Helmholtz Analytics Framework (HAF)

Proposal for a Pilot Project to the IuVF – June 12, 2017

Coordinators: Thomas Lippert - FZJ, Achim Streit - KIT

HELMHOLTZ | GEMEINSCHAFT

HAF – Goals and Objectives

Scientific big data analytics
- methodologies and tools for problems of highest data and compute complexity

Helmholtz Analytics Framework
- foster data science in Helmholtz
- develop and exploit the Helmholtz Data Federation (HDF)

Use case driven co-design between
- domain scientists
- data experts
- infrastructure professionals

Create data analytics techniques
- in a systematic manner
- domain-specific as well as generalizable and standardized

HELMHOLTZ | GEMEINSCHAFT

HAF – Contributing Centres

- Deutsches Elektronen Synchrotron

- Deutsches Krebsforschungszentrum

- Deutsches Zentrum für Luft- und Raumfahrt

- Forschungszentrum Jülich (Coord.)

- Helmholtz Zentrum München

- Karlsruhe Institute for Technology (Coord.)

HAF – Use Cases Addressing Key Areas in Helmholtz

Earth System Modelling – Energy
- Terrestrial monitoring and forecasting
- Cloud and solar power prediction
- Stratospheric impact on surface climate

Health – Neuroscience
- High-throughput image-based cohort phenotyping
- Multi-scale/area interaction in cortical networks

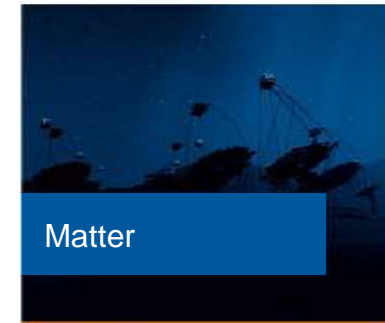**Space and Transport – Aeronautics**
- Virtual aircraft

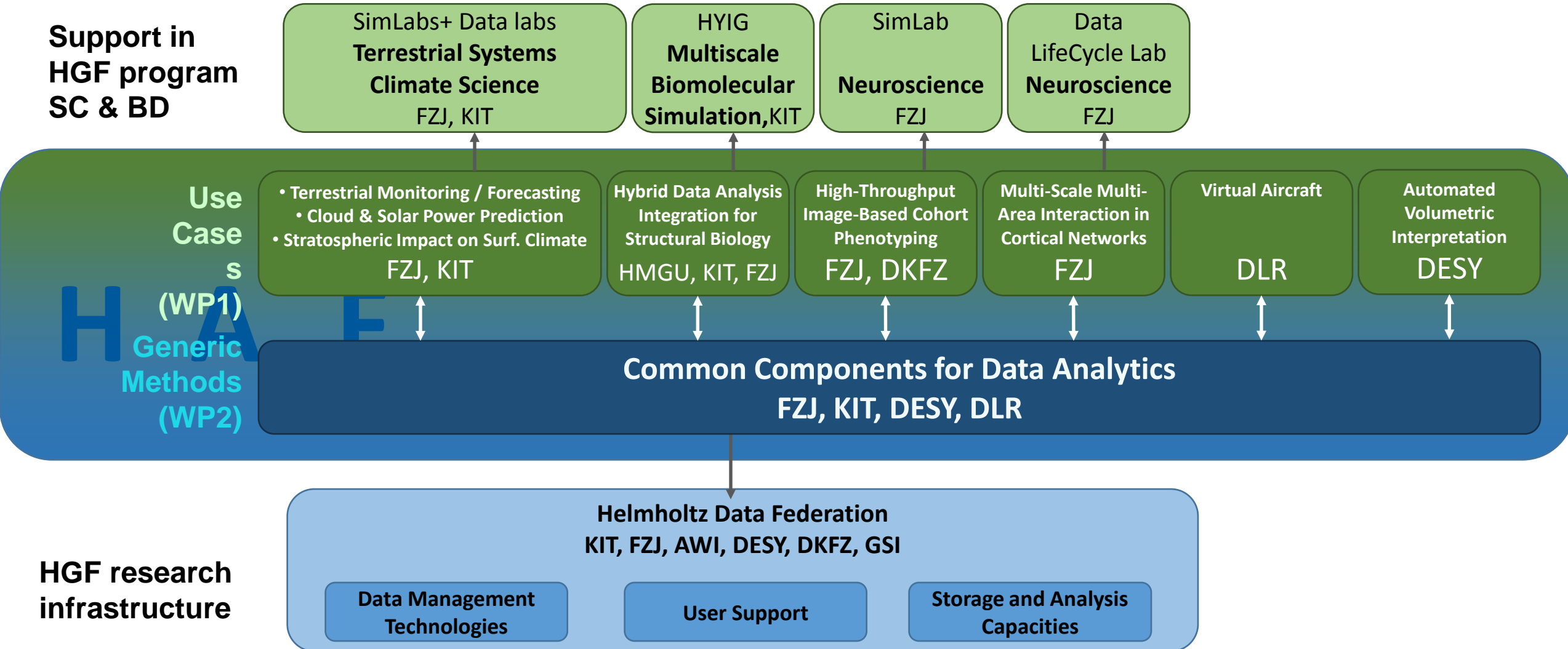**Matter – Research with Photons**
- Automated volumetric interpretation

Key Technologies – Information
- Big data analytics methods

Energy

Earth and Environment

Health

Aeronautics, Space and Transport

Matter

Key Technologies

HELMHOLTZ | GEMEINSCHAFT

HAF – Integration in Helmholtz Activities

Support in HGF program SC & BD

SimLabs+ Data labs **Terrestrial Systems Climate Science** FZJ, KIT	HYIG **Multiscale Biomolecular Simulation,** KIT	SimLab **Neuroscience** FZJ	Data LifeCycle Lab **Neuroscience** FZJ

H A F

Use Cases (WP1)

• Terrestrial Monitoring / Forecasting • Cloud & Solar Power Prediction • Stratospheric Impact on Surf. Climate FZJ, KIT	Hybrid Data Analysis Integration for Structural Biology HMGU, KIT, FZJ	High-Throughput Image-Based Cohort Phenotyping FZJ, DKFZ	Multi-Scale Multi-Area Interaction in Cortical Networks FZJ	Virtual Aircraft DLR	Automated Volumetric Interpretation DESY

Generic Methods (WP2)

**Common Components for Data Analytics
FZJ, KIT, DESY, DLR**

HGF research infrastructure

**Helmholtz Data Federation
KIT, FZJ, AWI, DESY, DKFZ, GSI**

Data Management Technologies	**User Support**	**Storage and Analysis Capacities**

HAF – Uniqueness

Foster trans-disciplinarity by strict co-design and bring together
– domain scientists
– data experts
– infrastructure professionals

Exemplary use cases chosen
– following the principle of scientific excellence
– as seed activities in data sciences in Helmholtz mission critical sciences
– to be generalized towards all Helmholtz fields
– to demonstrate added-value in scientific discovery

Boosting Helmholtz Data Federation in
– data management technologies: retrieval – assimilation – processing - federation
– world class user support
– giant federated storage and analysis capacities

HELMHOLTZ | GEMEINSCHAFT

HAF – Use Case "Earth System Modelling"

Stefan Kollet
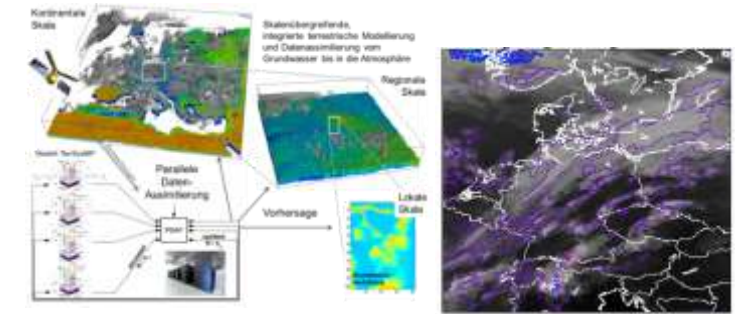
We improve

- state and predictions of the terrestrial water and energy cycle from groundwater to atmosphere;
- assessment and forecasts of radiative cloud properties / photovoltaics;
- forecasts of seasonal evolution of Arctic weather and climate focused on composition-circulation interactions.

We are concerned with

- water management, natural hazards warning (floods, droughts), and global change impact assessments;
- cloud formation simulations;
- the reliability of earth system forecasts and projections.

We use highly advanced

- massively parallel data assimilation technologies to merge observations with terrestrial models;
- signal processing and advanced parallel workflows for large ensemble screening;
- novel combination of Big Data Analytics (SVM) with satellite/cloud imagery, super-ensemble particle filtering.

We contribute

- generic data assimilation technologies for high-dimensional, non-linear inverse problems;
- prediction technologies for e.g. photovoltaic power, air quality, and precipitation;
- improved generalizable methodologies for big data analytics;
- significantly to the "Zukunftsthema" Advanced Earth System Modelling in HGF.

HAF – Use Case "Structural Biology"

Alexander Schug

We enhance

biomolecular structure determination to understand their function

We provide

new insights for essential structures inaccessible to single techniques

We use

- hybrid data management & integration via common analysis methods
- data from (sparse) NMR, X-ray, cryo-EM, SAXS/ SANS, genomic libraries, biomolecular simulations

We leverage

full interdisciplinary potential in Helmholtz by synergetic fusion of data and methods with relevance for basic & applied life sciences
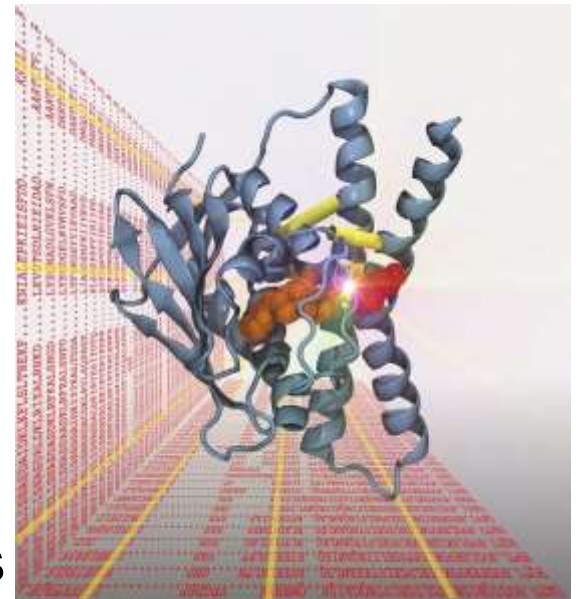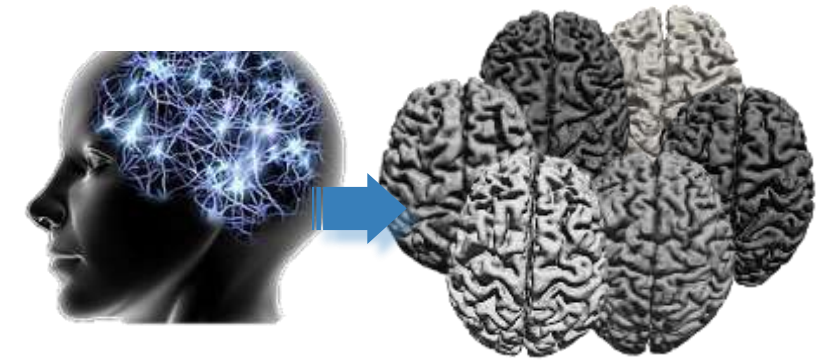
HAF – Use Case "Neuroscience"

Svenja Caspers

We extract

characteristic profiles of neuronal network activity and neuroimaging biomarkers in large population-based cohorts, linked to performance, environmental and genetic data;

We enhance

the deep understanding of inter-individual variability and synchronized network activity in the human brain as a basis for delimiting disease-related alterations;

We implement

optimized high-throughput HPC workflows for analyzing neuron ensembles to ten thousands of subjects, enabling machine learning and data mining on high-dimensional multi-level data;

We provide

HPC workflows for neuroimaging and neuronal network analytics as workbench for basic and clinical neuroscience users within HGF and beyond, expandable to other imaging domains.

HAF – Use Case "Research with Photons"

We focus on

photon sources as multidisciplinary instruments (biology, chemistry, medcine, material science, art, etc.)
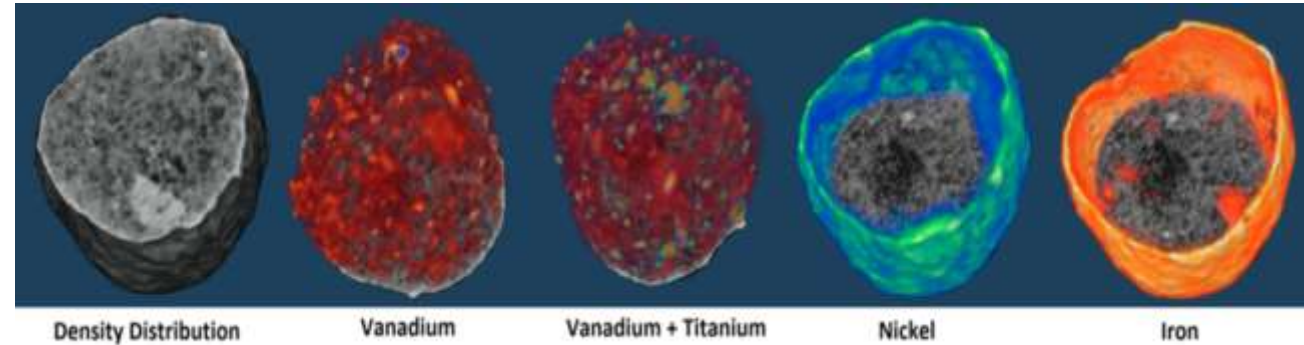
We provide

a data management and analytics framework for (often small) groups of scientists

We jointly build and validate

highly sophisticated data management solutions and analysis methods

We transfer

common solutions to other large (photon-) sources and increase scientific output of users

Density Distribution   Vanadium   Vanadium + Titanium   Nickel   Iron

HAF – Common Tools for Data Analytics

Björn Hagemeier

We identify

methodical commonalities from use cases and develop generic solution packages.

We work for

highest code quality and excellent quality of scientific results, and we spread profound methodical competence.

We will perform

method swaps for benchmarking, translate excellent methods between the fields and roll them out on the HDF.

We aim at

cross-fertilization among use cases, across disciplines and dissemination all over HGF.

Summary

The **Helmholtz Analytics Framework** will become a **new asset** in the Helmholtz Association which is of **strategic relevance.**

Use case-based development guarantees **excellence**, the extraction, development and deployment of **common tools** exploits **synergy** leveraged by a **critical mass** of users.

The integration of use cases and tool development into the future **programme oriented funding** guarantees **sustainability** and **broader perspectives.**

Pushing the **Helmholtz Data Federation** by the **HAF**, a nation- and Europe-wide **leadership** position in data analytics can be achieved.

With **HAF**+HDF Helmholtz can provide a scientific **seed activity and a nucleus** for the "**Nationale Forschungsdateninfrastruktur**" (NFDI).