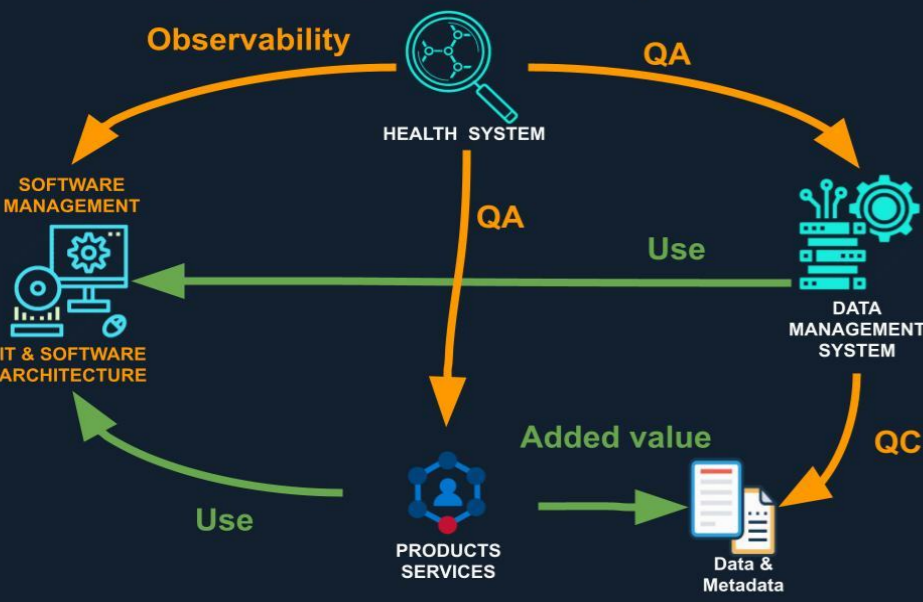


SOCIB Data Infrastructure

Improving the Data Management Program

01 Our Goal: Re-addressing Data Quality

SOCIB e-infrastructure components



The SOCIB e-infrastructure includes several key components including a Data Management System, the SOCIB Products & Services (accessed through Virtual Access) and the IT & Software system. The so-called Health System is also part of the e-infrastructure and it aims to ensure that everything works as expected. In this scenario, the SOCIB Data Quality is being re-addressed to better meet the new requirements from the coastal ocean community. These new Data Quality objectives focuses on the improvement of both (1) **data assets FAIRness** and (2) **data quality processes and procedures**. To guarantee that SOCIB benefits from the new Data Quality objectives, the Data Management Program is being updated accordingly.

02 Data Management Program update



The SOCIB **Data Management System (DMS)** is the set of processes, services and documents aiming to effectively manage the SOCIB data assets lifecycle. The SOCIB Data Management Program is the framework in which the SOCIB DMS is managed, assessed and improved. The SOCIB **Data Management Program** is being implemented based on the **Data Management Maturity model (CMMI institute)**, which is providing a best practices roadmap and framework to improve the SOCIB data management function, including data strategy, data governance, data operations, data quality and data preservation among other areas.

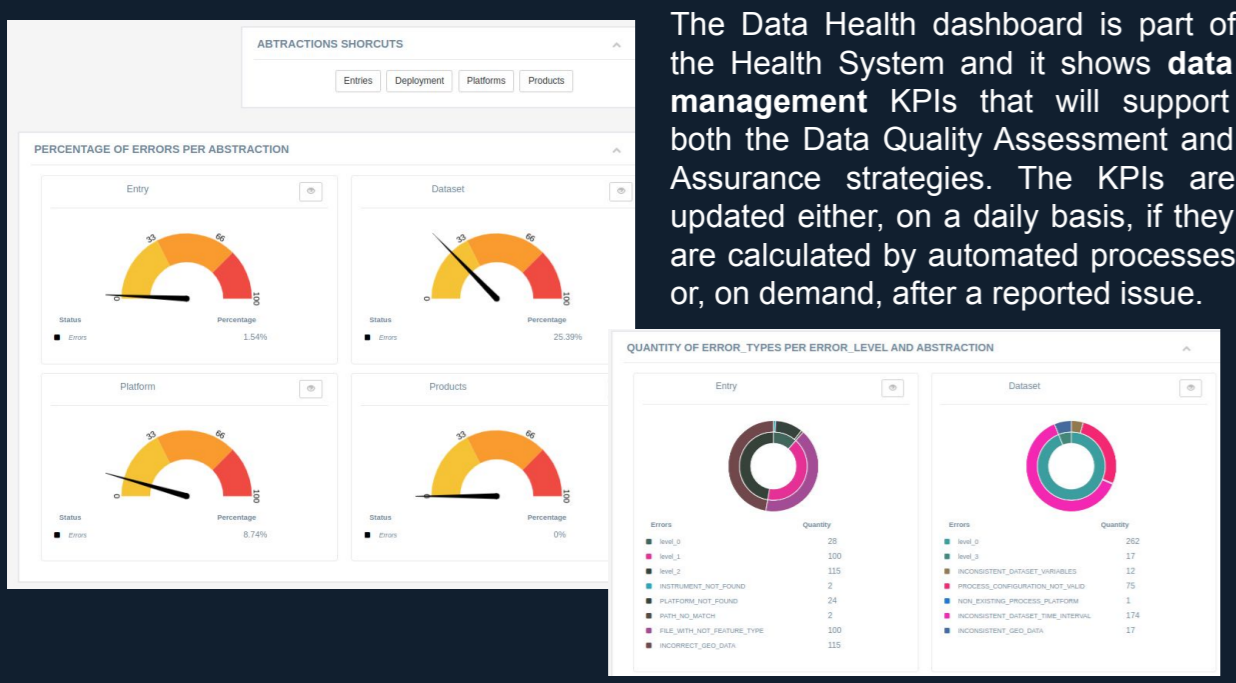
03 Action lines

- 1 **Quality Assurance, Quality Control and Quality Assessment**
- 2 **Upgrading the IT architecture**
From: Isolated and not measured
To: Integrated, scalable and measured

FAIRness increased

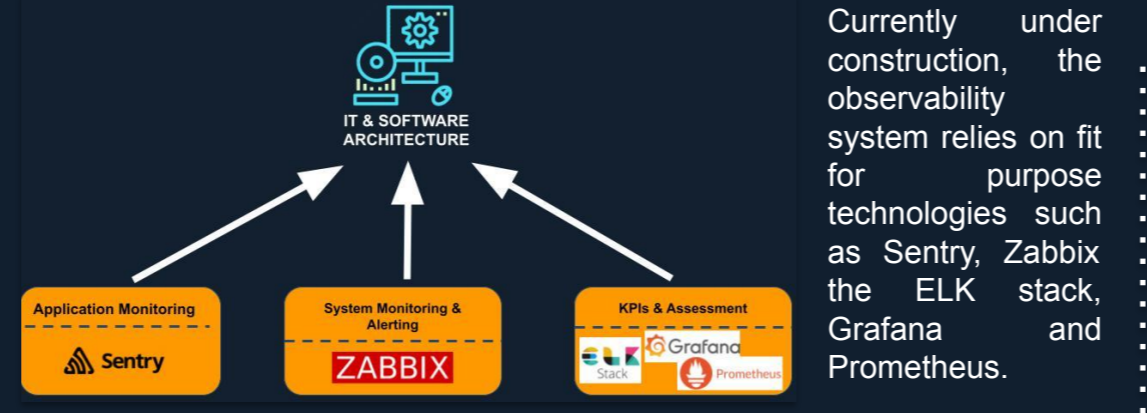
Specific actions

04 4.1 Data Health dashboard



The Data Health dashboard is part of the Health System and it shows **data management KPIs** that will support both the Data Quality Assessment and Assurance strategies. The KPIs are updated either, on a daily basis, if they are calculated by automated processes or, on demand, after a reported issue.

04 4.2 e-infrastructure observability

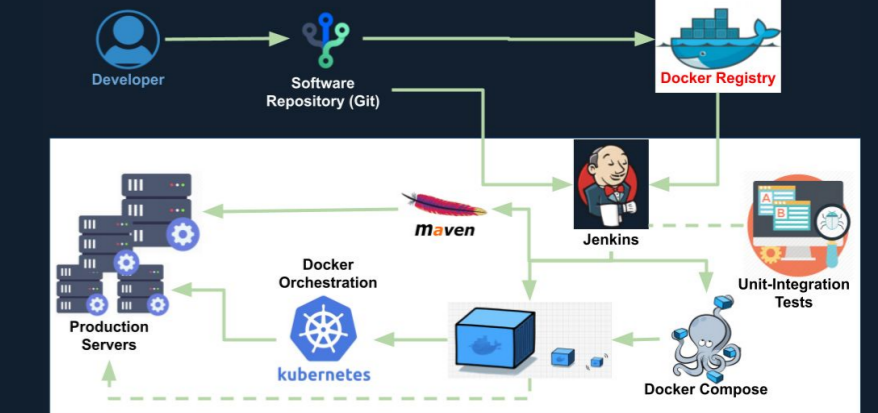


Currently under construction, the observability system relies on fit for purpose technologies such as Sentry, Zabbix the ELK stack, Grafana and Prometheus. The IT infrastructure contains all the software processes and services that powers the e-infrastructure. Observability provide us with key information of the behaviour and health of the system, and allow us to be informed quickly in respect to: SOCIB's software issues and system's services issues. In addition it is providing performance and access metrics to further assessment and refinement.

04 4.3 Data lifecycle observability

SOCIB's data assets include raw data, metadata, processed data (including QC flags) and data products (with DOIs). The underlying data lifecycle can be inspected (observed) to infer issues and information related to a variety of topics such as: the behaviour of the SOCIB multiplatform observing system (eg. sensors malfunctioning), both communication and data flow, archiving and preservation, exploitation of access metrics (aggregators, scholarly, social networks), among others. So far, some focussed pilots have been developed (eg. glider data delay of arrival monitoring), and plans are in place to extend the observability scope of the data lifecycle during 2021 and beyond. This new observability capacity will be key for supporting data stewardship and Data Quality overall.

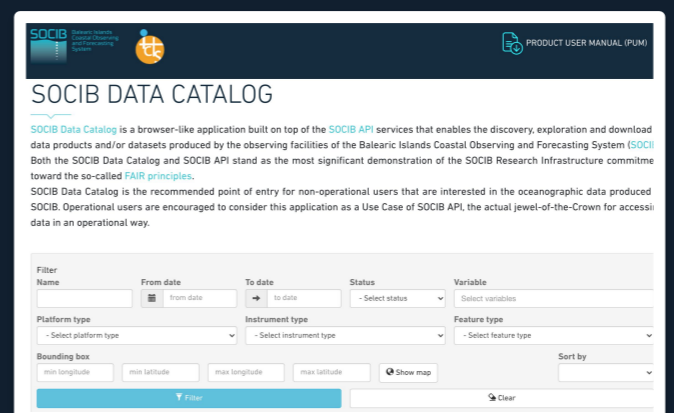
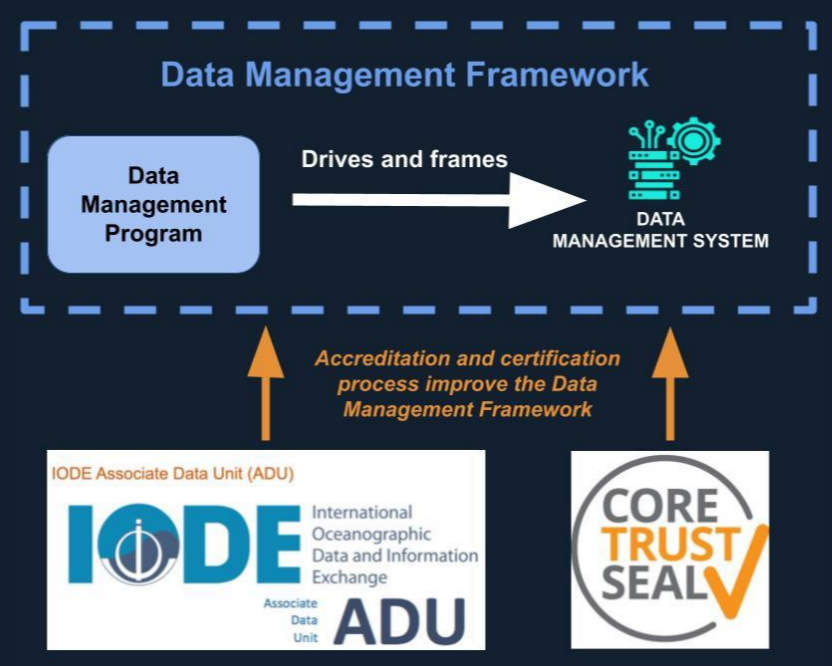
04 4.3 DevOps methodologies



SOCIB is upgrading the whole IT & software system focusing on adopting DevOps methodology and the Kubernetes tool. This is a major improvement that will be key for enhancing both the **Data Management Program** and the **Products & Services** that SOCIB offers.

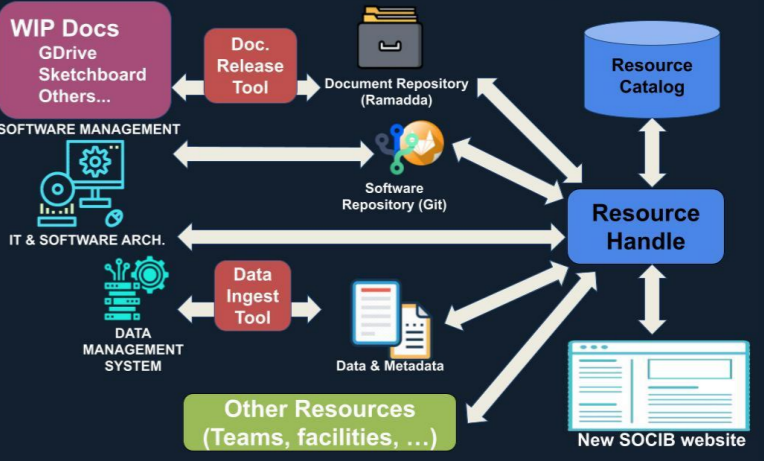
05 Supporting activities

The SOCIB Data Infrastructure will benefit from 2 certifications: check out IMDIS 2021 poster 127 "How certification process is helping SOCIB to improve Data Quality Management"



The Data Catalog portal, which includes embedded metadata in JSON-LD format (compliant with the DataCite metadata scheme), allowing Google Dataset to interoperate and discover the SOCIB data products (ie. the "I" of FAIR).

06 Outcomes: downstream services



A linked data resource catalog to be integrated in the new SOCIB corporate website. There are also future plans to implement a linked data node.