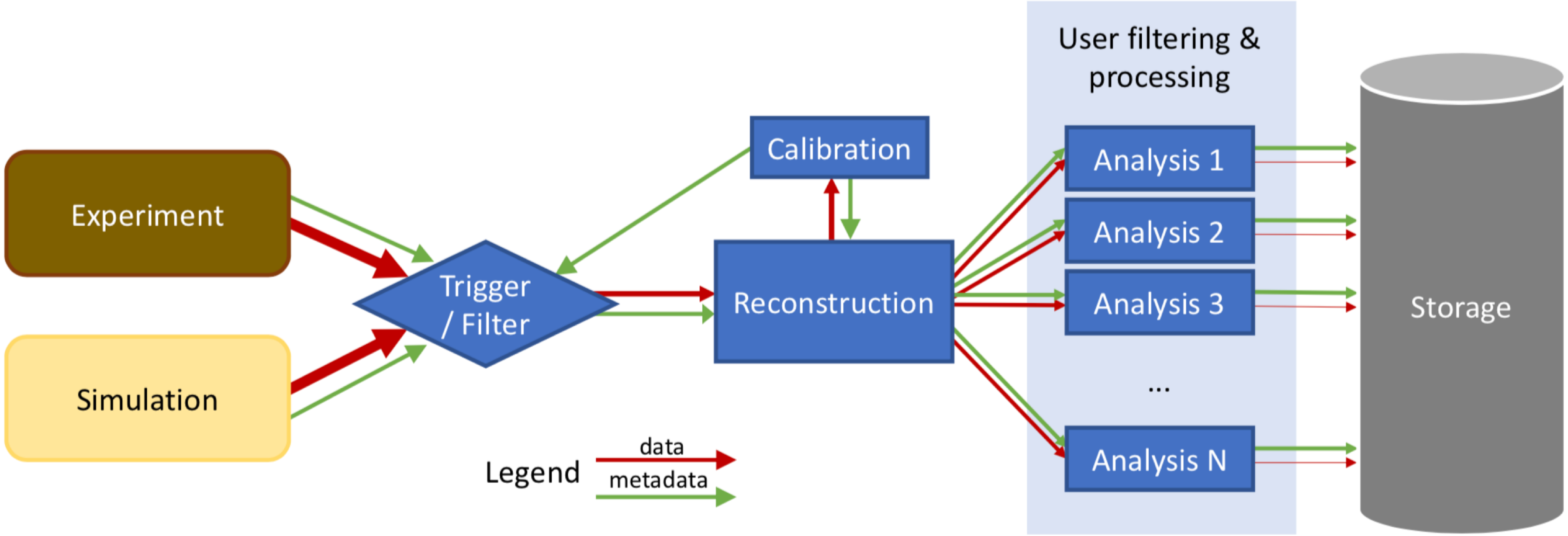
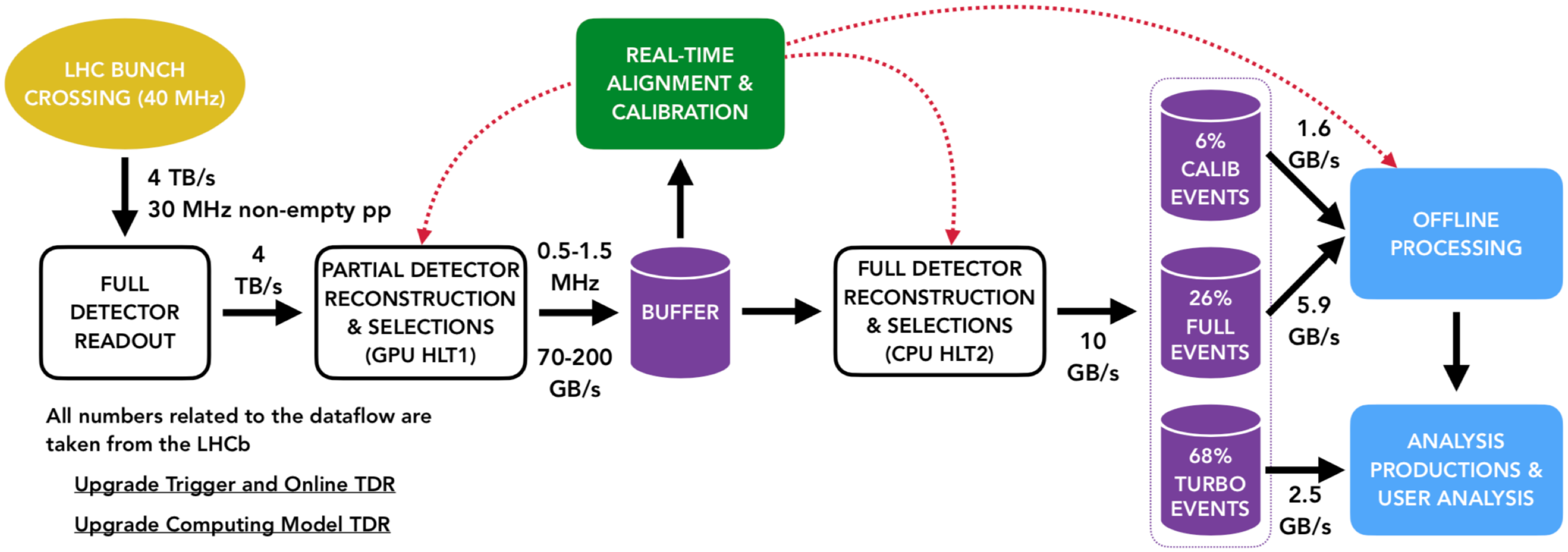


Figure 2: General data processing graph for particle and astroparticle experiments. Variations of the data flow and triggering scheme are possible. The arrow width qualitatively indicates the data rate.

TA5 Metadata document

Update: Figures illustrating workflows


LHC BUNCH CROSSING (40 MHz)
4 TB/s
30 MHz non-empty pp
FULL DETECTOR READOUT
4 TB/s
PARTIAL DETECTOR RECONSTRUCTION & SELECTIONS (GPU HLT1)
0.5-1.5 MHz
70-200 GB/s
BUFFER
REAL-TIME ALIGNMENT & CALIBRATION
FULL DETECTOR RECONSTRUCTION & SELECTIONS (CPU HLT2)
10 GB/s
6% CALIB EVENTS
26% FULL EVENTS
68% TURBO EVENTS
1.6 GB/s
5.9 GB/s
2.5 GB/s
OFFLINE PROCESSING
ANALYSIS PRODUCTIONS & USER ANALYSIS

All numbers related to the dataflow are taken from the LHCb

Upgrade Trigger and Online TDR

Upgrade Computing Model TDR

Figure 3: Current data processing pipeline of the LHCb experiment for proton-proton collisions [6, 7]. Arrows indicate data flow, which are annotated with event and data rates.

Inter-TA-Meeting on April 26

Coordinates:
Wednesday, 26.04.2023, 09:00h
Indico: https://indico.desy.de/event/38872/
Zoom: https://us02web.zoom.us/j/83495109516?pwd=SVEzR0xtejBJbnFpZHhQdWVuS0d4dz09

Preparation:
Uploading of document as pdf to Indico today
Feedback of other TAs requested

Our contributions:
Short overview of document and scope
Summary of included references to PUNCH or NFDI
Discussion of needs for further coordination
Steps for possible publication

→ Participation/contribution of many TA5 colleagues wanted