FAIR Big Data Management exploiting Cloud Computing resources (DataClouds)

Fernando Aguilar
Instituto de Física de Cantabria (UC-CSIC)
aguilarf@ifca.unican.es

Introduction

The growing number of **environmental data sources** provides new possibilities to understand the complex Earth system. Remote sensing data provided by satellites, historical and forecasting data provided by meteorological agencies or in-situ sensors gives a **lot of information** to **monitor** an **ecosystem**. However, they are very **heterogeneous** in terms of file formats and access and they can be tagged as **Big Data** in terms of volume. To enable the exploitation in a interoperable way, the produced data need to address the four **FAIR** principles.

Use Case

LifeWatch ERIC aims to address a complete solution to monitor and model the **water quality** by **integrating** different and heterogeneous **data sources** (satellite data, real-time monitoring systems, repositories, and meteorological data). Supported by **DataCloud** solutions, the proposed architecture provides a set of services for the user to **manage** the **D**ata **L**ife **C**ycle completely with several features: **automatic metadata** attachment and discovery, **data gathering** selecting locations and dates, different types of **analysis** and **curation**, **PID** minting and preservation rules definition.

DataClouds

Cloud Computing-based services and tools to be integrated in systems to offer a complete environment for Data Life Cycle management and access to different types of resources.

Develop a **data computing platform**, targeting various communities and deployable on hybrid Cloud infrastructures
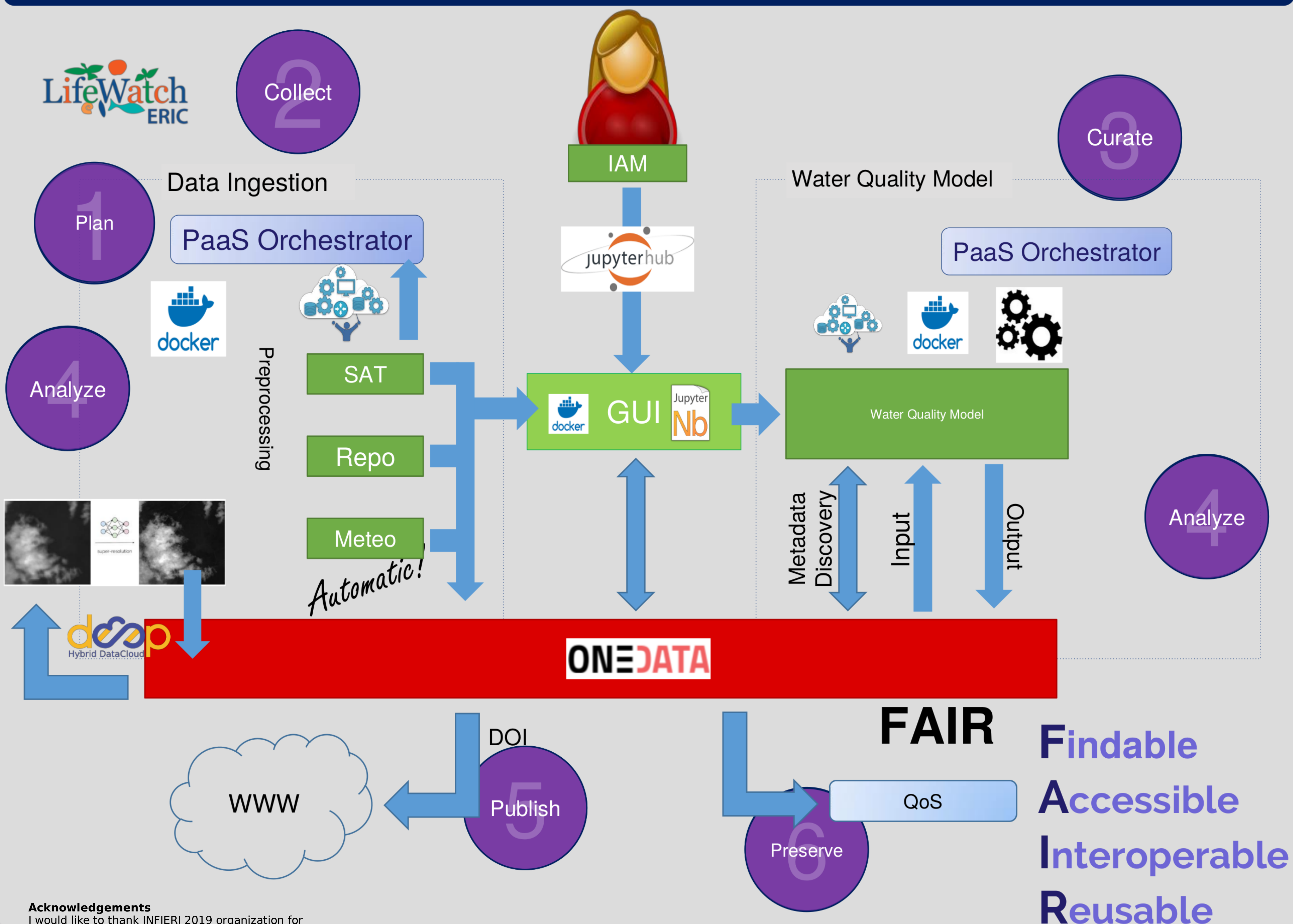
INDIGO - DataCloud

Promote the use of **intensive computing** services and the support by the corresponding e-Infrastructure providers

Dealing with **extreme-large** and heterogeneous **datasets** on the cloud.

Environmental Data Life Cycle

FAIR

Findable
Accessible
Interoperable
Reusable