# ENEA and
# the first Software Heritage Mirror

" software source code is a unique form of knowledge which is designed to be understood by a human being, the developer, and at the same time is easily converted into machine executable form"

[R.Di Cosmo, S.Zacchiroli, iPRES 2017: 14th Inter. Conf. on Digital Preservation, Kyoto]

*Annual symposium & summit on Software Heritage  2023*

*UNESCO Headquarters, Paris*

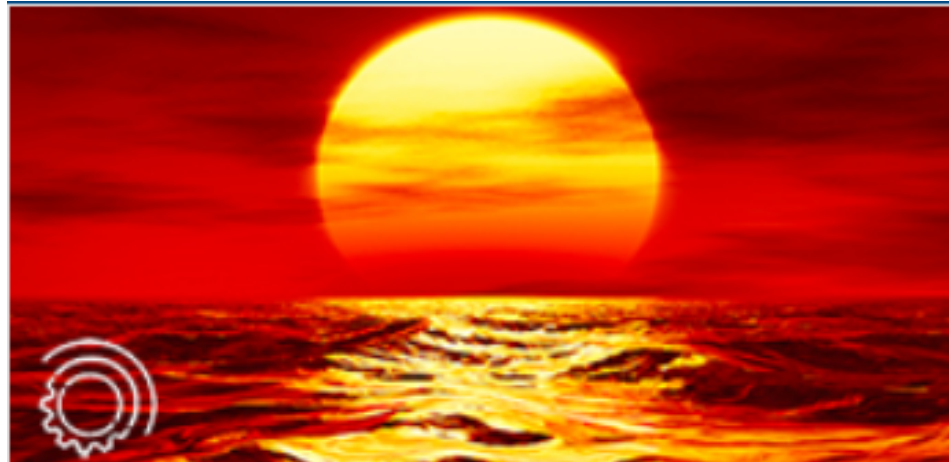Giovanni Ponti / ENEA-TERIN-ICT /                    February 7th 2023

# ENEA   National Agency for New Technologies, Energy and Sustainable Economic Development



ENVIRONMENT and SUSTAINABILITY
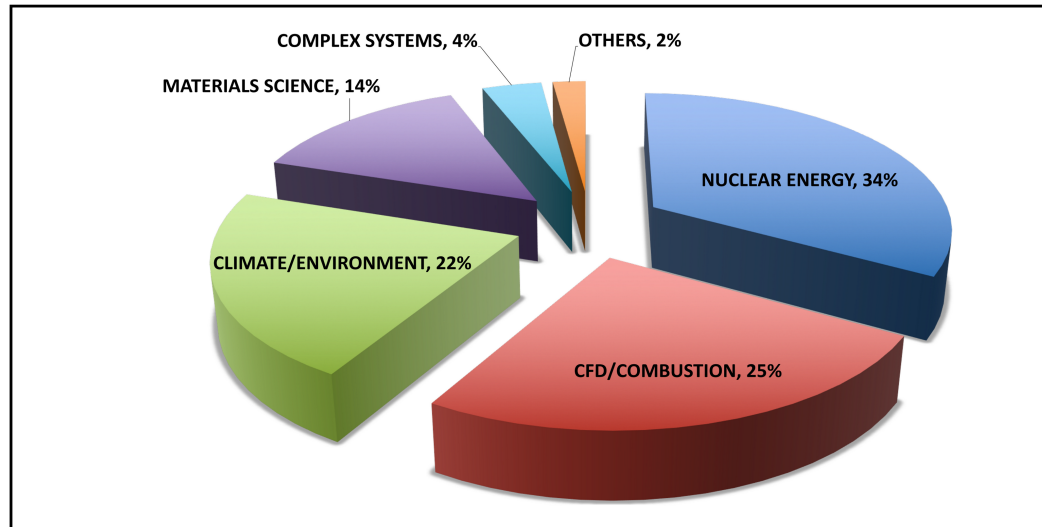


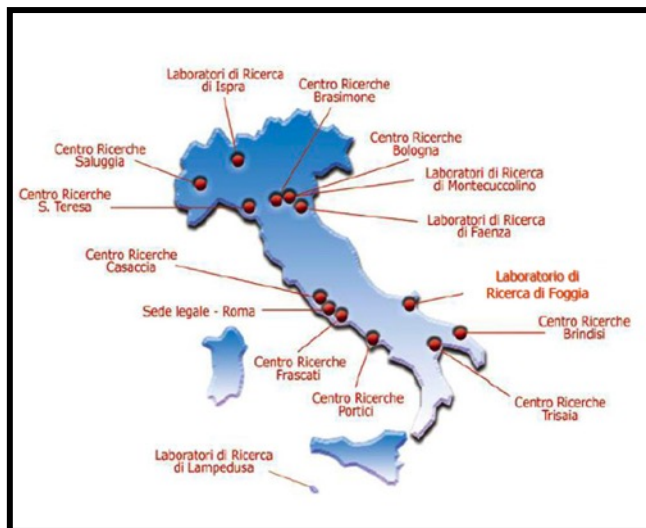ENERGY TECHNOLOGY



ENERGY EFFICIENCY
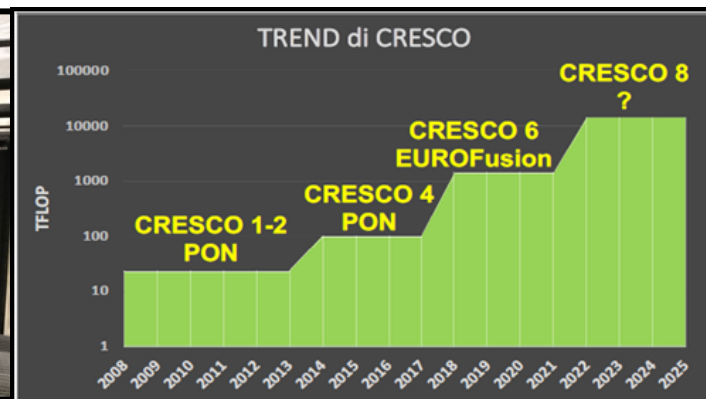


FUSION and NUCLEAR SAFETY

# ENEA - since 1960 a long tradition in computational science and HPC infrastructures





*CRESCO6 – Italian Tier1 HPC*

- *Cores 20832, Nodes 434*
- *Peak performance 1.4 PF*

# ENEA and the mirror of Software Heritage Archive



**Software Heritage**
A revolutionary infrastructure for software source code

**+**

**CRESCO6**

**ENEA
since 1960
a long tradition in
computational science and
code development**

**=**

**INRIA - ENEA Cooperation Agreement
a great adventure has started**

# ENEA: the [experimental] mirror

## A mirror for

- improving project sustainability

- reducing data loss risk

- experimeting new solutions

## based on 8 nodes

- 3 Dockers swarm nodes

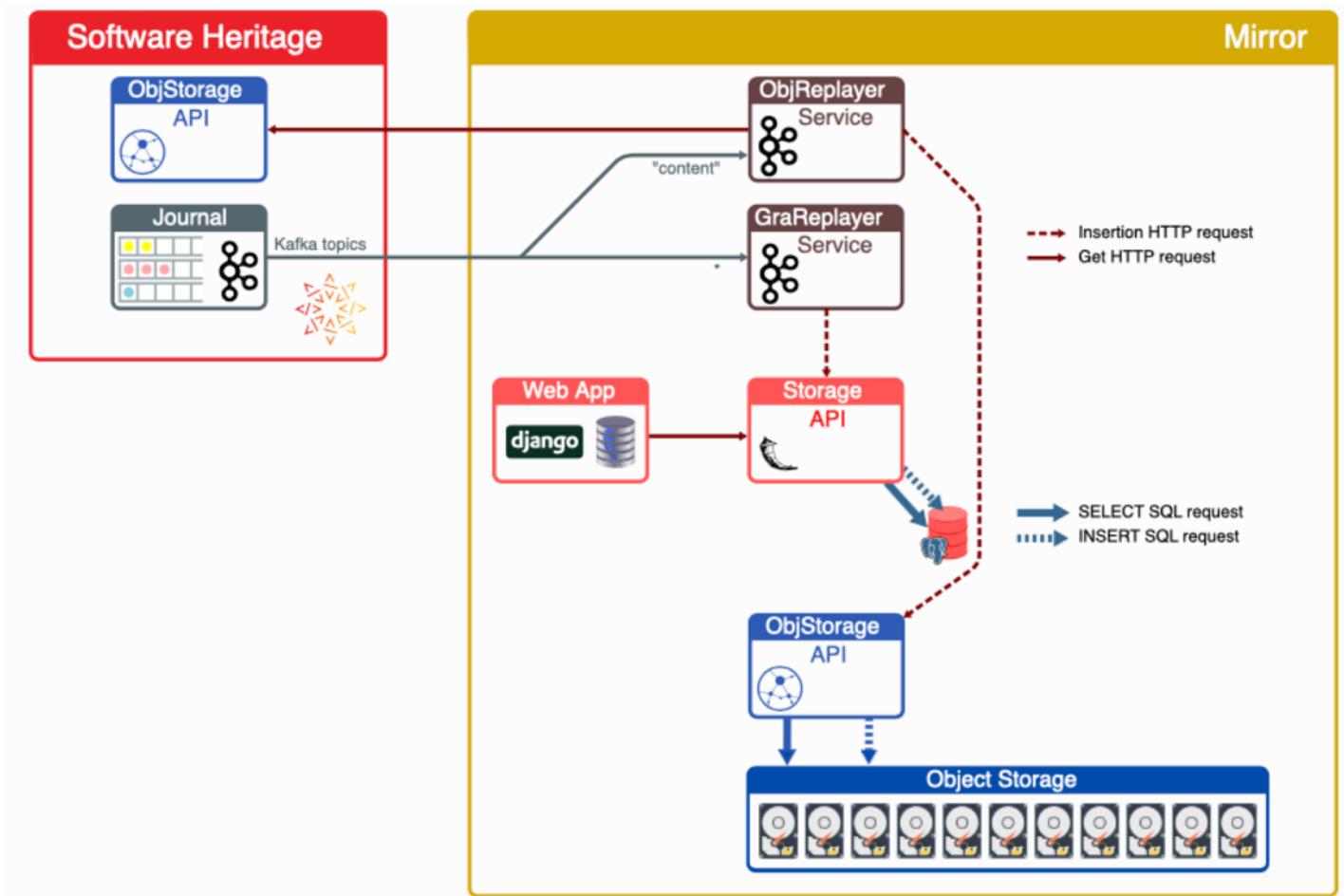- 1 DB server  [34 TB]

- 4 storage servers [755 TB + 1PB Q1 ]

## specific goal: testing SEAWEEDFS

- highly scalable distributed file system

- to store billions of files
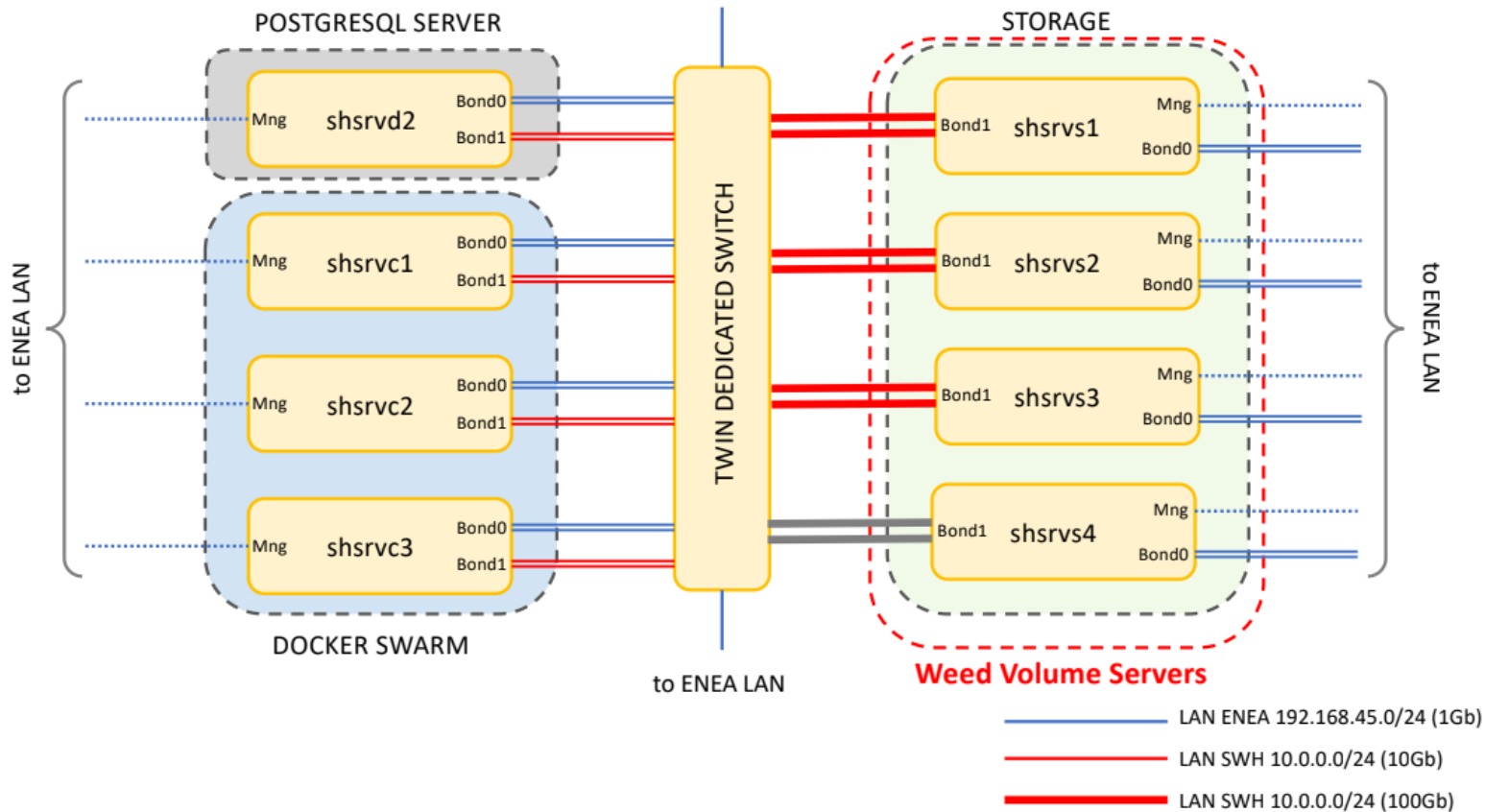
- to serve the files fast

## SWH mirror components

## SWH ENEA Mirror Architecture
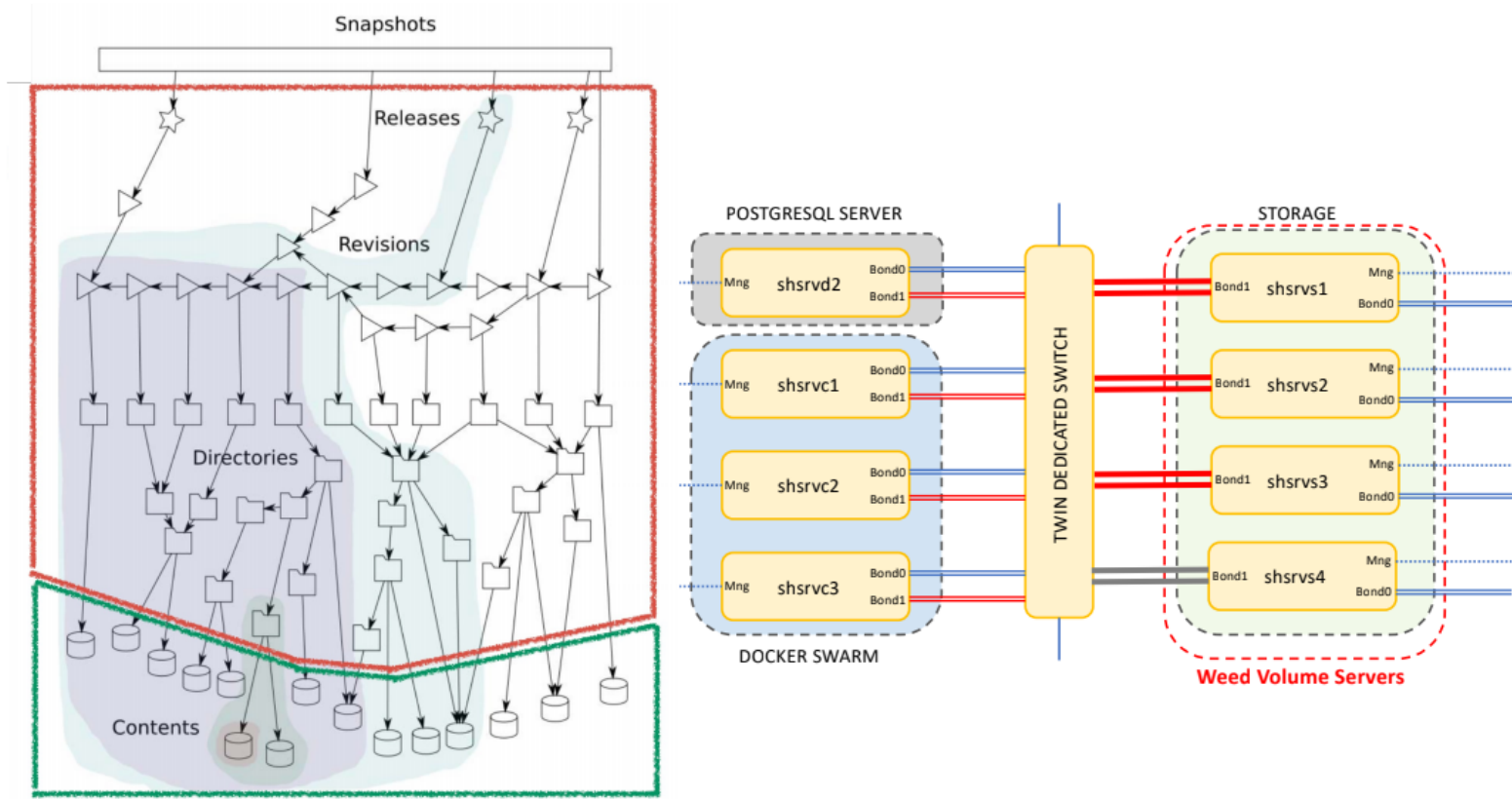
# ENEA: the [experimental] mirror

**S**

## SWH ENEA Mirror Architecture

## SWH ENEA Mirror Architecture

## SWH ENEA Mirror Architecture



MERKLE DIRECT ACYCLIC GRAPH
**STORAGE** **TRANSFER COMPLETE**

**OBJECT STORAGE**
**TRANSFER IN PROGRESS**
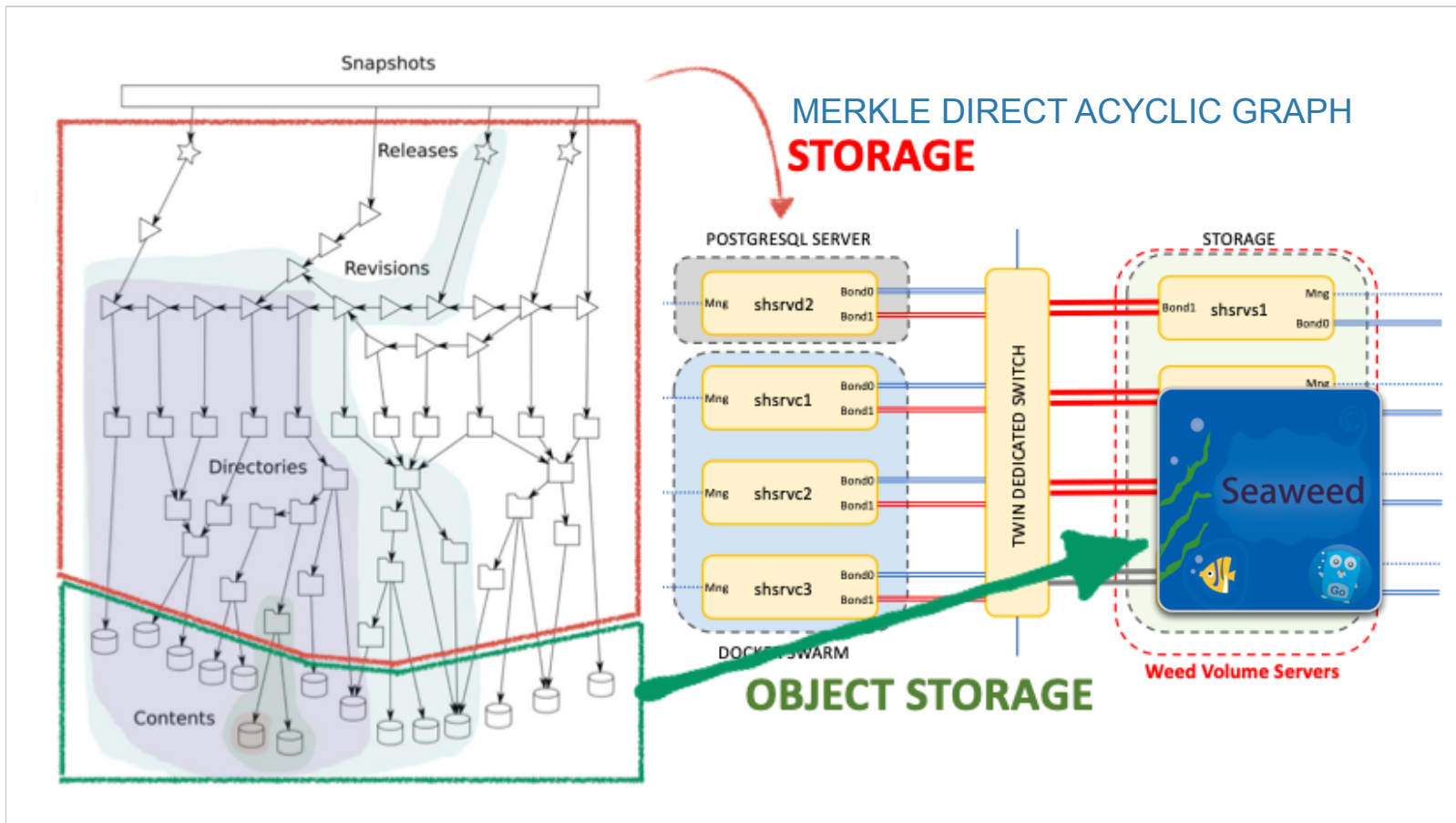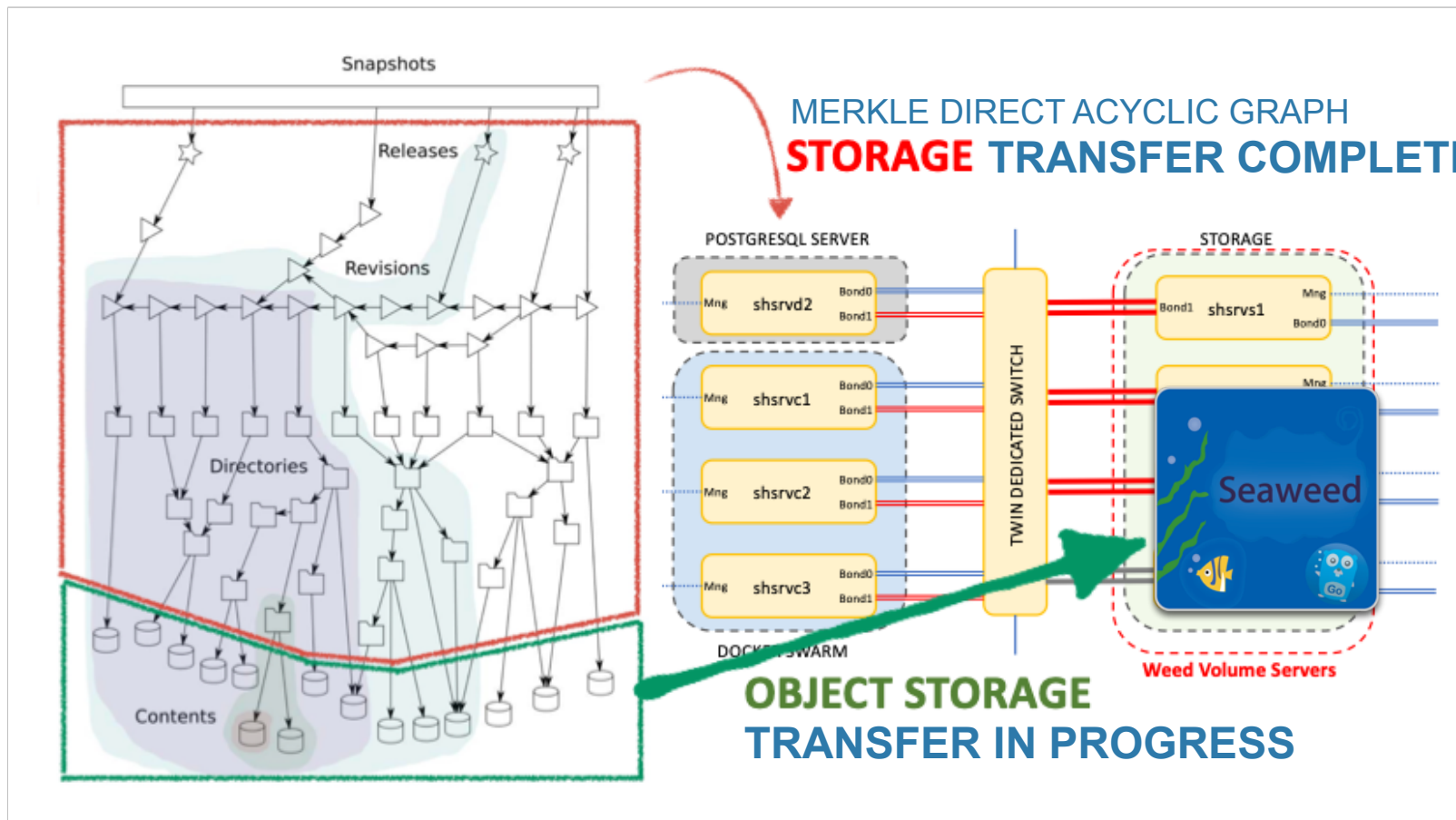
**THE MIRROR WILL BE IN PRODUCTION ON JUNE 2023**

# ENEA & SH: the mirror as a research infrastructure

**The Software Heritage's mirror as a starting point for research collaboration to:**

- experiment the application of Machine Learning and AI for software development and programming languages practice
  - Big code as Big data

- develop new solutions for Big Archival Storage

- tackling the problem of "resilient digital heritage" against catastrophic loss of data and obsolescence of archiving tools

- involve young researchers in a relatively new field of investigation (Big Code/Big Data/BigStorage)

# ENEA's SH mirror
## within the Big Data Technopole in Bologna

an area of 100.000 m2 - 1 billion € of investments for HPC - BigData - AI - QC



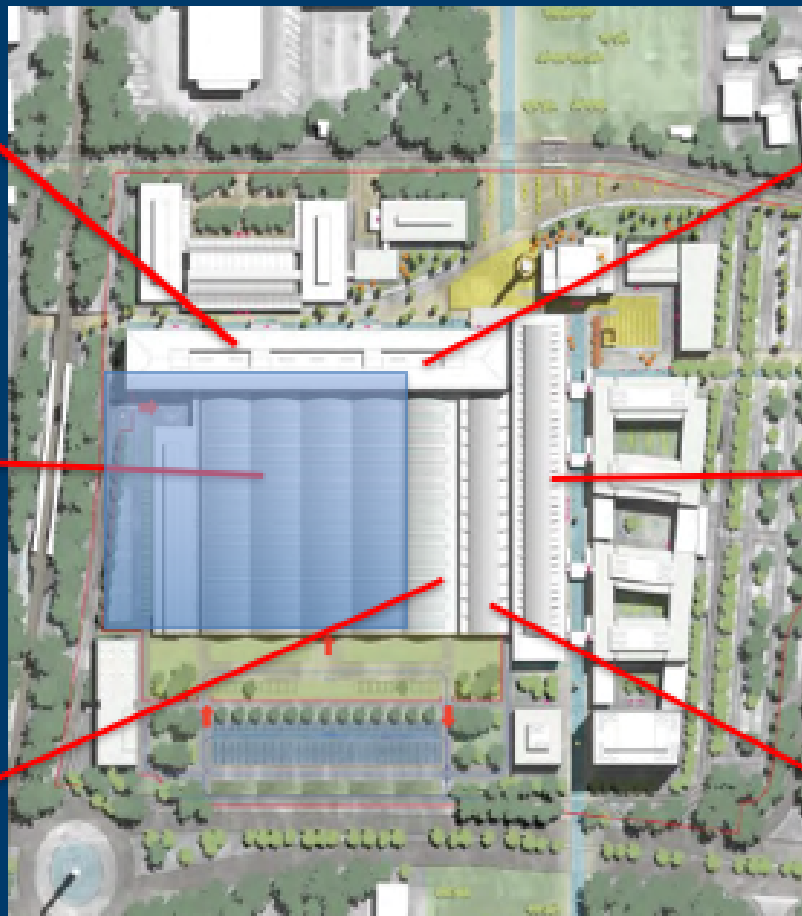**iFAB** International Foundation Big Data and AI for Human Development

**ICSC** Foundation *Research Center in HPC, Big Data, Quantum Computing*

**ECMWF** Data centre

**ENEA** Research Center

**INFN** Data centre

**CINECA** *LEONARDO* *supercomputer*

# Emilia Romagna Region and the SH Mirror



The warmly welcome of Stefano Bonaccini the President of the Emilia Romagna Region to the Software Heritage's Mirror, entrusted to a post on Twitter.

*"Here is the world software archive.*
*Our data valley to serve people and to facing future challenges"*

Giovanni Ponti

giovanni.ponti@enea.it