WESkit

GA4GH-compliant workflow execution

Sven Twardziok

Online training

29.11.2021

Aus Forschung wird Gesundheit

Contents

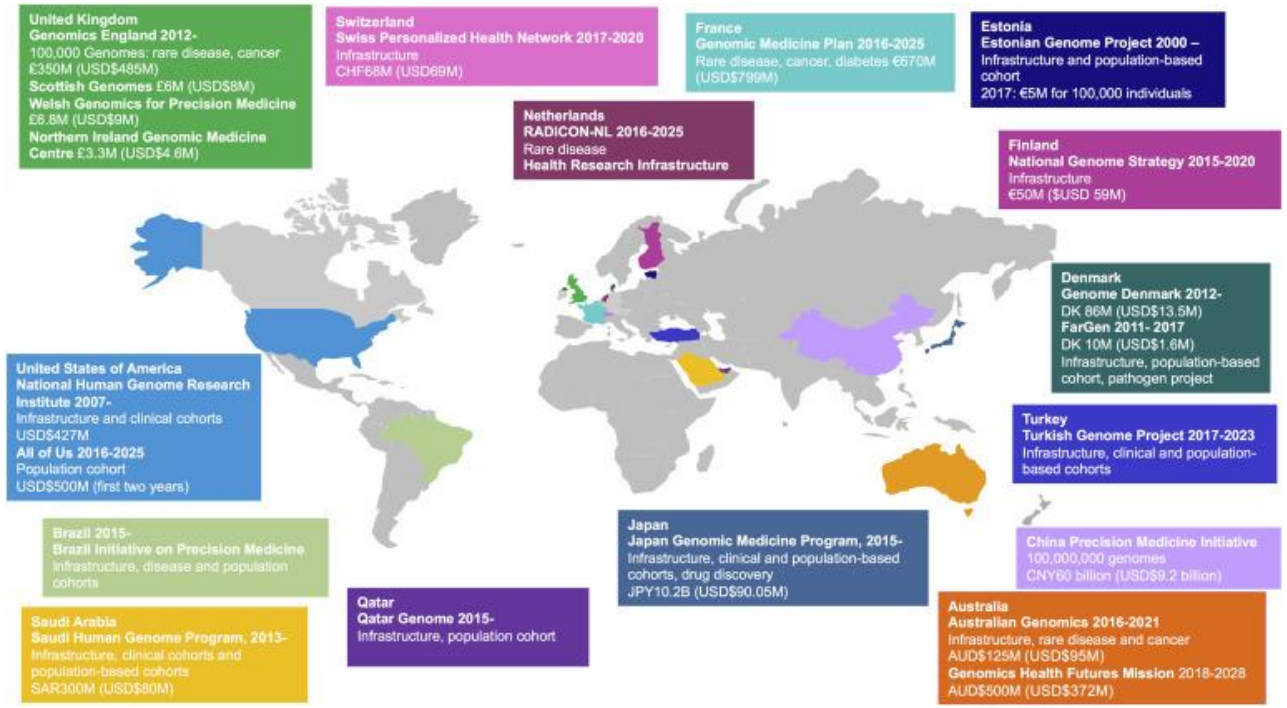
1. Introduction to GA4GH-WES specification
2. Introduction to WESkit
3. Demo: deploying WESkit
4. Exercise: using WESkit
5. WESkit use cases
6. Summary

GA4GH - WES

BIH Berlin Institute of Health @Charité

Workflows & Cloud
Genomics in healthcare

Expected 60 millions of human genomes sequenced within global healthcare systems **by 2025** - this will transform whole healthcare systems[1]



National sequencing initiatives[2]

Challenges[2]:

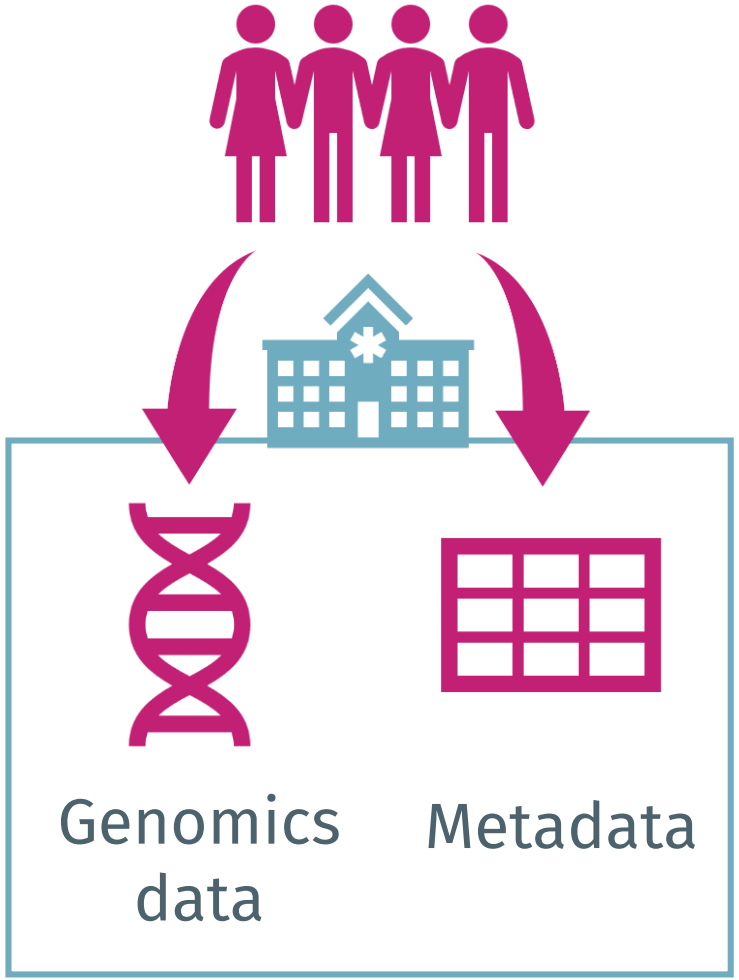
- **Technical**: infrastructure and bioinformatics capacity
- **Research**: data integration and interpretation
- **Social:** public acceptability, and ethical and legislative issues

[1] Genomics in healthcare: GA4GH looks to 2022 (2017) | bioRxiv.,
[2] Integrating Genomics into Healthcare: A Global Responsibility (2019) | AJHG

Berlin Institute of Health @Charité

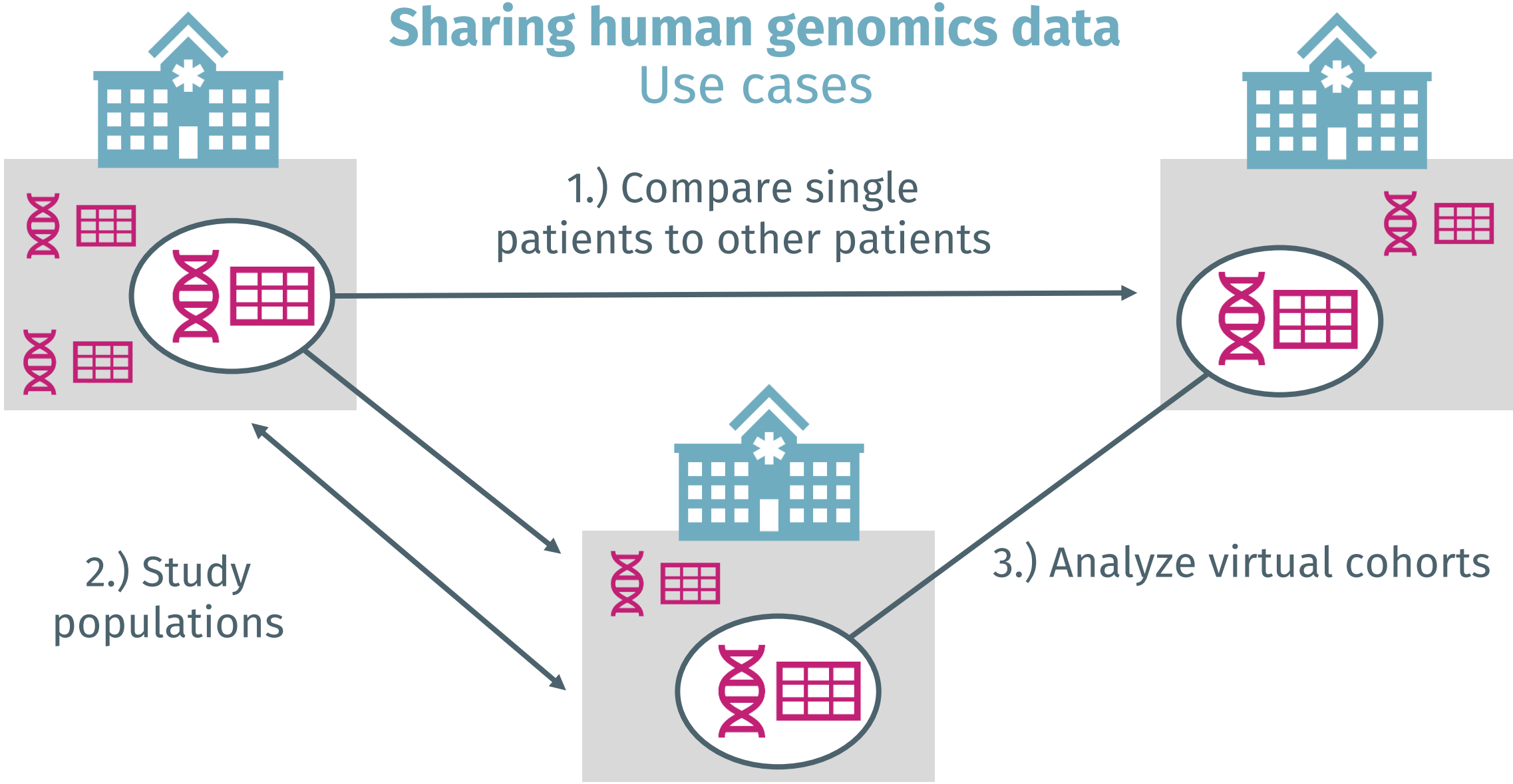
Genomics data in healthcare
Features

- **Big:** hundreds of gigabytes per patient
- **Sensitive:** genetic data is specifically protected by GDPR
- **Structured:** linked with many metadata such as clinical data
- **Decentralized:** located in sequencing facilities, (university) hospitals, private laboratories



BIH Berlin Institute of Health @Charité

Sharing human genomics data
Use cases



Berlin Institute of Health @Charité

Science Cloud
Sending analysis workflows to data

Requirements for standardization

- Organize a decentralized cloud infrastructure including a decentralized user access management.
- Send and process workflows in the cloud.
- Provide access to data and data discovery.
- Define good data/workflow management.

Global Alliance for Genomics & Health (GA4GH)

Mission

The Global Alliance for Genomics and Health aims to accelerate progress in genomic research and human health by cultivating a **common framework of standards** and harmonized approaches for effective and responsible genomic and health-related data sharing.

GA4GH **Driver Projects** are real-world genomic data initiatives that provide input on the standards most needed for the international genomics community to share data. Driver Projects put finished standards and frameworks to immediate use.

GA4GH Driver Projects

National Initiatives

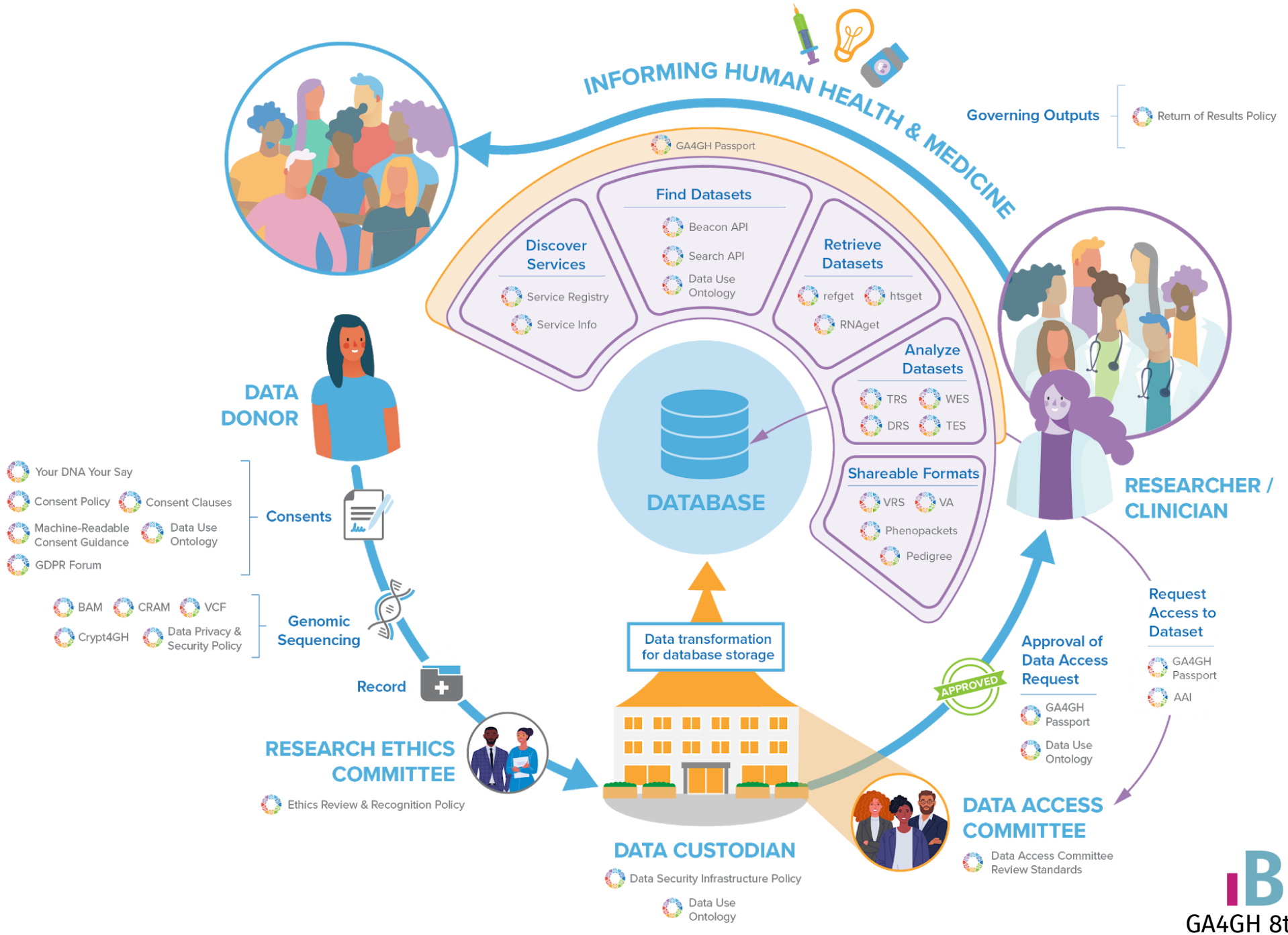
Pan-African

Pan-European

International

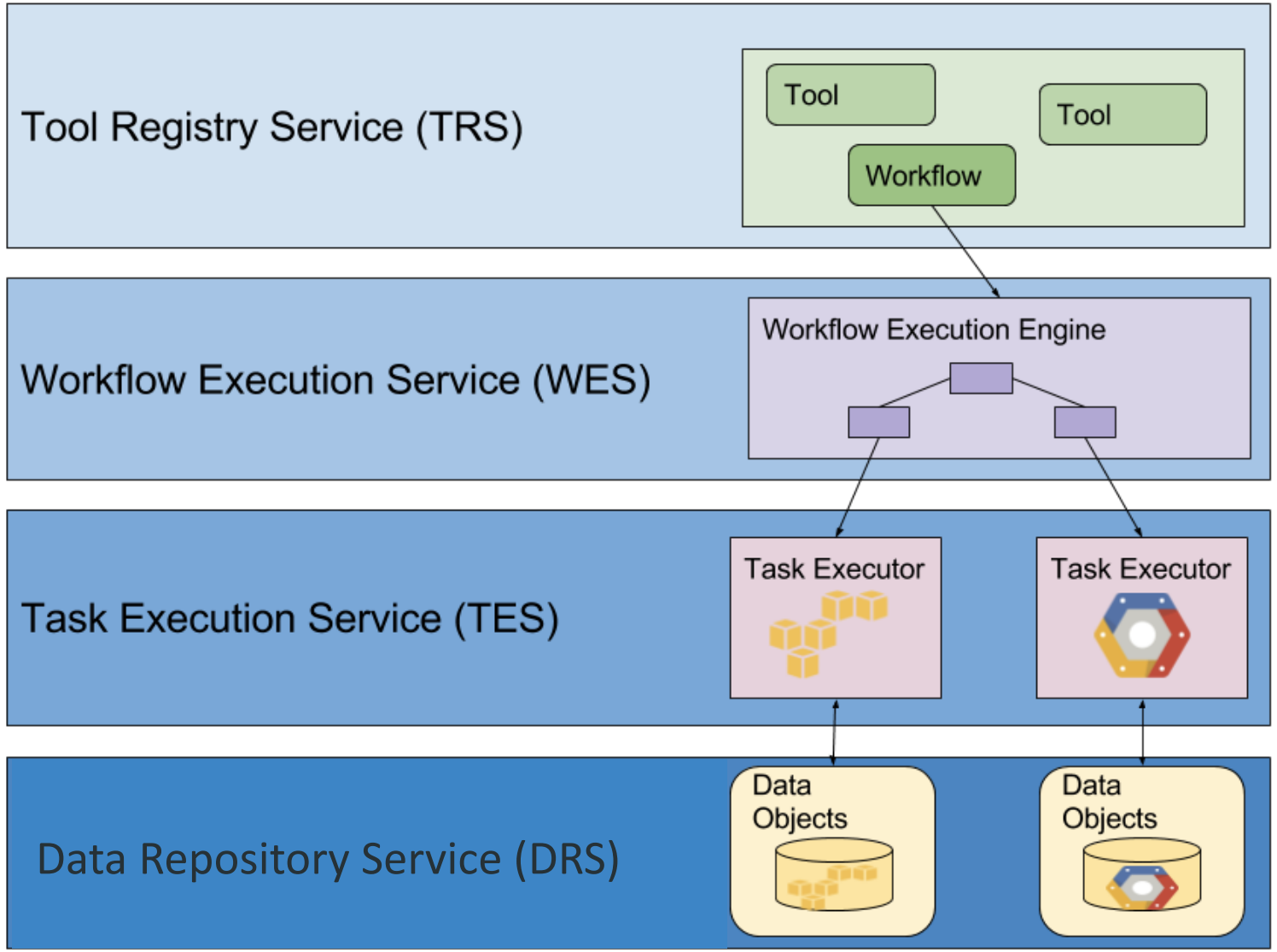
https://www.ga4gh.org/wp-content/uploads/GA4GH-onepager-v6.pdf

Berlin Institute of Health @Charité


INFORMING HUMAN HEALTH & MEDICINE
Governing Outputs
Return of Results Policy
GA4GH Passport
Find Datasets
Beacon API
Search API
Data Use Ontology
Discover Services
Service Registry
Service Info
Retrieve Datasets
refget
htsget
RNAget
Analyze Datasets
TRS
WES
DRS
TES
Shareable Formats
VRS
VA
Phenopackets
Pedigree
DATABASE
DATA DONOR
RESEARCHER / CLINICIAN
Your DNA Your Say
Consent Policy
Consent Clauses
Machine-Readable Consent Guidance
Data Use Ontology
GDPR Forum
Consents
BAM
CRAM
VCF
Crypt4GH
Data Privacy & Security Policy
Genomic Sequencing
Record
Data transformation for database storage
APPROVED
Approval of Data Access Request
GA4GH Passport
Data Use Ontology
Request Access to Dataset
GA4GH Passport
AAI
RESEARCH ETHICS COMMITTEE
Ethics Review & Recognition Policy
DATA CUSTODIAN
Data Security Infrastructure Policy
Data Use Ontology
DATA ACCESS COMMITTEE
Data Access Committee Review Standards

Cloud Work Stream APIs

Global Alliance for Genomics & Health



Sharing Tools and Workflows

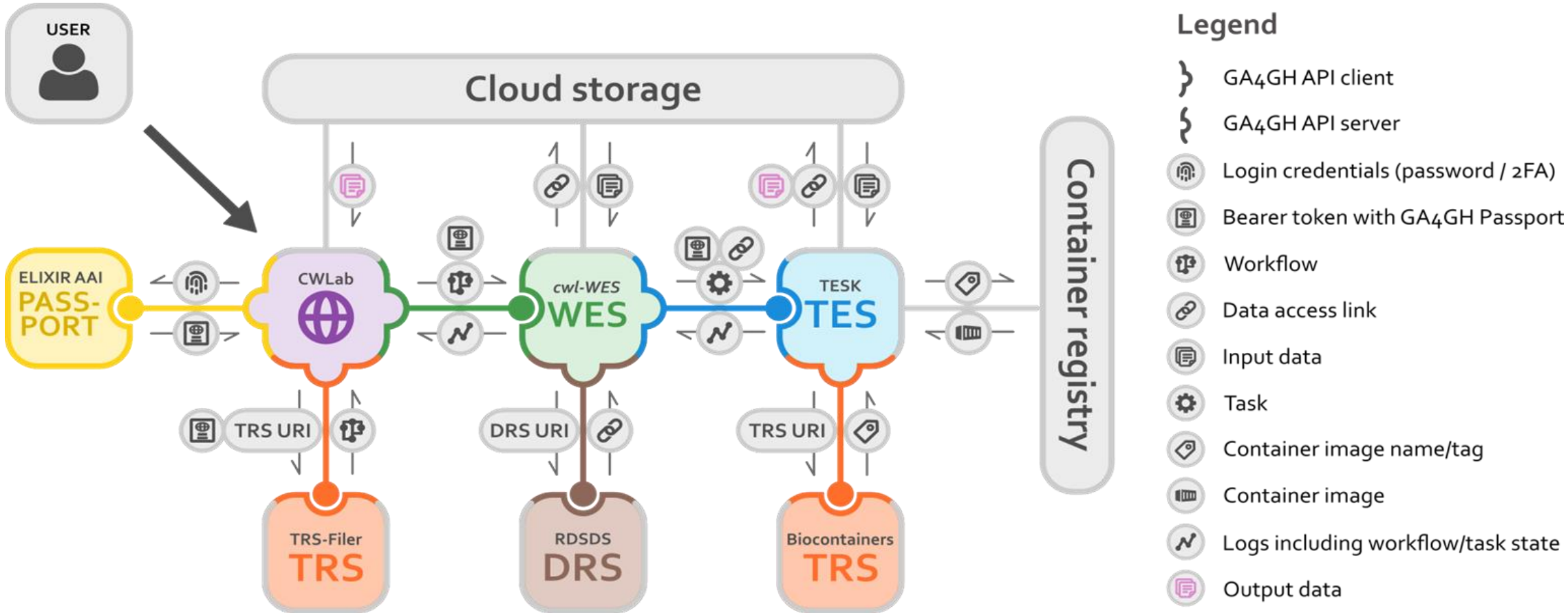
Executing Workflows

Executing Individual Tasks

Accessing Data

BIH Berlin Institute of Health @Charité

ELIXIR Cloud & AAI
Integration of services



2020-09-30 ELIXIR Cloud & AAI @ 8th GA4GH Plenary

GA4GH Workflow Execution Service (WES)
Endpoints

Method	Path	Description
GET	/service-info	Get information about Workflow Execution Service.
GET	/runs	List the workflow runs.
POST	/runs	Run a workflow.
GET	/runs/{run_id}	Get detailed info about a workflow run.
GET	/runs/{run_id}/status	Get quick status info about a workflow run.
POST	/runs/{run_id}/cancel	Cancel a running workflow.

https://github.com/ga4gh/workflow-execution-service-schemas/

BIH Berlin Institute of Health @Charité

Workflow Execution Service (WES)

POST /run

```
requests.post(
   url = WES_SERVER + "/ga4gh/wes/v1/runs",
   header = {"Authorization" : "Bearer " + ACCESS_TOKEN},
   data = {
      "workflow_params": CONFIG_JSON,
      "workflow_type": "Snakemake",
      "workflow_type_version": "5",
      "workflow_url": "/workflows/wf1/Snakefile"
      }
   }
)
{"run_id": "id_123"}
```


BIH Berlin Institute of Health @Charité

Workflow Execution Service (WES)
GET /runs/{run_id}

```
requests.get(url = WES_SERVER + "/ga4gh/wes/v1/runs/id_123")
{
  "run_id": "id_123",
  "request": {...},
  "state": "COMPLETE",
  "run_log":{
    "name": "my_workflow",
    "cmd": "snakemake …",
    "start_time": "2021-01-12T12:00:00Z",
    "end_time": "2021-01-12T12:05:00Z",
    "stdout": "/PATH/TO/STDOUT"
    "stderr": "/PATH/TO/STDERR",
    "exit_code": 0
  },
  "task_logs": [...],
  "outputs": {...}
}
```


BIH Berlin Institute of Health @Charité

WESkit

BIH Berlin Institute of Health @Charité

WESKIT

GA4GH WES implementation

https://gitlab.com/one-touch-pipeline/weskit

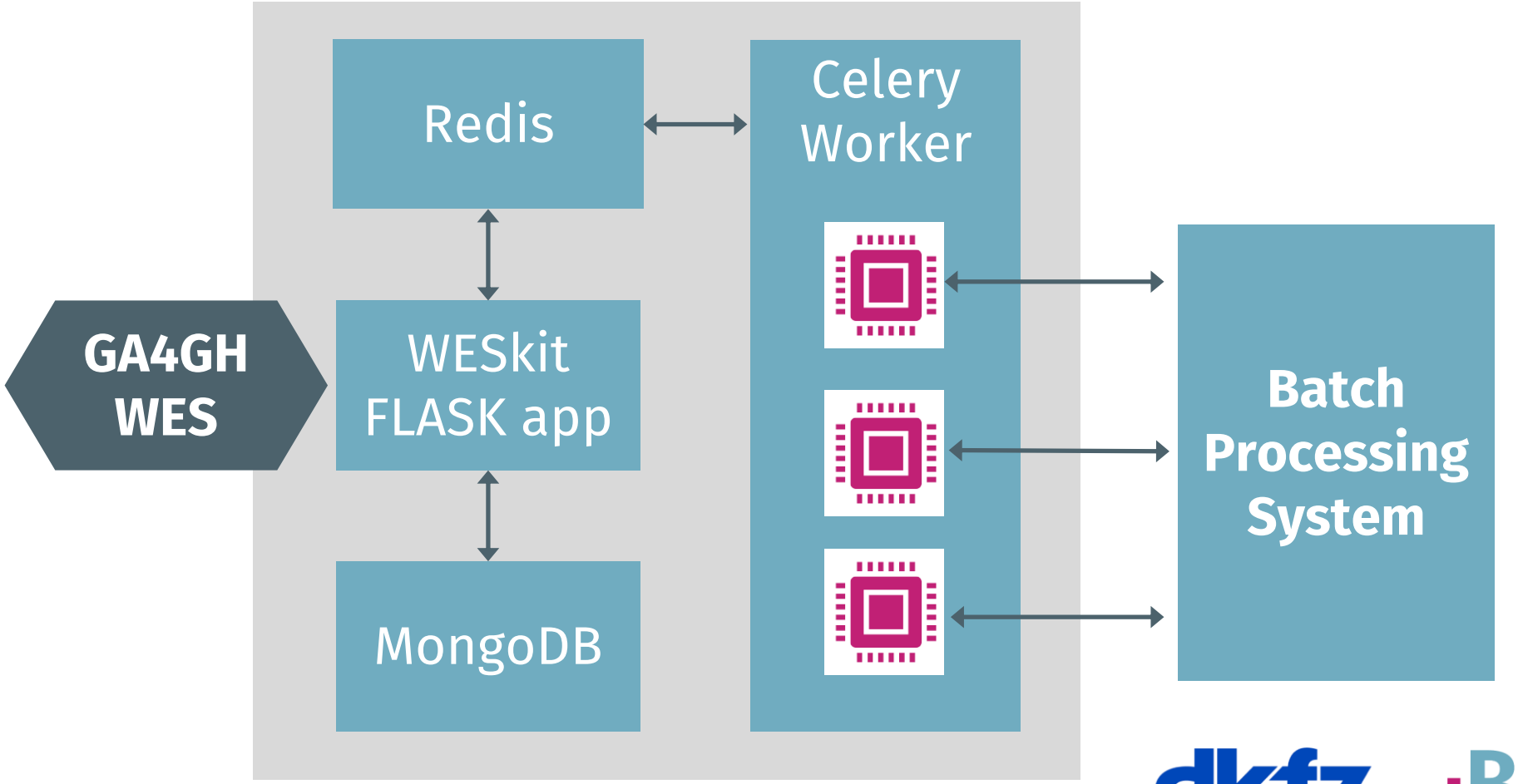
Features

- GA4GH WES implementation supporting Snakemake and Nextflow
- Developed for stability and high data throughput at BIH and DKFZ
- HPC and cloud deployment supported
- OAuth2.0 support
- Developed in cooperation with ELIXIR compute platform

WESKIT

GA4GH WES implementation

https://gitlab.com/one-touch-pipeline/weskit

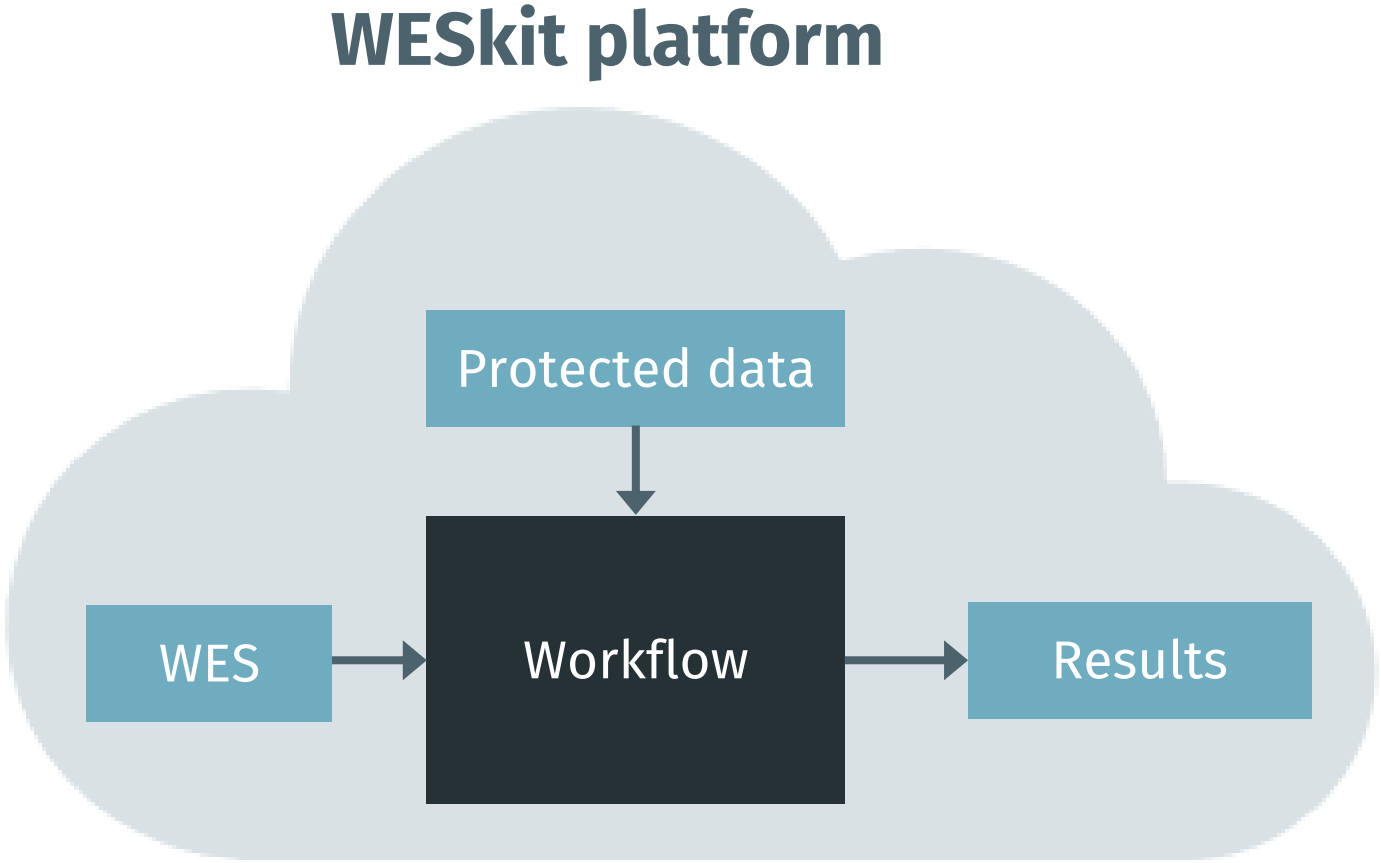


Containerized deployment

Demonstration

BIH Berlin Institute of Health @Charité

Demo: deploying WESkit



https://gitlab.com/one-touch-pipeline/weskit/documentation

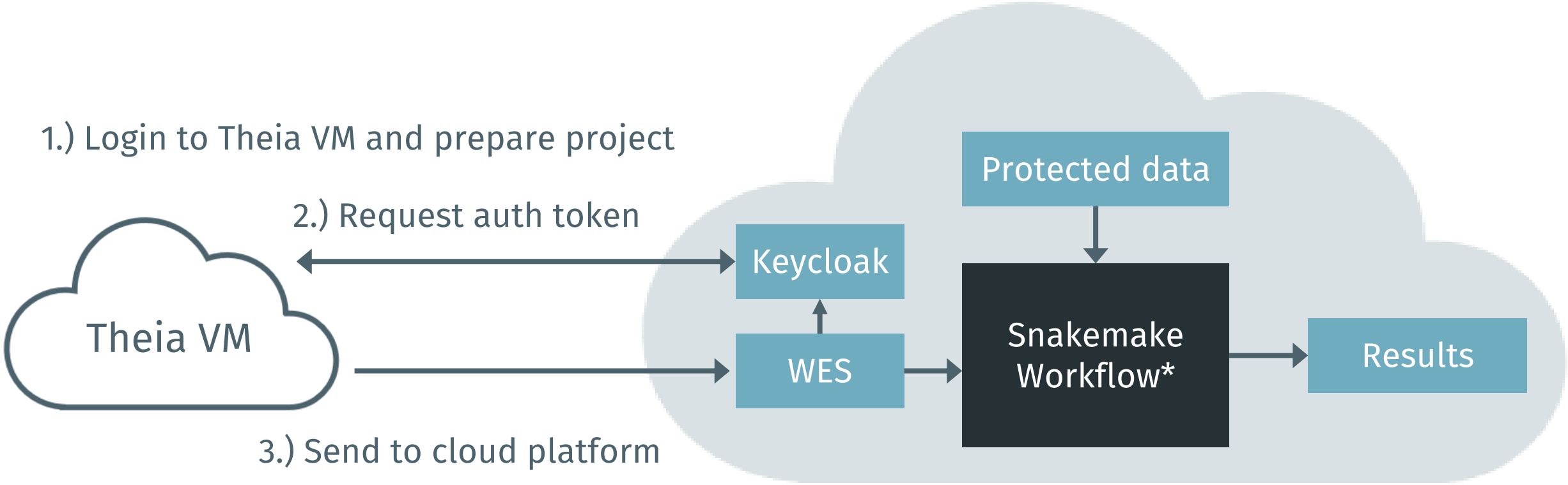
BIH Berlin Institute of Health @Charité

Exercise

BIH Berlin Institute of Health @Charité

Task: send workflow to cloud

https://gitlab.com/twardzso/weskit-elixir-workflow-workshop-2021



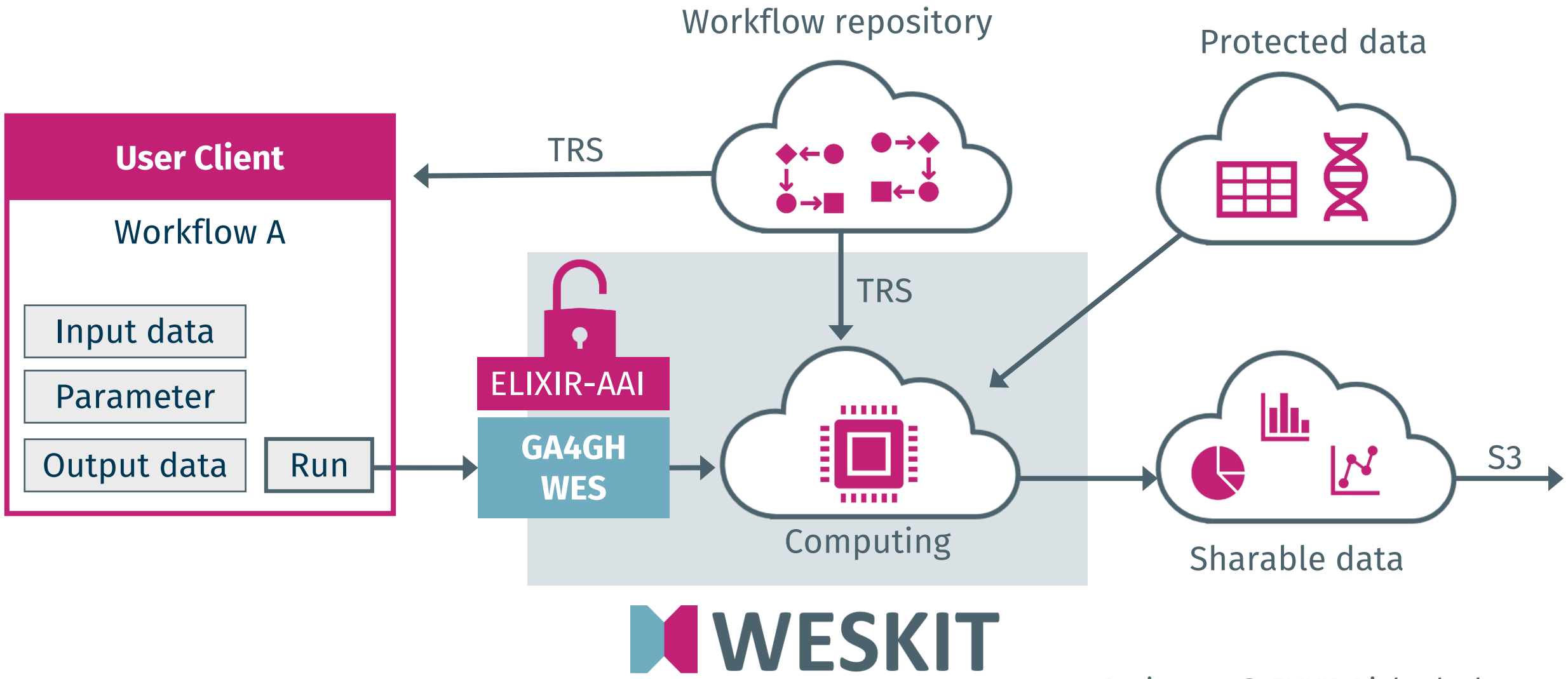
*https://snakemake.readthedocs.io/en/stable/tutorial/tutorial.html

BIH Berlin Institute of Health @Charité

WESkit use-cases

BIH Berlin Institute of Health @Charité

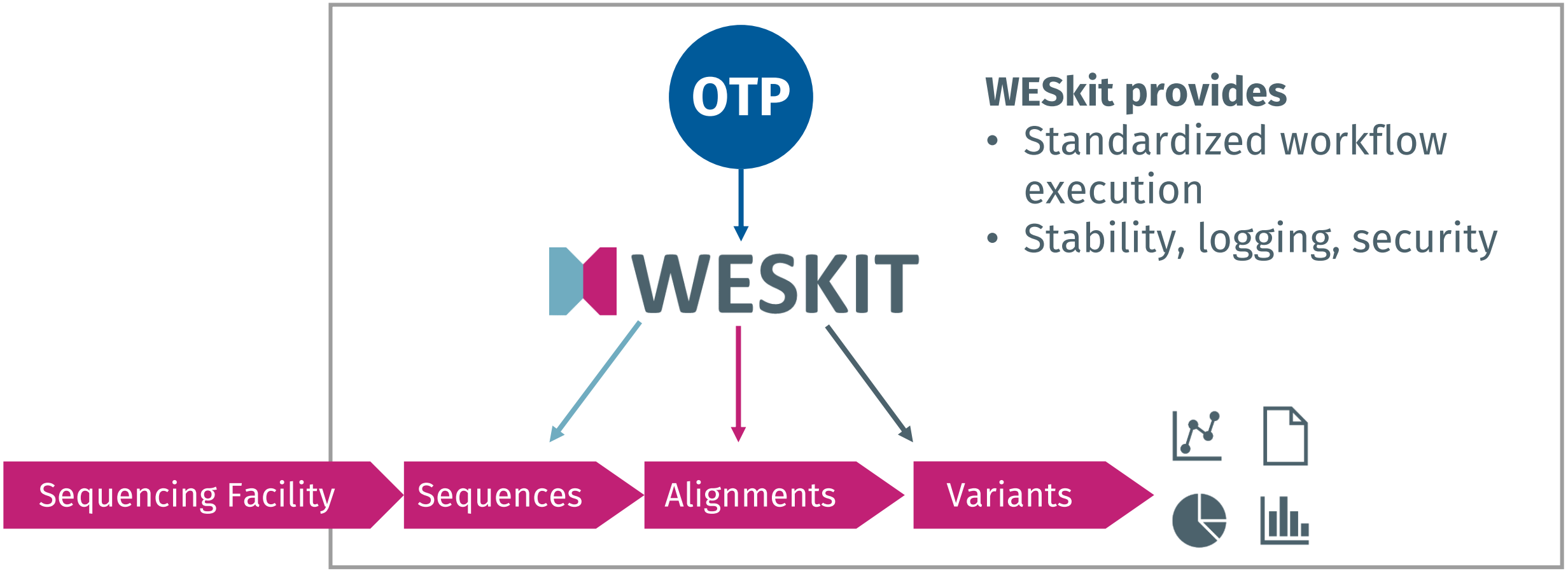
WESkit Cloud Use Case



Project 08 @ ELIXIR Biohackathon 2021

WESkit Automation Use Case

https://gitlab.com/one-touch-pipeline



WESkit provides

- Standardized workflow execution
- Stability, logging, security

Summary

- GA4GH-WES: specification for sending workflows to execution environment
- WESkit: implementation of GA4GH-WES for Snakemake and Nextflow
- Cloud use case: build platform to provide access to sensitive data
- Automation use case: provide standardized interface for repetitive execution of workflows

BIH Berlin Institute of Health @Charité

Thank you

Questions:
sven.twardziok@bih-charite.de

www.bihealth.org

Charité - BIH

Roland Eils

Andreas Kling

Valentin Schneider-Lunitz

Florian Kärcher

Harald Wagener

Jürgen Eils

Martin Braun

Matthias Bieg

Naveed Ishaque

Philipp Strubel

HD-Hub

Chris Lawerenz

Elisabeth Bürger

Ivo Buchhalter

Kersten Breuer

Pavlo Lutsik

Philip Kensche

ELIXIR Cloud & AAI

Alexander Kanitz

Alvaro Gonzalez

Ania Niewielska

Jonathan Tedds

Justin Clark-Casey

Kostis Zagganas

Lukas Hehtmanek

Marius Diekmann

elixir

de.NBI
GERMAN NETWORK FOR BIOINFORMATICS INFRASTRUCTURE

BIH Berlin Institute of Health Charité & MDC

Aus Forschung wird Gesundheit

Questions?

Berlin Institute of Health
Charité & MDC

Aus Forschung wird Gesundheit