

Project Title	Online Data Intensive Solutions for Science in the Exabytes Era
Project Acronym	ODISSEE
Grant Agreement No.	101188332
Start Date of Project	01/01/2025
Duration of Project	36 Months
Project Website	odissee-project.eu

D1.1 – Converged Science Programs Specifications Capture Document

Work Package	WP1 – Converged Science Programs
Lead Authors (Org)	Vladimir Gligorov (CNRS)
Contributing Author(s) (Org)	Chris Broekema (ASTRON) Maria Girone (CERN) Damien Gratadour (CNRS) Niko Neufeld (CERN) Emma Tolley (EPFL) Bernat Xandri (BSC)
Due Date	30.06.2025
Date	17.07.2025
Version	Final

Dissemination Level

<input checked="" type="checkbox"/>	PU: Public
<input type="checkbox"/>	PP: Restricted to other programme participants (including the Commission)
<input type="checkbox"/>	RE: Restricted to a group specified by the consortium (including the Commission)
<input type="checkbox"/>	CO: Confidential, only for members of the consortium (including the Commission)

Versioning and contribution history

Version	Date	Author	Notes
0.9	16.05.2025	WP1 members	Version circulated to consortium
1.0	17.07.2025	WP1 members	Version taking into account comments from the consortium
Final	21/07/2025	Clémentine FERRE (Neovia Innovation)	Final edition for submission

Disclaimer

This document contains information which is proprietary to the ODISSEE Consortium. Neither this document nor the information contained herein shall be used, duplicated or communicated by any means to a third party, in whole or parts, except with the prior consent of the ODISSEE Consortium.

AI tools may have been used to produce this deliverable. The ODISSEE consortium is ultimately responsible for its content and ensured that every AI generated content has been proof-read and checked if necessary. AI tools have been used within the boundaries of existing laws, in particular regarding the respect of privacy, confidentiality and Intellectual Property Rights.

Table of Contents

Versioning and contribution history	2
Table of Contents	3
List of Figures	3
Executive Summary	4
1. Introduction	5
2. Specifications capture - LHC	6
3. Specifications capture - SKAO	8
4. Similarities between the LHCb and SKAO specifications	11
5. Access and interoperability	11
6. Detailed definition of LHC scientific use-cases	12
7. Detailed definition of SKAO scientific use-cases	13
8. Conclusion	14
Acknowledgments	15
References	15

List of Figures

Figure 1 LHCb data processing facility overview (modified from an original diagram in [1])	7
Figure 2 A high-level overview of the SKAO instrument (from SDP architecture team)	9
Figure 3 Block diagram of the SKAO Science Data Processor (from [4])	10

Executive Summary

This document outlines a converged science program between the LHCb experiment at CERN and the Square Kilometre Array Observatory (SKAO), focusing on the search for Dark Matter. It details how both facilities face similar extreme data processing challenges due to the unprecedented volume of data generated, requiring real-time processing and efficient data management. The paper highlights the potential for knowledge and methodological transfer between the two distinct scientific communities, proposing the sharing of simulated data samples and the development of modular, interoperable software components to maximize their scientific output and efficiently address the persistent mystery of Dark Matter.

1.Introduction

Many astrophysical and cosmological observations are consistent with the existence of non-luminous, non-interacting Dark Matter (DM). The Standard Model (SM) of particle physics describes all known fundamental particles and their interactions and has successfully predicted a wide range of experimental results. However, no known particle can explain the behavior of Dark Matter, which we now believe comprises 84% of all matter in the Universe.

Several extensions to the Standard Model have been proposed which provide particle candidates for DM. In particular, the favored theory over the past few years has been the Weakly Interacting Massive Particle (WIMP) DM, a heavy neutral particle which interacts weakly with the SM. Stable WIMPs are predicted by supersymmetry and several other BSM theories. WIMPs would be created during thermal equilibrium in the early Universe when the temperature is much higher than the DM mass $T \gg m$. Predictions for the relic density of thermally produced WIMP DM almost perfectly match the observed relic density, a coincidence dubbed the “WIMP Miracle”. Although all current evidence for dark matter is gravitational, if WIMP DM was produced in thermal equilibrium with Standard Model particles in the early universe, it must have some nonzero coupling to ordinary matter.

The LHC can reproduce the hot conditions of the early universe in high-energy proton–proton collisions. If the collision energy is high enough compared to the DM mass, it may be possible to directly produce DM at the LHC. Such directly produced dark matter particles can be searched for either as an excess of “missing energy” or by searching for their decays into Standard Model particles. Because dark matter is so weakly coupled to ordinary matter these decays typically manifest as vertices displaced far from the proton-proton interaction region and inconsistent with originating from interactions between known Standard Model particles and the detector material. These “long-lived” searches are also sensitive to a whole range of dark matter and other beyond-Standard-Model particles.

Cosmological WIMPs may pair–annihilate into unstable SM particles, which themselves decay into stable SM particles and produce smooth (continuum) energy spectra, or directly into electrically neutral SM particles, generating spectra with prominent features (monochromatic lines) and high signal-to-background ratios. Indirect detection experiments search for evidence of these energetic particles created by DM annihilation. In particular, DM decays into charged particles (electrons/positrons or protons/anti-protons) may produce synchrotron emission due to propagation in magnetic fields such as the magnetic field of our own Galaxy. Synchrotron emission is the primary energy-loss process for energetic electrons, and as such synchrotron emission places the strongest constraints from secondary emission at the Galactic Center. We can search for evidence of this indirect DM synchrotron emission in other galaxies using the SKAO.

Current collider and astrophysical searches have set stringent limits on the existence of DM, which are complementary with each other as well as with direct searches for interactions between DM particles and ordinary matter targets carried out by dedicated terrestrial experiments. There nevertheless remain vast areas of parameter space which DM can occupy while being compatible with all existing constraints. Given the fact that modern particle and astrophysical experiments are designed and operate on timescales of decades it is critical to ensure that they can be coherently exploited to elucidate the nature of dark matter. In particular, the current and next generation of these experiments (will) produce datasets which are orders of magnitude too large to store

and distribute to physics analysts. The bulk of this data must therefore be processed and discarded in real-time, retaining only a permille (or less) of the most interesting data for the final analysis. It is equally necessary to derive analysis-quality detector calibrations and calibration samples in real or quasi-real time, both to apply them in the real-time processing itself and to maximise their precision.

As we will see in the remainder of this document, the physical data processing constraints faced by one of the LHC's experiments – LHCb [1] – and the SKAO [2] are remarkably similar. This motivates us to explore how their real-time data processings can be converged and made more interoperable. This will in turn facilitate an experimental convergence of their DM science programmes and ensure that the collaborations can learn from each other and react in an agile way to any hints of DM emanating from one or the other experiment in order to maximise sensitivity and ensure that any putative signals are not missed because of a misaligned real-time processing.

2.Specifications capture - LHC

The LHC scientific use-case is represented by the LHCb experiment, a general purpose detector in the forward direction optimised for the study of heavy flavour processes containing beauty and charm quarks. The upgraded LHCb detector [1], which has been taking data since 2022, uses a set of custom FPGA cards to read the detector data out and a cluster of commercial GPU and CPU processors to perform the full physics-analysis-quality data processing in real- and near-real-time. In particular the ability to find charged particles throughout the detector volume at the full LHC collision rate, with a good efficiency across the full range of particle momenta of interest, gives LHCb unique capabilities when it comes to searching for very light and long-lived DM candidates. The LHCb dataflow is illustrated in Figure 1.

All LHCb processing software is publicly freely available under the MIT and GPL licences. The codebase is divided into the following principal components specific to the LHCb collaboration. Its external dependencies are too numerous to list in this document but all relevant dependencies are themselves publicly available.

- Allen: LHCb's framework for heterogeneous real-time data processing. Allen follows a "write CUDA, compile for any architecture" philosophy in which a thin layer of header files assure the conversion between CUDA statements and equivalent codes which can be compiled for e.g. x86, PowerPC, ARM, and other architectures. Allen is primarily used by LHCb to implement its first real-time processing level which operates on every non-empty crossing of the LHC proton beams.
- Moore: LHCb's framework for the full physics-analysis-quality data processing on CPU architectures. The Moore package itself implements the pythonic configuration of LHCb's CPU data processing, while the C++ code which performs the compute-intensive tasks lives in a set of separate, dependent, packages. They can however be considered as a conceptual entity for the purposes of ODISSEE.

LHCb's processing pipelines are built around the concept of "algorithms" which define input and output data types and are composed into processing sequences by following their declared data dependencies and additional control flow specifications. Allen and Moore algorithms are not generally interoperable, and while blocks of Allen algorithms can be chained together and used in a Moore sequence the reverse is not true. Software

algorithms are not interoperable with processing deployed on the FPGA cards which receive the detector data.

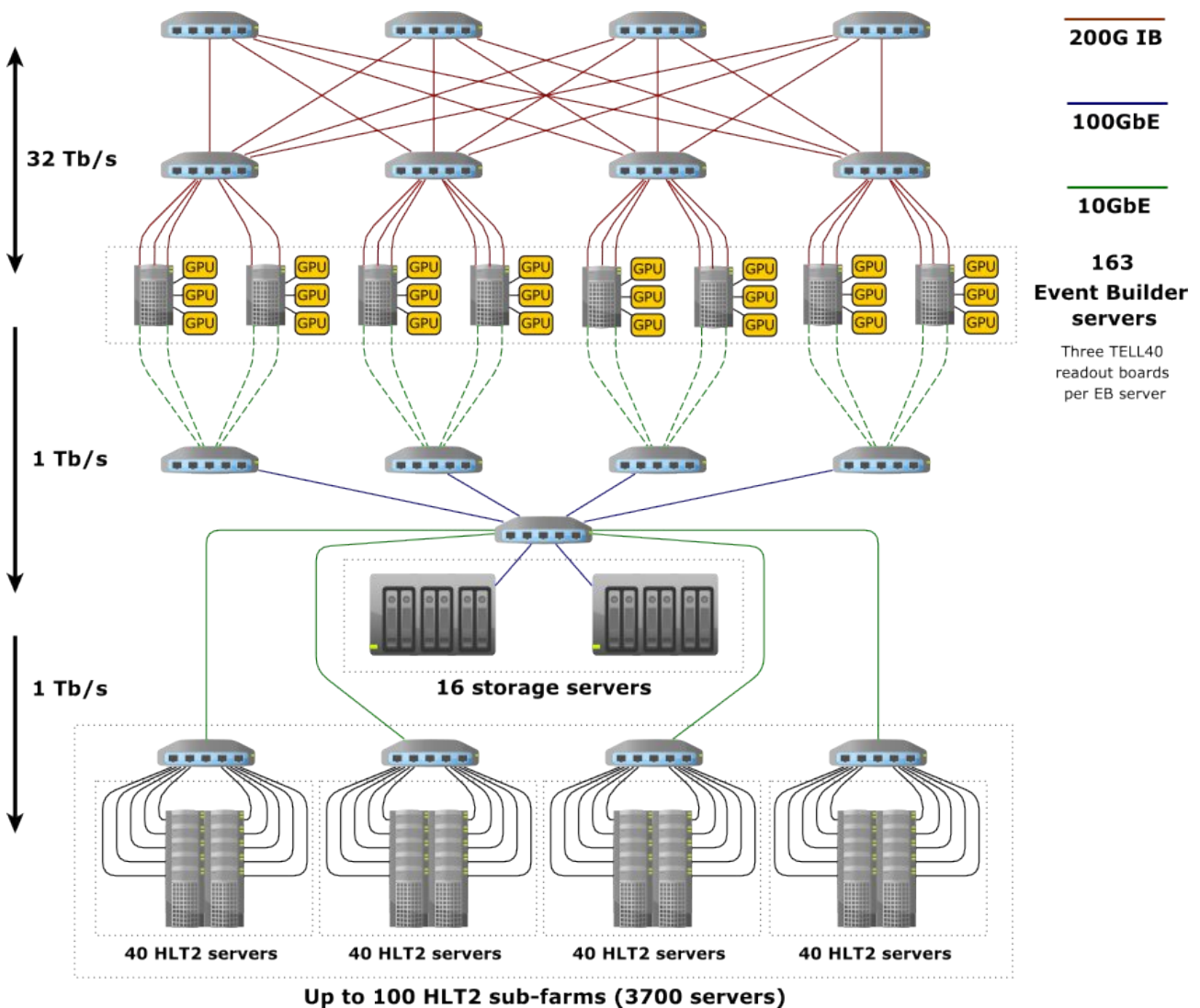


Figure 1 LHCb data processing facility overview (modified from an original diagram in [1])

LHCb's software is not presently usable by other collaborations, mostly because the framework components are mixed with experiment-specific algorithms and sequences in the same repositories. For example, a large part of Allen consists of a custom memory manager, algorithm scheduler, and portability layer which could in principle be used by any experiment which needs a heterogeneous data processing optimised for extreme throughputs – 30 million “events” or over two Terabytes of data per second. However at present these components are fully mixed with highly LHCb-specific reconstruction and selection algorithms optimised for GPU processing, such that the reusable components are not, in practice, composable with non-LHCb codes. Similarly Moore implements a fully pythonic algorithm scheduler which is able to resolve complex data- and control-flow specifications and take into account the predicted computational cost of algorithms to compose them into a computationally optimal sequence giving a specific physics outcome. But this scheduling is currently embedded within the LHCb-specific processing and is therefore similarly unusable outside LHCb. One of the ambitions of ODISSEE is to ease or fully remove these constraints on the reusability of these codes,

while simultaneously making it easier to integrate non-LHCb codes in LHCb processing pipelines. This work is being carried out as part of the ODISSEE WP2 activities.

The LHCb collaboration uses simulated data samples to develop and benchmark its data processing pipelines, and a set of such samples has been made available by LHCb as detailed later in this document. The LHCb collaboration uses a set of dedicated servers to benchmark the computational performance of its data processing in realistic conditions. The physics performance can be validated on any computer where the code can be compiled and executed on the simulated data. In particular the collaboration has the ability to measure both computational cost and energy efficiency and to compare *ceteris paribus* the cost of a given processing on different architectures. New architectures can be introduced and benchmarked together with existing ones so long as they are able to receive data and output their results in publicly documented formats.

The LHCb processing pipeline is independent of whether data or simulation are being processed, aside from detector conditions and calibrations which only apply to data or to simulation. The processing is also independent of whether it occurs in (quasi-)real time or asynchronously in data centres. At present the cost-benefit balance means that LHCb performs the vast majority of its data processing in (quasi-)real time in a data centre located near the detector, but there is no conceptual barrier to some or all of this processing occurring elsewhere if that balance changes. It should however be emphasized that trends within High Energy Physics are to perform more and more of the data processing as close to the detector as possible, and in practice the importance of data centre processing lies precisely in the ability to execute the same data processing pipelines on simulation.

One specific case of interest to ODISSEE is to what extent processing cost and energy efficiency can be improved by moving certain parts of the LHCb processing pipeline to the FPGAs which receive data from the detector. This will be evaluated by performing a full end-to-end benchmarking, injecting simulated data samples into the FPGA boards using servers which mimic as closely as possible those used for LHCb data taking.

3. Specifications capture - SKAO

Radio telescopes in general, and the SKAO [2] in particular, are highly data- and compute-intensive instruments. The SKAO is a large-scale distributed telescope with two distinct instruments in the South African Karoo desert (the mid-frequency SKA1 MID) and the low-frequency SKA1 LOW in the western Australian desert. While these instruments are very different in implementation, the data flow is similar enough that we can here describe them as one.

Data is generated at the site by receivers sampling the electromagnetic spectrum, either by using a steerable dish with a receiver in the focal point in the mid-frequency case, or by combining large numbers of omni-directional dipole antennas in the low-frequency instrument.

The Science Data Processor (there are two SDPs, one for each instrument, located in Perth and Cape Town respectively), the focus of this study, continuously receives raw visibilities at around 10 TBits/s. This data stream is received and some initial conditioning of the data is done in real-time. Data is temporarily stored in a high-performance buffer for calibration and imaging, which is currently an iterative process. This buffer has a capacity of about 50 PB and requires a write performance of about 1 TB/s and a read performance of about 10 TB/s to keep up with the processing

requirements while still able to receive data from the correlator beamformer (CBF). Compute capacity is expected to be around 130 PFlop peak at full operation. The resulting data stream to the SKA Regional Centres (SRCs) will measure about 100 GBits/s (aggregated).

In addition to this online processing capacity, each SDP is expected to host one copy of the Science Archive, which is expected to grow with about 300 PB/year. Duplicate copies will be distributed through the world-wide SRCNet.

Scientific processing of radio astronomy data, in particular for large scale distributed telescopes like LOFAR and SKAO, are computationally expensive.

Recent publications show that 8 hour observations using the European LOFAR array require ca. 140,000 core hours to produce a scientific image cube for analysis. [3] The use of advanced composable modular processing components accelerated using e.g. GPUs are expected to significantly improve the computational efficiency of such tasks. First targets for acceleration are potential computational hotspots, based on early profiling of LOFAR and prototype SKAO codes, in particular:

- Predict
- Calibration
- Gridding

These components are also attractive targets for evaluation on European digital IP. Composable modular components are likely more easily ported to accelerators, provided overheads can be minimised.

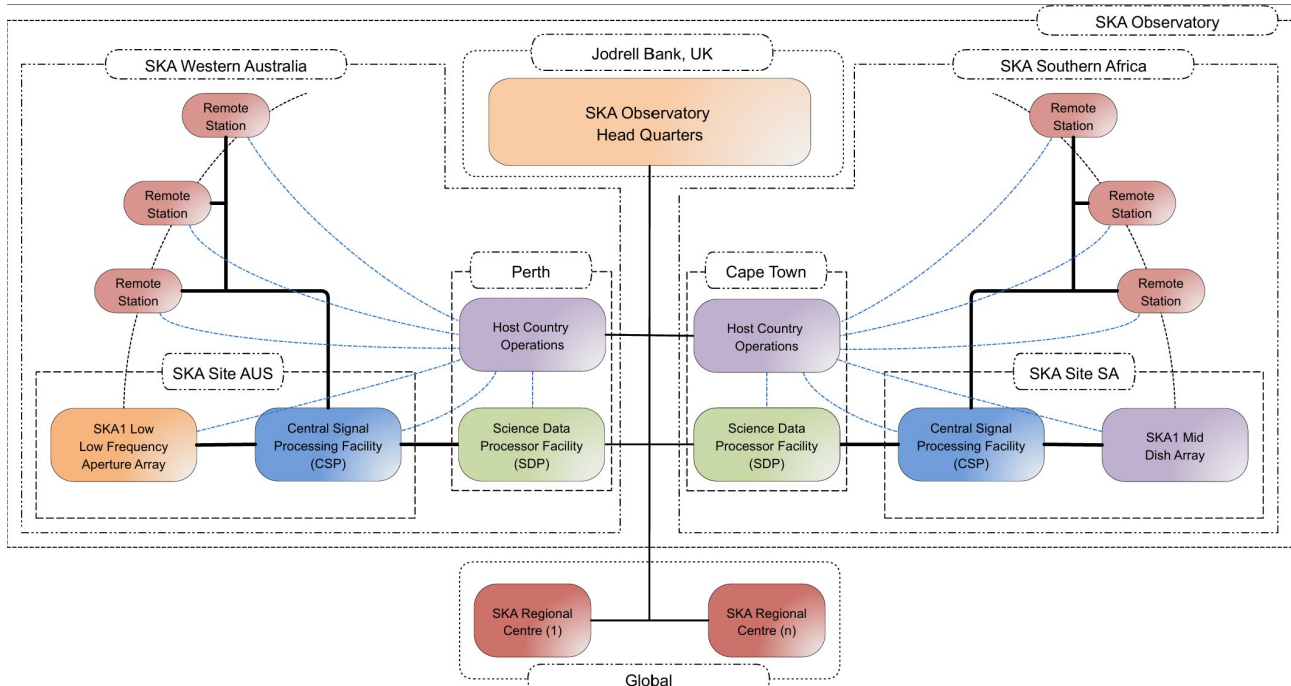


Figure 2 A high-level overview of the SKAO instrument (from SDP architecture team)

Generally, radio astronomy data is not compressed, with the exception of the LOFAR array, which uses DYSCO¹ compression on more recent data products. Compressing and decompressing such data is not very expensive, but the use of modular composable

¹ <https://arxiv.org/abs/1609.02019>

GPU accelerated components allows the introduction of transparent in-flight compression and decompression on the GPU. This could, in combination with technologies like GPUDirect, theoretically allow the streaming processing of telescope data from Central Signal Processor (CSP) to GPU to storage, without CPU intervention. This in turn would be much more efficient, at a slight security cost (due to the Operating system being bypassed by RDMA technologies). In a highly controlled environment like SKAO or LOFAR this is not considered disqualifying.

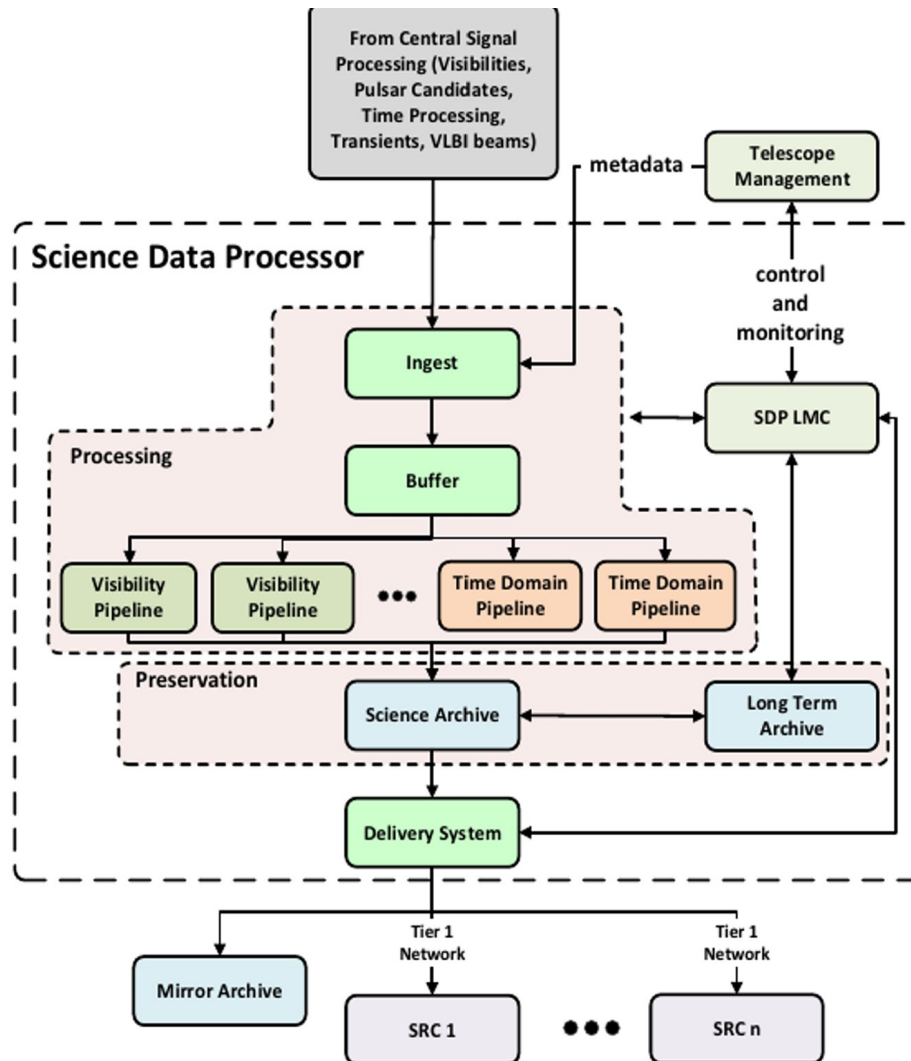


Figure 3 Block diagram of the SKAO Science Data Processor (from [4])

In terms of interfaces within the data center (or in this specific case the Science Data Processor), these differ depending on the stage of the processing. The high-speed, pseudo-real time data streams from the correlator/beamformer in the central signal processor (CSP) are UDP/IP based streams over high-speed Ethernet with a self-describing light-weight protocol (SPEAD) on top to ensure data quality and provenance. As data arrives, it is checked for missing and out-of-order data. Some initial light-weight processing may be done before it is stored in a high-performance buffer. This will use any of the multitude of available storage protocols and interfaces, no selection has been made at this time.

Further batch processing is done using common of the shelf processing frameworks, using their own communications interfaces. The current de-facto software of choice for this is Dask, but this may change during the course of SKAO construction. This may use

either Ethernet, or a dedicated low-latency network if provided, but this is hidden from the software and handled by the processing framework.

Final data products are both exported to the global science data archive using of the shelf cloud tools on high-bandwidth WAN connections. A local copy is retained in the self-hosted archive. The latter is likely shared with the second SKAO site, although this may be limited due to capacity and cost concerns, since the global science archive hosts a full copy as well.

4. Similarities between the LHCb and SKAO specifications

As we can see from the previous two sections, there are a number of key similarities between the LHCb and SKAO data processing specifications. First of all, the overall data volumes per unit time are of the same order of magnitude. Secondly, both instruments perform a partial preprocessing of the data near the detectors themselves, whether this is zero-suppression of the detector data in the LHCb case or frequency selection for SKAO. Third, both instruments rely on a deep buffering of the data while calibrations are performed in order to enable the full physics analysis quality reconstruction of the data to finally be performed in quasi-real-time. While the scientific content of the calibrations, reconstructions, and selections performed on the data is of course very different in the two cases, these similarities in the boundary conditions which govern the framework design of the data processing mean that there is great potential for two-way knowledge and methodological transfer between the two communities.

5. Access and interoperability

From a high level, to facilitate optimal re-use of code and data between high energy physics and radio astronomy, this first needs to be achieved within these sciences. Many of the codes currently in production use have evolved over time to be monolithic and complex. Modularisation and composability is essential, facilitating easy exchange of algorithmic components between instruments and science cases and easing commissioning by offering an easy roll-back option to the users. This is a key prerequisite.

The LHCb experiment has agreed to make a limited quantity of simulated data samples available to all ODISSEE researchers. These samples are significantly larger than those used in nightly continuous integration testing by LHCb and will allow a detailed qualification of the real-time and reconstruction performance. Together with the publicly accessible processing software any ODISSEE researcher will be able to reproduce LHCb's real-time processing workflow at the software level and benchmark the impact of substituting novel algorithms or computing architectures. The LHCb collaboration will facilitate the validation and benchmarking of workflows developed as part of the ODISSEE project by making the relevant validation hardware described earlier available, so long as this does not conflict with the collaboration's data taking or other priorities.

The LHCb collaboration will allow LHCb researchers and non-LHCb ODISSEE colleagues to publish technical papers based on these simulated samples, subject to a case-by-case collaboration review and approval prior to publication. A standard set of acknowledgements has been agreed which the ODISSEE authors will add to any publications using these simulated LHCb data samples or which use the publicly available LHCb processing software.

In contrast to the LHC, which has been in operation since 2008, the SKAO is still under construction. Consequently, while initial data products can be expected to appear during the course of the ODISSEE project, these are unlikely to represent the full capabilities of the instrument. It is therefore prudent to look for representative alternative data sources in pathfinder and precursor instruments. In addition, simulated data can be used as appropriate analogs for actual data.

The LOFAR radio telescope has been in operation since 2011. This telescope is a SKA pathfinder and its design very closely resembles key parts of the SKA1 low instrument. All data from this instrument is archived in the LOFAR Long Term Archive (LTA) and is freely available to the scientific community after a short proprietary period (usually 12 months). This volume and complexity of this data is very representative of expected SKAO data and has been used to test and validate SKAO software components during development. ASTRON, which is part of the ODISSEE project, designed and built the LOFAR telescope and currently hosts the International LOFAR Telescope European Research Infrastructure Consortium (ERIC). This makes access to both LOFAR data and expertise easy and convenient.

Similarly, data from the SKA precursor instruments MeerKAT, ASKAP and MWA are available for use by the ODISSEE consortium. Finally, simulated data specifically generated to emulate SKAO capabilities has been made publicly available for the SKAO data challenges. These data sets can also be used as substitute for real SKAO data until that instrument comes online.

6. Detailed definition of LHC scientific use-cases

The primary LHC scientific use-case relevant to ODISSEE is the reconstruction of putative dark matter signals at the full LHC collision rate. At present LHCb is already able to reconstruct charged particles produced throughout the detector volume at the full LHC collision rate, with an efficiency which depends on the particle momentum. LHCb is also able to reconstruct neutral particles as clusters in its Electromagnetic Calorimeter (ECAL) under the same conditions. Both the charged and neutral particle reconstructions lose performance compared to the full physics-analysis-quality reconstructions available at later LHCb processing stages, particularly for low-momentum particles. The primary scientific use-cases of ODISSEE for what concerns LHCb are to

1. Quantify the cost, both in terms of computing and energy efficiency, of improving the performance of the full collision rate processing to match the physics-analysis-quality processing;
2. Develop full collision rate processing based on new architectures or AI/ML algorithms and evaluate to what extent these can help to reduce the aforementioned cost;
3. Improve the modularity and composability of LHCb's data processing frameworks both so that they can be used outside LHCb and to make it easier for LHCb to integrate state-of-the-art algorithms and processing pipelines developed outside the collaboration;
4. Quantify the cost of deploying the same full collision rate processing in the second upgrade of the LHCb experiment, which is expected to have to process a roughly ten times higher data rate

The software packages and datasets which will be used to implement these scientific use-cases were described earlier in this document. The evaluation of LHCb's processing pipelines on HPC infrastructures is already part of the collaboration's workflow for the Moore/CPU processing. It will be expanded in the first instance to a combined GPU + CPU pipeline, resources which are readily available in HPC centres, and benchmarked against a purely CPU-based workflow. Differences in the efficiency of deploying GPU resources between the custom LHCb datacentre and the HPC centre will be evaluated and documented.

7. Detailed definition of SKAO scientific use-cases

The SKAO is expected to provide break-through astrophysical science on several important topics, including:

Dark energy: Our Universe is expanding, and this expansion is accelerating. An unknown force called "dark energy" is responsible for this acceleration. SKA will study dark energy by surveying tens of millions of galaxies, allowing for the study of subtle cosmological effects. SKA will also observe the gravitational effects of galaxies and galaxy clusters on the path of radio waves through the Universe.

Galaxy evolution: It is still a mystery as to how the early galaxies, in the millions of years following the Big Bang, began to evolve: where did they get their material? What drives their rotation? What has shaped them? SKA's unrivalled sensitivity and resolution will be able to track young, newly forming galaxies at cosmological distances and help us unravel these key mysteries.

Astrobiology: Astrobiologists will use the SKA to search for amino acids, the building blocks of life, by identifying their spectral signatures at specific frequencies.

Cosmic magnetism: SKA will revolutionize the study of magnetic fields in space. When radio emissions of distant galaxies pass through a cosmic magnetic field, the emissions will be transformed by Faraday rotation. SKA will be sensitive enough to measure the signals from these individual distant galaxies and map the invisible magnetic fields, allowing scientists to study the shape and strength of the magnetic field in the Milky Way, the magnetism of the Universe, and where and how these cosmic magnetic fields originate.

The sensitivity of the SKAO will also allow us to search for evidence of the particle nature of **Dark Matter**. Radio searches for WIMPs focus on detecting synchrotron radiation originating from the products of DM annihilation, and may be evident in continuum observations. Competitive constraints on sub-TeV WIMPs have already been derived using SKAO precursors looking at dwarf galaxies, galaxy clusters, and the Large Magellanic Cloud. The superior continuum sensitivity of the SKAO will allow us to progressively close in on the WIMP parameter space.

Dark Matter may also be comprised of the hypothetical Quantum Chromodynamics (QCD) axion or other axion-like particles (ALPs) predicted by string theory. There are many different terrestrial experiments for axions and ALPs (see the Snowmass 2021 White Paper on Axion Dark Matter for a review) which have strong synergy with constraints from astrophysical observations. The ALP signal arises from its decay or conversion into photon(s), and typically consists of a nearly monochromatic signature. The spectral resolution and sensitivity of the SKA AA4 can be leveraged to constrain the ALP-photon coupling through spectral line studies.

The specific nature of Dark Matter will also have a impact on the distribution of matter in the Universe due to Gravity. For example, “cold” Dark Matter made up of WIMPs should form many small haloes. In axion or ALP Dark Matter, these small-scale haloes will be suppressed due to the wave dynamics of axions and ALPs. This distribution of Dark Matter will dominate the gravitational evolution of matter in the universe. A useful tracer of the matter distribution is the distinct 21cm / 1.42 GHz emission of neutral hydrogen. The SKAO should be sensitive enough to measure 21cm emission even at extremely high redshift $6 < z < 30$. Observations of the matter distribution in the early universe will be useful to disentangle the effect of dark matter from Baryonic effects (active galactic nuclei, supernova feedback, etc) which dominate the growth of small-scale structure in the late Universe.

The detection of any of these signals requires extremely high sensitivity. High spatial resolution, high spectral resolution, and potentially long integration times are essential elements in any successful search strategy.

We have described the Dark matter use-case for the SKAO in some detail here to emphasize the similarities with the LHC use-case. However, SKAO is designed to be a very flexible observatory, and the search for dark matter is only one of many scientific use-cases. The flexibility of such an iconic instrument is highly beneficial, but also brings additional challenges to both the software and hardware designed for such an instrument. Exploring these challenges goes well beyond the scope of this document, but ODISSEE is one step on the process to addressing some of these challenges in the future.

The SKAO software project is developing the baseline software stack for the instrument. As such it must follow a fairly conservative approach. The ODISSEE project allows us to supplement that effort, exploring potentially more efficient but more risky and experimental technologies and solutions.

Within the ODISSEE project we will improve the modularity and composability of key radio astronomy pipelines. This in turn allows us to target computational hotspots for alternative implementation with well defined interfaces for easy and convenient drop-in replacements.

Furthermore, we will instrument and measure SKAO-like workflows for energy-efficiency, both at macro- (using EAR) and micro-level (using powersensor).

Finally, select pipeline components will be targeted for analysis and optimisation. For this we will focus primarily on the analysis of European digital IP for our use-case, but to ensure a fair comparison, we will also target optimisation for best-of-breed more conventional hardware and software.

One key opportunity from the two above sections is a realisation that both LHC and SKAO use-cases aim for modularisation and composability. While this does not guarantee inter-operability, it does open that opportunity in the software domain, in addition to the science cross-over in dark matter that was already discussed.

8. Conclusion

The search for dark matter remains one of the most important priorities in fundamental science. The next generation of scientific infrastructures have the potential to push this search into previously unexplored territory, but in order to do so they will have to process datasets of unprecedented size in real time. We have shown that the physical

data processing constraints faced by two such facilities: LHCb the SKAO are remarkably similar. We have documented a specification capture for a converged science programme between these instruments, driven by the ODISSEE project, which will ensure that the collaborations can learn from each other and react in an agile way to any hints of DM emanating from one or the other experiment. We have also described the ways in which the LHCb and SKAO collaborations have agreed to share samples of simulated data in order to facilitate this converged science programme, and the detailed scientific use-cases which will be enabled by it. The collaborations established around this programme are also a best-practice pathfinder for the future exploitation of shared scientific infrastructures and maximization of their scientific potentials.

Acknowledgments

The authors thank the LHCb and SKAO collaborations for their constructive engagement with the ODISSEE project and for facilitating our collaboration.

References

- [1] [The LHCb Upgrade I](#). LHCb Collaboration (Roel Aaij et al.) JINST 19 (2024) 05, P05065 DOI: [10.1088/1748-0221/19/05/P05065](https://doi.org/10.1088/1748-0221/19/05/P05065)
- [2] SKA Phase 1 Executive Summary. The SKAO Collaboration [PDF](#)
- [3] Into the depths: Unveiling ELAIS-N1 with LOFAR's deepest sub-arcsecond wide-field images. J. M. G. H. J. de Jong et al. A&A, 689 (2024) A80 DOI: <https://doi.org/10.1051/0004-6361/202450595>
- [4] Highlights of the SKA1-Mid telescope architecture. G. P. Swart et al. JATIS (2021) DOI: 10.1117/1.JATIS.8.1.011021