



# Data harmonization of Medical Records for federated search

Radovan Tomášik

BBMRI.cz

# Structure



**15min presentation**

Context  
Harmonization



**5min for Q&A**

Introduce czech biobanks to the 21st century

Part of the BBMRI-ERIC

Provide support and infrastructure

# BBMRI-ERIC



European research infrastructure  
for biobanking



Provide tools and software  
solutions

# Problems with sharing Medical data

---

Hoarding opposed to Sharing

---

Convenience opposed to Standardization

---

Conservatism opposed to Innovation

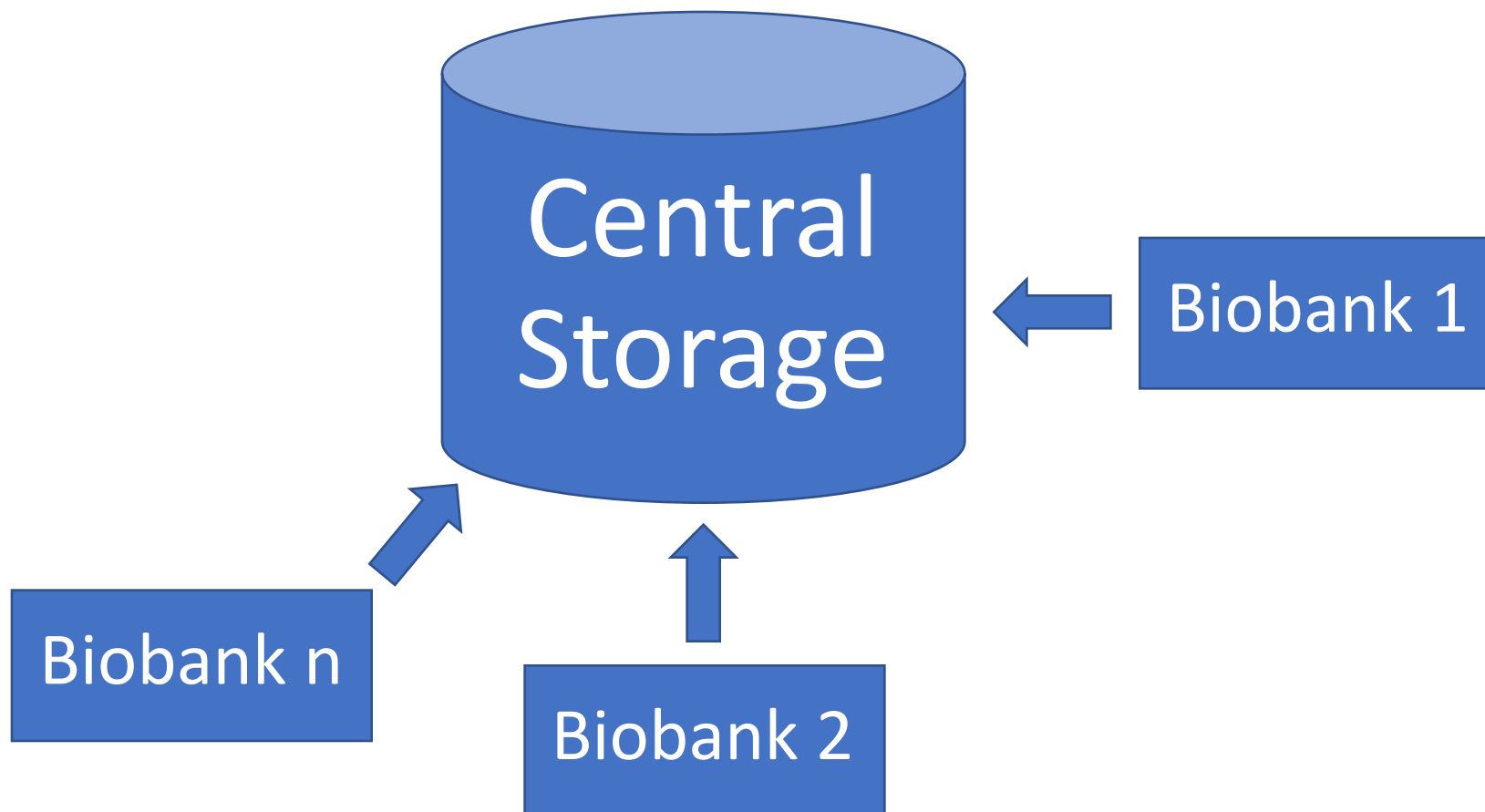
# Sharing samples

Publicly list  
samples available  
for research

What other data  
sets are available  
for those samples

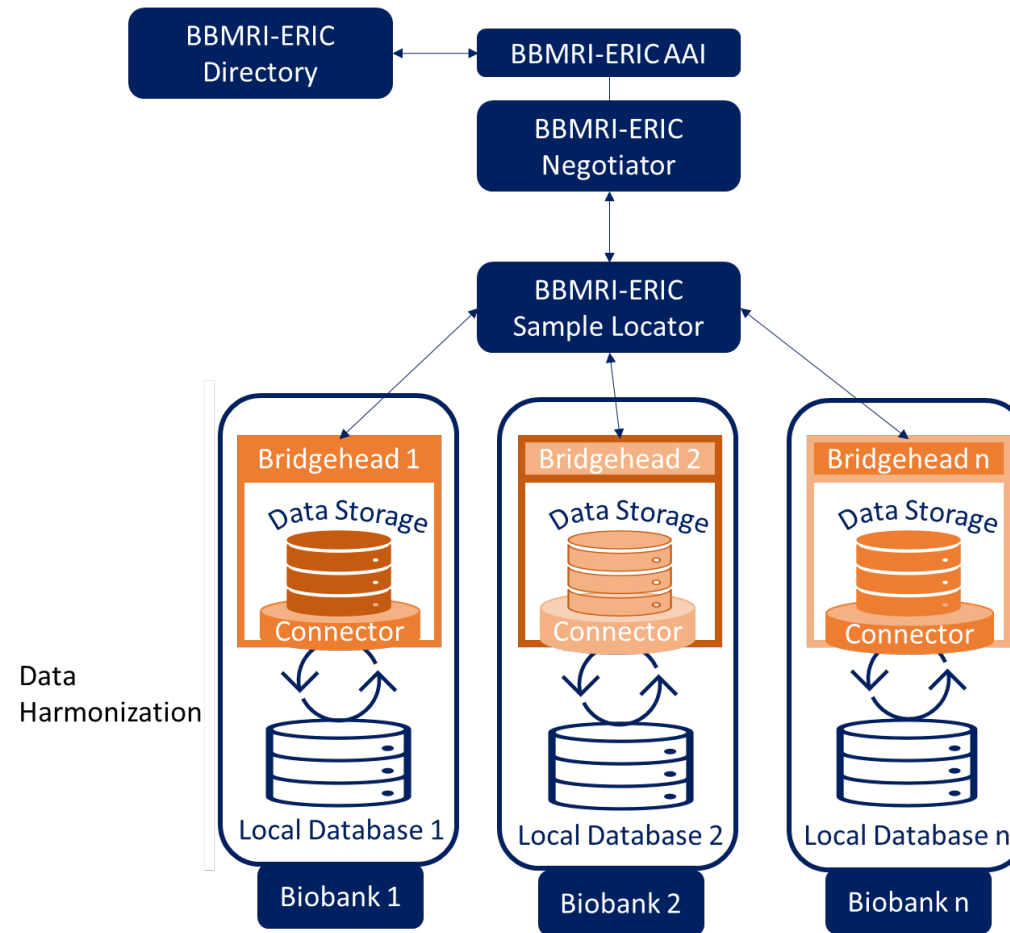
# Centralized vs Federated Search

# Centralized search





# Federated Search



# Central component



About Us   Negotiator   Login

+++ BBMRI Locator +++ search for biosamples and associated data across multiple biobanks +++

## Search

Diagnosis ICD-10

## Result

	6		134970		423200
--	---	--	--------	--	--------

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.

About Us   Negotiator   Logout  
(Radovan Tomášik)

		8		150320		514540
<input type="checkbox"/>	Hannover(DE)	41230		134200		
<input type="checkbox"/>	Heidelberg(DE)	31460		94500		
<input type="checkbox"/>	Brno(CZ)	28570		58520		
<input type="checkbox"/>	Leipzig(DE)	14960		89680		
<input checked="" type="checkbox"/>	Aachen(DE)	14080		58180		
<input checked="" type="checkbox"/>	Frankfurt(DE)	12180		46230		
<input type="checkbox"/>	Würzburg(DE)	7450		31570		
<input type="checkbox"/>	Göttingen(DE)	390		1660		

The result of your query represents only potentially available biosamples and data. Further information regarding access can be retrieved by direct communication with the respective biobanks activating the checkboxes and clicking the "Negotiate" button.

NEGOTIATE

# Local components of federated search

01

Install local components behind the institution's firewall

02

Open specific ports

03

Limit access from within the institution

# Extract Transform Load

Extract

Create export  
from (usually 2)  
databases

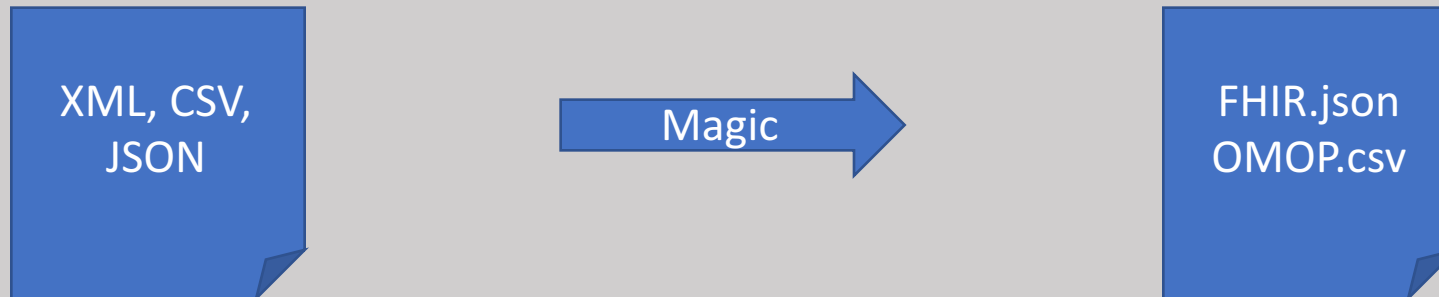
Transform

Harmonize the  
data into a  
standard format (  
OMOP, FHIR)

Load

Upload the data  
into the local  
component

# Transform/Harmonize



# Example OMOP File

.....  
.....  
.....

person\_2.csv

×

```
PERSON_ID,GENDER_CONCEPT_ID,YEAR_OF_BIRTH,MONTH_OF_BIRTH,DAY_OF_BIRTH,BIRTH_DATETIME,RACE_CONCEPT_ID,ETHNICITY_CONCEPT_ID,LOCATION_ID,PROVIDER_ID,CARE_SITE_ID,PERSON_SOURCE_VALUE,GENDER_SOURCE_VALUE,GENDER_SOURCE_CONCEPT_ID,RACE_SOURCE_VALUE,RACE_SOURCE_CONCEPT_ID,ETHNICITY_SOURCE_VALUE,ETHNICITY_SOURCE_CONCEPT_ID
1,8507,1923,5,1,,8527,38003564,1,,00013D2EFD8E45D1,1,,1,,1,
10,8532,1938,10,1,,8516,38003564,10,,00036A21B65B0206,2,,2,,2,
100,8532,1934,4,1,,8527,38003564,90,,002DF56C161DA9BF,2,,1,,1,
1000,8532,1918,4,1,,8527,38003564,504,,0212C0C0E325698E,2,,1,,1,
10000,8532,1939,7,1,,8527,38003564,1511,,159C97CBA5C3027F,2,,1,,1,
100000,8532,1941,2,1,,8527,38003564,618,,DBFB26CD49DC0E24,2,,1,,1,
100001,8532,1943,9,1,,8527,38003564,30,,DBFB408E1476556B,2,,1,,1,
```

.....  
.....  
.....

.....  
.....  
.....  
.....



# OMOP CDM

The Observational Medical Outcomes Partnership Common Data Model

The logo for Fast Healthcare Interoperability Resources (FHIR) is a large, dark grey circle with the letters 'FHIR' in white, sans-serif font. It is positioned on the left side of the slide. There are also several faint, overlapping circles and dotted lines around it, creating a layered effect.

FHIR

Fast Healthcare  
Interoperability Resources

Framework comprised of 92  
so called “Resources”

Usable for a variety of  
medical data



# Example XML patient file

```
<patient id="16872164005" consent="true" year="1980" month="--02" sex="female" b
  xmlns="http://www.bbMRI.cz/schemas/biobank/data">
  <LTS>
    <!--Poradove cislo v ramci roku -->
    <!-- sampleId = rok + number + typ-->
    <tissue year="2008"
      number="4028"
      sampleId="200840281">
      <!--Pocet alikvotu -->
      <samplesNo>4</samplesNo>
      <!--Pocet vydatelnych alikvotu -->
      <availableSamplesNo>2</availableSamplesNo>
      <!--Ciselnik - 1 znamena nador v ramci MOU-->
      <materialType>1</materialType>
      <!--TNM klasifikace -->
      <TNM>T1aN0M0</TNM>
      <!--pTNM klasifikace -->
      <pTNM>T1aN0M0</pTNM>
      <!--morfologicka klasifikace -->
      <morphology>0000/11</morphology>
      <!--cas odebrani -->
      <cutTime>2012-05-15T08:30:00</cutTime>
      <!--cas zamrazeni -->
      <freezeTime>2012-05-21T10:30:00</freezeTime>
      <!--vzorek odebran pri operaci -->
      <retrieved>operational</retrieved>
    </tissue>
  </LTS>
  <STS>
    <diagnosisMaterial year="2010"
      number=""
      sampleId="20084029P">
      <!-- P = plasma -->
      <materialType>P</materialType>
      <diagnosis>N26</diagnosis>
      <takingDate>2012-04-21T10:30:00</takingDate>
      <retrieved>unknown</retrieved>
    </diagnosisMaterial>
  </STS>
</patient>
```

# Example FHIR File

```
{
  "entry": [
    {
      "fullUrl": "http://example.com/Patient/bbmri-69",
      "request": {
        "method": "PUT",
        "url": "Patient/bbmri-69"
      },
      "resource": {
        "birthDate": "1999-05-01",
        "gender": "male",
        "id": "bbmri-69",
        "meta": {
          "profile": [
            "https://fhir.bbmri.de/StructureDefinition/Patient"
          ]
        },
        "resourceType": "Patient"
      }
    },
    {
      "fullUrl": "http://example.com/Condition/bbmri-69-condition-0",
      "request": {
        "method": "PUT",
        "url": "Condition/bbmri-69-condition-0"
      },
      "resource": {
        "code": {
          "coding": [
            {
              "code": "C541",
              "system": "http://hl7.org/fhir/sid/icd-10",
              "version": "2016"
            }
          ]
        }
      }
    }
  ]
}
```

Issues that  
need to be  
addressed

Mapping

Extending the data model

Updating available samples

Updating patient consent

Correct formats

# Mapping

Impossible to do a  
general mapping guide

1:1, 1:2 or even 1:3  
mapping

Biobanks don't have the  
IT resources needed

## Leave the biobanks to do the mapping themselves

- Expensive for the biobanks
- hard to maintain
- No benefits for others
- Off-putting
- Takes too long

## Support the biobanks by providing a set of tools for setting up the ETL pipeline

- Documentation!
- GUI app for less tech savvy biobanks
- Libraries for more demanding cases

Questions?



Thank you for your  
attention

Radovan Tomášik