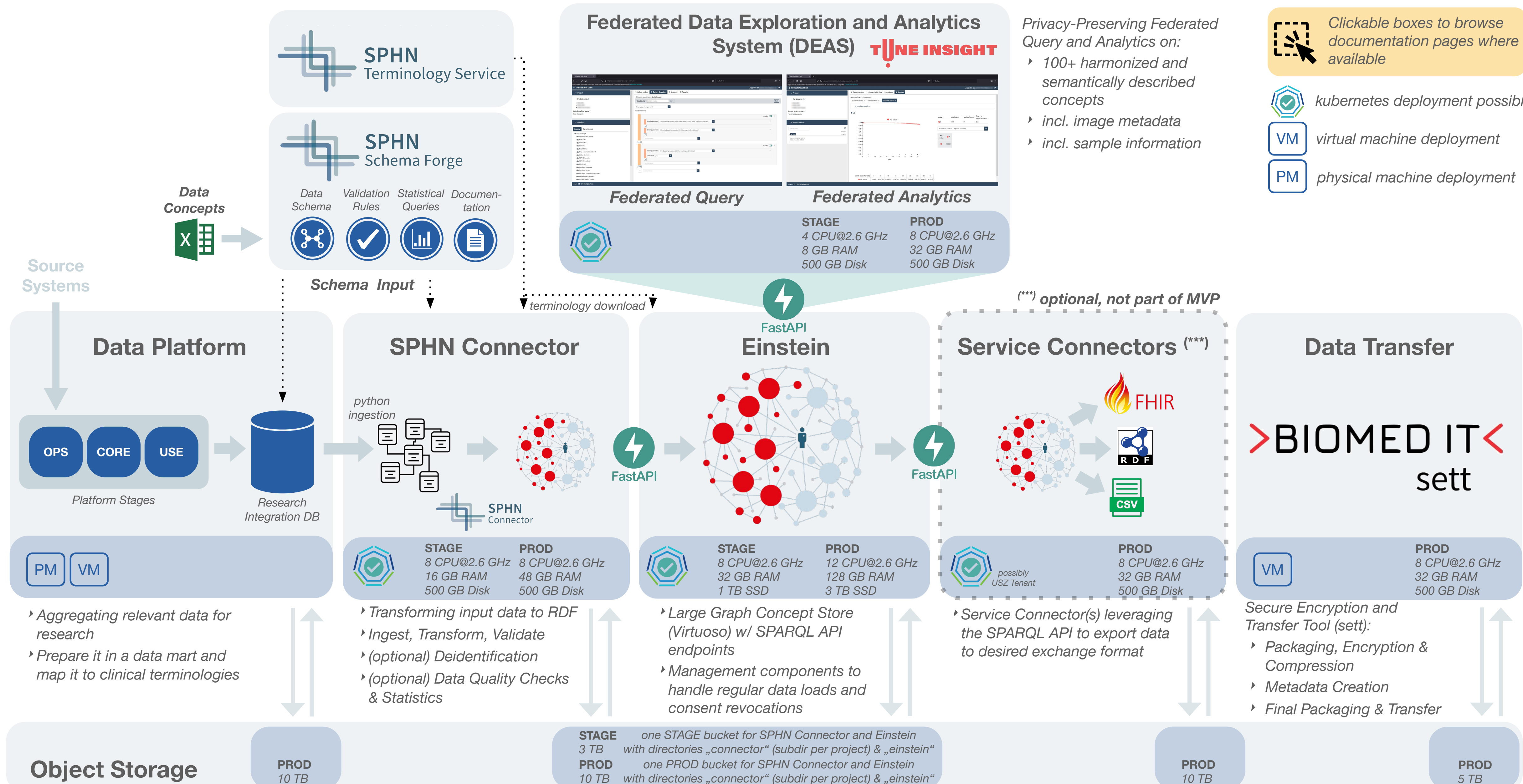# Data for Health: Short Talk

Integration, Operation and Usage of the SPHN Toolset in a University Hospital

Patrick Hirschi, 31.10.2024

Federated Data Exploration and Analytics System (DEAS) — TUNE INSIGHT

Federated Query — *Federated Analytics*

Privacy-Preserving Federated Query and Analytics on:
- 100+ harmonized and semantically described concepts
- incl. image metadata
- incl. sample information

	STAGE	PROD
	4 CPU@2.6 GHz	8 CPU@2.6 GHz
	8 GB RAM	32 GB RAM
	500 GB Disk	500 GB Disk

Clickable boxes to browse documentation pages where available

- kubernetes deployment possible
- **VM** virtual machine deployment
- **PM** physical machine deployment

SPHN Terminology Service

SPHN Schema Forge

Data Schema · Validation Rules · Statistical Queries · Documentation

Data Concepts → (Excel) → *Schema Input*

terminology download

Source Systems

Data Platform

OPS · CORE · USE

Platform Stages

Research Integration DB

PM	**VM**

- Aggregating relevant data for research
- Prepare it in a data mart and map it to clinical terminologies

SPHN Connector

python ingestion

SPHN Connector

FastAPI

	STAGE	PROD
	8 CPU@2.6 GHz	8 CPU@2.6 GHz
	16 GB RAM	48 GB RAM
	500 GB Disk	500 GB Disk

- Transforming input data to RDF
- Ingest, Transform, Validate
- (optional) Deidentification
- (optional) Data Quality Checks & Statistics

Einstein

FastAPI

	STAGE	PROD
	8 CPU@2.6 GHz	12 CPU@2.6 GHz
	32 GB RAM	128 GB RAM
	1 TB SSD	3 TB SSD

- Large Graph Concept Store (Virtuoso) w/ SPARQL API endpoints
- Management components to handle regular data loads and consent revocations

Service Connectors (***)

(***) optional, not part of MVP

FHIR · RDF · CSV

	PROD
possibly USZ Tenant	8 CPU@2.6 GHz
	32 GB RAM
	500 GB Disk

- Service Connector(s) leveraging the SPARQL API to export data to desired exchange format

Data Transfer

>BIOMED IT< sett

VM	PROD
	8 CPU@2.6 GHz
	32 GB RAM
	500 GB Disk

Secure Encryption and Transfer Tool (sett):
- Packaging, Encryption & Compression
- Metadata Creation
- Final Packaging & Transfer

Object Storage

	PROD 10 TB

STAGE 3 TB	one STAGE bucket for SPHN Connector and Einstein with directories „connector" (subdir per project) & „einstein"
PROD 10 TB	one PROD bucket for SPHN Connector and Einstein with directories „connector" (subdir per project) & „einstein"

PROD 10 TB

PROD 5 TB

USZ Universitäts Spital Zürich

Main Challenges

- data quality is an issue and best-effort solutions from data platforms are not enough - it affects the entire organization, and the processes to measure and improve it must be continuous and well-integrated hospital wide

- designing solutions that can work in all the hospitals with different IT architectures and still be compliant to all the ethical, data protection and information security aspects is very hard and time consuming

- linking what's there, consolidating, coming up with a common vision and convince everyone to get the resources (people and infrastructure) needed

There still seem to be no real incentives to work on good quality data and let other projects benefit from it.

we need…

QUALITY
BEFORE QUANTITY

Whatever we build during funded initiatives: at some point the funding will be reduced/gone, but the solutions need to live on.

we need…

SUSTAINABLE
SOLUTIONS

In a highly competitive field it is hard to work on common goals and stop people from constantly trying to reinvent the wheel.

we need…

COMMITMENT &
COLLABORATION