Development of the Program Execution System Architecture (PESA) for MSE

C. Surace*[a], J. Marschal [b], J. Sobeck[c], K. Szeto[b,c], , K. Ho[c], T Vermeulen[c], L. Wells [c], C. Wipper[c], K. Withington[c] and the MSE PESA team

[a]Aix Marseille Université, CNRS, Laboratoire d'Astrophysique de Marseille, France.

[b]National Research Council Canada, Herzberg Astronomy and Astrophysics, 5071 West Saanich Road, Victoria, BC, Canada, V9E 2E7

[c]CFHT Corporation, 65-1238 Mamalahoa Hwy, Kamuela, Hawaii 96743, USA

ABSTRACT

Maunakea Spectroscopic Explorer (MSE) is a telescope dedicated to multi-fibers spectroscopy and IFUs observations of the sky. Program Execution System Architecture (PESA) is one of the systems of MSE, responsible for planning, executing, reducing, and distributing science products from survey programs. Work is being done to design PESA in a modular way to include several sophisticated software tools, organized into an operational framework. This paper describes the first step of its organization and the concepts that will be used in the development of PESA.

Keywords: MSE, data processing pipelines, Computing Infrastructure, Observation organization

1. INTRODUCTION

Maunakea Spectroscopic Explorer (MSE) is the first of the future generation of massively multiplexed spectroscopic facilities. MSE is designed to enable transformative science, being completely dedicated to large-scale multi-object spectroscopic surveys, each studying thousands to millions of astrophysical objects. At a minimum, MSE will use an 11.25 m aperture telescope to feed 4,332 fibers over a wide 1.52 square degree field of view. It will have the capabilities to observe at a range of spectral resolutions, from R~3,000 to R~40,000, with all spectral resolutions available at all times across the entire field. Alternate facility architectures are under evaluation with insight from participants' national strategic planning priorities along with technical feasibility. Engineering development is supported by a culturally and geographically diverse design team that is centrally coordinated and managed by the Project Office. We are cognizant that the decisions we make today are intertwined with the future of Maunakea and its cherished summit. The MSE Project deeply respects its cultural importance and storied past.

Program Execution System Architecture (PESA) is being developed to a level of design where the essential functions are identified, and interfaces have been defined to enable smooth, repeatable, and observable operation of the pipeline. Requirements have been analyzed and quality/product assurance procedures are defined to monitor the increasing quality and maturity of the codes and the reliability of the pipelines. The preliminary study has been carried out by T. Flagey (Flagey et al., 2017 [1**Erreur ! Source du renvoi introuvable.**]) The project is assembling a collaborative team to plan this software effort.

PESA should provide four main functionalities: provide access to proposals submission and modification, provide optimization steps for surveys and observations, provide access to data, and offer data processing services and provide access to high level data throughout a specific platform.

Moreover, PESA will interface with different external systems as OESA, and on-site Infrastructure. Indeed, the PESA System will run inside the CFH infrastructure and will have to be compliant to the infrastructure requirements. The structure, products, and architecture of PESA are described in Section 2

* christiansurace@mkspectroscopicexplorer.onmicrosoft.com

2. PESA ARCHITECTURE

The PESA structure has been defined to cover the pre- and post-observation phases. It is linked to the Observatory Execution System Architecture (OESA) which contains the different observatory control systems, hardware, and software, and the MSE IT infrastructure.

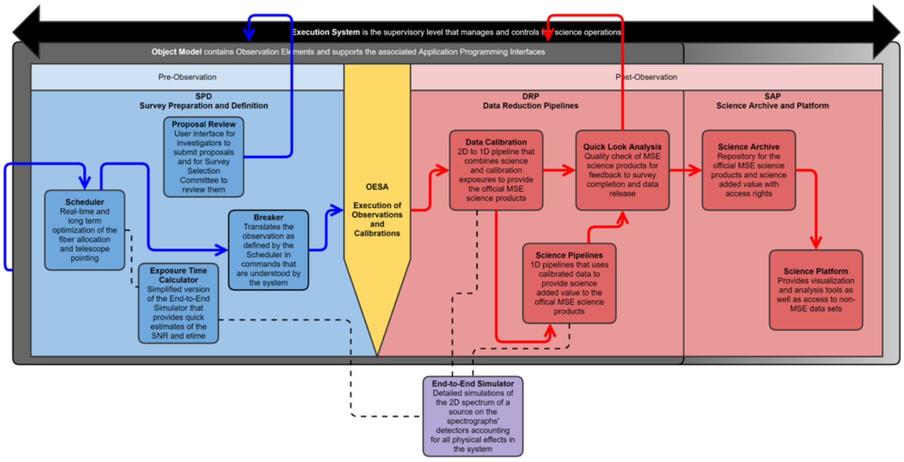

Figure 1: PESA block diagram describing the workflow of PESA. Arrows indicate the flow of data and interfaces between SPD, DRP, SAP and OESA. The Object Model is taking care of the data interfaces.

As shown in Figure 1, PESA is organized into 3 groups defined by functionality: Survey Preparation and Definition (SPD), Data Reduction Pipelines (DRP) and Scientific Platform and Archive (SAP). The object model supports the interactions between each of these groups.

The SPD group products complete the pre-observation operations where the selected survey proposals are converted into observation commands that are understood by the OESA observatory control sequencer. MSE will observe the survey programs simultaneously from an integrated target list, which contains all target definitions, managed by the object model (ObjMod). The DRP encompasses the proposal selection work of phase 1 and the target definition of phase 2. The products of the DRP and SAP groups generate, validate, and then deliver the final science data products of MSE by processing the science detector readout from the spectrographs collected in the observation phase via OESA. Individually, DRP is taking care of Reduction/Validation work, and SAP is responsible for the data distribution. The functionality of each PESA product is presented in the following sections.

2.1 PESA Product Breakdown

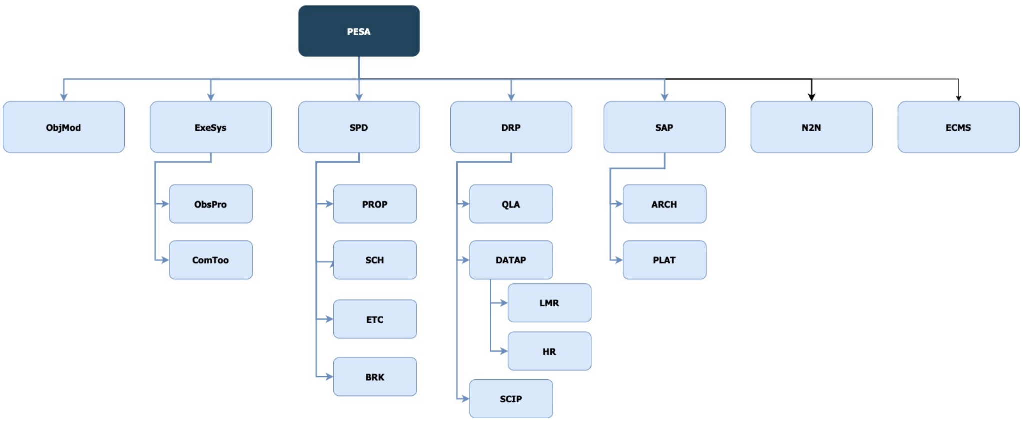

Figure 2: MSE PESA Product Breakdown Structure. The main systems are split in subsystems that could be developed and deployed independently.

PESA has been split into subsystems that are defined to deal with different levels of data, infrastructure, and interfaces. Each subsystem will deal with different level of data product. Data products produced by MSE are referred to as Level 0 (raw data), 1 (early calibrated data ready for quick look analysis) and 2 (calibrated data ready for Scientific analysis), and those by external teams are referred to as Level 3 (scientific data with added values from MSE teams) and 4 (scientific data with added values from astrophysical community). Figure 2 shows the products breakdown structure of PESA. They are:

- PESA.ExeSys: Execution system to organize the overall infrastructure of PESA
  - PESA.ExeSys.ObsPro: Observation Production to define, install, host the infrastructure of development and production
  - PESA.SPD.ComToo: Common tools to provide APIs, data access and developing and tests facilities
- PESA.ObjMod: Object Model to structure the interfaces and data exchanges
- PESA.SPD: Survey Preparation and Definition to organize the pré-observations facilities
  - PESA.SPD.PROP: Proposal Review to organize the proposal reception, validation ans estimation
  - PESA.SPD.SCH: Scheduler to optimize the organization of the surveys
  - PESA.SPD.ETC: Exposure Time Calculator to compute Signal to noise ratio
  - PESA.SPD.BRK: Breaker to prepare deliverables to OESA for Fiber allocation
- PESA.DRP: Data Reduction Pipelines, to organize the reduction of raw data to L2 data
  - PESA.DRP.DATAP: Data Pipeline, to organize the L1 to L2 data pipeline
    - PESA.DRP.DATAP.LMR: Low Medium Resolution Pipelines, to define and develop LMR data reduction pipeline
    - PESA.DRP.DATAP.HR: High Resolution Pipelines, to define and develop HR data reduction pipeline
  - PESA.DRP.SCIP: Science Pipelines, to define and develop L2 to L3 data data reduction pipeline
- PESA.SAP: Science Archive and Platform, to organize development on data dissemination
  - PESA.SAP.ARCH: Science Archive, to define infrastructure and databases development for archiving data
  - PESA.SAP.PLAT: Science Platform, to define and develop the interfaces to data dissemination, including links to data processing tools
- PESA.N2N: End-to-end simulator, to develop simulations for technical and scientific validations.
- PESA.ECMS: Environment Controller, to develop and interface with data from Environment sensors.

2.2 PESA subsystems

2.2.1 ExeSYS.ObsPRO

ExeSYS.ObsPRO is responsible for creating the production infrastructure and logical environment. It will define the constraints and the computing resources available to run PESA. He will take care of the access and rights to be granted to observers and scientists. He will also collect logs and ensure the high availability of the computing facilities.

2.2.2 ExeSYS.ComToo

ExeSYS.Com brings together all the common tools that will be useful to develop and operate PESA pipelines. It includes the creation of the development infrastructure and the provision of the virtual test environment. It also includes the development of the API for external interfaces to the OESA data management system, and high-level APIs for internal interfaces to the SAP.ARCH archive and the object model ObjMod. It finally gathers as well tools for quality assessment.

2.2.3 Object Model (ObjMod)

The purpose of the object model is first to define the content of the data and associated metadata to be exchanged and to connect the pre-observation and post-observation domains. Having defined the content, ObjMod is also responsible for defining the data format and its serialization. It also defines the links to the environmental and technical data in the OESA Data Management System (DMS), its status at system level and the performance of the system during scientific operations.

Figure 3 shows the primary design of the object model. Depending on the subsystem that uses the ObjMod, one or more classes will be instantiated to store the information to be exchanged. The object model will not be serialized before each exchange, however a process to serialize the ObjMod in case of a crash and a "recover from last save" function must be implemented.

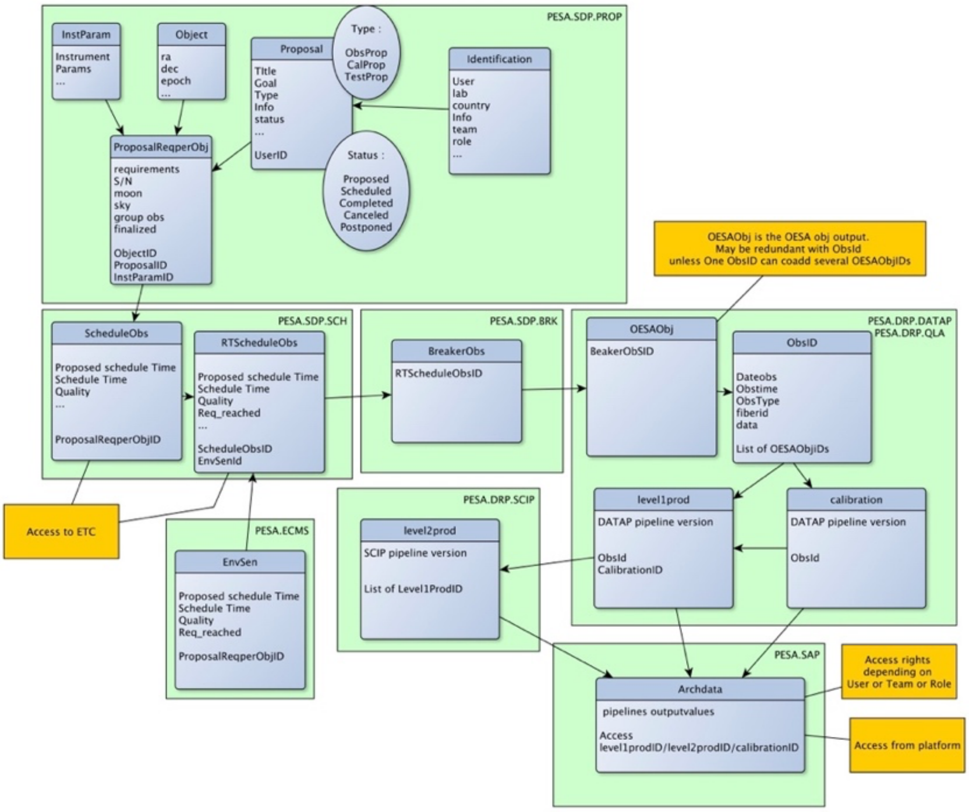

Figure 3 : MSE PESA preliminary design of the Object Model. Each blue rectangle box is a class of the ObjMod and arrows define the links between each class. Each underlying green box is the associated functionality defined in the product tree. Orange box is a comment.

2.2.4 SPD.PROP

SPD.PROP is the subsystem responsible for the submission, review, acceptance of observation proposals to be operated with MSE. It allows scientists to define and prepare the survey proposal during the specific call for proposals period. The SPD.PROP collects the requests and provides all the necessary tools for the Survey Selection Committee to select, modify, reject or accept the proposal. It provides interfaces to facilitate and automate the selection process. Proposals can be interfaced with the object model (ObjMod) to gather all the information needed to finalize the observations plan for the proposed survey. A facility should be provided by SPD.PROP to select target lists based on nature, magnitude, sky regio. Depending on the goal of the proposal, alternative targets may be proposed or added to the obsever list.

SPD.PROP will also validate the feasibility of observing the list of targets and will deduce the probability of carrying out the project while respecting all the constraints given by the observer.

SPD.PROP will be based on the new proposal system called Kealahou to conduct Queued Service Observations at CFH (C. Wipper et al. [7])

The outputs of the SPD.PROP will be stored in the archive SAP.ARCH

2.2.5 SPD.SCH

One of the main keys to the success of the MSE is the efficient scheduling of observations. The goal of the scheduler (SCH) is to optimize science observations based on time, sky coverage, and proposal targets. It will need to consider the priorities and requirements of each survey program and prioritize targets across multiple proposals. It will consider all observational constraints and plan for the long term (year ahead) and short term (night ahead). The algorithms are complex because of the large number of variables to be taken into account: sky position, sky brightness, observability of objects, homogeneity of magnitudes of the selected samples to be observed, priority of the studies and constraints on the assignment of MSE fiber to target allocation simulator described in (J. Marshall et al., 2022 [6]). Such algorithms have already been tested for 4MOST (Tempel et al., 2020 **Erreur ! Source du renvoi introuvable.Erreur ! Source du renvoi introuvable.**) or Rubin telescope (Naghib et al., 2019) and could be reused for optimization. During the operation, SCH is also responsible for updating the observing plan continuously. It interacts with ETC and ECMS, during current observation to compute the characteristics of the observations and maintain optimal scheduling. An interface to the observer is also planned to modify the ranking of targets and surveys priorities.

2.2.6 SPD.ETC

The Exposure Time Calculator (ETC) is a subsystem, the algorithm of which estimates system performance, SNR and integration time. It provides Graphical interface to ease the use and computation of SNR of observations depending on sky conditions, observation characteristics and nature of the source. It has an interface with the Scheduler (SPD.SCH) as described in previous section. The ETC is being developed by the KHU University (Tae-Geun Ji et al., SPIE 2022 [3]).

SPD.ETC is also described in (J. Marshall et al., 2022 [6])

2.2.7 SPD.BRK

Once the SPD.SCH has defined an observation, the Breaker, (SPD. BRK) formats the output extracted from the ObjectModel with all the information needed to set up the observation. It is the interface to OESA and strongly follows the definition of the external interface for data exchange between PESA and OESA. It monitors the status of the fiber allocation and forwards any messages (warning, error, success) from OESA to the PESA logging system.

2.2.8 DRP.QLA

DRP.QLA is run on demand and uses some of the algorithms of DRP.DATAP. The purpose of this subsystem is to derive an initial analysis of the observed data overnight. It will use a non-optimal calibration to perform a quick reduction of the observation data to perform a quality assessment of the observations. Some statistical analysis can be performed like SNR calculation, spectral extraction, visualization.

2.2.9 DRP.DATAP

The raw Level 0 2D spectra data extracted from the OESA data management system are processed using the DRP.DATAP data pipelines. DRP.DATAP uses either a fast calibration (Level 1 data 1D Spectra) or the optimized overnight flux and wavelength calibration, the sky corrections that will be implemented, and the cosmic corrections (Level 2 data 1D Spectra). If the Level 1 data are processed and delivered the day after the observation, the Level 2 data can be delivered after the appropriate calibration and processing have been completed. It can be delivered later, after several observations and several nights. DRP.DATAP delivers the Level 1 and Level 2 products stored in the ObjModel to the science pipelines (DRP.SCIP). A statistical analysis will be used to return information to the planner (SPD.SCH) and to populate specific quality assurance metadata for each spectrum. The object model will provide more information about the environment of the observation, which may help identify problems that occurred during data acquisition. It includes the specific calibration pipelines as well that will run in between scientific observations.

DRP.DATAP has been divided into two subsystems: DRP.DATAP.LMR which processes LMR data and DRP.DATAP.HR which processes HR data. Although some of the processing methods should be the same, the advanced processing of LMR and HR data has some specificities that require separate development and execution.

The outputs of the DRP.DATAP will be stored in the archive SAP.ARCH

2.2.10 DRP.SCIP

DRP.SCIP is the second part of the pipeline process. From Level 2 1D spectroscopic data, it will produce Level 3 and 4 scientific products that are value-added products such as redshift determination, line fluxes, line profile, clustering indices... it is foreseen that the architecture of the DSP.SCIP will be open enough to allow the observer and research team to add their own pipelines. This impacts the PESA architecture and must be validated against the constraints of the production architecture defined by ExeSYS.ObsPRO. There is no separation between HR and LMR modes as DRP.SCIP could combine both observations for specific data fusion optimization.

The outputs of the DRP.SCIP will be stored in the archive SAP.ARCH.

2.2.11 SAP.ARCH

SAP.ARCH is the archiving facility where all data and metadata are serialized, from level 0 to level 4 products. SAP.ARCH is developed to store and easily retrieve information acquired during the data life cycle. A subsample of environmental data and quality control metadata will be available. Although the archive is intended to be accessible using the SAP.PLAT scientific platform, authorized individuals will also be able to access it through direct queries.

During observation, access to the overnight data will be granted once the data have been validated by the MSE team.

Concerning the technical point of view, it has not yet been defined whether the architecture will be a relational database or a NOSQL database. Both solutions are plausible and further work is needed to make a final decision. Moreover this decision will have an impact on the architecture provide by ExeSys.ObsPro.

2.2.12 SAP.PLAT

The scientific platform (SAP.PLAT) is the interface with the archives. It will provide access facilities to the data available in the archive. In cae of restricted access It will give access only to the PI of the surveys and to the research team responsible for the data processing. The data will be private until the end of the ownership period. SAP.PLAT will provide visualization and statistical tools to validate the data and analysis tools to the community.

2.2.13 PESA.N2N

It will provide 2D and 1D spectra (LMR and HR) to validate the pipelines and provide a test sample to facilitate comparison of different stages of pipeline development. The N2N goes into more detail than the SPD.ETC. It will create an artificial pixel-by-pixel image of the detector and provide data for all stages of the pipeline. It will then be able to simulate a complete run with observation constraints and environmental data. Some extra noise or uncertainties can be added to the output data in order to simulate unknown effects.

2.2.14 PESA.ECMS

The Environmental Conditions Monitoring System (ECMS) is providing to the pipelines all information concerning the external sensors. It is built to interfaces with the OESA sensors and will provide information to be store into the archive (SAP.ARCH). The Environmental Sensors (EnSe) product is comprised of sensors for monitoring system-level conditions such as temperature probes, anemometers, humidity and precipitation sensors, seeing cameras, cloud/sky cameras, water column density monitor, dust particles, etc. The environmental data are stored in the OESA Data Management System and its status server

2.3 Interfaces

Most of the internal interfaces will be provided by the object model and APIs provided by the common tools. Figure 4 shows the interfaces between each subsystem. From observation proposals to value-added data dissemination, processing history and product provenance must be available. The use of the ObjectModel allows this information to be maintained and global dependencies and history to be serialized when necessary.

External interfaces to OESA and DMS systems are crucial for the execution of PESA. ExerSYS.ComTool will provide APIs to interface with them and will also retrieve information about the status of the various processes.

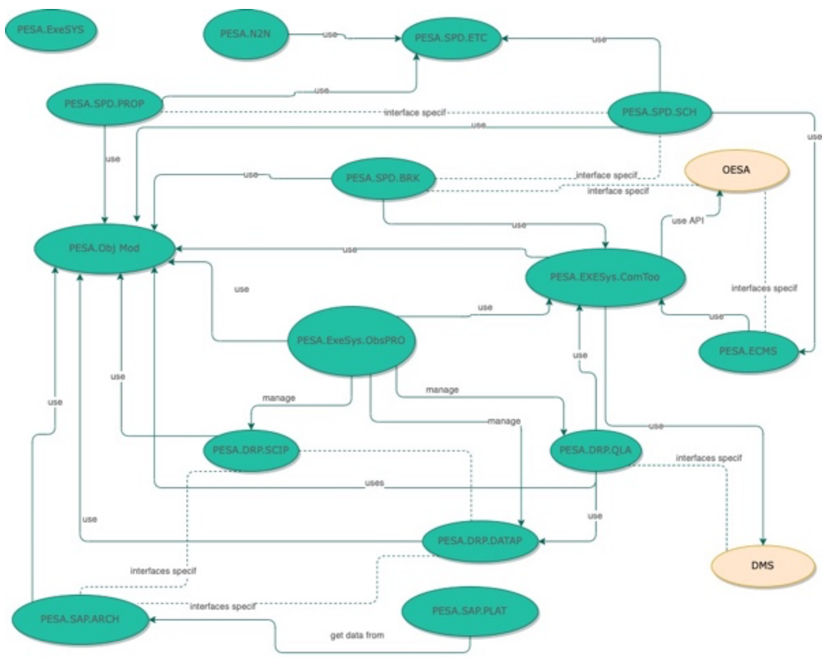

Figure 4: MSE PESA Products interfaces. PESA products are fille with green color, CFH products are fille with light orange color. Relations between each product are defined with straight lines, while dashed lines show the logical and interfaces specifications links between 2 PESA products.

3. DEVELOPMENT PLAN

3.1 General considerations

PESA will be developed with distributed teams dedicated to the development and integration of specific work packages. Each work package has been assigned to teams based on their expertise and experience in this type of development. An advisory board has been set up to help the PESA teams solve specific problems. Each work package has a list of deliverables, and each work package manager is responsible for the development and deliveries due to their work package.

Each delivery will be tested and validated in specific meetings. It is expected that these meetings will be linked to specific stages of the project. The development of the PESA system and sub-systems is managed by progress meetings dedicated to each PESA sub-system, and a periodic global meeting is organized with all PESA contributors and members of the Advisory Board are invited to participate. Most of these meetings take place mainly in the form of remote video conferences. To monitor actions, organize development and share information, a dedicated platform has been set up.

3.2 Technical organization

PESA will use several facilities and technical implementations to facilitate the development, registration, information, and monitoring of the subsystems.

It will provide a virtual environment as well as a remote environment to help subsystems to integrate their developments as well as possible. This will allow some new developments to be tested in local mode before any official release. This virtual environment will bring together all the necessary common tools and libraries dedicated to the development and validation of the subsystems' functionalities. This environment is still to be defined with the parties to converge on a list of tools and libraries. It will include the version control system (like GITLAB), associated with a continuous deployment (GITLAB CI/CD), and a container registry (like GITLAB's Docker container registry) to facilitate delivery and a specific validation and verification environment.

This option will allow testing of developments with local and specific test data. However, this environment must mimic the production system to facilitate the integration of developments into the production system.

The production facilities will be dedicated to the execution of PESA developments and will be designed to ensure the performance of PESA according to the technical requirements. The architecture and capacity of storage, processing power and networking will be evaluated.

3.3 Development approach

The development of PESA will aim to be implemented through an agile approach. This approach will define an iterative development and build, test and deliver modules to be deployed with periodic deadlines. The idea is to increase the maturity of the code to be delivered and to deliver a functional PESA with increasing complexity.

However, each development phase will be divided into several iterative sprints which include:

- Requirement analysis and requirement modification analysis
- Software architecture definition and modification
- Candidate algorithms prototyping if needed
- Development and tests and deployment of prototypes and/or Development, tests, and deployment of production algorithms (depending on the phase of PESA development)

3.4 Development planification

With a test-driven approach, the development of PESA has been divided into several challenges to validate each new delivery and functionality. We have identified several challenges that should take place every year: the system challenge, the data challenge, and the validation test plan (see Figure 5: Overview of the MSE PESA development plan divided into 5 technical or scientific themes.. While the system challenge and the data challenge are technical challenges, the validation test plan is a more scientific test against the requirements. A data challenge whose goal is to simulate and analyze with the pipeline developed up to this event. The goal is to verify if the overall pipeline can provide the necessary functionality. This will be an indicator of the speed of development.

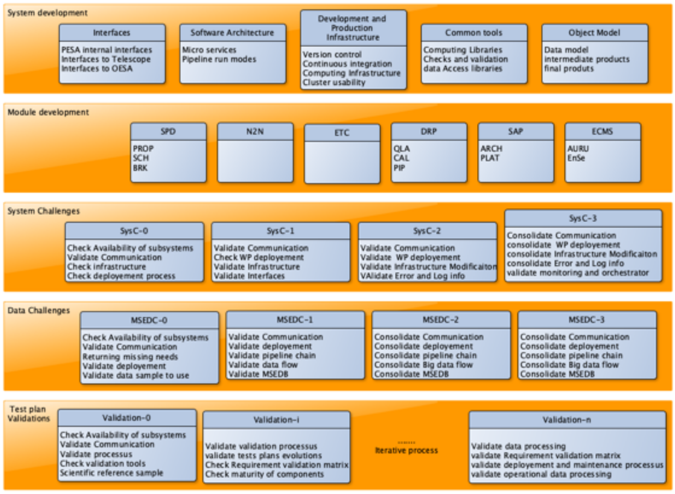

Figure 5: Overview of the MSE PESA development plan divided into 5 technical or scientific themes.

4. FUTURE DEVELOPMENT

MSE has evolved smoothly, the last 3 years have been impacted by COVID but PESA will be part of the next MSE development process. A design review (CoDR) is scheduled before the end of the year. For the coming year, there are well identified needs such as ETC, SCH, DRP, N2N to better help define science projects. These are the subsystems that will start to be developed. PROP is still under development as Kealahou version 1 has been developed and deployed in production mode in Q3 2022. The MSE change study will have little impact on the definition of the co-DR, but the requirements flow will certainly change some of the functionality and constraints.

The development plan has been written but still needs to be refined to be the baseline to follow.

Acknowledgements:

The mission of the MSE Project is to realize a dedicated facility that enables a diverse suite of large-scale spectroscopic surveys of millions of astrophysical objects at a range of wavelengths, spectral resolutions, redshifts, and spatial scales.

The MSE Project is hosted by the Canada-France- Hawaii Telescope Corporation, and supported by contributing organizations in Canada, France, Hawaii, Australia, China, India, South Korea, Texas, the UK, and the US. The MSE collaboration recognizes the cultural importance of the Maunakea summit to a broad cross-section of the Native Hawaiian community, and is committed to equity, diversity, and inclusion. Statements of MSE's mission, cultural respect, and equity, diversity and inclusion are available on https://mse.cfht.hawaii.edu.

REFERENCES

1. Flagey N., "MSE Operations Concept Document", MSE.PO.ENG.DOC-REQ-002, MSE Project document (November 2017
2. Vermeulen T., Isani S., Withington K., Ho K., Szeto K., Murowinski R, "Observatory software for the Maunakea Spectroscopic Explorer", Software and Cyberinfrastructure for Astronomy, Proc. SPIE. 9913 (July 2016).
3. E. Tempel, T. Tuvikene, M. Muru, et al.; An optimized tiling pattern for multi-object spectroscopic surveys: application to the 4MOST survey, MNRAS, 497, 4, 4626–4643 (October 2020 (https://doi.org/10.1093/mnras/staa2285)
4. Elahesadat Naghib, Peter Yoachim, Robert J. Vanderbei, Andrew J. Connolly, and R. Lynne Jones; A Framework for Telescope Schedulers: With Applications to the Large Synoptic Survey Telescope, , The Astronomical Journal, 157, 4, 2019
5. Tae-Geun Ji, Taeeun Kim, Changgon Kim, Hojae Ahn, Mingyeong Yang, Soojong Pak, et al., SPIE 2022, this volume, 2022
6. Marshall, J. L., Petric, A. O., Sobeck, J., Flagey, N. "Planning scientific operations for the Maunakea Spectroscopic Explorer" 2022, SPIE Astronomical Telescopes+Instrumentation, 12186-7
7. C. Wipper et al., https://www.cfht.hawaii.edu/~wipper/CASCA/IntroducingKealahou.pdf ,2020