Hackathon Registers of Cultural Entities

Reprex non-competing pitch

No statisticaly valid data

You cannot send a survey questionnaire if you do not know who should get it.

No coordination of treasures

You cannot lend books across libraries because they do not know other libraries and cannot send a copy there.

Low visibility & efficiency

It is impossible to support the visibility and popularity of Slovak cultural products if you do not know whom to support.

No policy control

The Slovak state, regional and municipal bodies do not know who needs to be helped.

No business controlling

Cultural institution managers do not know if they are doing a good or bad job, no clear comparisons.

No cooperation

If you do not know the other peers, then you cannot solve problems together.

Building the Slovak Music Dataspace

We realised these problems in the Slovak Music sector and we are close to a comprehensive solution

2020

First quantitative data mapping of the Slovak music industry; first in-depth CCI report in Slovakia with extensive data collection.

2020

Feasibility Study: Identification of serious data problems, demo version of the Slovak Comprehensive Music Database with a registry.

2021-2022

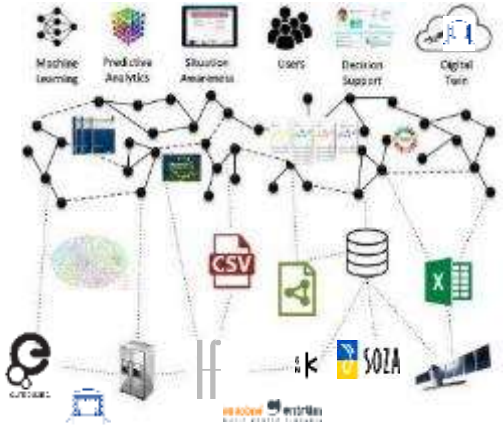
Listen Local App Demo

Various data collections

Open Music Europe Horizon Europe project

2023

Memorandum of Understranding Cultural ministry, IKP, EUBA, Reprex, SOZA & Consortium



2024

Slovak Music Dataspace Wikipedia/Wikidata

Extendible via federation to many countries in Europe

We considered the best practices in the world

We build a system from the best information (conceptual) models and open-source linked database components

Satellite Business Registers

Luxembourg Shared Authority File + Wikibase

University of Luxembourg
Competence Centre

WIKIBASE

WIKIDATA

EU Open Data Portal Statistical Data & Metadata eXchange

EUROPEAN DATA PORTAL

interoperable europe

StatDCAT Europe

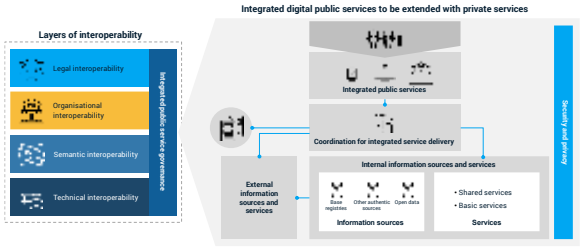
sdmx

Cultural Information Models

europeana

CIDOC CRM
CONCEPTUAL REFERENCE MODEL

European Interoperability Framework



BVDA/Gaia-X Data (Sharing) Space

BDV gaia-x

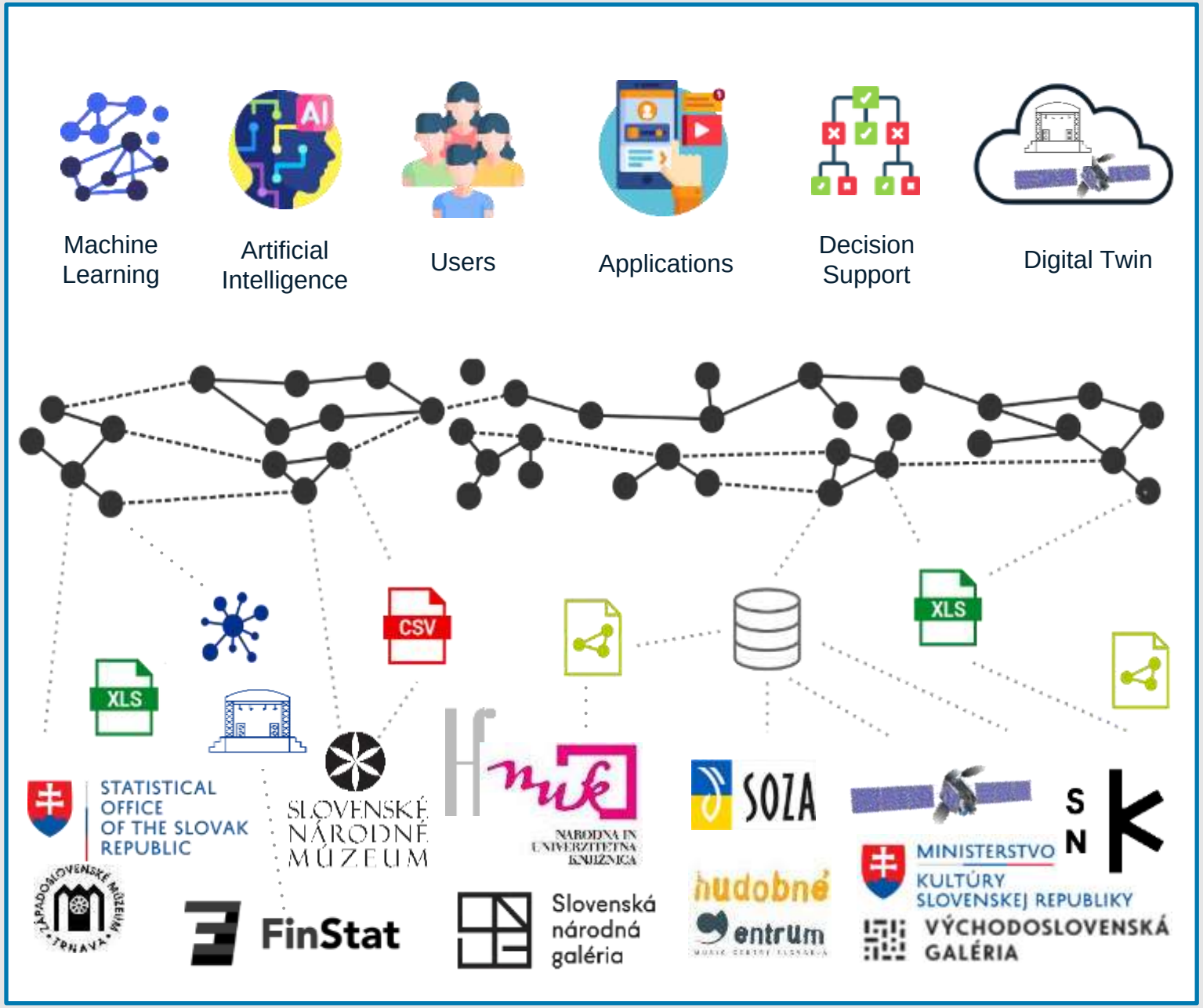
Slovak Cultural Dataspace

Application layer: Listen Local Apps for radio or educational playlist editing. Name Entity Matching and Name Entity Disambiguation for rights management.

Semantic layer: data coordination among public, private, local and international data owners.

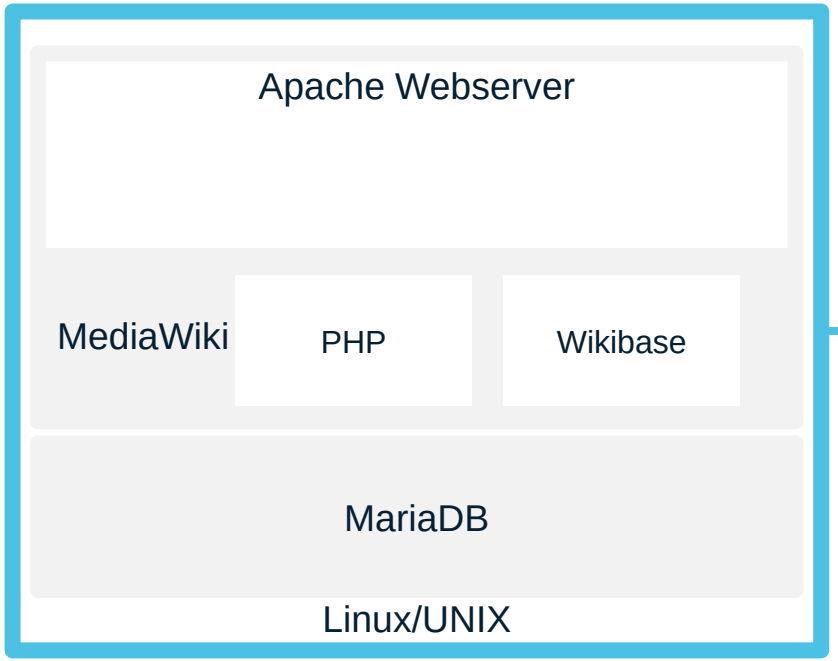
Database layer: databases of participating organisations, shared only as needed and as permitted.

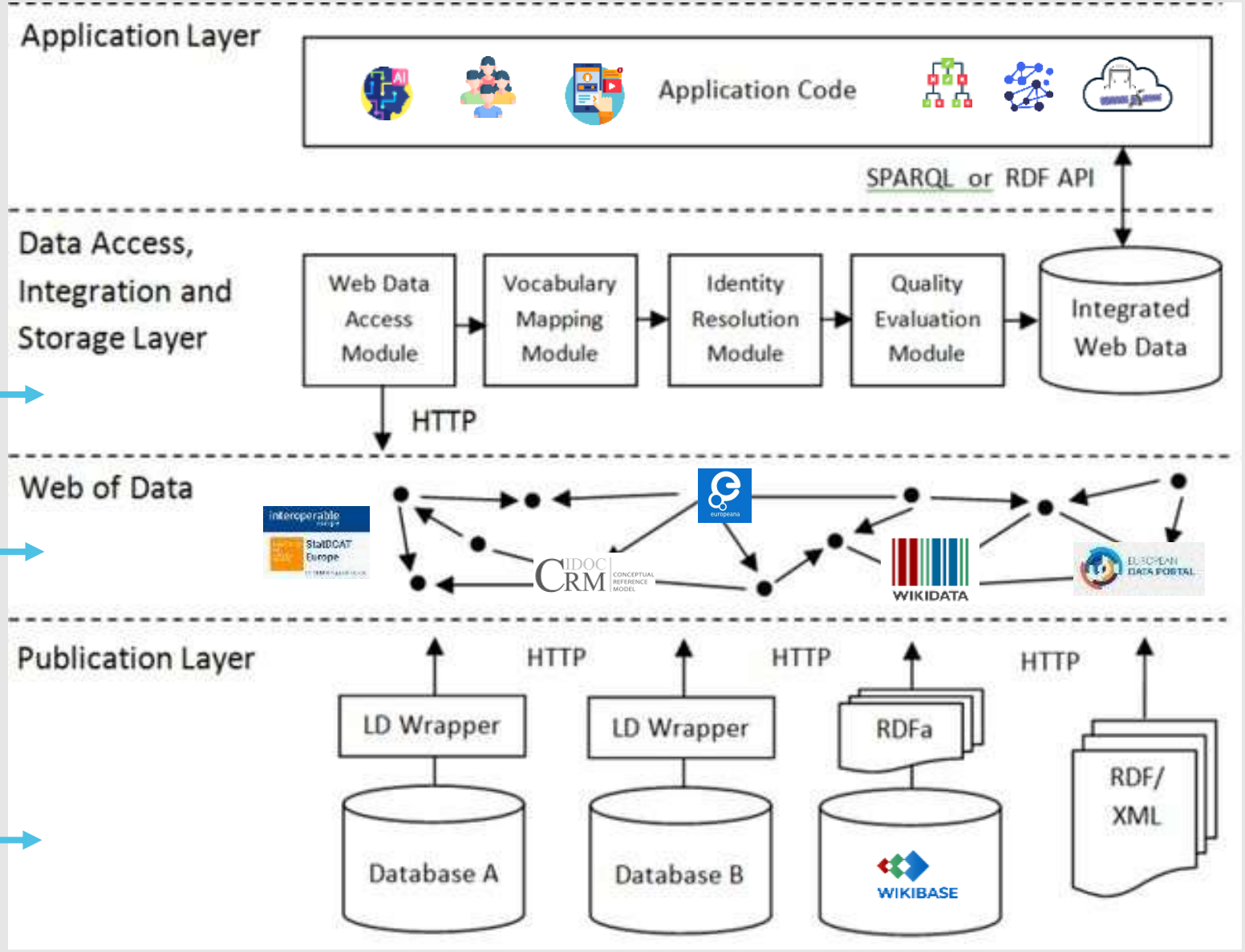
Collection layer: data collection internally (existing databases, registers, membership lists) and external data sources.



System Components

Open source software and interoperable, open information models





We are able to meet most requirements in the first 2 months

Most features will be available in the first months, some will be deployed around M3 or later

Public and private sections	We create a private Wikibase instance with strict hierarchical administration, and train ministry administrators.	
Automated data collection	We start first with automated data collection in music, and gradually extend the scope to other sectors.	
Manual registration of entities	Manual data entry will be possible from deployment day.	
Categorization of entities by cultural sectors	We will create an initial taxonomy that we will align with international ontology standards in several iterations.	
Import and export of data	Initially we support Excel, CSV, RDF, export/import, and import from music databases. In subsequent months we build tools for existing fixed-schema RMDBS systems in SQL as needed.	
Output data for managerial decision-making	Managerial decisions data will be available once the system is reasonably well populated, around M3.	
Dynamic pages, widgets	Dynamic pages, dashboards, widgets can be developed in early phases but we believe that sufficient experience with managerial data is needed to design lasting, user-friendly dynamic applications.	
Automated processing of statistical results	Our systems use the novel EU/UN methodology for extending official governmental statistics with satellites. We create “official” quality statistical output with our own peer-reviewed, open-source software.	
Trustworthy AI	Our system works as an explicit knowledge base (as defined by the EU AI Act) and we extend its functionality with trustworthy AI functions to improve data quality and knowledge coverage.	

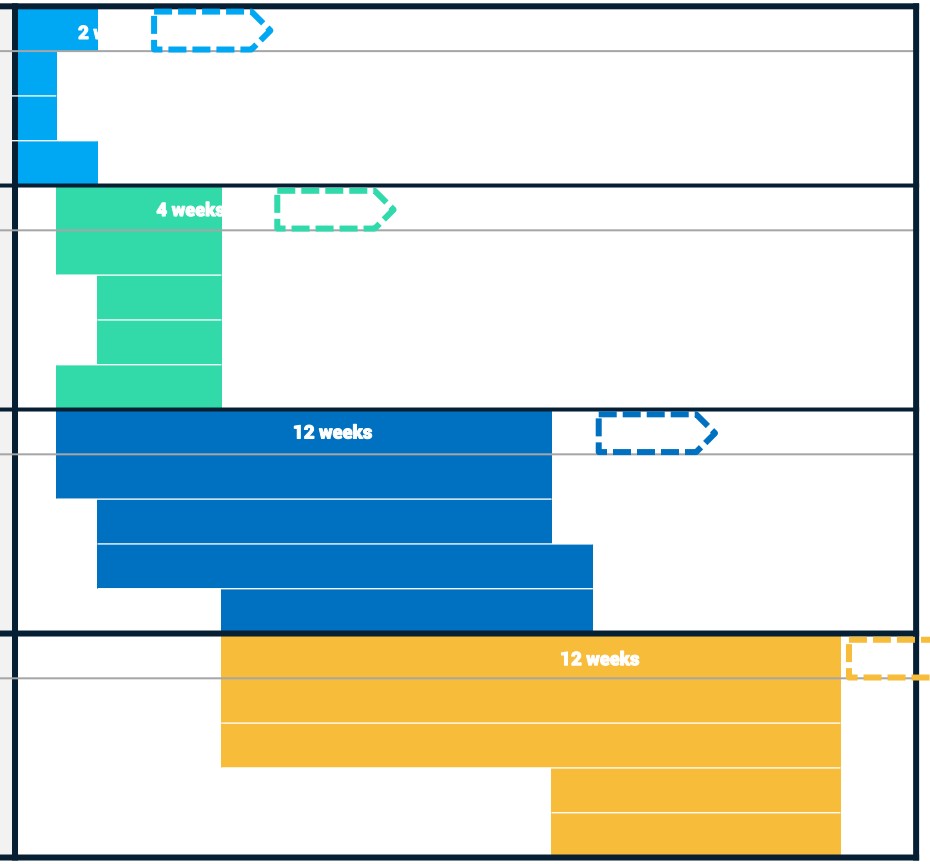
Source: Based on open call

Phase 1 Phase 2 Phase 3 Phase 4

System is usable in about 12 weeks and extendible with new services

The most resource-intensive tasks are workflow planning, education, onboarding, the system itself is made of open-source components that are widely used worldwide for similar problems.

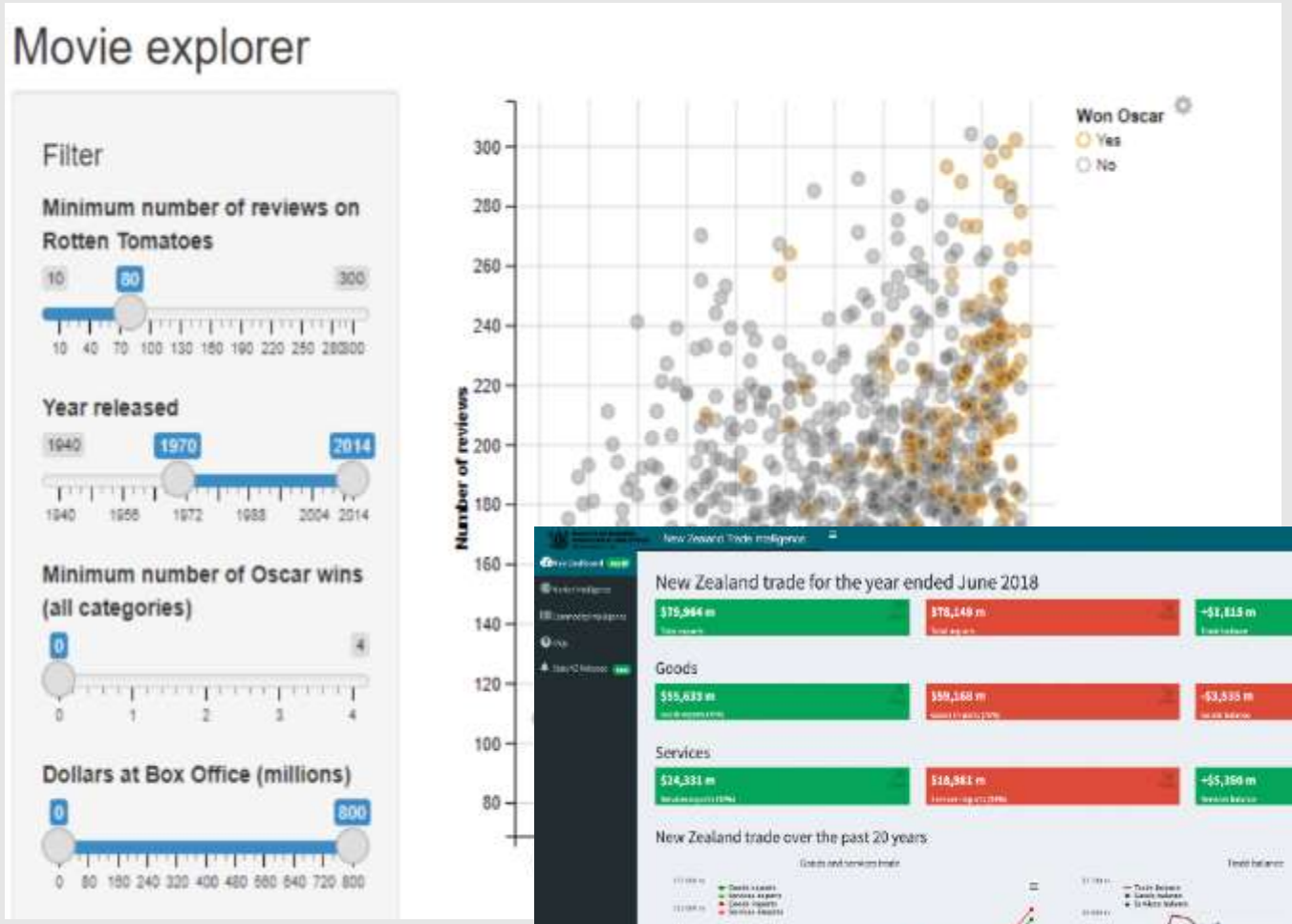
1. System Deployment
1.1. All data recording, exporting functions work; dedicated cloud server or local server
1.2. Internationally interoperable taxonomy for organisations
1.3. Automated import from Ministry of Culture, music works, Wikidata, VIAF
2. Maintenance workflow and training
2.1. Training registry administrators for onboarding new organisations
2.2. Defining management information outputs
2.3. GDPR and data protection training, legal interoperability
2.4. Onboarding case study with one cultural field (for example, libraries)
3. Onboarding and partner organisation education
3.1. Onboarding more and more private registries, or other public registries
3.2. Importing historical data for comparison and change tracking, ideally since 1993
3.3. Writing open-source connector to legacy systems, partner cultural organisation systems
3.4. Web-scraping information
4. Application development
4.1. Onboarding more and more private registries, or other public registries
4.2. NER, NED entity, link matching with machine learning and deductive engines
4.3. Statistical application that crates satellite statistics (well aligned with KULT microdata)
4.4. Dynamic dashboards



Shiny Apps & Dashboards

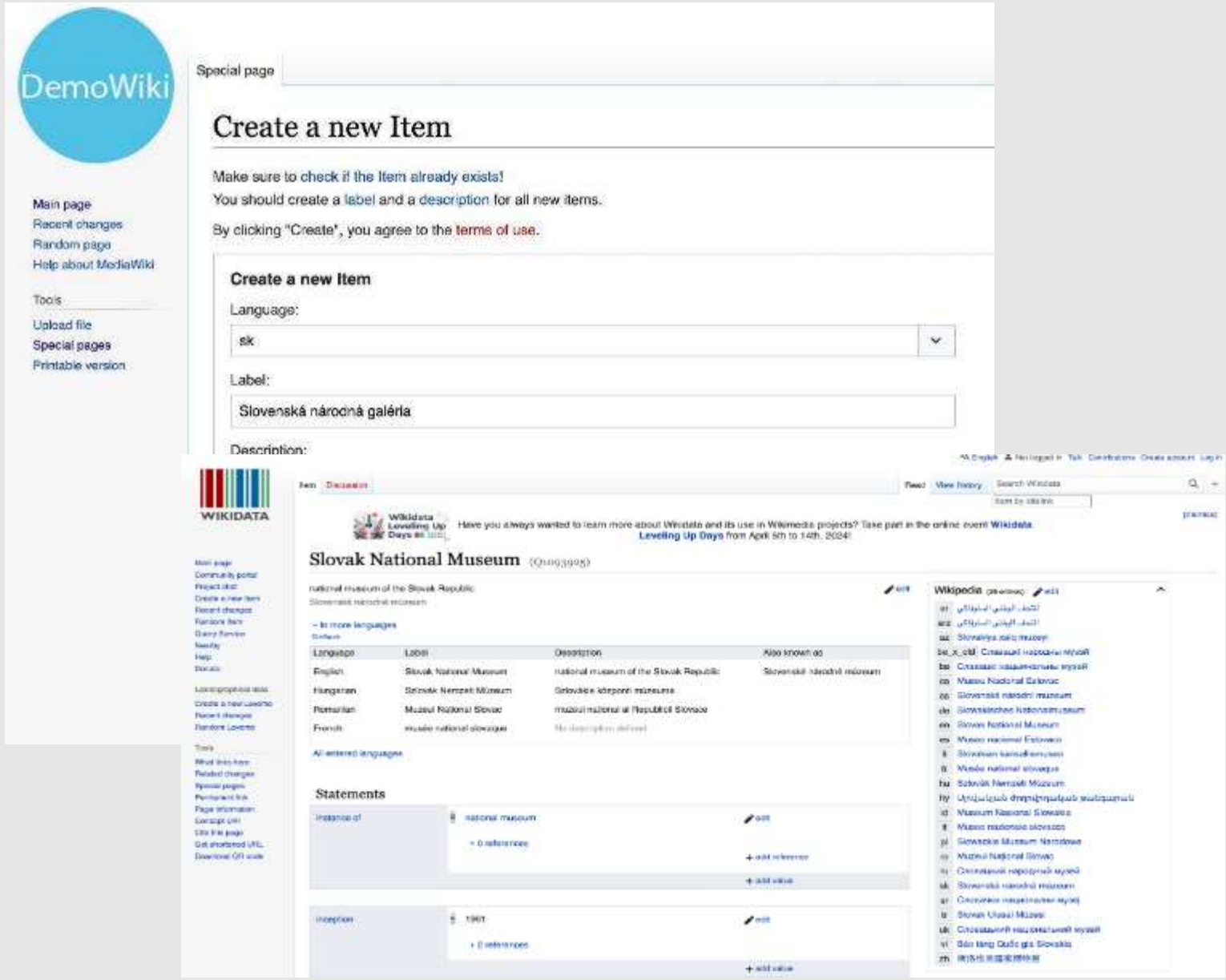
Shiny Apps are extending the popular, open-source R statistical environment and they can be deployed in hours.

Reprex developed several official extensions to this globally used system; some of them were developed with Slovak data.



Well-known, easy to use GUI

Wikidata and Wikipedia were designed for citizen scientists and



Value proposition

A.
Ready to deploy in a month

B.
Open source, open knowledge, modular, trustworthy AI

C.
Sound data coordination and validation workflows used in other countries

Ready to deploy in a month

Open-source solutions have a very high **Technical Readiness Level**, and deployment in music is already under way with much more granular data

A.
Ready to deploy in a month

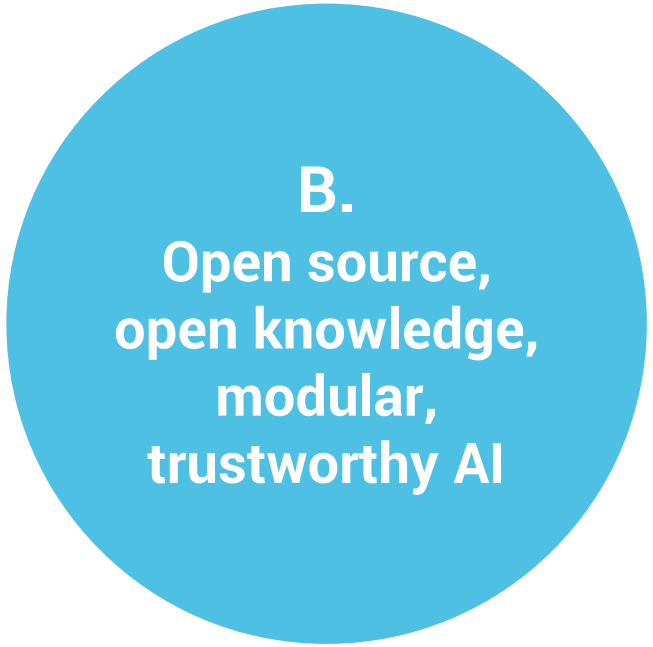
Very fast deployment

Our system is modular and uses trusted open-source systems that other EU member states use for similar purposes. We will launch our music dataspace in June, and it can include other sectors from Day 1.

Our legal interoperability model can be used in other sectors, can go well beyond the needs of this hackathon objectives.7.3.

Open for collaboration with other hackathon contestants

We use only transparent, open-source solutions, which are easy to replace or fix, and non-sensitive data can be synchronized with public knowledge systems like Wikipedia



Open Collaboration model

1. We have designed our solution for open collaboration from the onset, and prepared a cooperation with **Wikipedia Slovensko** and the Wikipedia Foundation.
2. Because we are working in open collaboration, **we can host the solutions of other competing teams**!
3. Algorithm only uses verifiable, trustworthy, explicit knowledge, **no black box guesswork**.

Sound data coordination and validation workflows used in other countries

Introduce the **European Interoperability Framework** into the cultural organisations of the Slovak Republic, extending the data standards to non-state entities like art galleries or orchestras, too

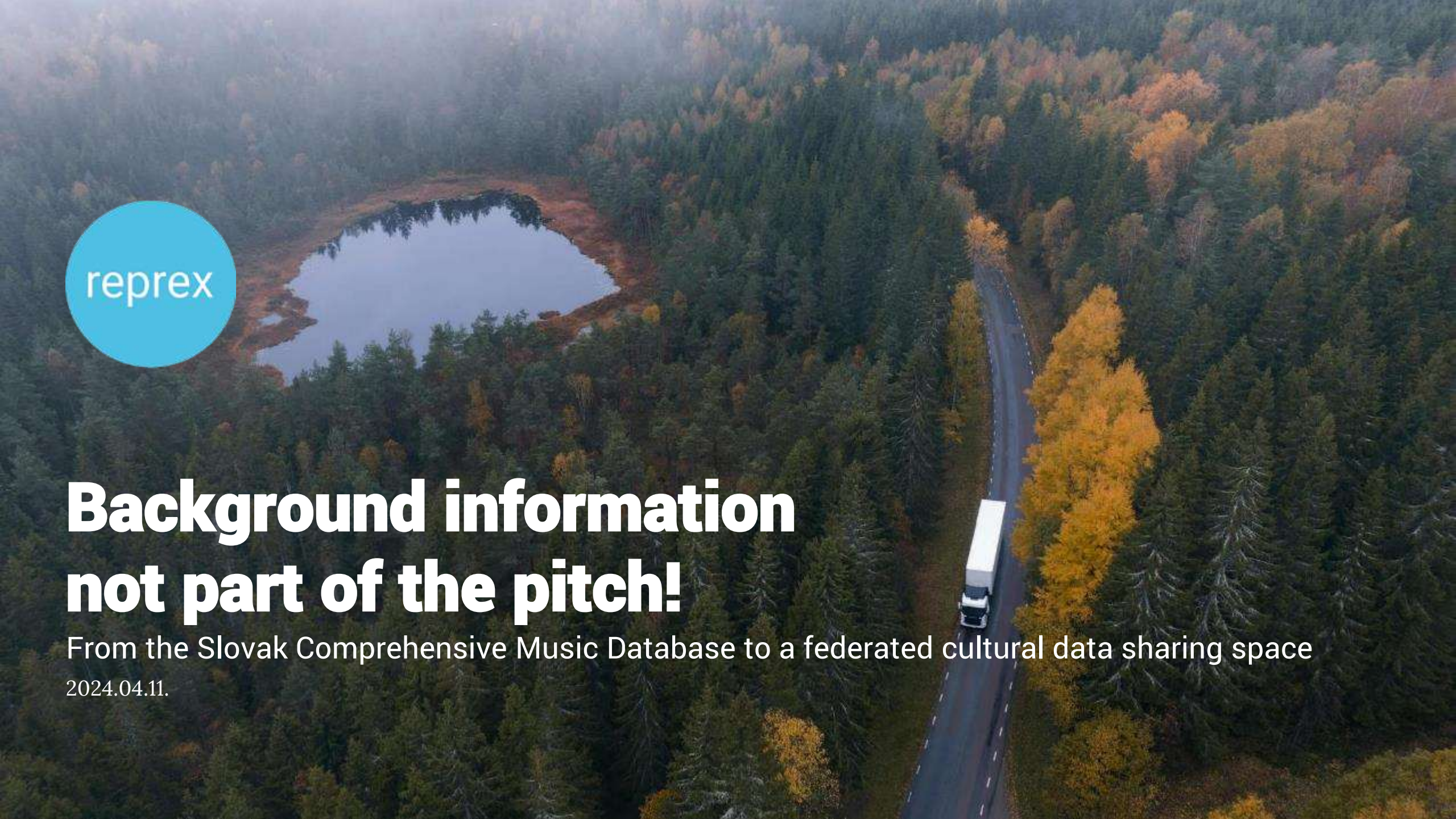
Internationally interoperable worfklow

Our system automates data collection, data exchange and validation workflows that are considered valid and sound by statistical authorities, international museum and heritage organisations, libraries, and other domain-specific standards. We know how to translate their knowledge into the ministry's data systems in the correct form and semantic meaning.

Thank you for your attention

Questions? Get in touch:

reprex.nl/contact
www.linkedin.com/in/antaldaniel



Background information not part of the pitch!

From the Slovak Comprehensive Music Database to a federated cultural data sharing space

2024.04.11.

Problems with data

01

Hackathon - Registers of Cultural Entities

Initial Situation

The Ministry of Culture of the Slovak Republic does not have a register of all existing legal entities in the field of art, culture, and the creative industries, making it difficult to implement measures aimed at the entire sector. Its significance became fully apparent after the outbreak of the Covid-19 pandemic when the ministry needed to target governmental support, but it did not have a complete list of eligible entities

Existing registers

- Sectors of cultural heritage (museums, galleries, libraries) – nearing 100% repletion
- Art sector (visual arts, theaters, musical entities) – low repletion
- Sectors falling under KKP (architecture, design, gaming industry) – no registers

Expected solution

A method of creating an information system (IS) that will contain registers of legal entities whose main activities have an artistic or cultural character, as well as the method of filling it with data and how to obtain, process, and display this data.

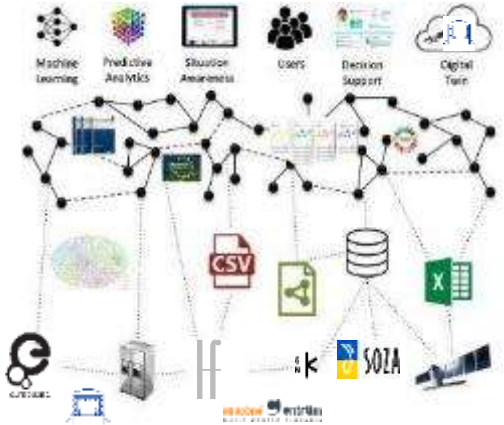
Source: [Sanitized]

History of Innovation in Slovakia

02

Building the Slovak Music Dataspace

Our roadmap is replicable, but requires many conversations that build trust among institutions



2020

Feasibility Study: how to support the policy goals of the local content regulations (from radio to streaming) to support the local music ecosystem. Identification of serious data problems.

2021-2022

Listen Local App Demo

Various data collections

Open Music Europe Horizon Europe projekt

2023

Memorandum of Understranding Cultural ministry, IKP, EUBA, Reprex, SOZA & Consortium

2024

Slovak Music Dataspace Wikipedia/Wikidata

Extendible via federation to many countries in Europe

2025

Listen Local app for radio and streaming playlisting HearDis! in-store demonstration Unlabel label services for the unlabeld

Sustainability

Why do they not recommend music from Slovakia in Slovakia for Slovak people?

How can we avoid that small repertoires lose their original market, their only market?

Findings

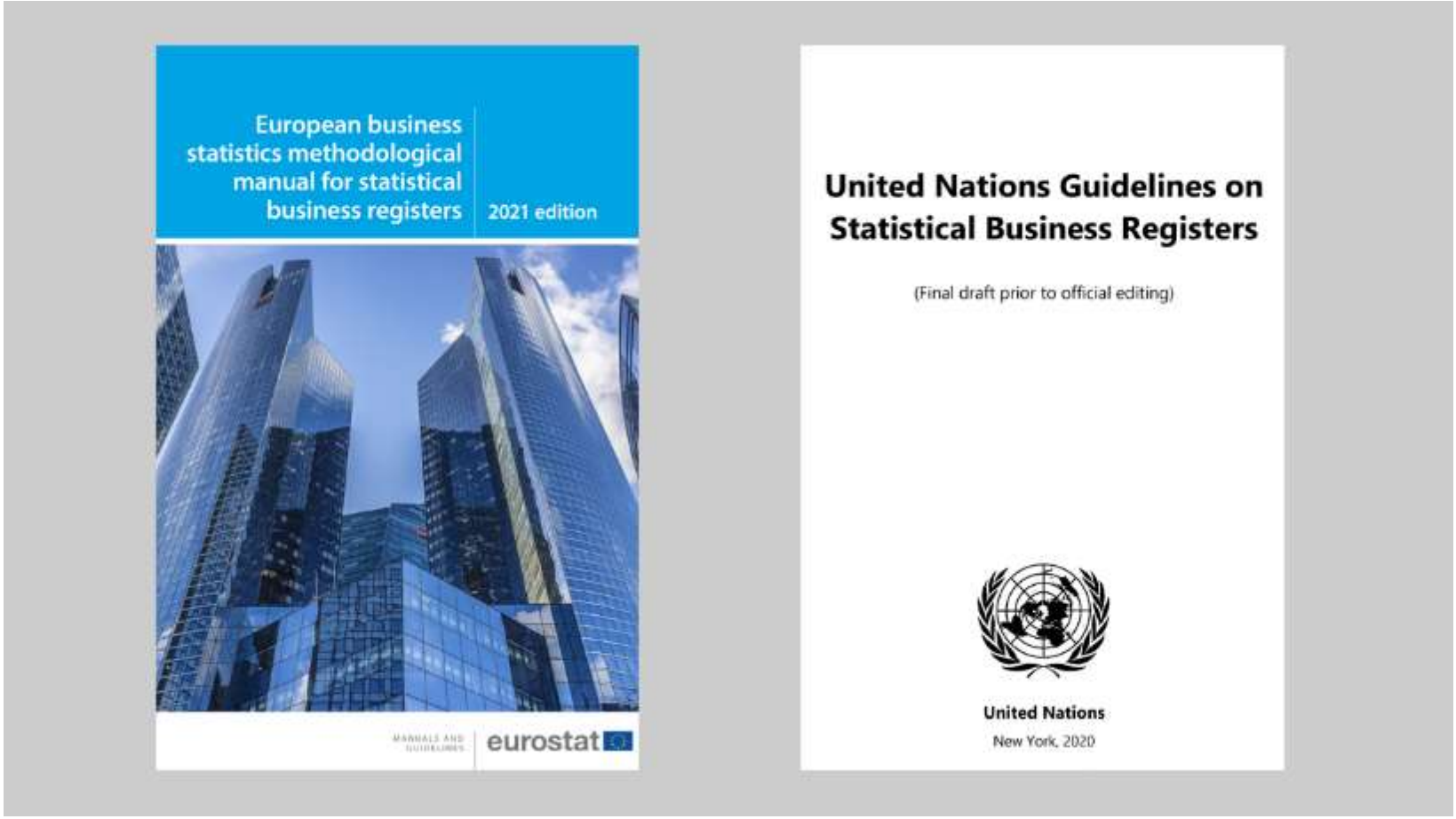
- Small repertoires have very low-quality data representation
- Because of low income, data improvement and documentation must not be expensive
- Public-Private Partnership is needed to solve data problems

Building blocks: the best practices

03

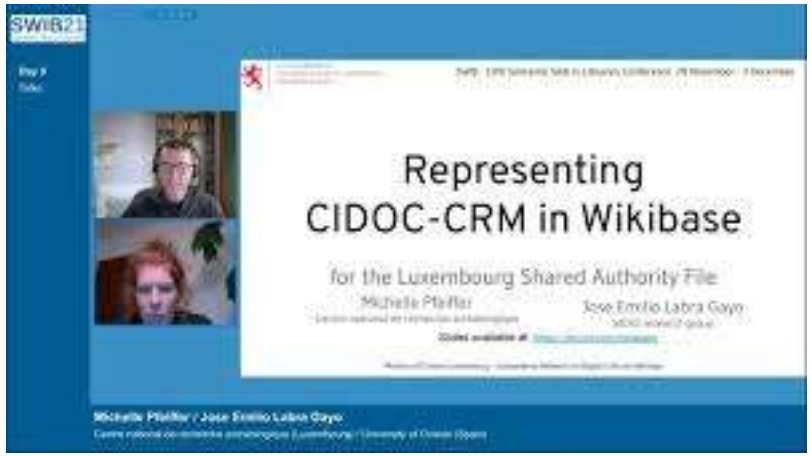
Satellite Business Registers

New methods and regulations to allow the coordination of state statistical registers with privately-held registers, such the registers of SOZA or the chamber of architects



The Luxembourg Shared Authority File & Wikibase

University of Luxembourg
Competence Centre



WIKIDATA

WIKIBASE

EU Open Data Portal, Statistical Data & Metadata eXchange

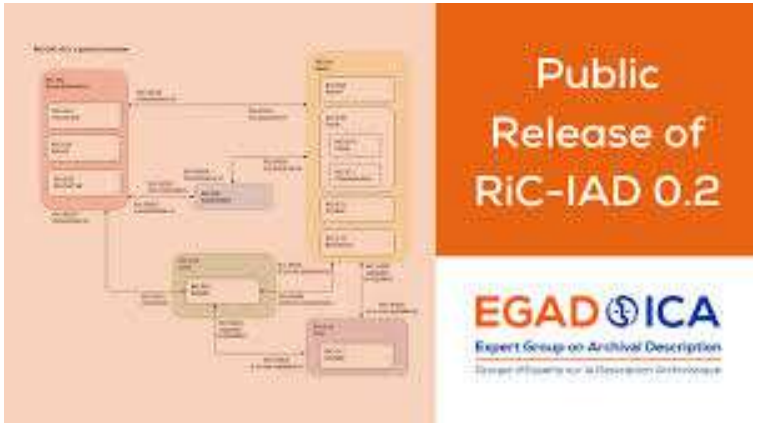
DCAT-AP and its statistical modification

Cultural Information Models

In the European cultural sector, the Europeana Data Model is a de-facto interoperability minimum standard, with museums and national archives using more complex conceptual models

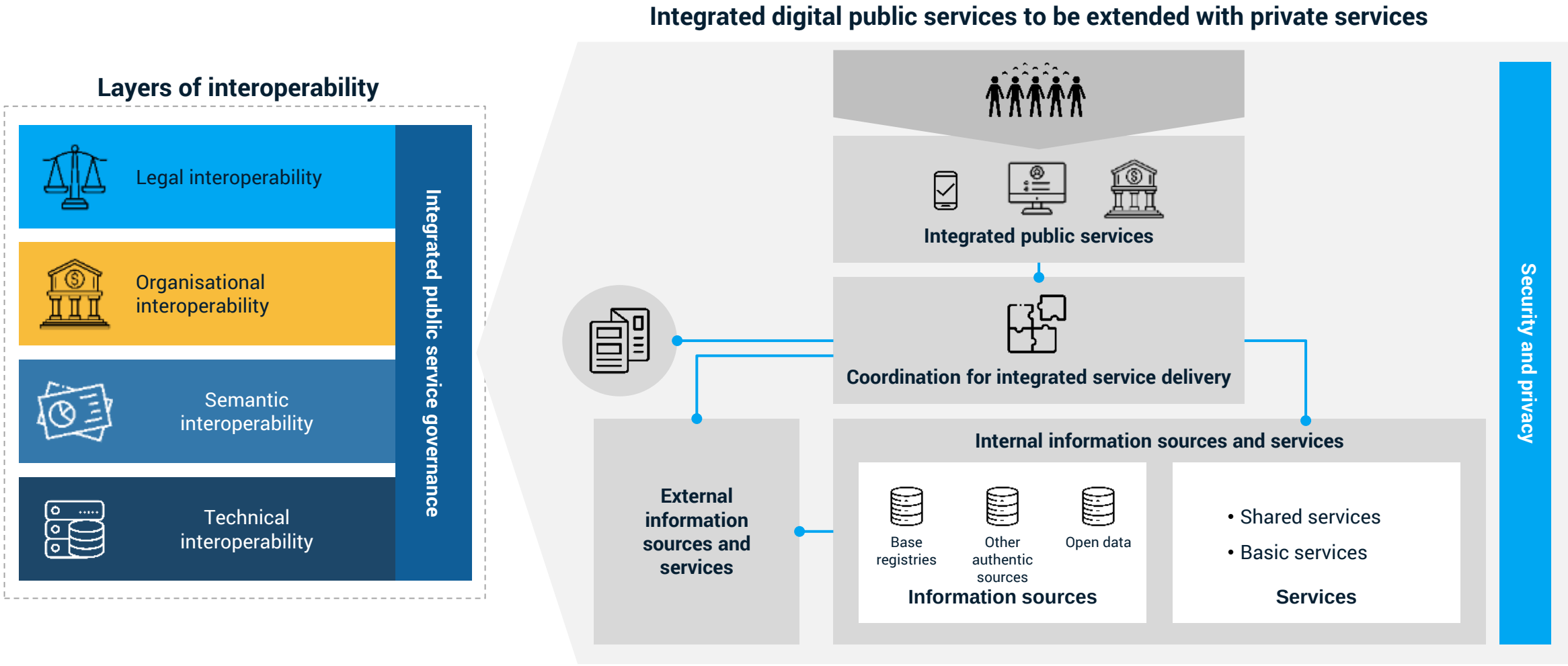
europeana

CIDOC CRM | CONCEPTUAL REFERENCE MODEL



Embracing the European Interoperability Framework

Extending the interoperability of digital services among public and private parties



BVDA/Gaia-X Data (Sharing) Space

A novel legal, organisational and technical institution introduced by the Data Governance Act, and developed by Gaia-X and the Big Data Value Association for the EU and its member states

Our solution: Slovak Cultural Dataspace

04

Slovak Cultural Dataspace

Maintainance process

A strong business process that ensures all data is found and updated when changes happen; proper information model to automate the process.

Knowledge graph with database GUI

We build on Wikibase, a graph database of Wikipedia, because it is aimed at citizen scientist level, and has an intuitive GUI.

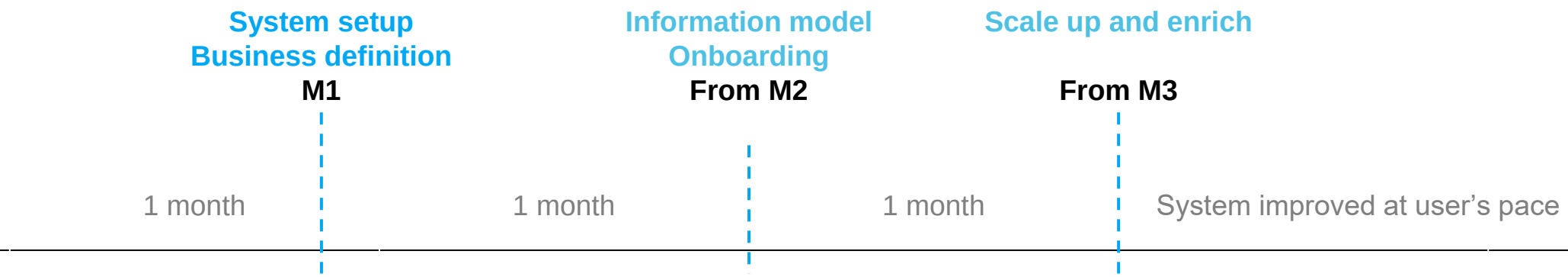
Apps for Analysis, Export/Import

We create dashboards, statistical applications, and export import tools to/from Excel, CSV, SQL and RDF.

Achieve strong, long-term future-proof solution

Source: [Sanitized]

Implementation Phases



Business process
Define the business process, and derive the supporting information model for high level of automation. The information model contains taxonomies of entities.

Slovak Shared Authority File: Wikibase with authority control
We adapt to Slovak institutional framework the Luxembourg Shared Authority File

Information model
Ontology and data scripts (OWL/RDF/R) to for automated data maintenance

Manual oversight and manual entry
Simple, widely used GUI with plenty of Slovak, English documentation: Wikipedia

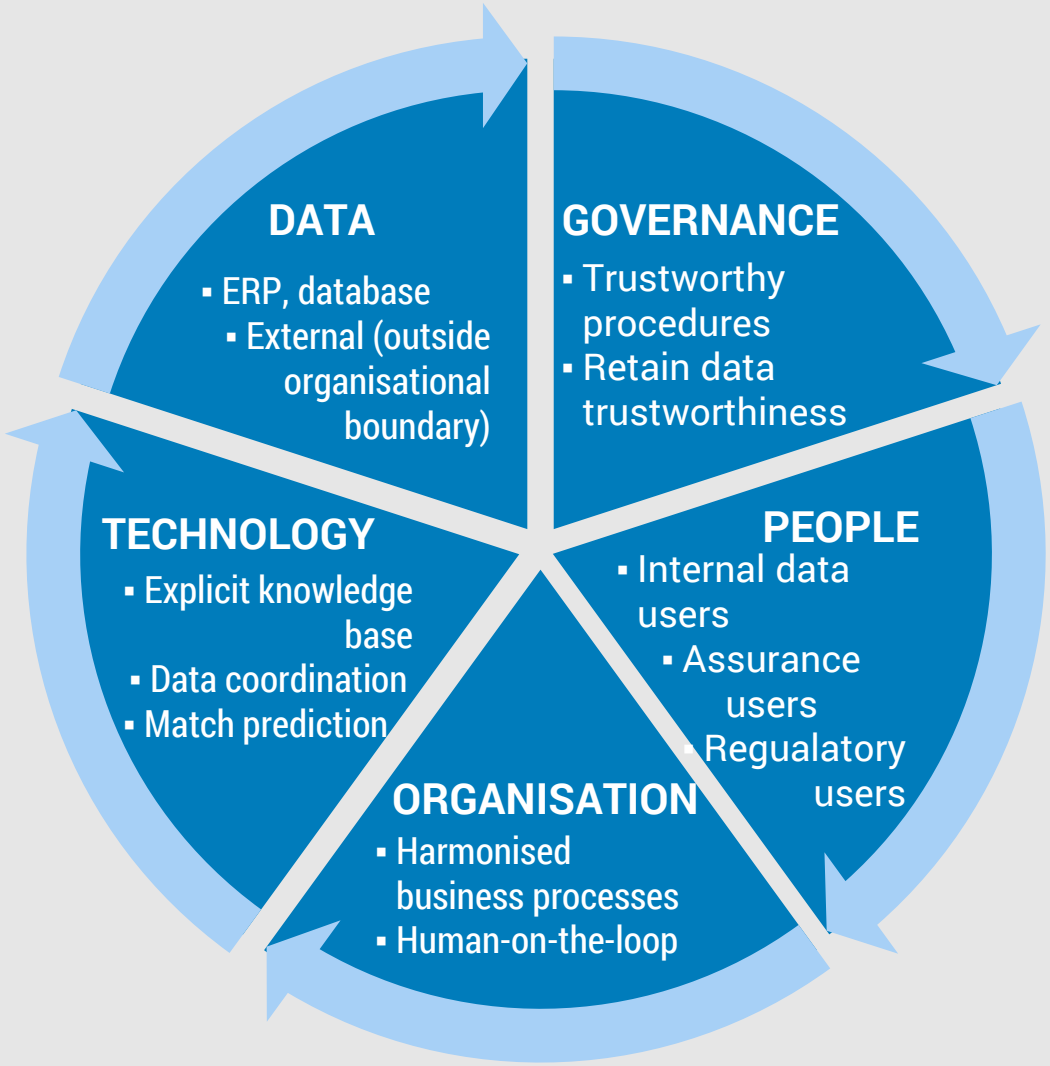
Pre-population with current and historical data
To initiate the change management, we will upload past data whenever possible since 1993.

Maintainance
Maintenance can include the iinclusion of new data, definition of new dashboard items, visualisations

Source: xx

Music Data Space

- A **data (sharing) space** is a legal, technical, organisational solution to trustworthy public-private data sharing.
- **Trustworthy public-private data sharing** is essential to solve big data problems like ESG reporting or trustworthy AI.
- Reprex is associated with the **Big Data Value Association** and follows its data (sharing) space model.
- As a member of the **Dutch AI Coalition**, we are committed to building **trustworthy AI** applications.



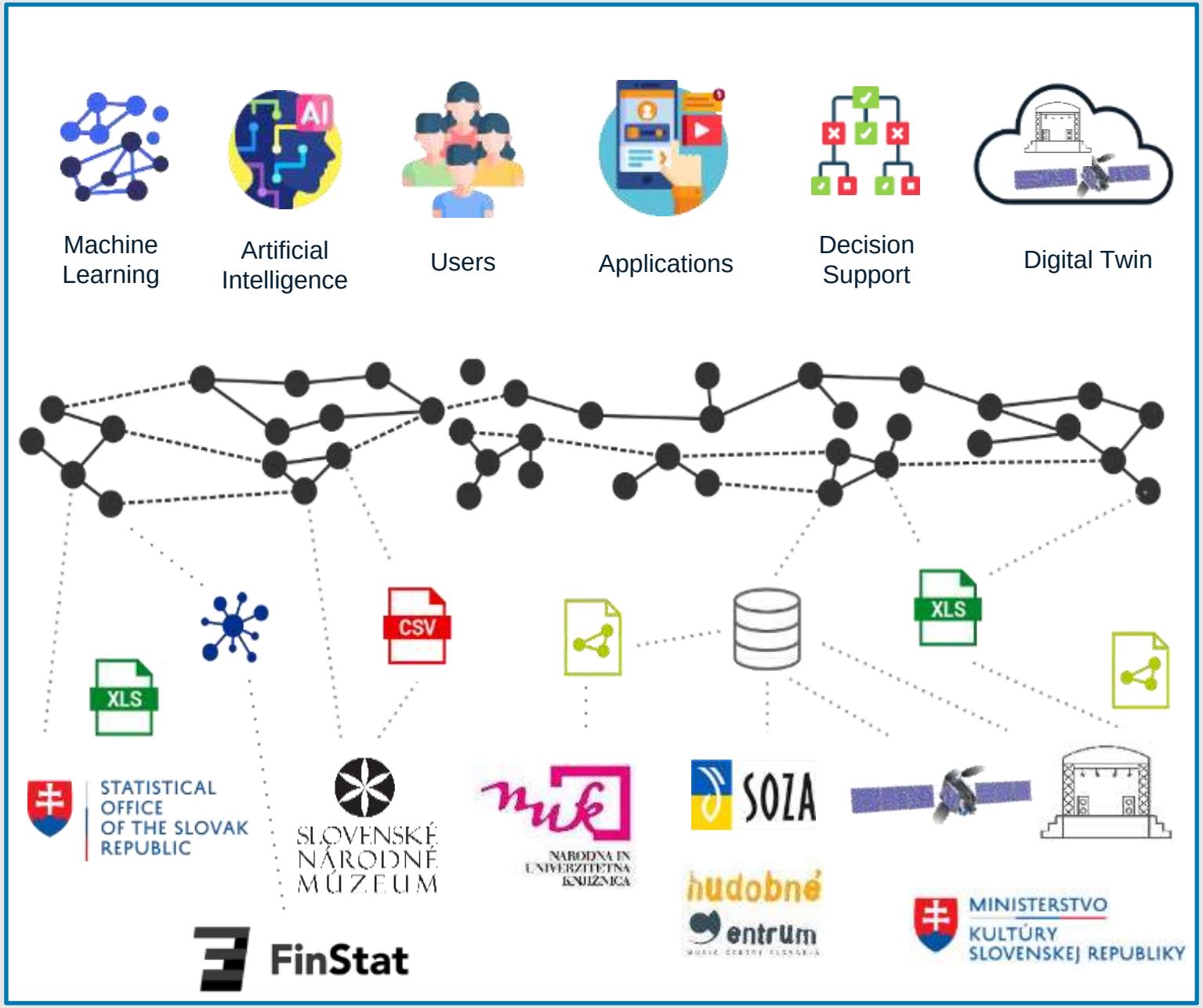
Slovak Cultural Dataspace

Application layer: Listen Local Apps for radio or educational playlist editing. Name Entity Matching and Name Entity Disambiguation for rights management.

Semantic layer: data coordination among public, private, local and international data owners.

Database layer: databases of participating organisations, shared only as needed and as permitted.

Collection layer: data collection internally (royalty accounts, work registers) and external data sources.

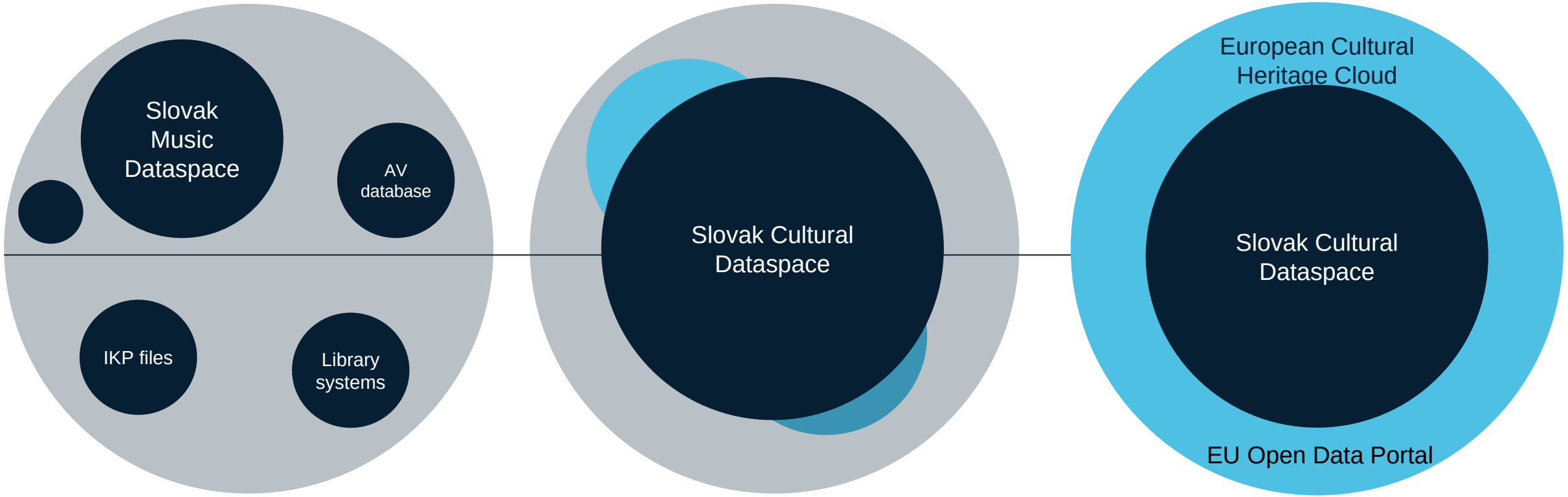


From the Slovak Music Dataspace to a federated cultural data space

Scalable national interoperability system using the infrastructure of the **Slovak Music Dataspace**

Via data federation we can add the gaming industry, galleries, libraries, museums, cinemas, theatres….

We are fully interoperable with European initiatives



The Slovak Music Datapsace connects the databases of SOZA, Hudonbné centrum, Hudobny Fond, Wikidata, Wikipedia, MusicBrainz and other international sources.

The data federation model of dataspaces allows onboarding the entities of other cultural sectors with no significant added cost.

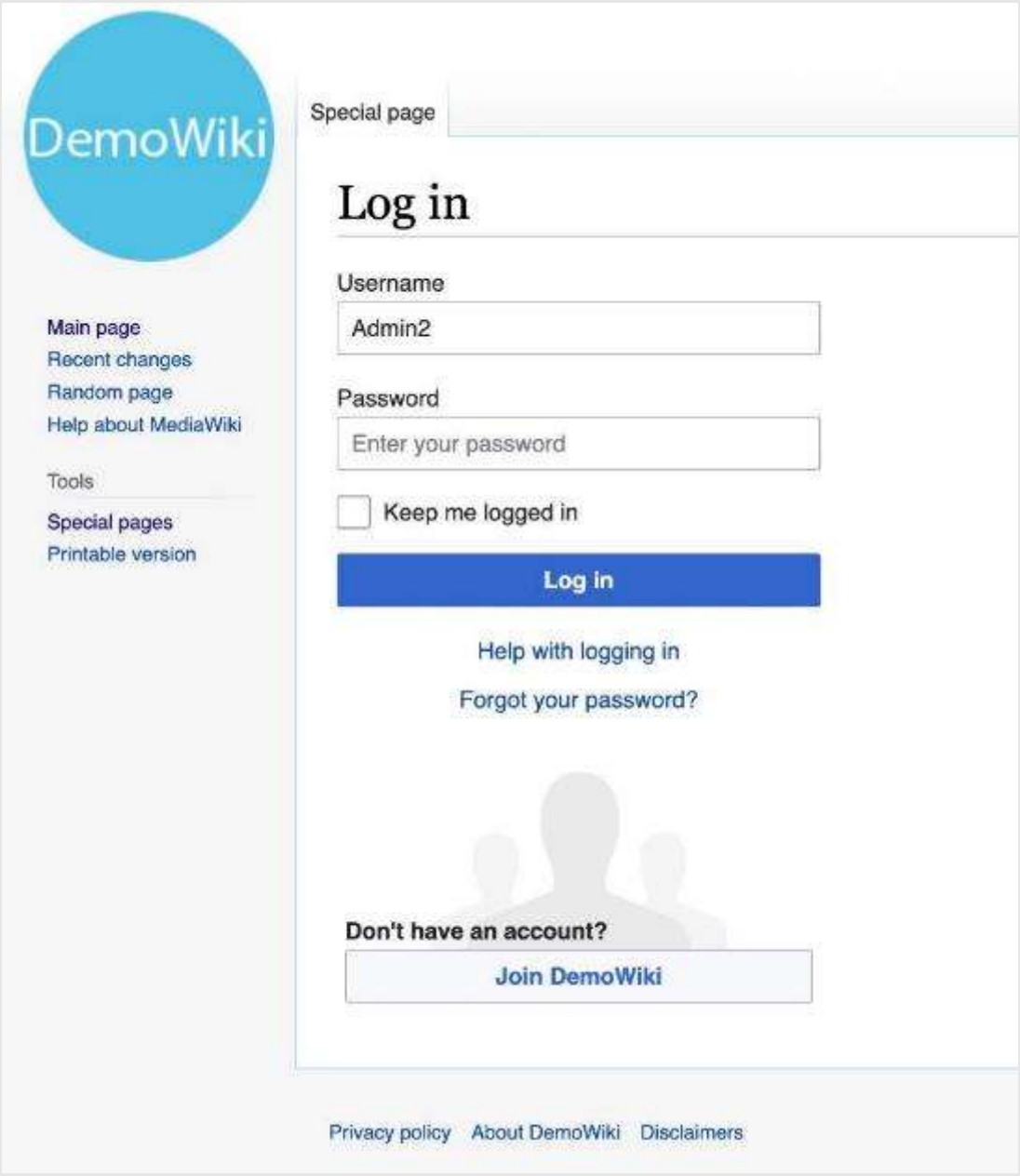
Our model follows the European Interoperability Framework and connects seamlessly with EU data integration projects.

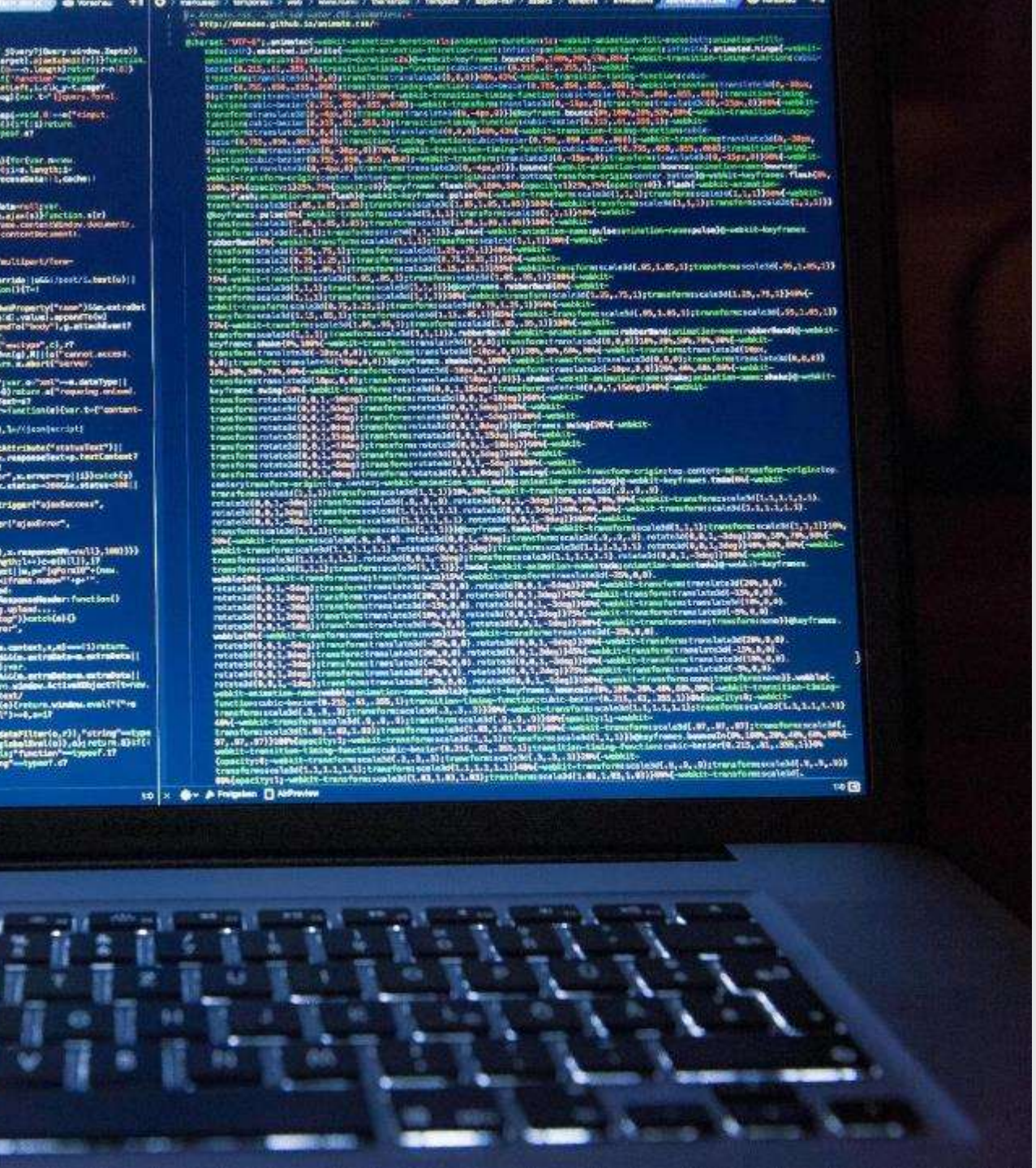
Features

04

Public and private sections

We create a private Wikibase instance with strict hierarchical administration, and train ministry administrators.



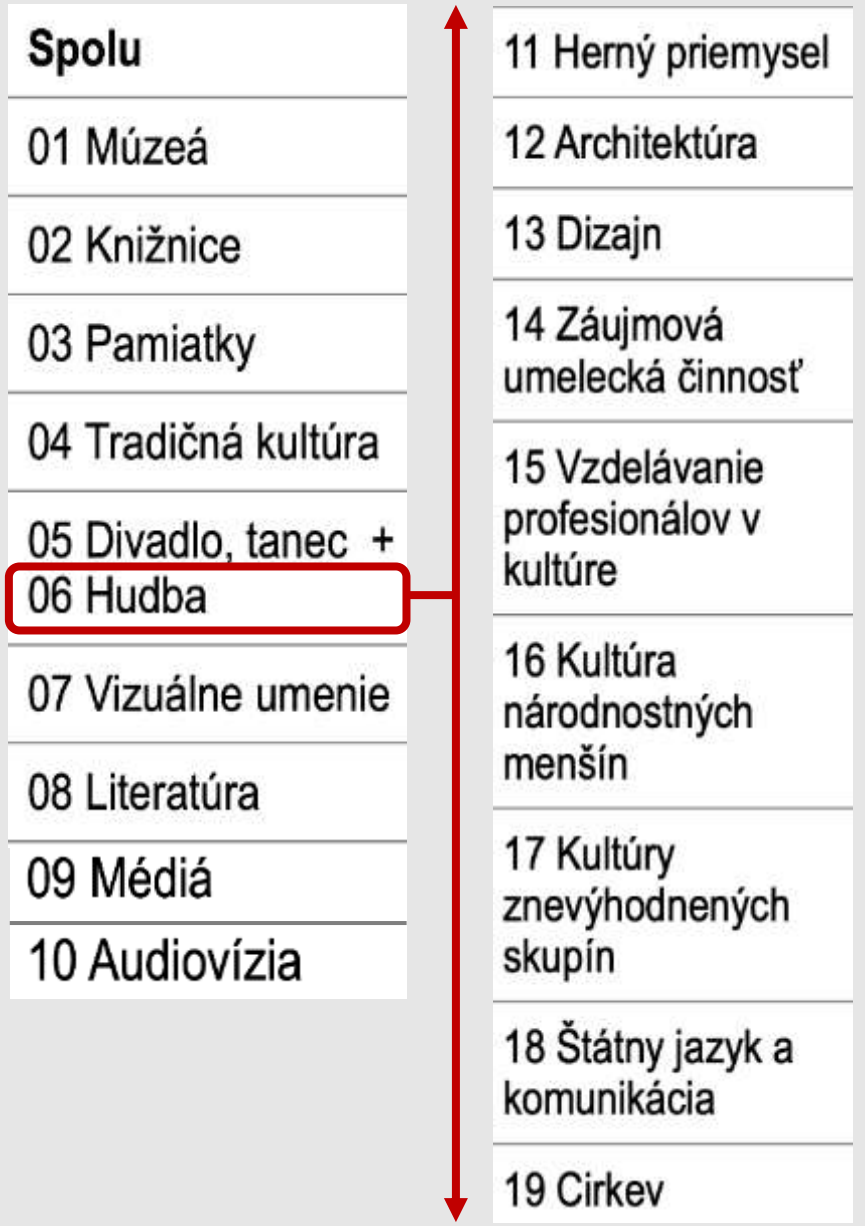


Automated data collection

We start first with automated data collection in music, and gradually extend the scope to other sectors.

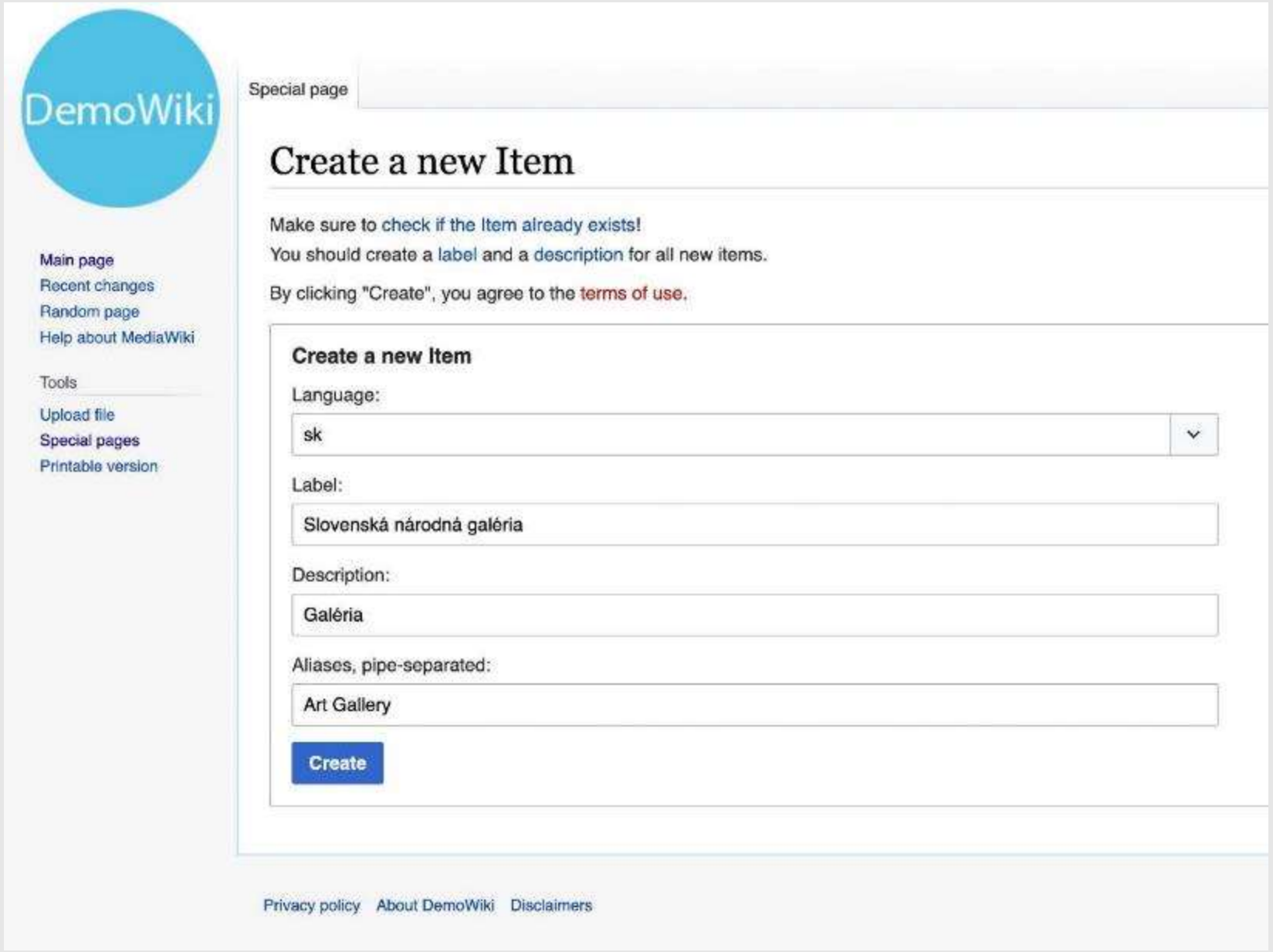
Automated data collection

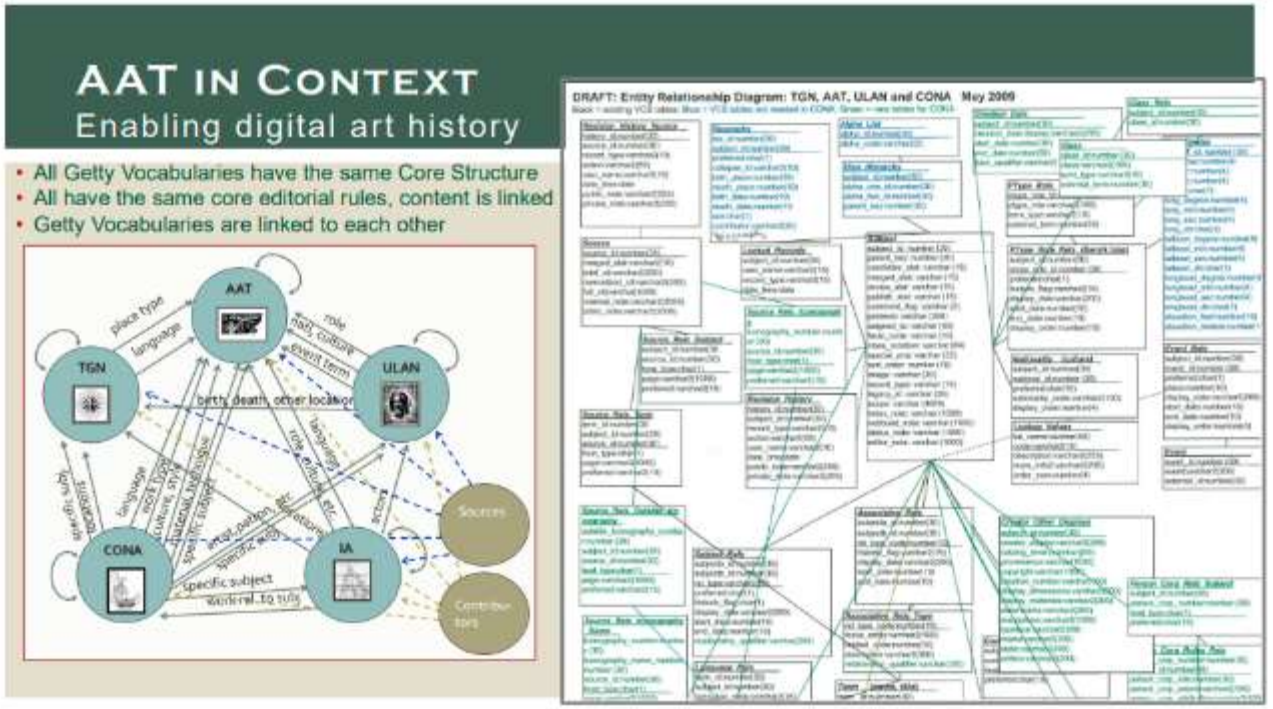
We start first with automated data collection in music, and gradually extend the scope to other sectors.

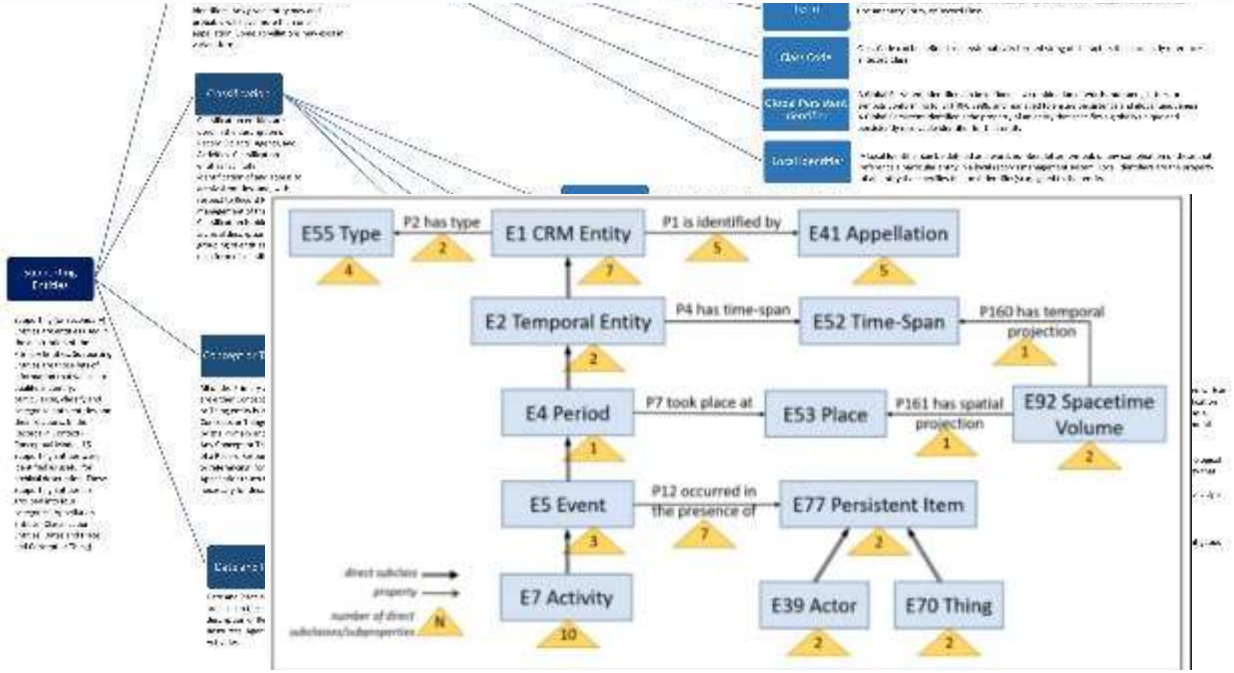


Manual registration of entries

Manual data entry will be possible from deployment day.







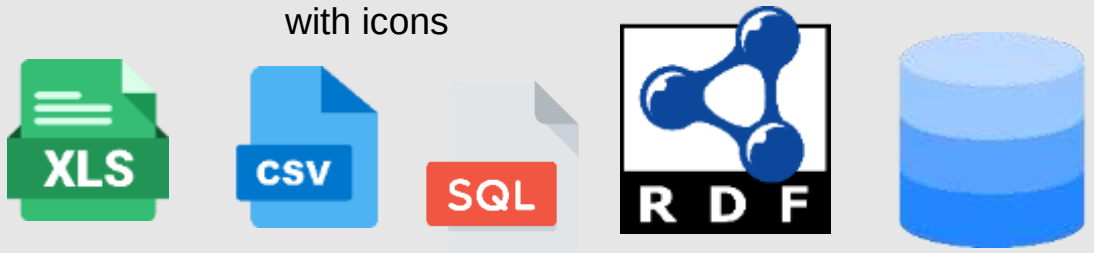
Categorization of entities by cultural sectors

We will create an initial taxonomy that we will align with international ontologies and thesauri, such as AAT, RiC, CIDOC, statistical taxonomies that provide a full interoperablity with existing state registries, international data, and existing museum, library and gallery management systems.

Import and export

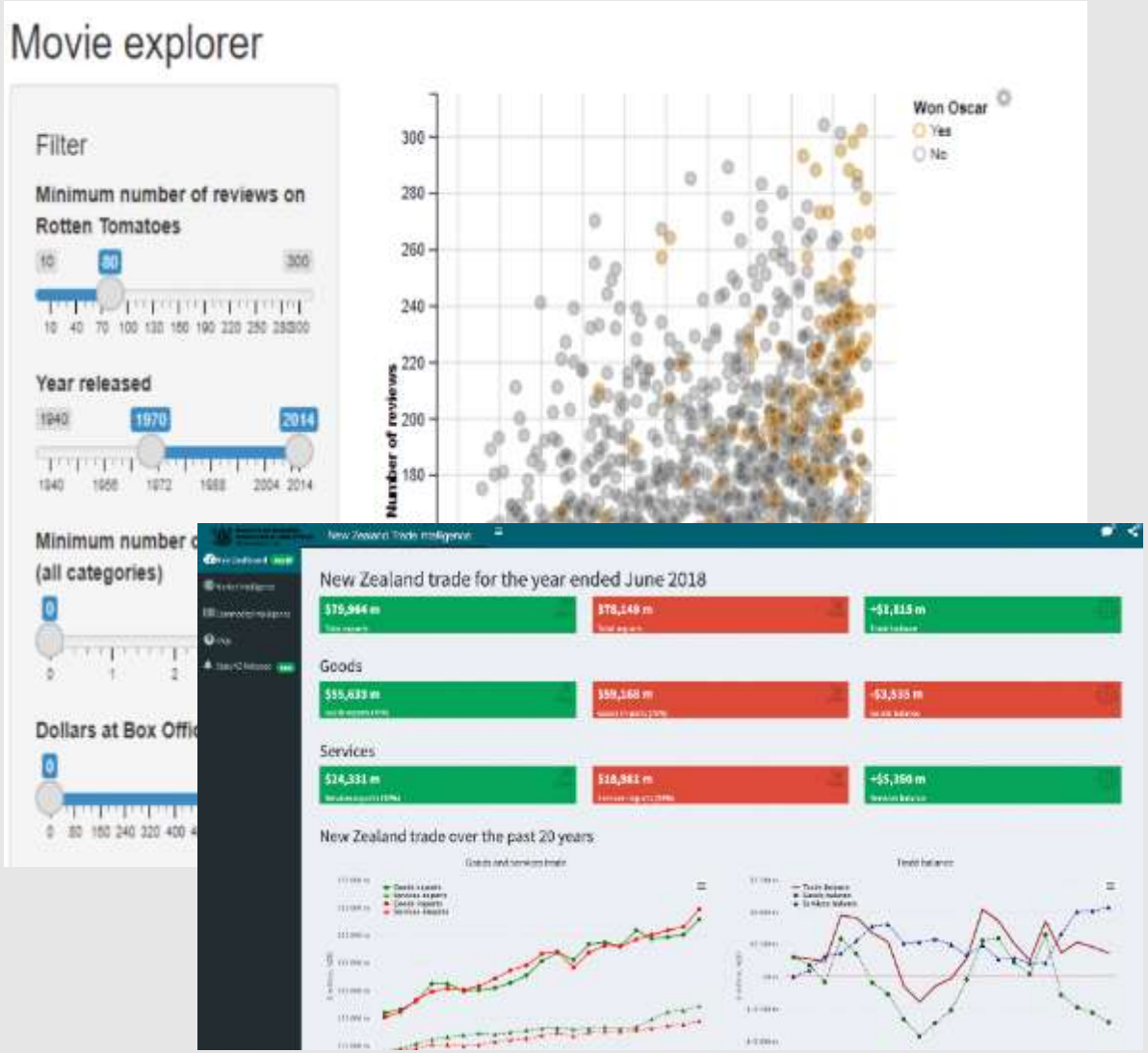
Initially, we support Excel, CSV, RDF, export/import, and import from music databases. In subsequent months we build tools for existing fixed-schema RMDBS systems in SQL as needed.

Only schematic reference to XLSX, CSV, SQL database, RDF graph, even with icons



Generation of output data for managerial decision-making

Managerial decisions data will be available once the system is reasonably well populated, around M3. They will be placed in to Shiny Apps, and training will be provided for their long-term customisation.

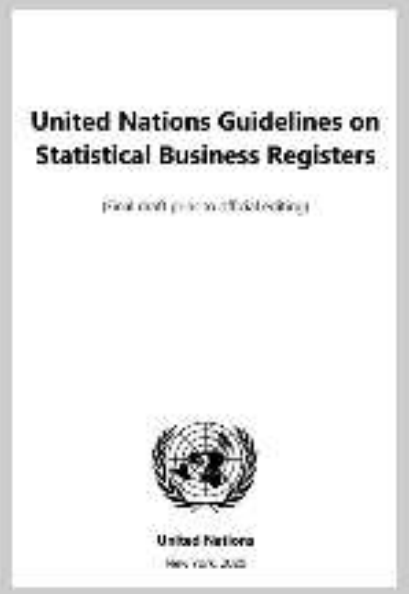


Dynamic pages & dashboards

Dynamic pages, dashboards, widgets can be developed in early phases, but we believe that sufficient experience with managerial data is needed to design lasting, user-friendly dynamic applications.

Automated statistical processes

Our systems use the novel EU/UN methodology for extending official governmental statistics with satellites. We create “official” quality statistical output with our own peer-reviewed, open-source software.



interoperable europe

StatDCAT-AP FOR STATISTICAL DATASETS

StatDCAT Europe

SEMIC Support Centre

sdmx

EUROPEAN DATA PORTAL

Features

Overview of system features

Public and private sections	[Description of risk]	
Automated data collection	[Description of risk]	
Manual registration of entities	[Description of risk]	
Categorization of entities by cultural sectors	[Description of risk]	
Import and export of data	[Description of risk]	
Generation of output data for managerial decision-making	[Description of risk]	
Creation of galleries, navigation menus, dynamic pages, and widgets	[Description of risk]	
Automated processing of statistical results	[Description of risk	

Source: Based on open call

Futureproof interoperability solutions

We are creating **future-proof systems** that can regenerate themselves. We create databases that contain the detailed and standardised knowledge on how to build up a database application that can import and use their data. We can provide support for inherited legacy systems, and make them future-proof. We want to avoid to create the next decade's legacy systems with utilising the world wide web / ISO RDF standards to describe data, metadata and knowledge to use them.

Illustrations: .伊勢神宮 Ise Grand Shrine, which had been rebuilt about every 20 years in the last 1600 years from its components.

Capitalizing on open knowledge and open science

Reprex creates open-source scientific components → to be used in business application with SGD impact

Public (P) Reprex creates extensions to the R Statistical Environment and Language, which is used in most statistical agencies and statistics-based scientific research organizations. Our key components go through scientific and data science peer review, and the facilitate **public** data access, coding, environmental and social impact calculations with approved EU/UN methodologies.

Private (P) We build our proprietary **Eviota** components in OWL/RDF/R to enable **private** partners to efficiently integrate trustworthy data from public sources (statistical, meteorological, environmental satellite…) with private data vendor, internal bookkeeping, ERP, and credit application/monitoring data.

Partnership (P) We offer more than a decade of experience in data **partnerships** that bridge organizational boundaries, public and private sector, and manage conflicts of interests, because in terms of big data, all corporations, state bodies and research universities are small.

Trustworthy AI

05

Knowledge-base design

1. Our overall aim is to show that we can locate more useful data and via increased.
2. Reprex utilises **eXtreme Design**, a UX-infused ontology pattern design method for interviewing internal and external stakeholders to set information needs, knowledge base competencies, and functional requirements for our new code.
3. The eXtreme Design method helps to translate the non-technical expectations of ministry analysist into technical specifications for software and AI.
4. With the help of eXtreme Design we can design those information model patterns and software patterns that solve repeting problems. This ensures that the new registry will have the right data from the correct sources in the right form.

Extension to inventory management

06

Public-Private Partnership for trustworthy music data

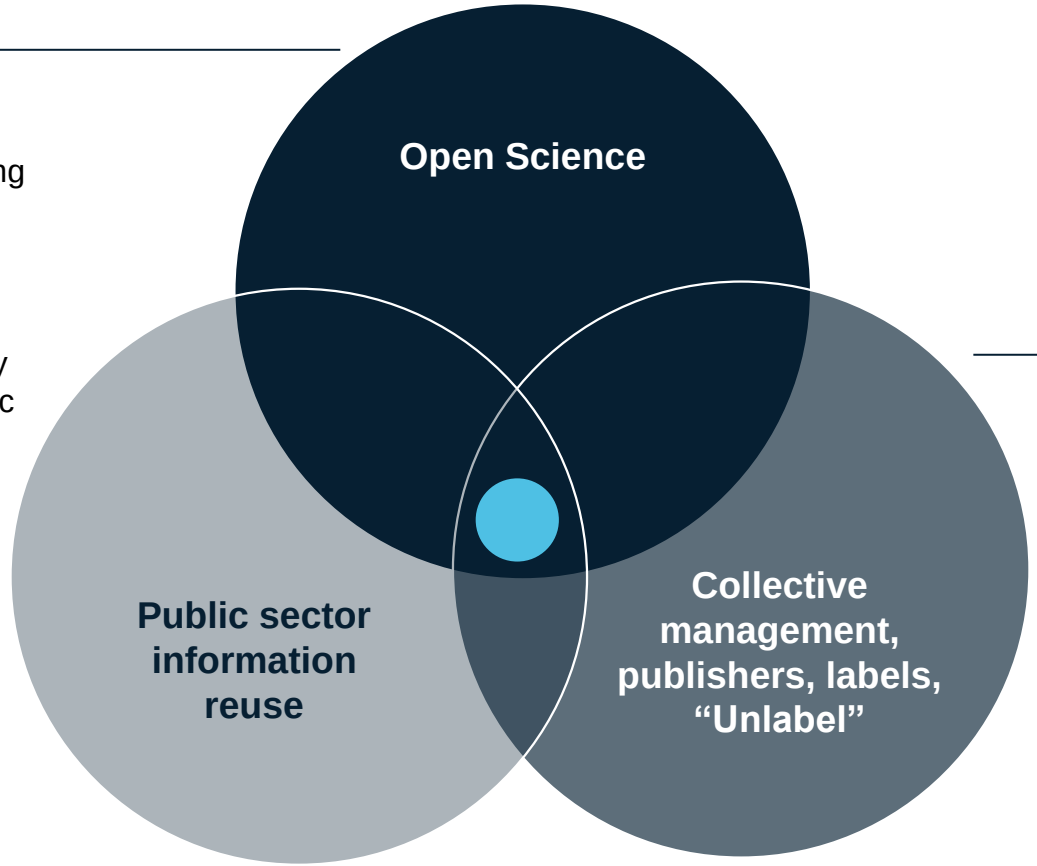
We create data (sharing) spaces that not only follow the architecture of the European Interoperability Framework but extend the use of these interoperability standards among private partners

Open Science

Many disciplines study music, including in Open Music Europe, and we would like to reuse their data. We see the reuse of information models and ontologies for metadata improvement the most promising, because ontology development is more suitable for basic research than business.

Public sector information reuse (open data)

The Open Data Directive (2019/1024/EU), the Data Governance Act (2022/868/EU) gives legal access to much data for free or at marginal cost, however, these valuable data assets need to be reprocessed to be useful for the music sector.



Coordination of privately-held and public sector data

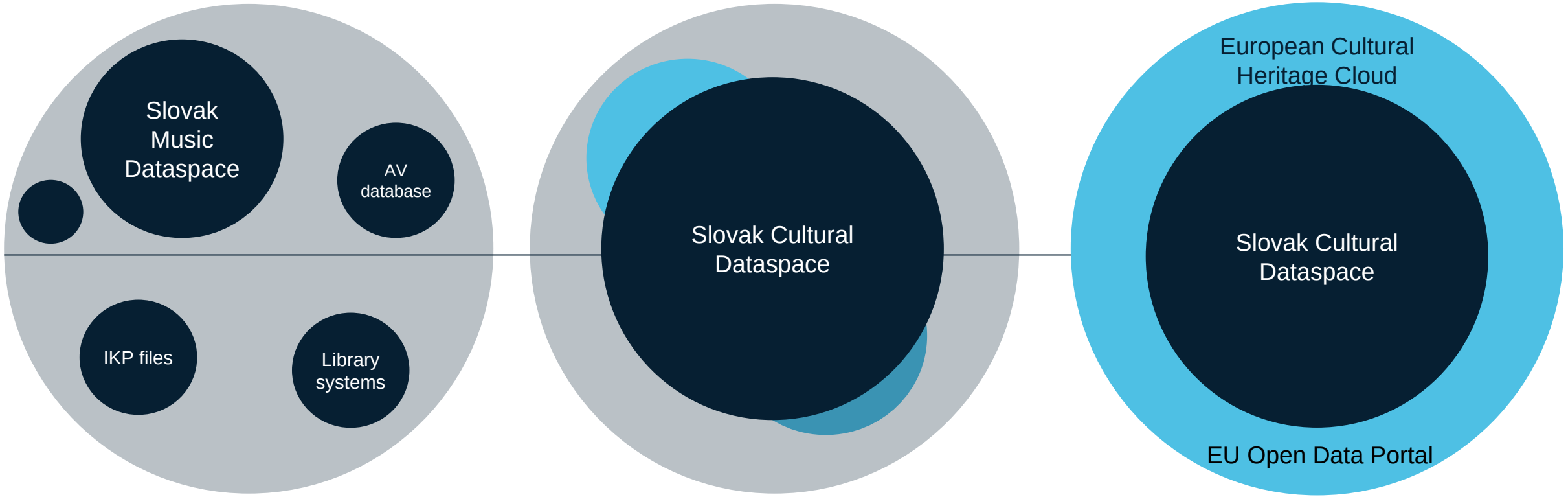
The Data Governance Act and various statistical regulations allowing novel data coordination between privately-held and governmental data. We want to facilitate “experimental statistics” and novel music library products to help metadata-poor rightsholders to become competitive in the digital space.

From the Slovak Music Dataspace to a federated cultural data space

Scalable national interoperability system using the infrastructure of the **Slovak Music Dataspace**

Via data federation we can add the gaming industry, galleries, libraries, museums, cinemas, theatres….

We are fully interoperable with European initiatives



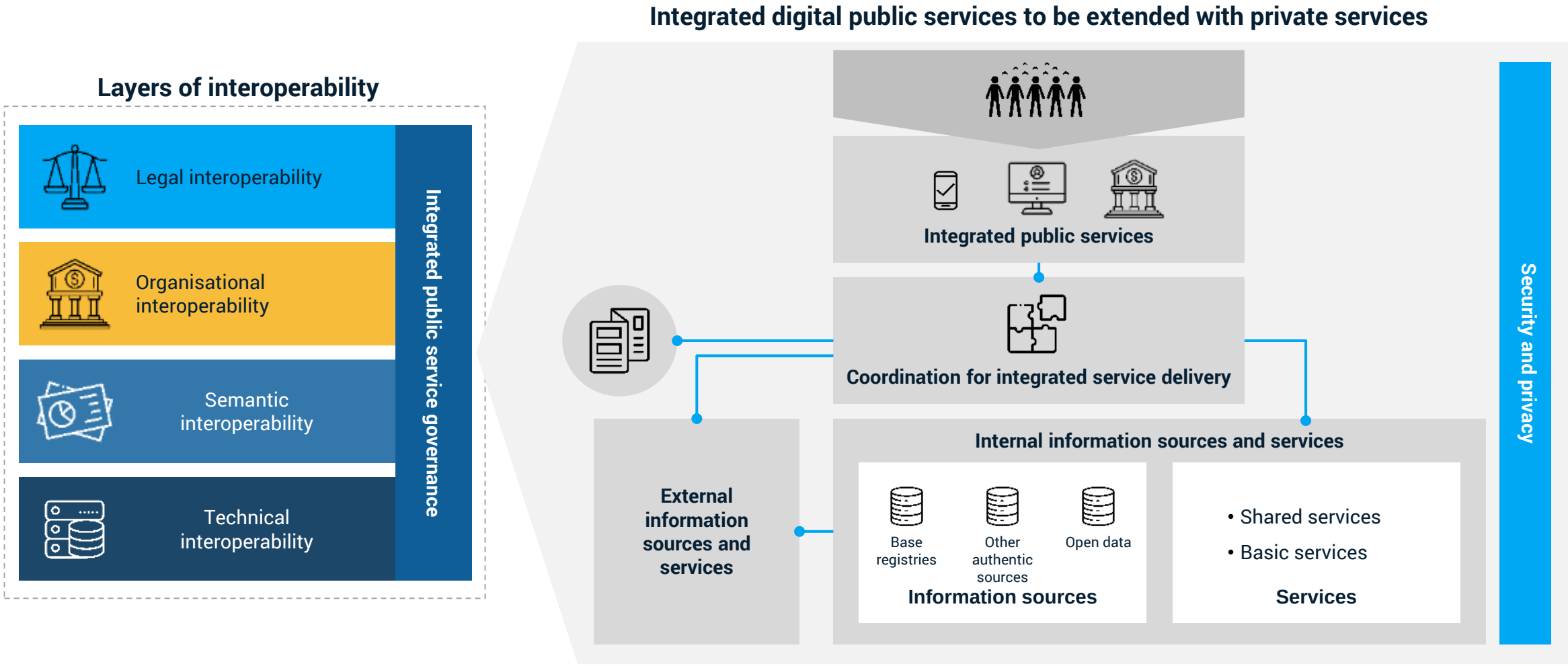
The Slovak Music Datapsace connects the databases of SOZA, Hudonbné centrum, Hudobny Fond, Wikidata, Wikipedia, MusicBrainz and other international sources.

The data federation model of dataspaces allows onboarding the entities of other cultural sectors with no significant added cost.

Our model follows the European Interoperability Framework and connects seamlessly with EU data integration projects.

Embracing the European Interoperability Framework

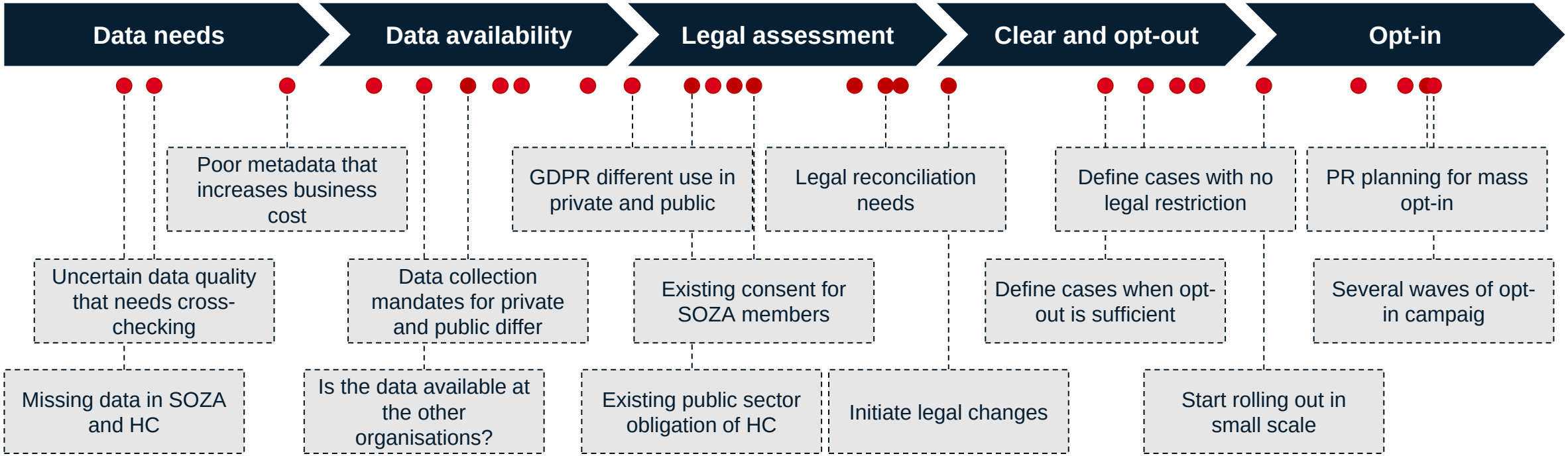
Extending the interoperability of digital services among public and private parties



Legal interoperability: focus on using GDPR in private and public institutions

Legal interoperability: data input and disclosure output automation (replicable from SOZA)

- New instruments: Open Data Directive, Data Govern Act, European AI Act, standardised licenses
- Cooperation with BVDA and the Data Space Support Centre
- Data protection for personal and business confidential data is standardised by SOZA.

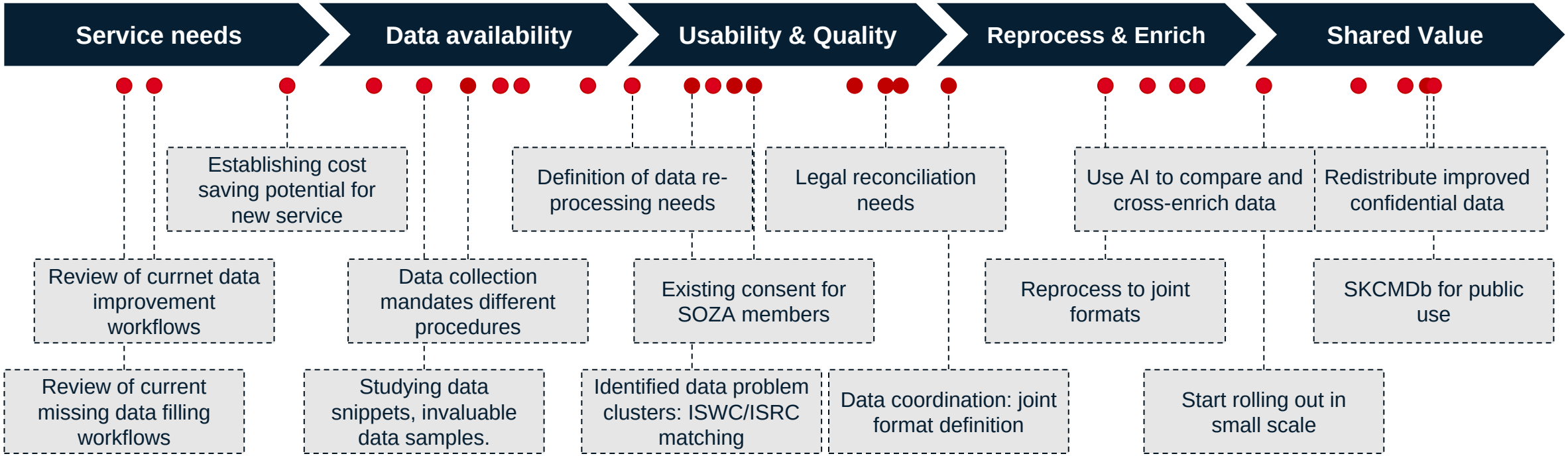


Source: European Interoperability Framework and relevant legal instruments

Organisational interoperability: public and private organisations have business processes that help each other

Organisational interoperability: management control and reporting automation

- Reprex is currently mapping CMO, IMIC music information center and IAML music library processes. Further processes can be designed to architecture, game industry, theatre, etc. Institutions throughout the project. Our competent, native Slovak-speaking colleague will assists the ministry for a year to build interoperability with all cultural organisations work-reporting processes.

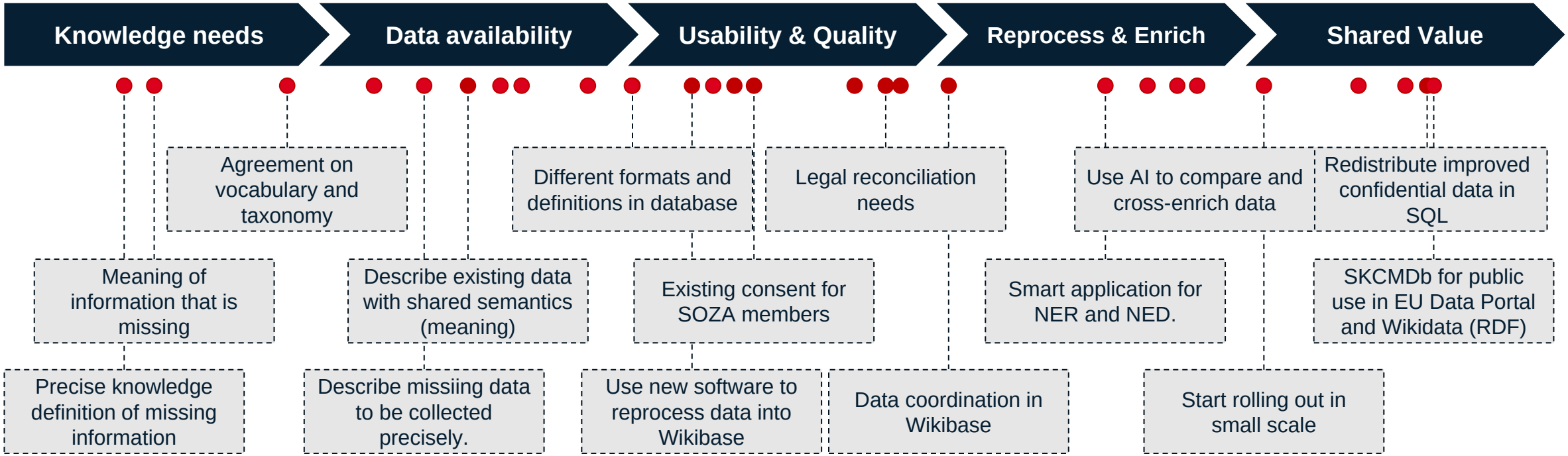


Source: European Interoperability Framework and relevant legal instruments

Semantic interoperability: public and private organisations share the same understanding of data, metadata, categories

Semantic interoperability: shared understanding of knowledge, translation among organisations

- New instruments: Polifonia Ontology Network, RiC, StaDCAT-AP and other semantic standards
- Software is needed to enrich existing datasets with semantics or to export/import to existing formats
- We support Excel/CSV, SQL and RDF (graph) formats.

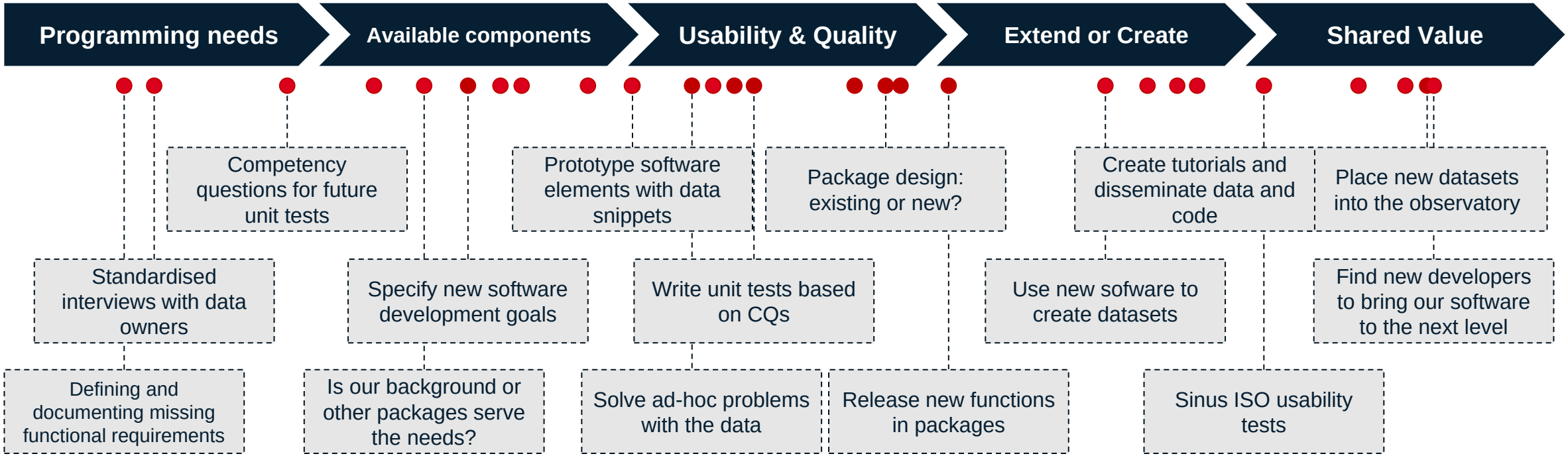


Source: European Interoperability Framework and relevant legal instruments

Technical interoperability: public and private organisations can transform their data into a format that facilitates lossless sharing

Technical interoperability: open source scripts, software, apps, to download and transform data to correct form

- User requirement interviews with Ministry and external data source stakeholders to set functional needs, find data samples for programming, and set Competency Questions for unit testing and application design.
- Programming with the re-use of high-quality OWL/RDF (metadata) and R language (data) components



Source: European Interoperability Framework and relevant legal instruments

Slovak National Museum (Q1093925)

national museum of the Slovak Republic

Slovenské národné múzeum

Identifiers

VIAF ID	129832625
ISNI	0000000110885330
CANTIC ID	981058527861406706
Bibliothèque nationale de France ID	12060890p
National Library of Israel J9U ID	987007268167805171
Library of Congress authority ID	n81101098
NL CR AUT ID	ko2002101443
IdRef ID	028861051
BHCL UUID	a51f6cb0-e295-4334-8540-7f500127d8b8
COURAGE ID	25013
CREPČ institution ID	4AC43E04191C7B6BD72C65AB
Encyclopædia Britannica Online ID	topic/Slovak-National-Museum subject named as
EU Knowledge Graph item ID	Q3122219
Facebook username	slovenske.narodne.muzeum
Freebase ID	/m/02qx6fh
Google Arts & Culture partner ID	slovak-national-museum
GRID ID	grid.455019.c

Slovak National Museum (Q1093925)

national museum of the Slovak Republic

Slovenské národné múzeum

Identifiers

VIAF ID	129832625
ISNI	0000000110885330
CANTIC ID	981058527861406706
Bibliothèque nationale de France ID	12060890p
National Library of Israel J9U ID	987007268167805171
Library of Congress authority ID	n81101098
NL CR AUT ID	ko2002101443
IdRef ID	028861051
BHCL UUID	a51f6cb0-e295-4334-8540-7f500127d8b8
COURAGE ID	25013
CREPČ institution ID	4AC43E04191C7B6BD72C65AB
Encyclopædia Britannica Online ID	topic/Slovak-National-Museum subject named as
EU Knowledge Graph item ID	Q3122219
Facebook username	slovenske.narodne.muzeum
Freebase ID	/m/02qx6fh
Google Arts & Culture partner ID	slovak-national-museum
GRID ID	grid.455019.c

East Slovak Gallery (Q3094652)

art museum in Košice, Slovakia

Identifiers

VIAF ID	148472503
ISNI	0000000122266094
NL CR AUT ID	ko2003195914 ko2003195915
WorldCat Identities ID (superseded)	lccn-n85084140
Athenaeum museum ID	5180
CREPČ institution ID	CABE63230C4296FD1D6141B36E
Freebase ID	/m/0blp47
Google Arts & Culture partner ID	east-slovak-gallery-kosice
OpenStreetMap node ID	2349271078 691083875

West Slovak Museum (Q98831205)

museum in Slovakia

Identifiers

VIAF ID	157545435
	0 references
	127326203
	0 references
	172378027
	0 references
	236933965
	0 references
NL CR AUT ID	ko2016909904
CREPČ institution ID	069BD78221CDCAB105A9F3E570
	AB574C4B6B0EDE306DB9177C
OpenStreetMap relation ID	2193450
SK cinema authority ID	0102251
SNK ID	114007

Cypronia (Q5200430)

Slovakian indie video game developer and publisher based in Michalovce

Ablaze Entertainment | Cypron Studios | Cypronia sro | Cypronia SRO | Cypronia s.r.o. | Cypronia S.R.O.

Identifiers

EU Knowledge Graph item ID	Q3122954
Freebase ID	/m/03mb4mm
GameFAQs company ID	102898
Gaming-History company ID	10560
GRY-Online company ID	9927
LaunchBox Games Database developer ID	20756
LaunchBox Games Database publisher ID	2735
Mod DB company ID	cypron-studios
OGDB company ID	11680
TheGamesDB developer ID	2043
TheGamesDB publisher ID	2603
UVL company ID	6232
VideoGameGeek developer ID	18025
	18017
	24815