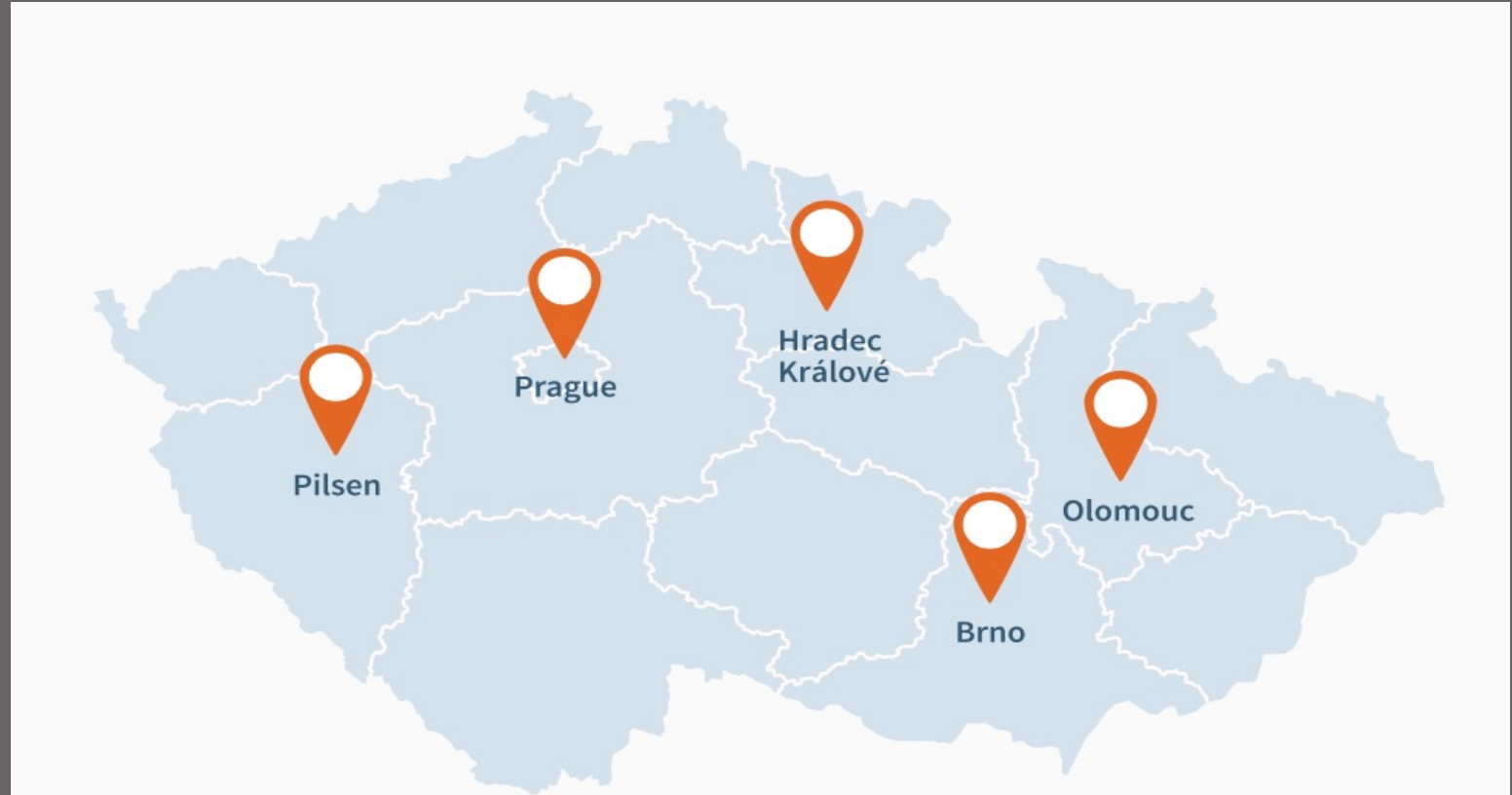# Implementation and development of Federated Search for biospecimen and data within BBMRI.cz
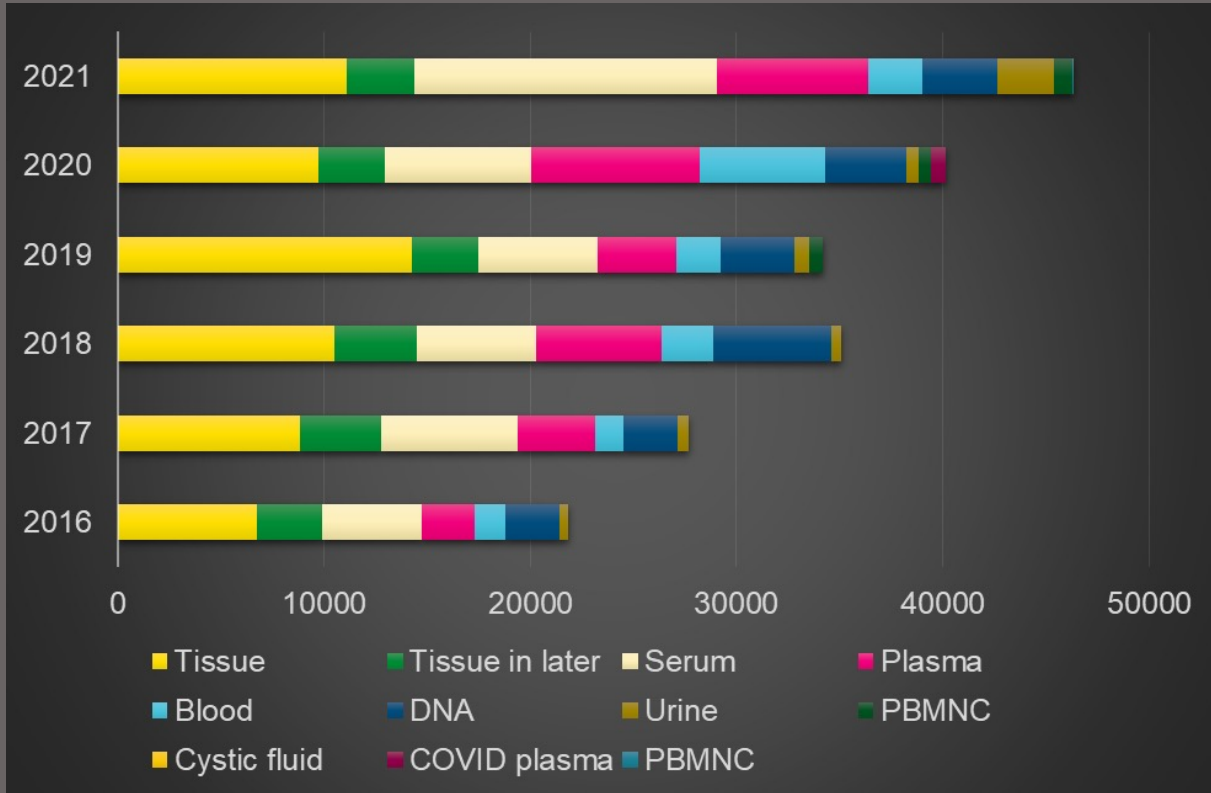
Radovan Tomasik

BBMRI.cz
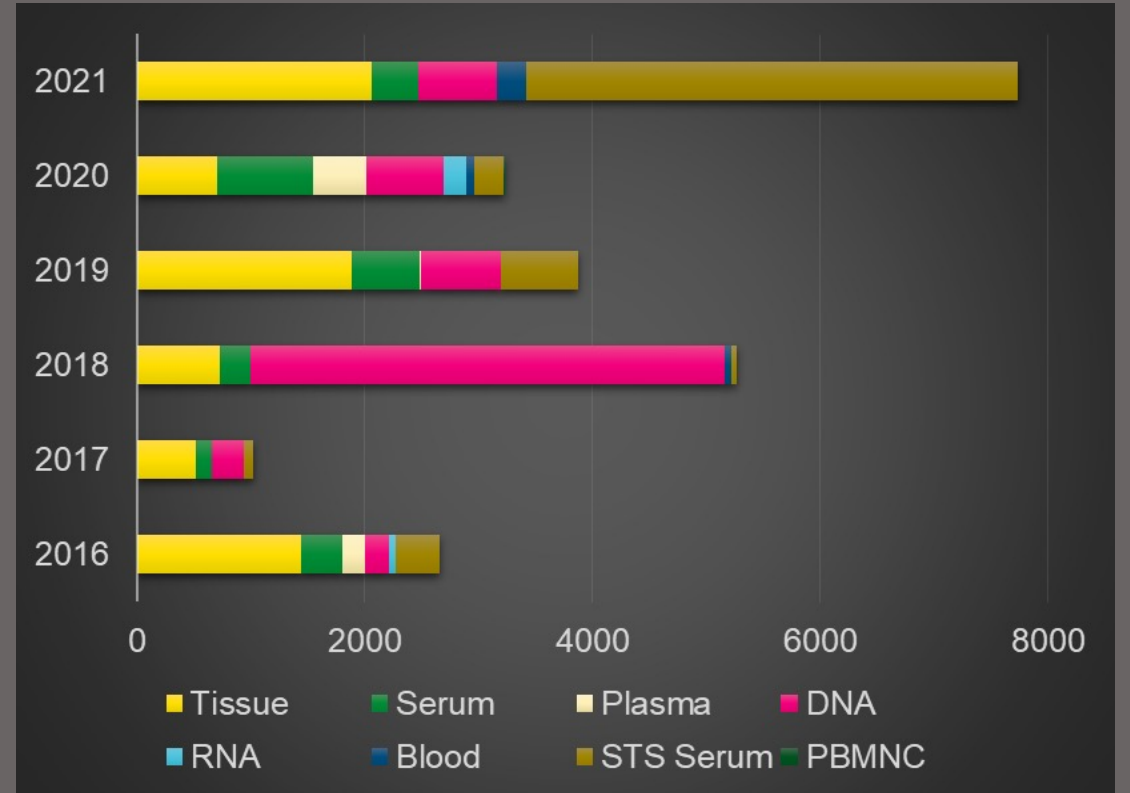
The Czech national node of BBMRI

- 6 Member biobanks
- 2 more to join
- (collected tumors of breast, pancreas, lung, liver, colorectal, ovaria, kidney, melanoma, brain, head & neck, bile duct,…)

Newly stored aliqotes:

Released Samples:
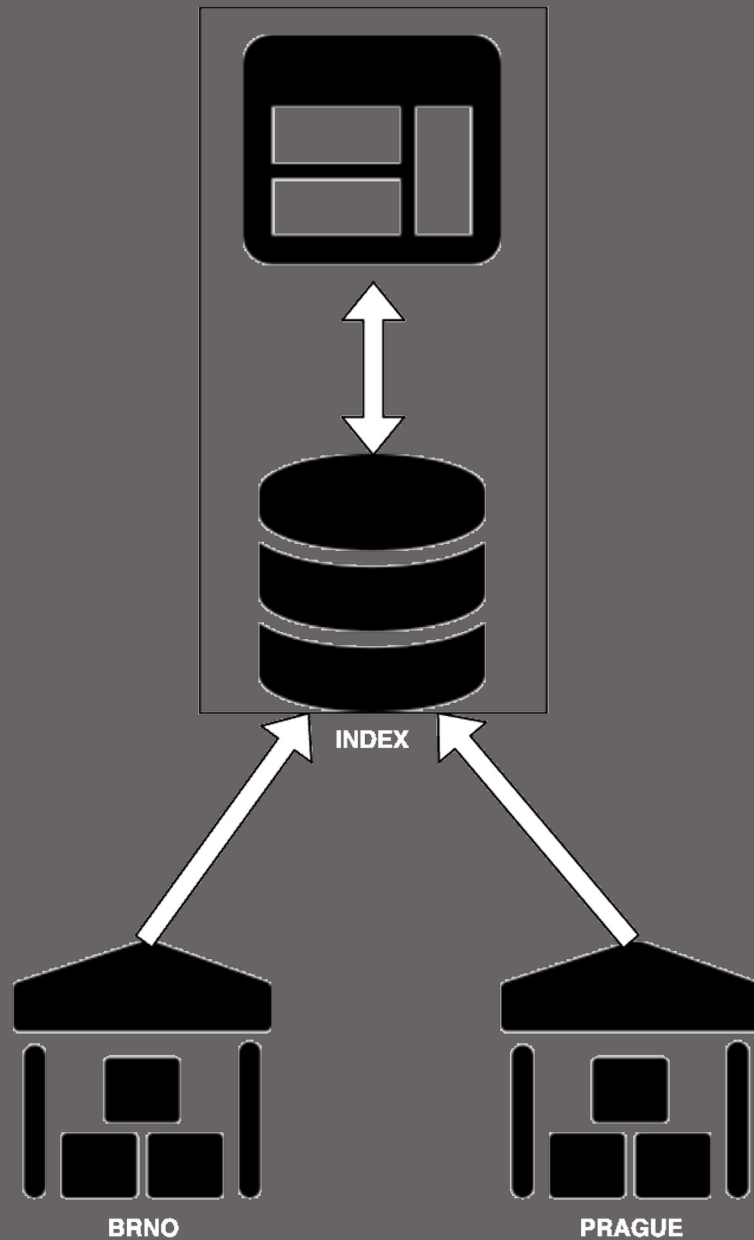
Supplier

Biobank

Keeps:

- Samples
  - Data

Customer (In a sense)

Researcher

WANTS:

- Specific **high quality** samples
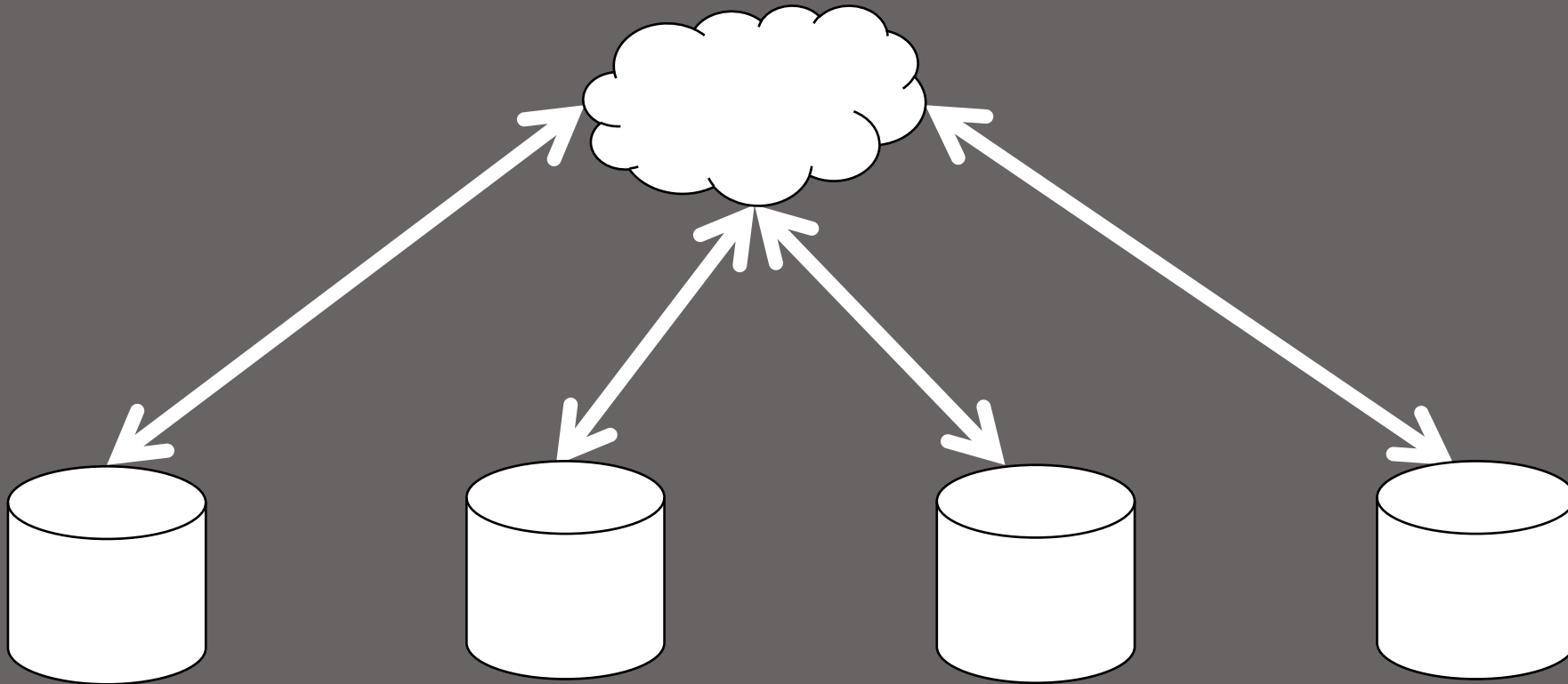- Relevant acompanying, structured data

BBMRI.cz Index

- Centralized search Project
- 2 biobanks: MMCI Brno and LF1 Prague
- Unfinished
  - Too costly
  - Difficult to extend

Federated/Distributed Search

- Searching or querying multiple "Collections" from a single entry point

Google

Google Search   I'm Feeling Lucky

Diagnosis ICD-10   = ⌄   e.g. C25.1   ⊖ ⊕

ADD FIELD ⌄

✕ CLEAR   🖉 EDIT   ⏀ SEND

🤝	🏢	15	👤	191310	🧪	923880
☐	Essen			42180		199750
☐	Hannover			41230		134200
☐	Heidelberg			34080		110810
☐	Aachen			15460		81090
☐	Leipzig			14960		89670
☐	Frankfurt			12880		50940
☐	Würzburg			7600		32600
☐	Dresden			7460		37170
☐	Jena			5860		123230
☐	Berlin			3810		34670
☐	München-HMGU			3080		18450
☐	Lübeck			1530		5440
☐	Marburg			520		3180
☐	Göttingen			400		1740
☐	Regensburg			260		940

Search

Diagnosis ICD-10   = ▾   e.g. C25.1   − +

ADD FIELD ▾

DONOR/CLINICAL INFORMATION
SEX
DIAGNOSIS ICD-10
DATE OF DIAGNOSIS
DIAGNOSIS AGE DONOR
SMOKING HABIT
BMI
BODY WEIGHT
FASTING STATUS
SAMPLE
DONOR AGE
SAMPLING DATE
SAMPLE TYPE TISSUE
SAMPLE TYPE LIQUID
STORAGE TEMPERATURE

✕ CLEAR   ✎ EDIT

| Privacy Policy | 2.1.0 (UI) | 8.6.0-SNAPSHOT (Backend)

dkfz. GERMAN CANCER RESEARCH CENTER IN THE HELMHOLTZ ASSOCIATION

Result

SPONSORED BY THE

Federal Ministry of Education and Research

Search

Diagnosis ICD-10   = ▾   e.g. C25.1   − +

ADD FIELD ▾

✕ CLEAR   ✎ EDIT   ➤ SEND

Result

| 🏢 | 1 | 👤 | 1010 | 🧪 | 10110 |

LOG IN FOR RESULTS PER BIOBANK

Most results have been received after one minute. However, in some cases it might take up to five minutes for the last results to appear.

Local Component Architecture:

Bridgehead:
- Connector (GUI)
- PSQL Database
- Blaze store (FHIR store)
- Operational Component (Under Development)

Biobanks with Bridgheads installed (5 out of 6):

	id	name				biobankid		active
1	4	Pilsen	...	...	...	bbmri-eric:ID:CZ_CUNI_PILS	...	• true
2	3	Prague (IRP)	...	...	...	bbmri-eric:ID:CZ_REVMA	...	• true
3	5	Olomouc	...	...	...	bbmri-eric:ID:CZ_UPOL_LF	...	• true
4	2	Prague (LF1)	...	...	...	bbmri-eric:ID:CZ_CUNI_LF1	...	• true
5	1	Brno	...	...	...	bbmri-eric:ID:CZ_MMCI	...	• true

Federated Search contract

- "You install the local component and fill it with data, we will take care of the rest."

Implementation Process of a FS infrastructure

1. Deploy a central SampleLocator

2. Deploy local components (Bridgeheads)

3. Set up ETL pipelines

4. Connect biobanks

Reality:

1. Convince the biobank management that FS is beneficial

2. Convince the IT dep. and ELSI teams that the system is safe

3. Convince the management and the IT dep. to export relevant data according to the given datamodel (MIABIS)

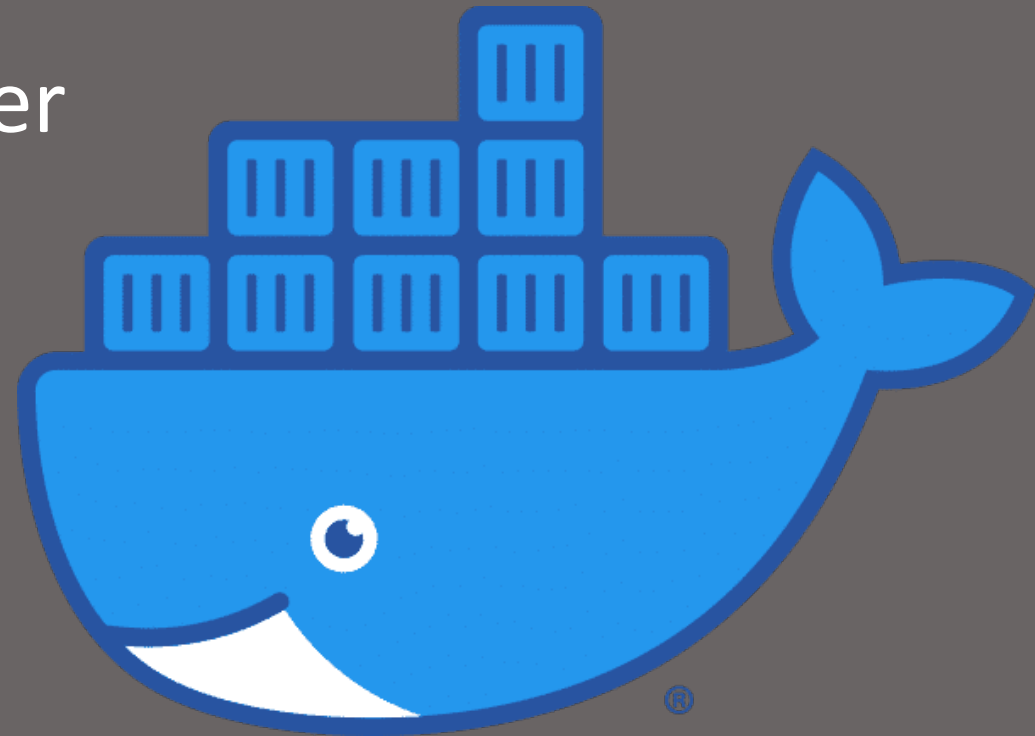4. Data Integration (HL7 FHIR)

5. Testing

Challenges

Data Model

- Usable by most Czech biobanks
- Informative enough
- MIABIS

Local Deployment

- Deployment of the Bridgehead => 3 console commands
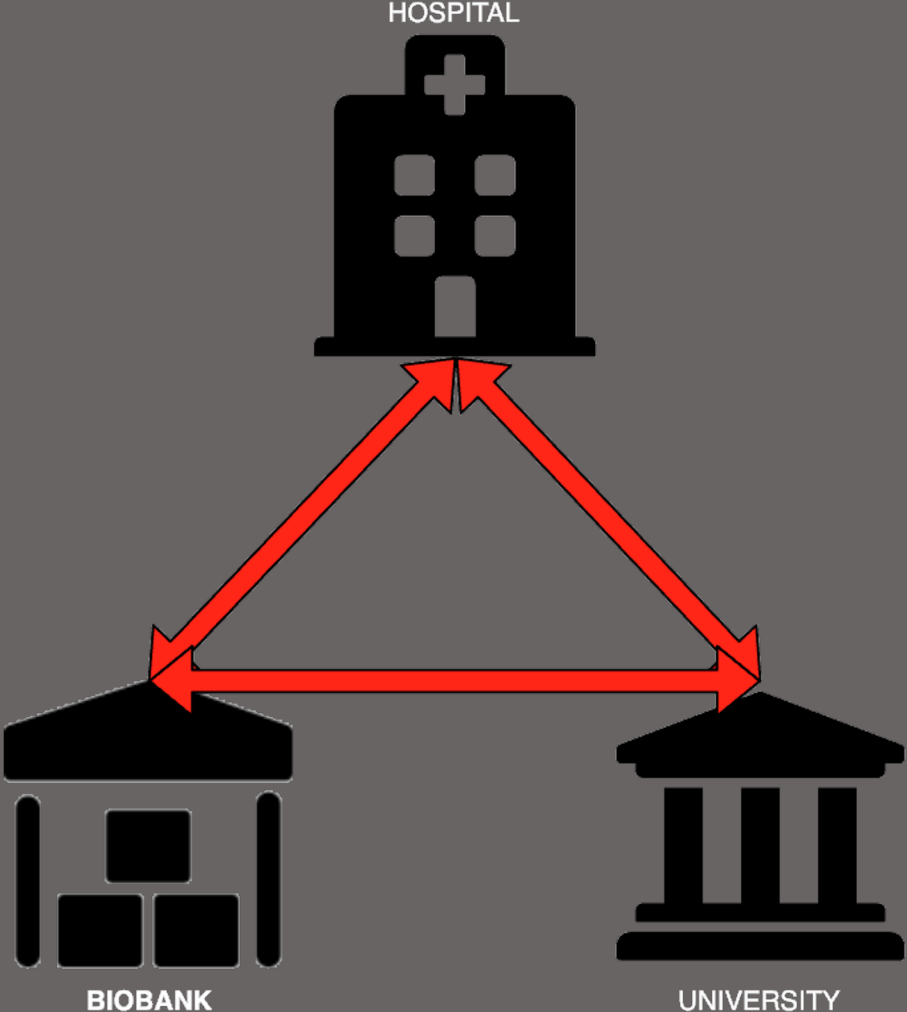- Docker
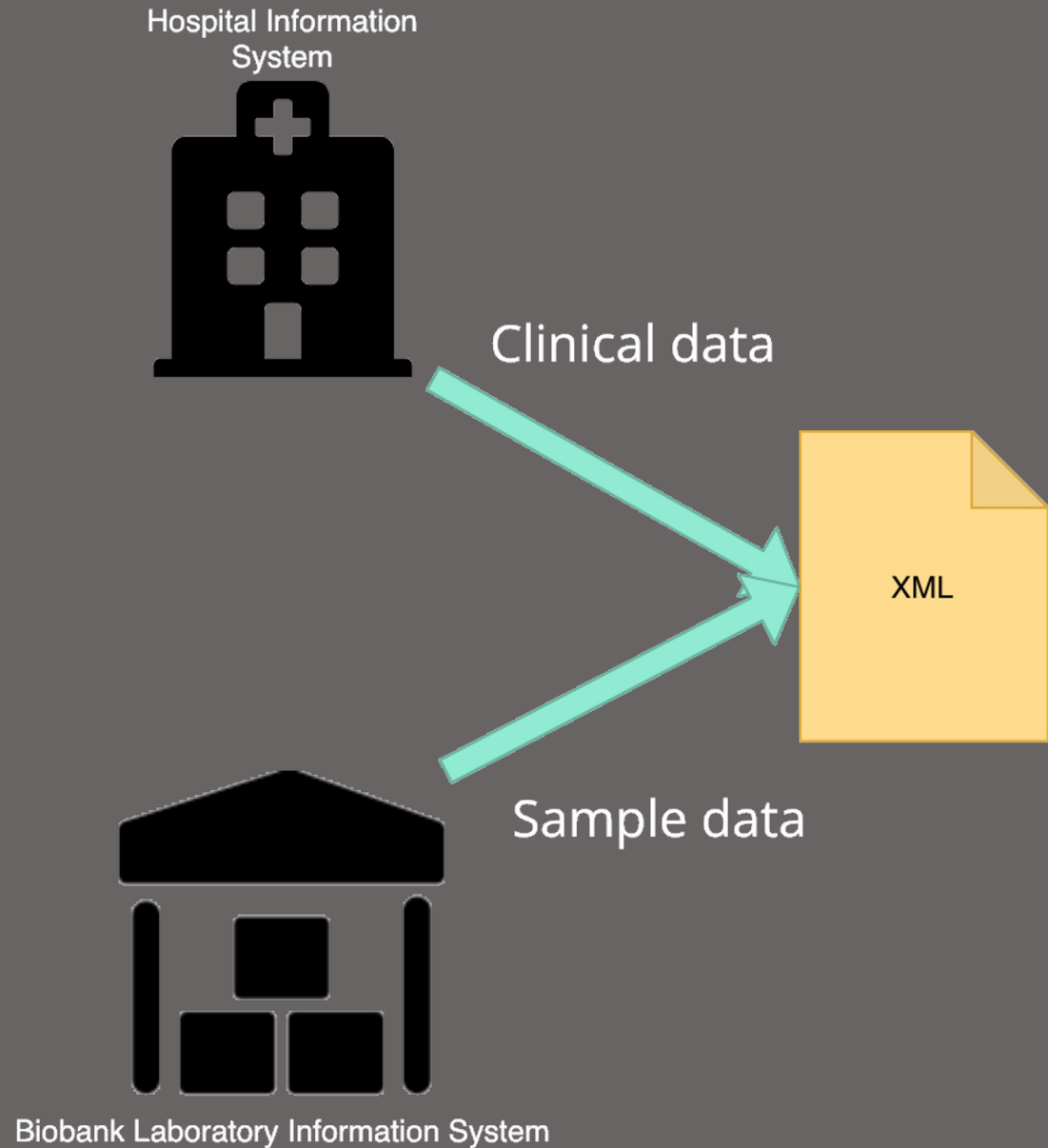- Deployable in a Kubernetes cluster
- Clear environment requirements

Security:

- Black Box environment
- No ingress from the internet
- Different user roles

Jurisdiction

Data extraction

HL7 FHIR

- Vast potential but very complex
- Few applications in the Czech Republic
- Over 90 different resources
- Few IT personnel with any experience
- Resource valiadtion

Specimen tree structure

📁 Specimen	I		Specimen
⊟ ⭐ extension	I	0..*	Extension
⭐ storageTemperature	**S** I	0..*	Extension(CodeableConcept)
⊞ ⭐ diagnosis	**S** I	0..*	Extension(CodeableConcept)
⭐ custodian	**S** I	0..1	Extension(Reference(Collection))
🔵 identifier	Σ	0..*	Identifier
🔵 accessionIdentifier	Σ	0..1	Identifier
▢ status	Σ ?!	0..1	code Binding
🔵 type	**S** Σ	0..1	CodeableConcept Binding
↗ subject	**S** Σ I	0..1	Reference(Patient)
▢ receivedTime	Σ	0..1	dateTime
↗ parent	I	0..*	Reference(Specimen)
↗ request	I	0..*	Reference(ServiceRequest)
⊟ 📁 collection	**S**	0..1	BackboneElement
↗ collector	Σ I	0..1	Reference(Practitioner | PractitionerRole)
⊞ ❓ collected[x]	**S** Σ	0..1	
🔵 duration	Σ I	0..1	Duration Binding
🔵 quantity	I	0..1	SimpleQuantity
🔵 method		0..1	CodeableConcept
⊞ 🔵 bodySite		0..1	CodeableConcept
⊞ ❓ fastingStatus[x]	**S** Σ	0..1	Binding
⊞ 📁 processing		0..*	BackboneElement
⊞ 📁 container		0..*	BackboneElement
🔵 condition	Σ	0..*	CodeableConcept Binding
🔵 note		0..*	Annotation

CORRECT:

```
{
    "url": "https://fhir.bbmri.de/StructureDefinition/Custodian",
    "valueReference": {
        "identifier": {
            "system": "http://www.bbmri-eric.eu/",
            "value": "bbmri-eric:ID:CZ_MMCI:collection:8"
        }
    }
}
```

WRONG:

```
    ],
    "collection": {
        "id": "9412"
    },
```

Development

- Additional tools and documentation for Data harmonization
- Documentation about data protection based on GBA data concept
- Bug fixing
- Improvement of Docker images for Kubernetes deployment
- Universal FHIR transformation tool

Benefits:

- The biobanks are now using structured and consistent data
- Experience with HL7 FHIR
- Promotion of the biobanks
- Tearing down misconceptions about data sharing
- Better overview of the national node's capabilities

Summary:

- Easy deployment and maintenance
- Data Integration support from the national node
- Secure data sharing
- Problematic bureaucracy

Questions?

Thank you!