Deployment of a flexible and secure computing environment with OpenStack

CERN OpenStack Day May 27, 2019

Jani Heikkinen

sciCORE, University of Basel

Stig Telfer

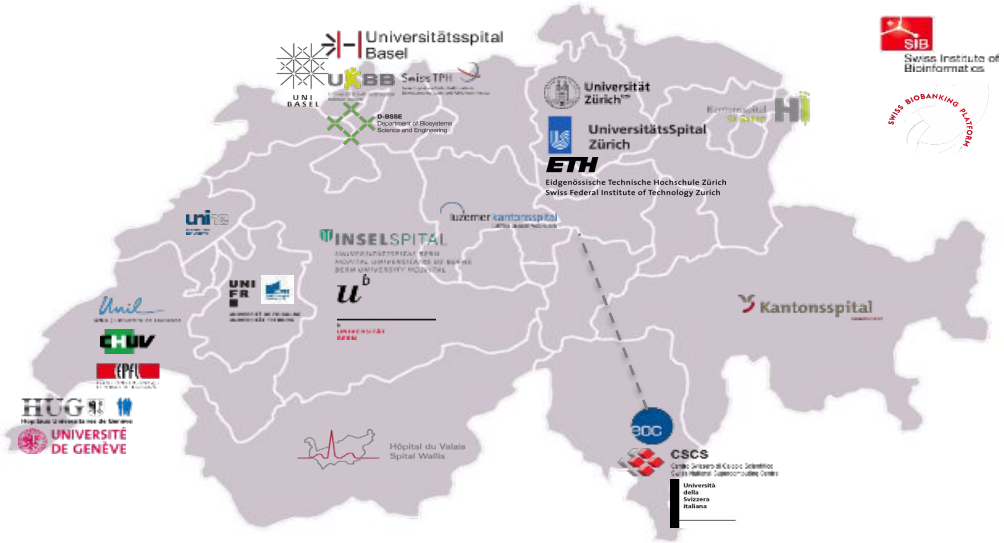
StackHPC

Main responsibilities in short:

- Providing and maintaining high-performance computing (HPC) infrastructures and services for researchers in the University of Basel
- Offer an up-to-date software stack (over 1’000 modules)
- Support and training to researchers in using HPC and services

Swiss Personalized Health Network SPHN

- A national initiative to promote research in Personalized Health
- Proposed in 2014, started in 2017
- Partners: university hospitals, schools of higher education, research institutes, and other organizations
- Mandated by the SERI (SBFI), co-mandated by the FOPH (BAG)
- **BioMedIT**: an SIB project with the aim to **create an IT environment for research with biomedical data in academia**



Swiss Personalized Health Network

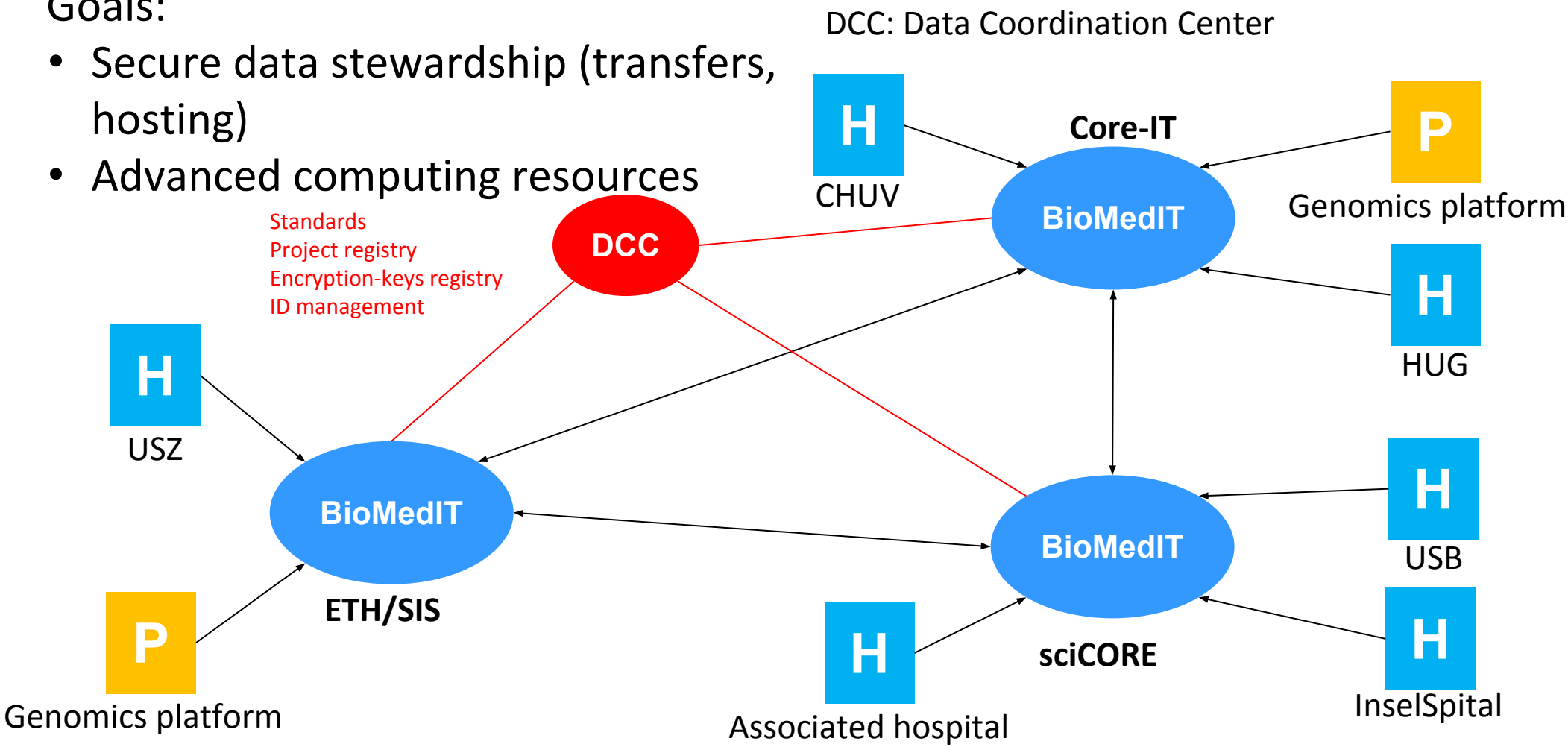
SAMW ASSM

SIB Swiss Institute of Bioinformatics

SPHN / BioMedIT interoperability

Goals:

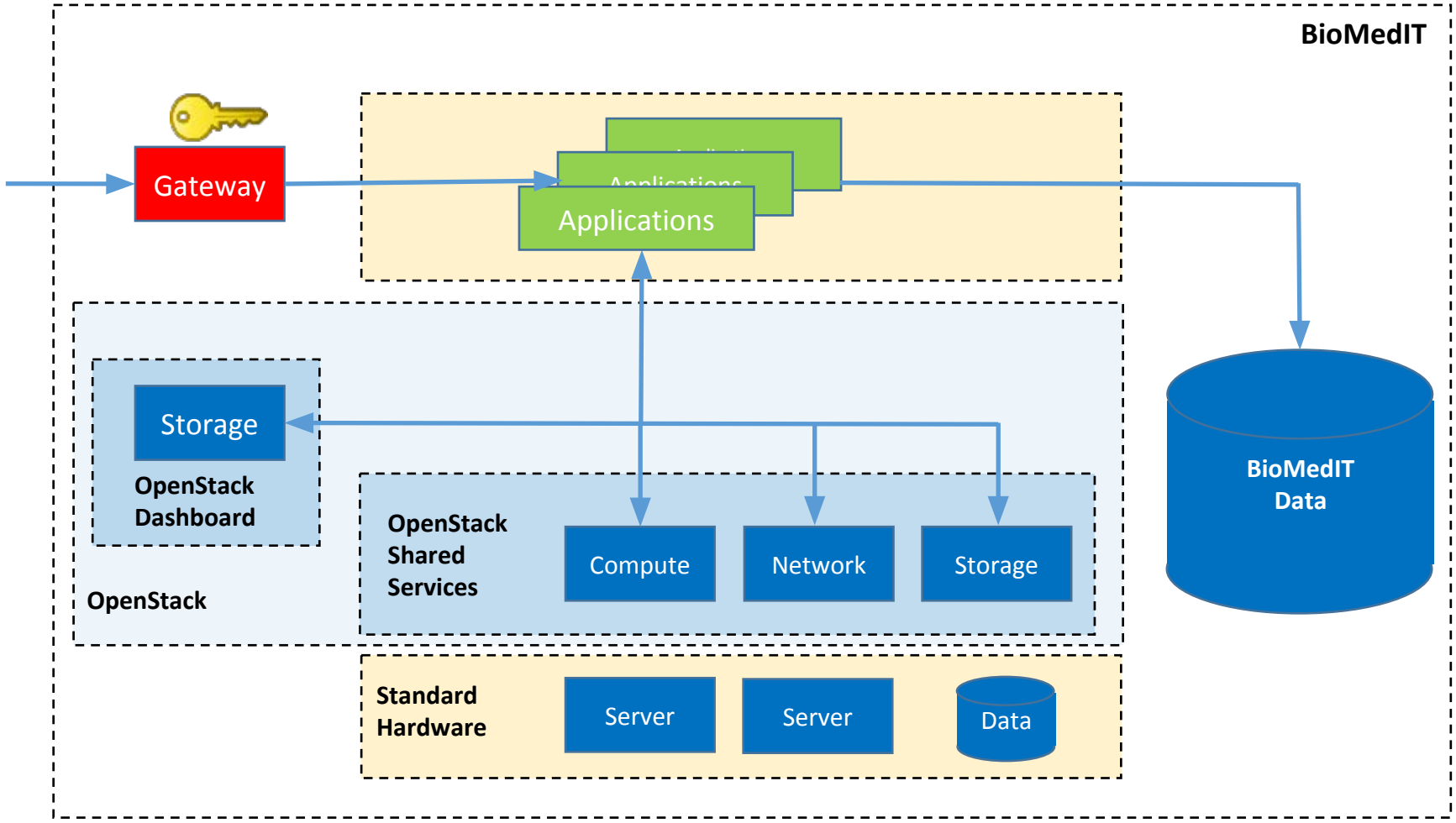
- Secure data stewardship (transfers, hosting)
- Advanced computing resources



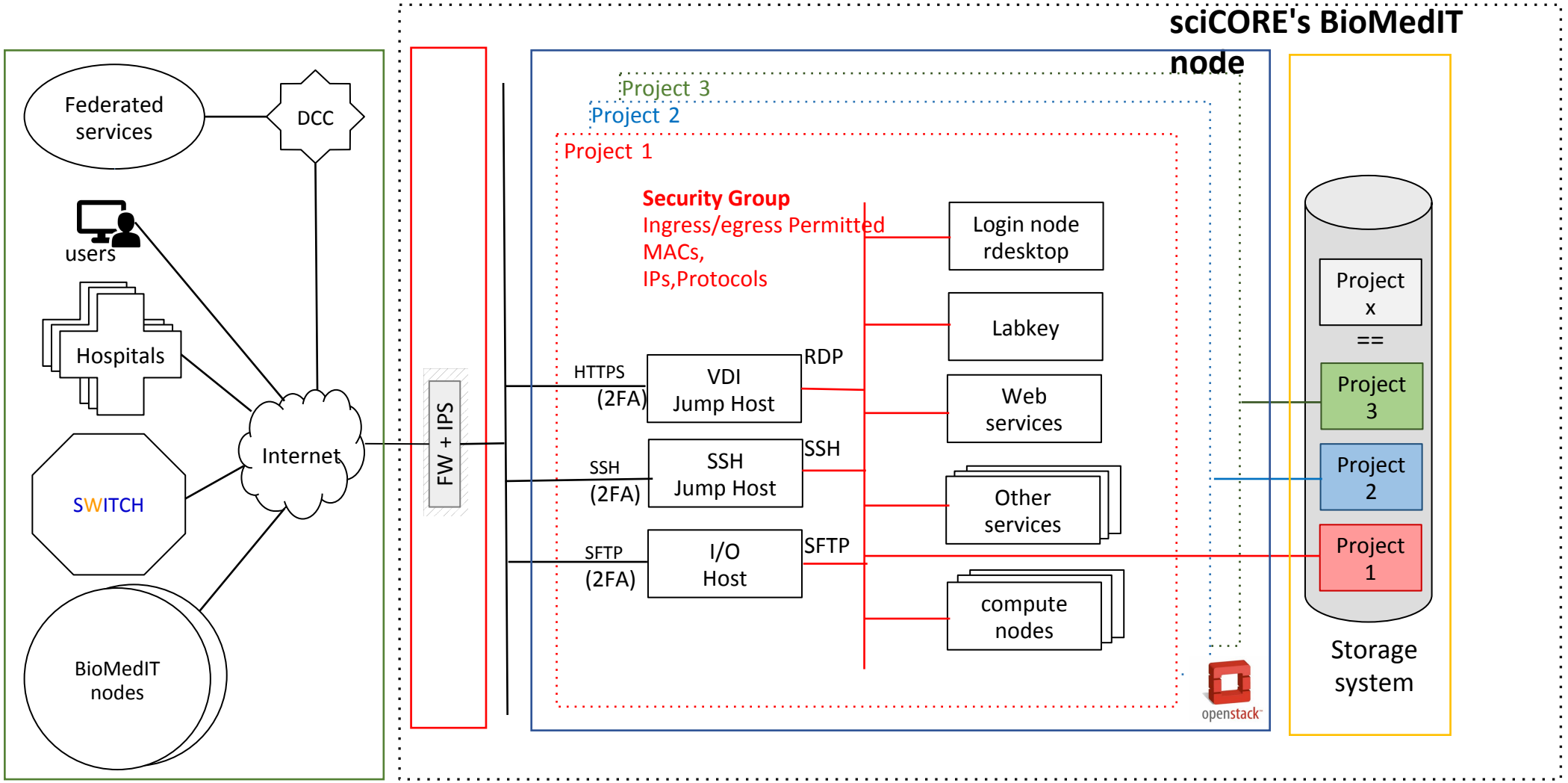
BioMedIT features and requirements

- BioMedIT features
  - Support current and future driver projects
  - Provide a secure data processing environment
  - Provide scientific and technical expertise in computational live science research
  - Provide pre-installed software and workflows
  - Provide large scale storage and reference data for applications
  - Provide high performance computing capacity
- BioMedIT requirements
  - Support containers, docker, singularity, ...
  - Support deep learning, machine learning (GPUs)
  - Support bare metal
  - Provide project isolation (multitenancy)
  - Support container orchestration (docker swarm, kubernetes, …)
  - Afford monitoring, logging and auditing (intrusion detection, controlling data access, incident handling)
  - **Provide flexibility, scalability, security**

BioMedIT Initial Design



BioMedIT Infrastructure on OpenStack
Project architecture



OpenStack deployment tool Kayobe

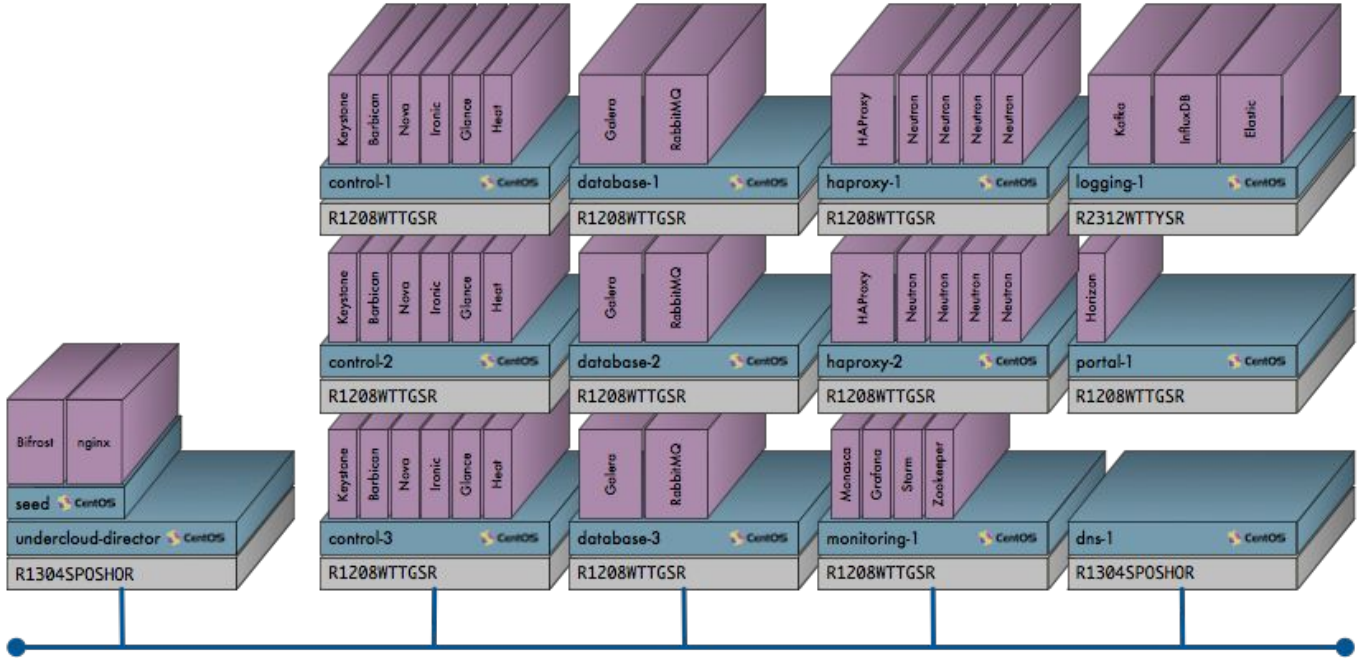
- Kayobe - Deployment of containerised OpenStack to bare metal.
  - Discovery, introspection and provisioning of control plane hardware using Bifrost
  - Configuration of control plane hosts for Kolla using Ansible
  - Deployment of an OpenStack control plane using Kolla-Ansible
  - Discovery, introspection and provisioning of bare metal compute hosts using Ironic

ANSIBLE

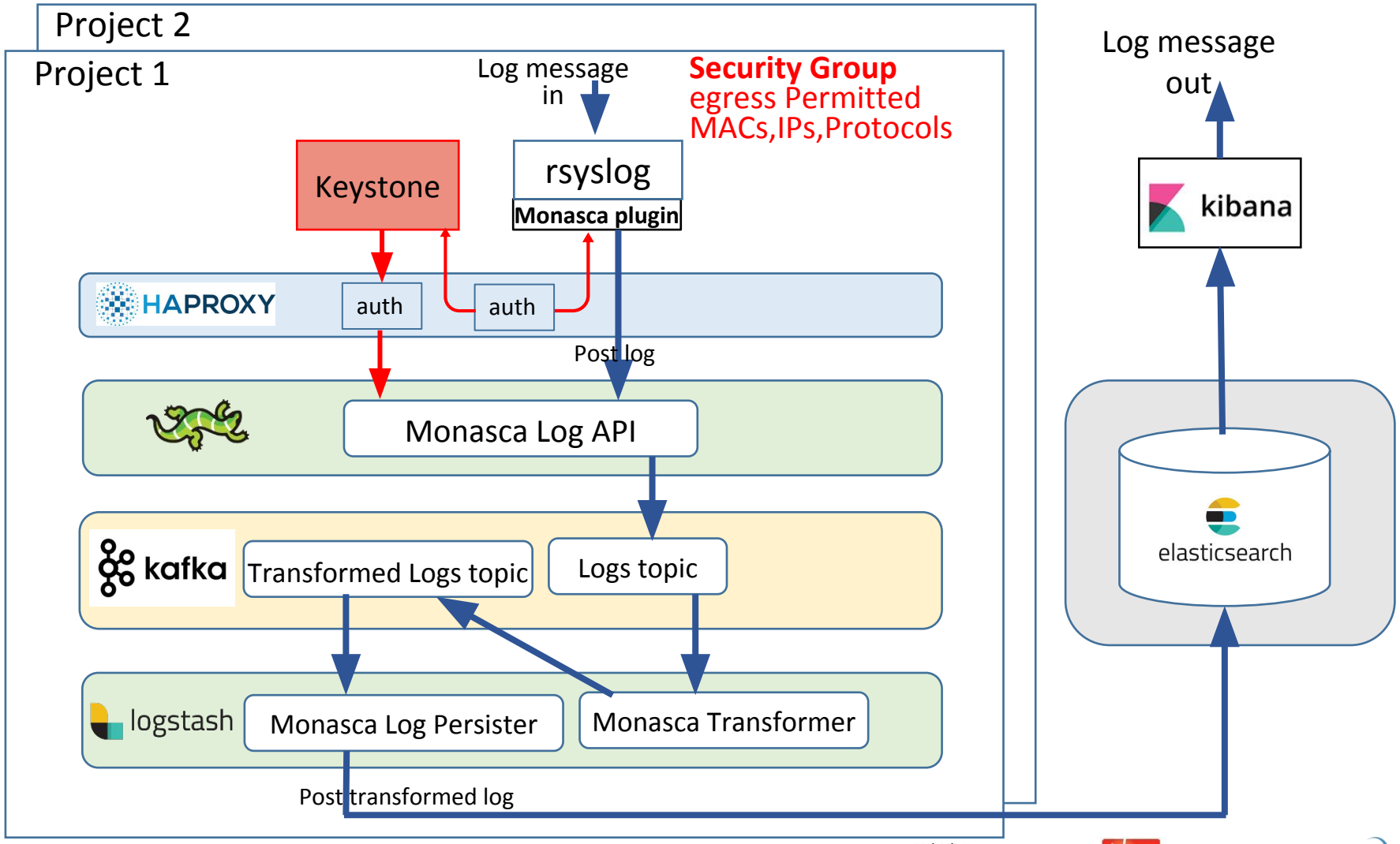
docker

KOLLA
an OpenStack Community Project

openstack.



Logging, Auditing architecture



Driver projects (call 2017)

	**Swiss Frailty Network and Repository** Heike Bischoff-Ferrari
	**Population-wide screens of the human immune repertoire: a reverse personalized-medicine approach (CCIP)** Adriano Aguzzi
	**Swiss Molecular Pathology Breakthrough Platform (SOCIBP)** Mark Rubin co-funded with Strategic Focus Area Personalized Health and Related Technologies
	**Swiss Personalized Oncology (SPO)** Olivier Michiélin and Mohamed Bentires-Alj
	**PRECISE: Identification of biomarkers and therapeutic targets in inflammatory disease immunotherapy by high-dimensional single cell analysis and cluster proteomics** Manfred Claassen co-funded with Strategic Focus Area Personalized Health and Related Technologies
	**PSSS: Personalized Swiss Sepsis Study**: Detection and modelling of sepsis using machine learning to analyse continuous ICU monitoring, laboratory, microbiology, and -omics data for personalized sepsis management Adrian Egli co-funded with Strategic Focus Area Personalized Health and Related Technologies

+ 8 infrastructure development projects

Domains: patient-data governance, e-consent management,
data de-identification, NLP processing of patient records, variants database (oncology)

Driver projects (call 2018)

	**SACR: The Swiss Ageing Citizen Reference** Nicole Probst
	**CREATE PRIMA: Clinical Research from multi-modality big data sources without proprietary interfaces in a multicenter approach** Joerg Leuppi
	**IMAGINE: Radiomics for comprehensive patient and disease phenotyping in personalized health** Matthias Guckenberger
	**SOIN: Swiss Ophthalmic Imaging Network** Thomas Wolffensberger
	**SHFN: SWISSHEART Failure Network** Christian Matter, Johachim Buhmann

co-funded with Strategic Focus Area Personalized Health and Related Technologies

+ 5 infrastructure development projects

Domains: variants database (non-oncology), reference datasets, QC metrics in radiology, secure distributed computing algorithms

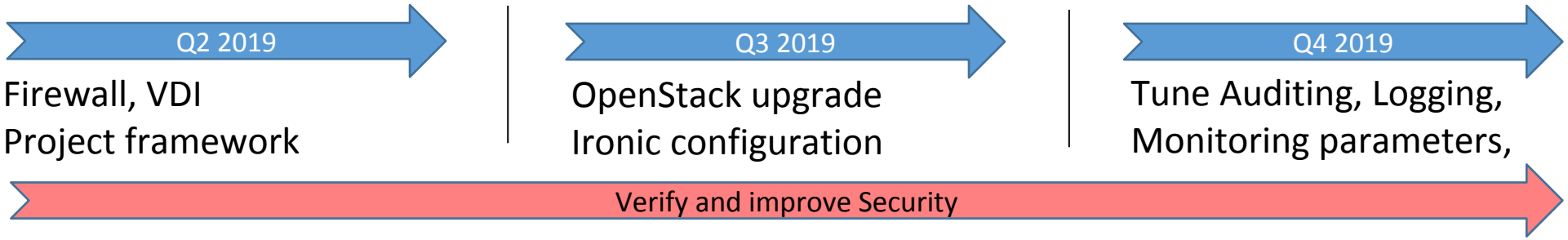
Use case - PSSS

- **PSSS: Personalized Swiss Sepsis Study:** Detection and modelling sepsis using machine learning to analyze continuous ICU monitoring, laboratory, microbiology and –omics data for personalized sepsis management
- About 400 patients to be enrolled in 5 university hospitals
- Data involved
  - ICU continuous data streams (blood pressure, heart rate, temperature,…)
  - Clinical assessments (diagnostics), outcome
  - Genomics (patient and host), whole-genome sequencing of pathogens
  - Metabolomics (patient and host)
  - Immuno-phenotyping
  - Multiple time points, physiological compartments (gut, lung, blood)
- Data analysis
  - Feature extraction from ICU data (HMM, Generative Adversarial Networks, …)
  - Iterative multi-omics analysis of host-pathogen interactions (machine learning)
- Data volume: 300 TB

BioMedIt deployment next steps

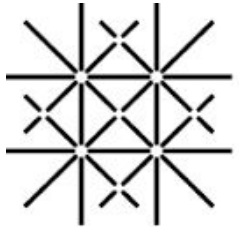
- Firewall and IPS installation (done) and configuration (ongoing)
- Monitoring, logging, auditing and alerting of the projects
- Project deployment framework
  - Ansible playbooks and roles to deploy a full project on OpenStack
- OpenStack update (Version: Stein), Ironic configuration (for Bare Metal supportm GPUs)
- Security (review of the security design by a external security company, penetration tests, developing workflows for incident response and/or emergency patching, finalising security hardening both hardware and software level)

- Timeline



Project-specific challenges

- High security requirements (by default no outgoing traffic)
- Handles sensitive data!
- Tenant isolation / security settings on perimeter / firewalls
- OpenStack security functionality
- User training / awareness
- High learning curve for sysadmins, sufficient resource allocation is paramount.
- Auditing, logging and monitoring user and sysadmin access



Universität Basel

StackHPC

Jani Heikkinen
(jani.heikkinen@unibas.ch, sciCORE)

Stig Telfer
(stig@stackhpc.com, StackHPC)

Thank You

Acknowledgments

- sciCORE colleagues: scicore.unibas.ch/about-scicore/people/
- StackHPC: Will Szumski, Bharat Kunwar, Doug Szumski, Nick Jones