TA5 Meeting

Michael Kramer (MPIfR Bonn)
Andreas Redelbach (FIAS, Uni Frankfurt)
22/09/2020

News from TA coordination

- Documents needed from Co-Applicants (and participants)

- Summary of resources / infrastructure:
  - Collected in TA5-GoogleDoc
  - To be included in section6/inkind.tex of PUNCH proposal
  - Details: Document whether these are shared, dedicated, etc

- Suggestion: Number of deliverables to be related to FTEs (approx. 1 del. Per 2 FTE request)

- All-hands meeting this Wednesday

- TA5 section for PUNCH proposal:
  - Max. 14 pages including FTE tables
  - To be re-integrated today
  - Deadline for editing this Thursday

- Definition of main objectives under discussion: Central role of Data Science Platform (DSP)

Input for section TA5

- 1-2 pages **general introduction** and motivation to the TA

- In doing so (and later in the WP descriptions) refer EXPLICITLY to
  - the **objectives** in section 2.2, (how do you relate to the objectives?)
  - the services in 4.4 (**contribution to services**),
  - the use cases or **use case classes** (in section 4.1)
  - the **research data management strategy** (esp in section 4.2 - which aspects of it do you address)

- Express our connection to the users in our proposal, and we ask you refer to processes that involve the users in your TA / WP descriptions.

- Summary of contributions of all institutes centrally (outside of TA5 section)

- Extra subsection: Sustainability, risks, and mitigations

- **References** in bibtex format

FTE numbers

See also spreadsheet:
https://docs.google.com/spreadsheets/d/15jxmB3Xw1Apszlgd9BDDP7WC5DzTKMhGw41qwW_LHMU/edit?usp=sharing

Requested

All WPs	Institute	Oct-Dec 2021	2022	2023	2024	2025	Jan-Sep 2026	Sum
	DESY HH	0.125	0.5	1	1	1	0.375	4
	DESY Z	0.0625	0.25	0.5	0.5	0.5	0.1875	2
	FIAS	0	1	1	1	1	0	4
	FZJ	0.2	0.8	0.8	0.8	0.8	0.6	4
	HTW Berlin	0.2	0.8	0.8	0.8	0.8	0.6	4
	MPIfR Bonn	0.215	0.46	1	1	0.9	0.425	4
	TU Dortmund	0.30625	0.7	1	1	1	0.49375	4.5
	TU Dresden	0.125	0.5	0.5	1	0.5	0.375	3
	U Bielefeld	0	0.0625	0.5	0.5	0.5	0.4375	2
	U Heidelberg	0.125	0.5	1	1	1	0.375	4
	U Mainz	0.06	0.24	1.05	1.05	1.05	1.05	4.5
	Total requeste	1.41875	5.8125	9.15	9.65	9.05	4.91875	40

Inkind

All WPs	Institute	Oct-Dec 2021	2022	2023	2024	2025	Jan-Sep 2026	Sum
	DESY HH	0.125	0.5	0.5	1	0.5	0.375	3
	DESY Z	0.0625	0.25	0.5	0.5	0.5	0.1875	2
	FIAS	0.1875	0.75	0.75	0.75	0.75	0.5625	3.75
	FZJ	0.2	0.8	0.8	0.8	0.8	0.6	4
	HTW Berlin	0.25	1	1	1	1	0.75	5
	MPIfR Bonn	0.28	1.12	0.7	0.7	0.6	0.6	4
	TU Dortmund	0.1375	0.65	1	1	1	0.7125	4.5
	TU Dresden	0.125	0.5	0.5	1	0.5	0.375	3
	U Bielefeld	0.025	0.1	0.1	0.1	0.1	0.075	0.5
	U Heidelberg	0.355	0.92	0.7	0.7	0.7	0.625	4
	U Mainz	0.06	0.24	1.05	1.05	1.05	1.05	4.5
	Total Inkind	1.8075	6.83	7.6	8.6	7.5	5.9125	38.25

Re-arranged time-slices

Question of postdocs versus ph.d.s:
Baseline assumption is that it is **ALL postdocs** - unless I am told explicitly differently for dedicated positions in a defined TA, WP, time slice.
Same funding for developer positions

Definition of deliverables

Some aspects to be checked:

- Relation to users
- Interfaces between WPs and other TAs
- FTE assignment
- Former On-line/Off-line scaling workflows merged to one WP5.4
- Connections to use cases (use case class 5) and services

Status of deliverables

WP1

- **D-TA5-WP1-1 —- by month 24:** Report of analysis on impact of on-line filtering on discovery potential
- **D-TA5-WP1-2 —- by month 48:** Report of analysis on impact of on-line filtering on FAIR principles, especially reproducibility
- **D-TA5-WP1-3 —- by month 60:** Demonstrator on an astronomical example

WP2

- **D-TA5-WP2-1 —- by Month 06:** Designing curation criteria and meta-data schemes for dynamical online filtering.
- **D-TA5-WP2-2 —- by Month 06:** Developing a strategy concept for identifying highly complex (multi-parametric) signals in huge data streams.
- **D-TA5-WP2-6 —- by Month 24:** Setting up a test environment for identifying highly complex (multi-parametric) signals in huge data streams using MeerKAT data.
- **D-TA5-WP2-3 —- by Month 36:** Developing a generic tool that converts trained neural networks into efficient High Level Synthesis (HLS) and VHDL firmware implementations for FPGAs optimized for a real-time, low-latency environment.
- **D-TA5-WP2-4 —- by Month 48:** Development of massively parallel algorithms used for real-time sorting, clustering algorithms, and pattern recognition on specialized hardware.
- **D-TA5-WP2-7 —- by Month 48:** Evaluating different algorithms and Machine Learning methods for filtering and selecting relevant transient/anomalous signals out of huge data streams.
- **D-TA5-WP2-5 —- by Month 60:** Development of generic anomaly detection based triggers. A first step is a proof of principle of a generic anomaly detection trigger in pp and ee colliders, which will then be generalised.
- **D-TA5-WP2-8 —- by Month 60:** Development of an anomalous signal detecting pipeline with low false-alarm probability for a multi-messenger follow-up with applications across the electromagnetic spectrum and beyond (e.g. LOFAR, MeerKAT, CTA, neutrino detectors).

WP3

- **D-TA5-WP3-1 —- by Month 06:** General protocol on capturing the decisions related to Dynamic Filtering in dynamic archives and how they are influence archive queries.
- **D-TA5-WP3-2 —- by Month 12:** Methods by which one or more dynamical archives can be jointly queried and interpreted in the presence of information loss
- **D-TA5-WP3-3 —- by Month 18:** Methods for transforming a dynamical archive query into a dynamic filter.
- **D-TA5-WP3-4 —- by Month 26:** Workflows and algorithms to search for anomalous/unexpected signals in (a set of) dynamical archives.
- **D-TA5-WP3-5 —- by Month 36:** Implement a test case setup for a high-throughput sensor archive
- **D-TA5-WP3-6 —- by Month 48:** Establishing a complete dynamical filter and dynamical feed-back loop

Status of deliverables

WP4

- **D-TA5-WP4-1 —- by 2022/Q1:** Porting CASA (Common Astronomy Software Application package) to a memory-based computing prototype in order to prepare an analysis of "data monster".
- **D-TA5-WP4-2 —- by 2024/Q2:** Work on the compatibility of CASA with Gen–Z. This is a core component of WP4 and technically challenging as CASA provides interfaces to the database *MeasurementSet*, whose tables are not stored in a single file but distributed over directories containing several files.
- **D-TA5-WP4-3 —- by 2026/Q2:** Developing feedback interfaces between CASA and (selected) real-time processes of the telescope MeerKAT (in cooperation with WP 5.3).
- **D-TA5-WP4-4 —- by 2025/Q1:** Optimizing the caching strategies for processing a set of benchmark files with the intermediate step of evaluating efficiencies and latencies of different caching strategies for processing a set of benchmark files.
- **D-TA5-WP4-5 —- by 2025/Q2:** Porting CASA to a HPC platform such that it scales reasonably well.
- **D-TA5-WP4-6 —- by 2025/Q2:** Development of concepts for the optimisation of the hard/software co-design for CPUs which include GPU or FPGA features.
- **D-TA5-WP4-7 —- by 2025/Q4:** Developments towards an efficient real-time data processing framework with the intermediate step of an initial definition and implementation of data formats and interfaces for high-performance online processing

WP5

- **D-TA5-WP5-1 —- by Month 36:** Development of machine learning solutions for anomaly detection in detector data, predictive maintenance and process control, and implementation of two concrete prototypes in the context of the real-time trigger systems of the LHCb and ATLAS experiments.
- **D-TA5-WP5-2 —- by Month 60:** Expansion of the concept to a generalized toolkit for predictive maintenance and anomaly detection.
- **D-TA5-WP5-3 —- by Month 36:** Interference recognition and mitigation schemes for transient discovery leading to a "robust" triggering system for multi-messenger follow-up
- **D-TA5-WP5-4 —- by Month 60:** Evaluation of false-alarm rates and improvements via machine learning, dynamic archive queries, on-line feedback and modification of archive metadata