# HELMHOLTZ

#### RESEARCH FOR GRAND CHALLENGES





### Federated data storage for Helmholtz Research & Friends

<u>Uwe Jandt</u>, Paul Millar, Tigran Mkrtchyan (DESY) on behalf of HIFIS and dCache teams



National and International Trends in Research Storage at Scale

Nov 12, 2020

www.helmholtz.de hifis.net



# Why HIFIS?

#### **HELMHOLTZ** RESEARCH FOR GRAND CHALLENGES

- Helmholtz aims for joint research & information environment for all Research Fields



# Why HIFIS?



- Helmholtz aims for **joint research & information environment** for all Research Fields
  - High performance + collaborative services
  - Shall connect all centres
    - ….and their world-wide collaboration partners!
  - Secure, simple access and easy-to-use
- Widely establish best-practices for development + use of research software:
  - high level of <u>quality</u>, visibility and sustainability





# Why HIFIS?



- Helmholtz aims for joint research & information environment for all Research Fields
  - High performance + collaborative services
  - > Shall connect all centre This year's pandemic
    - > ....and their world still and again ヽ(♂\_♂)ノ collaboration partners!
    - is a powerful reminder of Secure, simple access and easy-to-use
- Widely establish best-pr how important collaborative development-use of research scalable IT services are.
  - high level of <u>quality</u>, visibility and sustainability

# Scientific Advisory Board (SAB) – Feedback



#### HIFIS SAB Meeting + Report (May'20)

- Open discussions and written feedback report
- High visibility  $\rightarrow$  expert feedback

#### Major comments

- Integration with other platforms + internationally
- Integration into the proposal process
- Support scientists + tackle IT knowledge gap
- Appreciated HIFIS is a platform, not a project

#### **18 members**

from EU, USA, AUS:

Ari Asmi Rosa M Badia Magchiel Bijsterbosch Michael Brünig **Isabel Campos** Mirjam van Daalen Tiziana Ferrari Andy Götz **Christian Grimm** Marc Heron Jörg Herrmann Neil Chue Hong Christine Kirkpatrick Rupert Lück Pierre Etienne Macchi Wolfgang E. Nagel Davide Salomoni **Bruno Weikl** 

University of Helsinki BSC SURF University of Queensland CSIC PSI EGI ESRF DFN DLS MPG SSI NDS, SDSC, OSN EMBL **IN2P3 CNRS** TU Dresden INFN FhG

# **Getting domain scientists involved**

### **HIFIS conducted surveys throughout Helmholtz**

- Covering all research fields, all centres, scientists/users and ICT provider!
- <u>Software Services survey:</u>

Approximately **1000 replies** from nearly **all centers** 

<u>Cloud Service Survey:</u>

**18+1 centres visited**; Feedback from more than **110 scientists** + ca. 100 IT experts

Resulted in Initial Service Portfolio

management programming regular organi tool scientific know COde training cI COURSe language group create dataleast project good development working good development software development deve much knowledge infrastructure develop time Gitlah publication people



# **Cloud Services**





### <u>Done</u>

- Service Selection and Priorization finished
  - Process reviewed by Fed Board
  - Strong focus on benefit for researchers, and research platforms + International compatibility
  - ➢ Full list: <u>hifis.net → news</u>
- Top priority: Urgent interest in collaboration services and common access infrastructure

## <u>Next</u>

- 1<sup>st</sup> "MVP" Version of Helmholtz Cloud Portal
- Start / continue Integration of first services

Storage (HDF)			
Storage (HDE)		indifferent	
locoroge (mor)	Storage for HPC,	FZJ	
Storage (HDF)	Storage for HPC, general	DESY	
Storage (HDF)	Collaboration and Energy	КІТ	
Communitystorage		indifferen	
Community/ Science	ce Services		
		indifferen	
GitHub	GitHub		
GitLab, GitLab-Cl		indifferen	
GitLab, GitLab-Cl		HZDR	
louis out o	Isontware-code online,	Let T	
5.ct lude	Provile		

# **Backbone Services: Getting everybody connected**

### **Common Helmholtz Authentication and Authorization Infrastructure (AAI)**

- Enabling common access to collaborative and scientific services through credentials of the home institution
- Helmholtz AAI ready and in use, compatible to AARC
- Precondition to build Helmholtz Cloud
  - Various technologies: OIDC, SAML, SSH, LDAP
  - Group / VO Management
  - Secure SSH key management
  - 2FA
  - Deprovisioning (not yet in all modules)
- Documentation / Policies available:
  - https://hifis.net/doc/backbone-aai
  - https://hifis.net/mission/publications





## dCache: Areas of Interests





# dCache Motivation and Design

- Data never fits into a single disk/host
  - Multiple data servers
  - Tape off-load
- Number of clients only grows
- Control over HW/OS selection
  - Maximize local expertise
  - More flexibility on HW offers

- Single-rooted namespace, distributed data
- Clients talk to namespace only for metadata operations
- Bandwidth and capacity grows with the number of data servers
- Standard access protocols
- The same data available with any protocol, independent from Authentication schema



# **Distributed Data Management**

- dCache is designed to federate storage
- Flexible data placement policies allows to get most out of local resources
- HA-capability reduces operational overhead to achieve desired SLO
- Cache-only sites can be integrated with HSM-enabled sites into single deployment
- In-transit data protection provide data privacy in distributed environment

# **5** Countries One instance Uni of Bergen 150 Uni of Oslo National Supercomput er Center dCache head node Nordu Net

# dCache Identity Management



- gPlazma: a pluggable authentication and identity management system.
- Provides flexible framework for identity management, based on Pluggable Authentication Module (PAM).
- dCache comes with plugins to support common approaches: Kerberos, htpassword, OpenID-Connect, X.509, LDAP, ...
  - OpenID-Connect integration provides Helmholtz users with seamless access.
- OAuth2 supports delegation, allowing storage access to be delegated from user to an agent working on their behalf.



## dCache Integration with Transfer Service

- CERN File Transfer Service (FTS):
  - Deployed throughout the world for managing transfers between different storage endpoints.
- dCache includes integration with FTS where information provided by FTS is recorded
- Active party in third party copy (WIP!):
  - Third party can commission transfers between source and destination
  - Data is transferred **directly** between endpoints w/o third party
  - One endpoint needs to understand TPC-COPY extension (active party)
    - dCache
  - The other endpoint needs to enable PUT or GET requests for files (passive party)
    - Available extensions for Apache webserver (OIDC, write with specific user ACLs, ...)





# dCache Event Generation



- dCache trigger actions on data state changes.
  - Clients may consume events to react to changes
- Automated work-flows integration
  - metadata extraction
  - catalogue update
  - replication
  - data processing



- Kafka provides notification of uploads, deletions, media locality changes.
- Inotify-over-SSE provides notification on namespace changes.



# **Roadmap (Selection) / Next Steps**



- End Oct'20: Roadmap for Integration of Initial Cloud Services
- Nov'20: Integrate the Software Management Platform into the Helmholtz Cloud
- Jan'21: Set-up of Helmholtz Backbone Connections
- March'21: Reporting for the first two years of HIFIS: Including User-oriented KPIs
- April'21: Start evaluation of further Cloud Services
- End'21: Cloud Service Process Manual

	$\succ$	https://hifis.net/roadmap	)
--	---------	---------------------------	---

ROADMAP FOR THE INTEGRATION OF THE	E INITIAL SERVICES		<b>Cloud Services</b>
Having finished the list of the initial service po	rtfolio, the integration of	the services will start.	. As different
services need different time for their integration	on, a specific roadmap w	ili be established.	
- November 2020			
INTEGRATE THE SOFTWARE MANAGEMEN CLOUD	IT PLATFORM INTO TH	IE HELMHOLTZ	Software Services
The software management platform is planned technical platform is available. Continuous Inte	I to be integrated into the egration will be part of th	e Helmholtz Cloud as s ne software manageme	soon as the ent platform from
the very beginning.			
RECIPES FOR A SCALABLE CI SOLUTION	rement platform is a solu	tion for continuous int	egration (CI)
Reusable Ansible recipes for a GitLab CI infras	structure will be provided	I. The Ansible role is de	eveloped here.
	2021		
– January 2021			
SET-LID OF HELMHOLTZ DACKDONE CONN	IECTIONS		Sackbone Services





- Very positive and intensive feedback from surveys conducted by HIFIS.
  - Next round of survey on usage by scientists
- Highest interest in and outside of Helmholtz for commonly accessible collaboration tools, high performance IT and sustainable software development.
  - Numerous pilot projects throughout 2020
- Currently: Integration of initial set of Helmholtz Cloud IT services.
  - dCache: a science agnostic system which natively integrates into Helmholtz federated cloud environment.



#### https://hifis.net

- Subscribe for HIFIS announcement letter!
- General queries to: <u>support@hifis.net</u>
- Coordinators:
  <u>office@hifis.net</u>

