Implementation of On-premise Research Data Science and Systems Explorer (ODySSEy)

Singapore Healthcare Management 2023

Narayanan Ragavendran, SHHQ
Lam Shao Wei Sean, SHHQ
Aw Kok Leong, Singhealth
Chong Wai Keong, Singhealth
Yeo Khung Keong, Singhealth
Ong Eng Hock Marcus, Singhealth

INTRODUCTION

On-premise Research Data Science and Systems Explorer (**ODySSEy**) is a purpose-built data science platform for SingHealth Research and Innovation with cleaned Data Pipelines. The Platform is intended to transparently orchestrate and streamline the lifecycle, management, compliance and governance of SingHealth research data.

OBJECTIVES

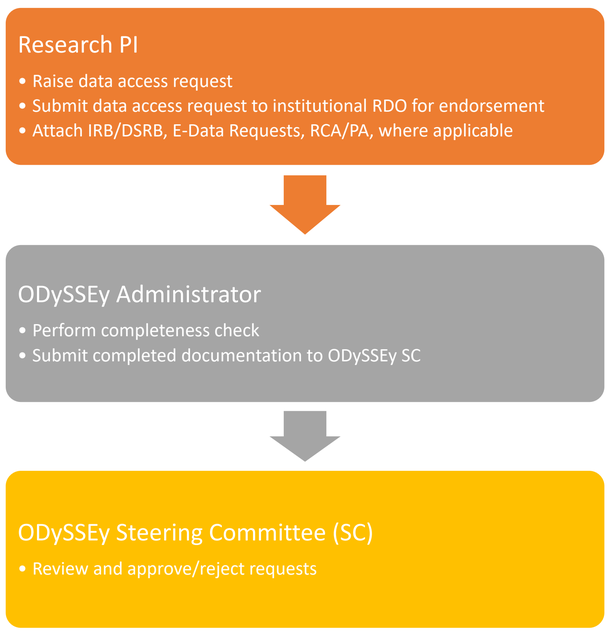
- To Provide Researchers with a robust data science platform with secure access to research data under the existing SingHealth IT and data governance policies, national legislation (e.g., Human Biomedical Research Act, Health Information Manual, Personal Data Protection Act).
- To enable linkages to other data sets approved for research (OMICS, Images, Hospice, external data sources)

METHODOLOGY

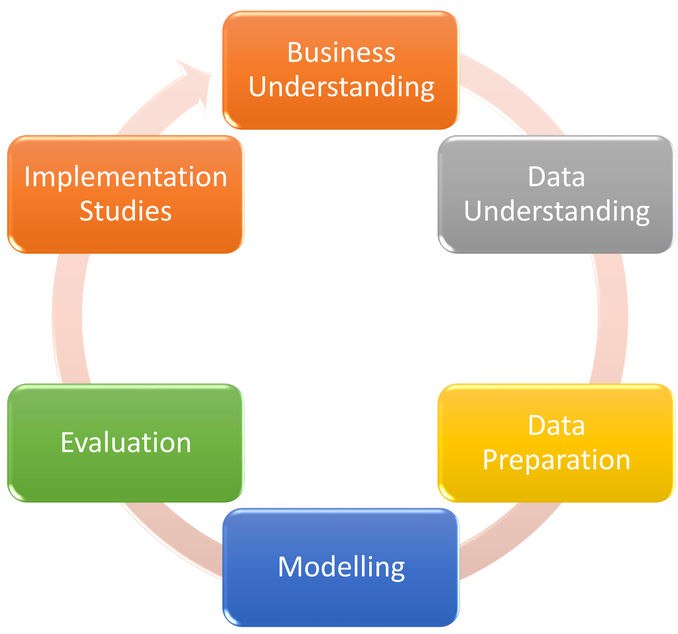
Successful Project implementation begins with an alignment to concrete business problems. ODySSEy platform is designed to provide services that involve stakeholders to ingest, process, store, access, analyze and present data and insights during the data science processes.

The significant ODySSEy lifecycle steps involve data ingestion, data engineering, modeling, evaluation and deployment. Data linkages via the ODySSEy platform to other data sets approved for research (OMICS, Images, Hospice, external data sources) enabling self-service technological infrastructure with scalable capabilities to extend to secure and approved hybrid cloud configurations in future phases. The data governance policies and procedures are applied at every stage of the ODySSEy data life cycle to ensure the quality, compliance and usability of data.

ODySSEy Governance (Ref:SHS-MI-206)



ODySSEy Support Data Science Process



The multidisciplinary project team comprised members from

- SingHealth Health Services Research Centre (HSRC)
- SingHealth Office of Deputy Group Chief Medical Informatics Office (Research) – DGCMIO (Research)
- SingHealth Office of Academic Informatics and
- Integrated Health Information Systems (IHiS)

The team worked directly with the ODySSEy System Integrator throughout the project.

The project implementation started during the unprecedented COVID-19 pandemic period. The project team was continuously monitoring the project demands and resources in the face of this unknown threat and prepare to cope with change in the work environments.

CONCLUSION

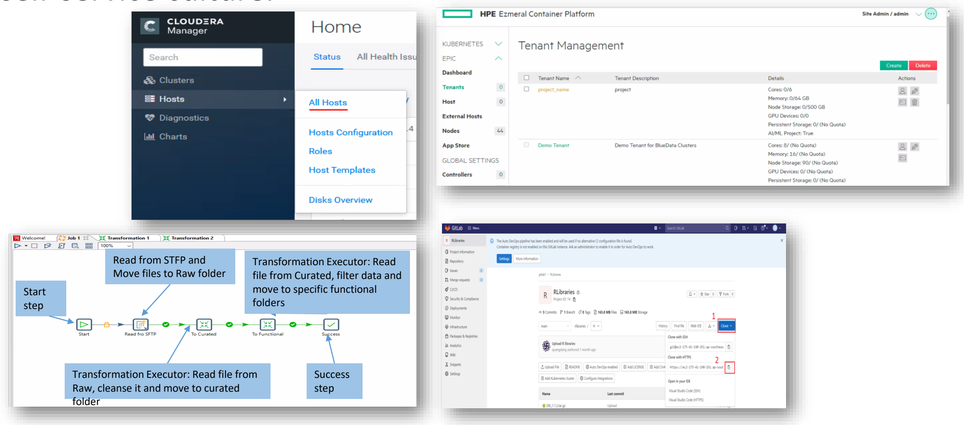
ODySSEy Data Science Platform is now operational and available for users. We have initiated process of on-boarding of approved Data Science use cases to ODySSEy Platform through DGCMIO (Research) office. Upon successful completion of this Phase 1 , enhancements to this platform will be taken up in subsequent phases.

ACKNOWLEDGEMENT

The project team would like to thank SingHealth Senior Management, System Integrator (Hitachi Vantara), and subsystem vendors (HPE and Cloudera).

RESULTS

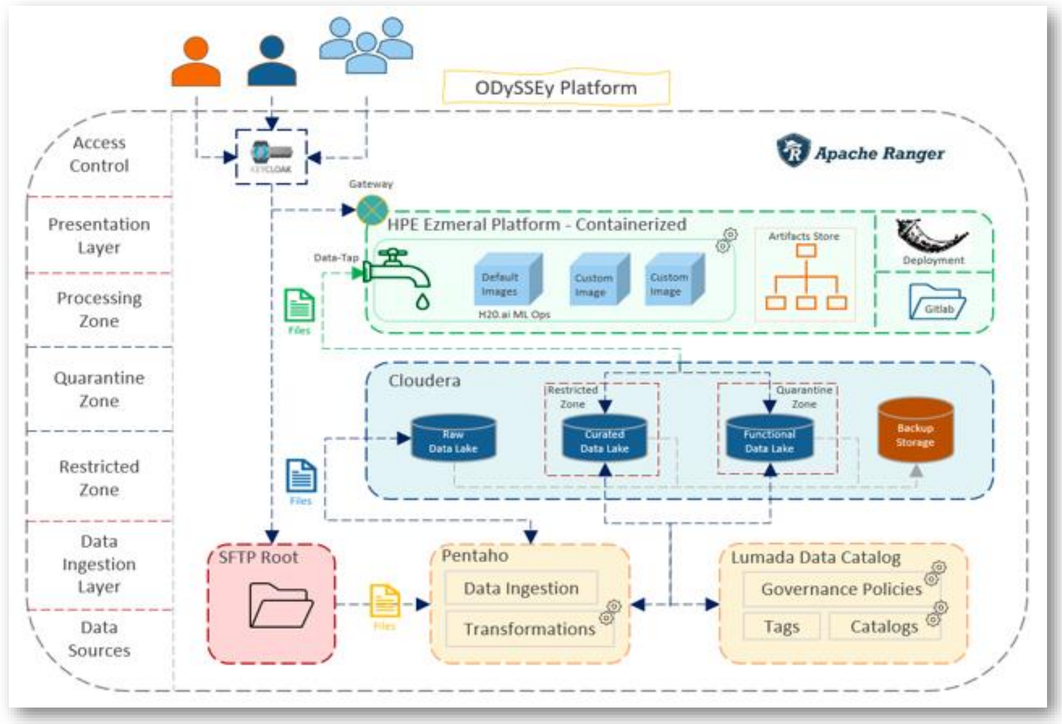
We have established a unified data science platform for SingHealth research needs. Consolidating data repository of all SingHealth research data to build a SingHealth cluster wide research platform that supports a self-service culture.



ODySSEy Platform - Tools & Softwares

Category	Name
Operating System	RHEL
Hadoop Software	Cloudera Data Platform
Data Processing Software	Pentaho Enterprise – PDI, BI, Hadoop Shim
Cataloging Software	Lumada Data Catalog
AI / ML Software	HPE Ezmeral Platform
Code Repository	GitLab
Artifacts Repository	JFrog Artifactory
MSAD integration Software	Keycloak
Security Software	Apache Ranger

ODySSEy – Functional Architecture



ODySSEy Data Architecture