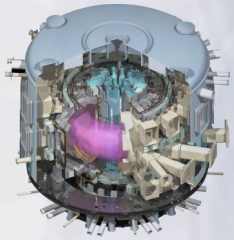
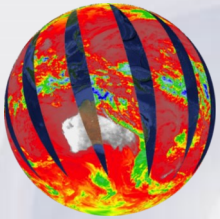
# CO-DESIGN

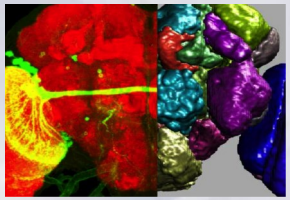
The SAGE Project deals with Extreme Computing I/O and Data Intensive Science use cases, indicative of trends as we move towards Exascale. The use cases signify cutting edge problems in a variety of domains including physics, space sciences, meteorology, genetics and biology - which will benefit tremendously from having the SAGE Percipient Storage platform. The SAGE platform will ultimately lead to more efficient science and HPDA (High Performance Data Analytics) for industry and commercial applications, contributing to societal progress .

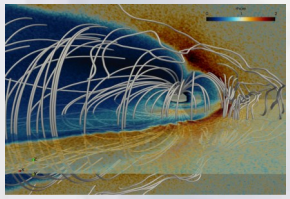
**Application Examples**

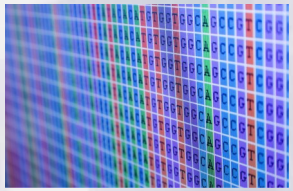
- Visualization
- Satellite Data processing
- Space Weather
- Bio-Informatics
- Nuclear Fusion
- Synchrotron Experiments











# SAGE AT A GLANCE

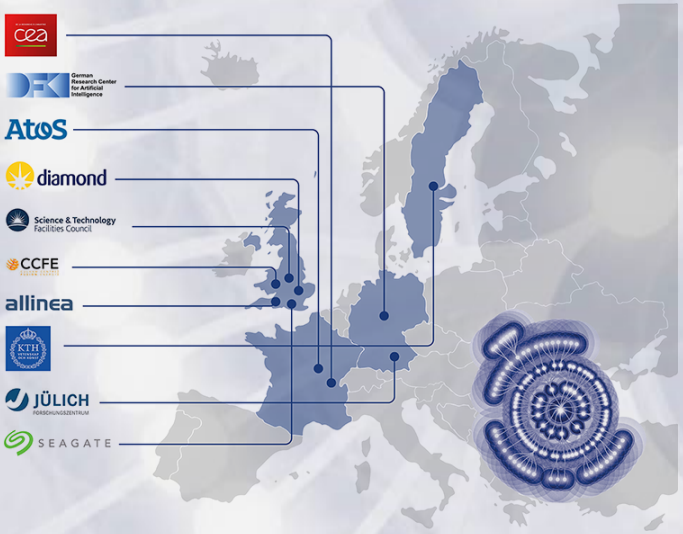
Worldwide data volumes are exploding and islands of storage remote from compute will not scale. SAGE will demonstrate the first instance of intelligent data storage, uniting data processing and storage as two sides of the same rich computational model. This will enable sophisticated, intention-aware data processing to be integrated within a storage systems infrastructure, combined with the potential for Exabyte scale deployment in future generations of extreme scale HPC systems.

**BUDGET:** € 7.9 million

**TIME PERIOD:** Sep 2015-Sep 2018

**CONSORTIUM:**

- Co-ordinator: Seagate
- 10 Partners
- 4 European Countries




SAGE

# Percipient StorAGE for Exascale Data Centric Computing

Web: www.sagestorage.eu
Email: Info@sagestorage.eu
Twitter: @Sagestorage

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement nº 671500

# RESEARCH

The SAGE project tasks itself with building a data centric infrastructure for handling extreme data in the Exascale/Exabyte era centred on a storage-oriented solution, 'Percipient Storage'. SAGE will work with I/O Intensive codes and data centric applications that enable the derivation of scientific or business insights from very large data sets. SAGE incorporates cutting edge innovations in an ecosystem of critical supporting research into highly distributed storage platforms, data analytics frameworks, data management tools and programming models

The 8 fields of research are shown diagrammatically in the image below


# SAGE ECOSYSTEM


**Extreme Data Management**
The highly tiered SAGE platform that can be exploited by data management tools and utilities (eg: Hierarchical Storage Management & Data Integrity checking tools). Adaptations and evolutions of existing management tools will be researched in SAGE along with backwards compatibility with existing data access interfaces.

**Extreme Data Analytics**
SAGE will research and Integrate advanced data analytic techniques exploiting the computation capability associated with tiered data and non-volatile memory available with Percipient Storage based around Apace Flink batch and stream data processing tools.

**Programming Methods**
SAGE will look into advancements in programming models for Percipient Storage as well as integration with existing programming models .

- MPI: Offloading computational kernels to the different tiers of SAGE I/O
- PGAS: Integrating the NVRAM tier to the virtual memory of the compute nodes

**Optimisation Tools**
SAGE will address the need of sophisticated tools that can expose the true cost of I/O and storage as well as providing debugging capabilities for computational offload to storage.

# PERCIPIENT STORAGE

We propose an advanced object based storage solution, termed *Percipient Storage,* with a very flexible new API enabling applications to achieve Exascale I/O loads exploiting deep I/O hierarchies. The solution will have the capability to run computations on data from any tier – with a homogenous view of data throughout the stack. A high-level schematic architecture for SAGE is depicted below.

The SAGE architecture reflects the need for reducing data movement in order to improve energy efficiency, as well as the technology trend towards new non-volatile memory technologies.