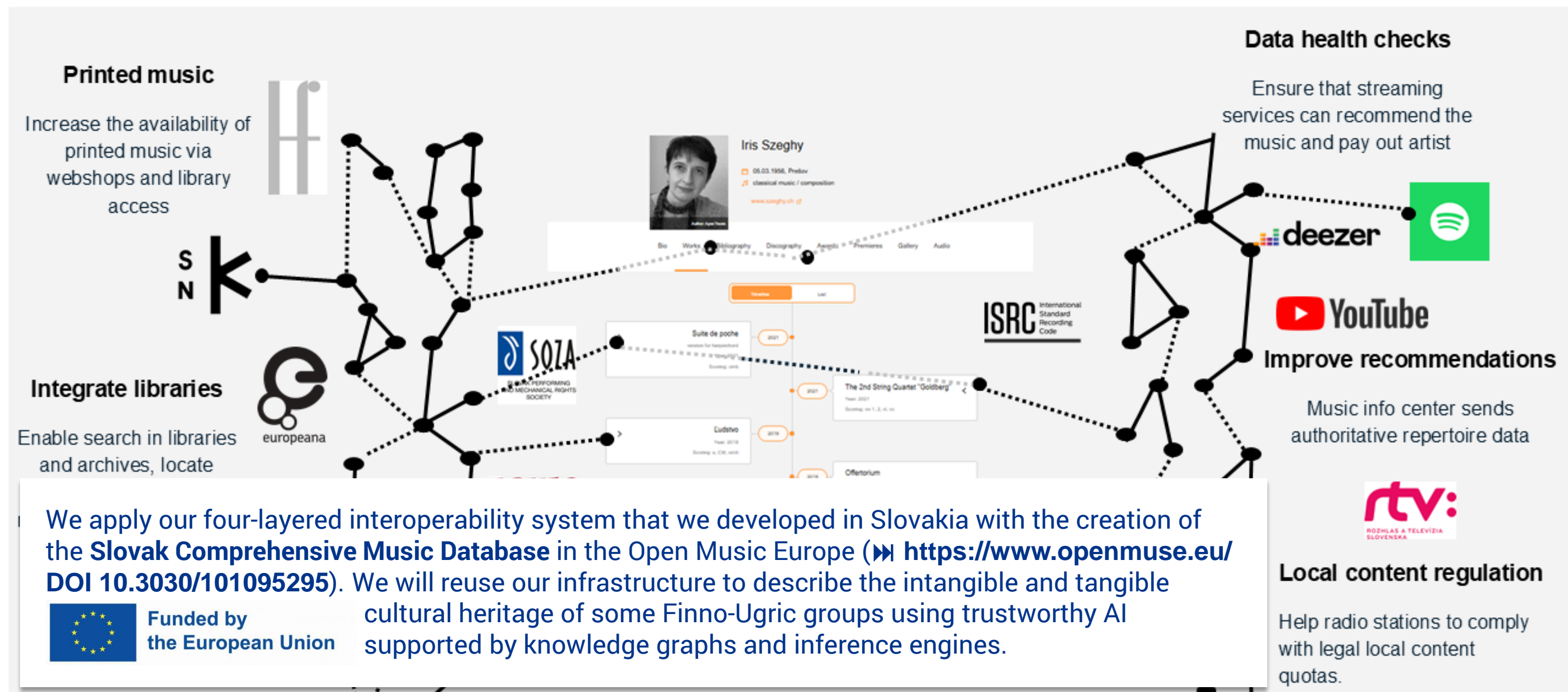


A Finno-Ugric Data Sharing Space

We started experimenting with the legal, organisational, semantic and technical challenges of creating a genuinely trustworthy, AI-supported data-sharing space that can find and connect tangible and intangible elements of the Finno-Ugric cultural universes. We were also seeking a better governance model for oversight for the custodians of these endangered, shrinking universes in their language and with little technical knowledge, partly as alternatives to the established Wikipedia to the open knowledge incubation method for small linguistic minorities.

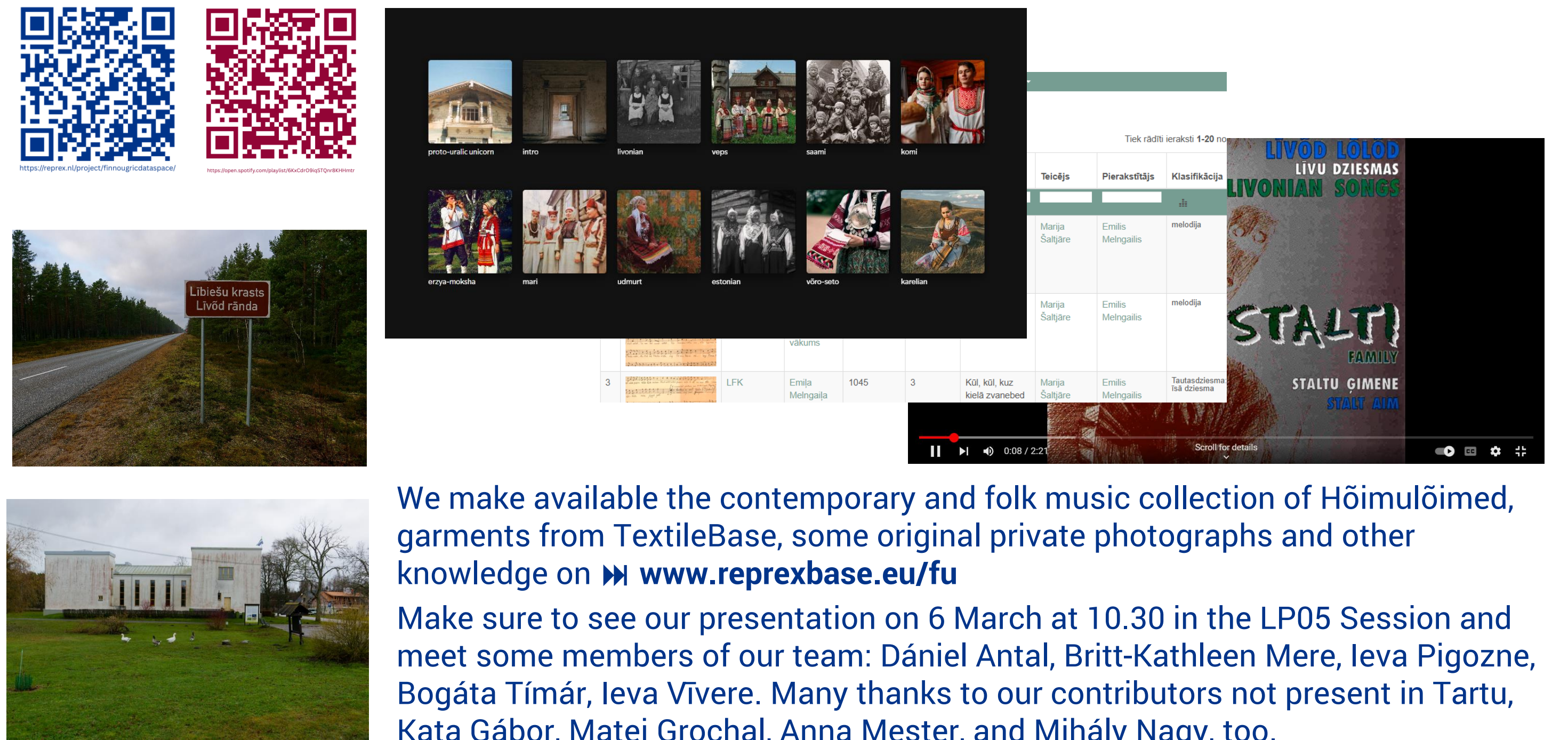
Q5 Interconnected private and public digital services in culture

We are designing a data sharing space that can confidently work with the metadata schemas, ontologies of all GLAM institutions (public libraries and archive, museums) as well as private services like Bandcamp, YouTube, Flickr, Spotify, name registration services, or radio playlisting services. Applying the European Interoperability Framework extended to privately-held data, we ensure that our users have an accurate 360° view of the digital heritage in their custody or interest, and we guarantee that research or streaming services properly use the elements of these cultural universes.



Find more about us

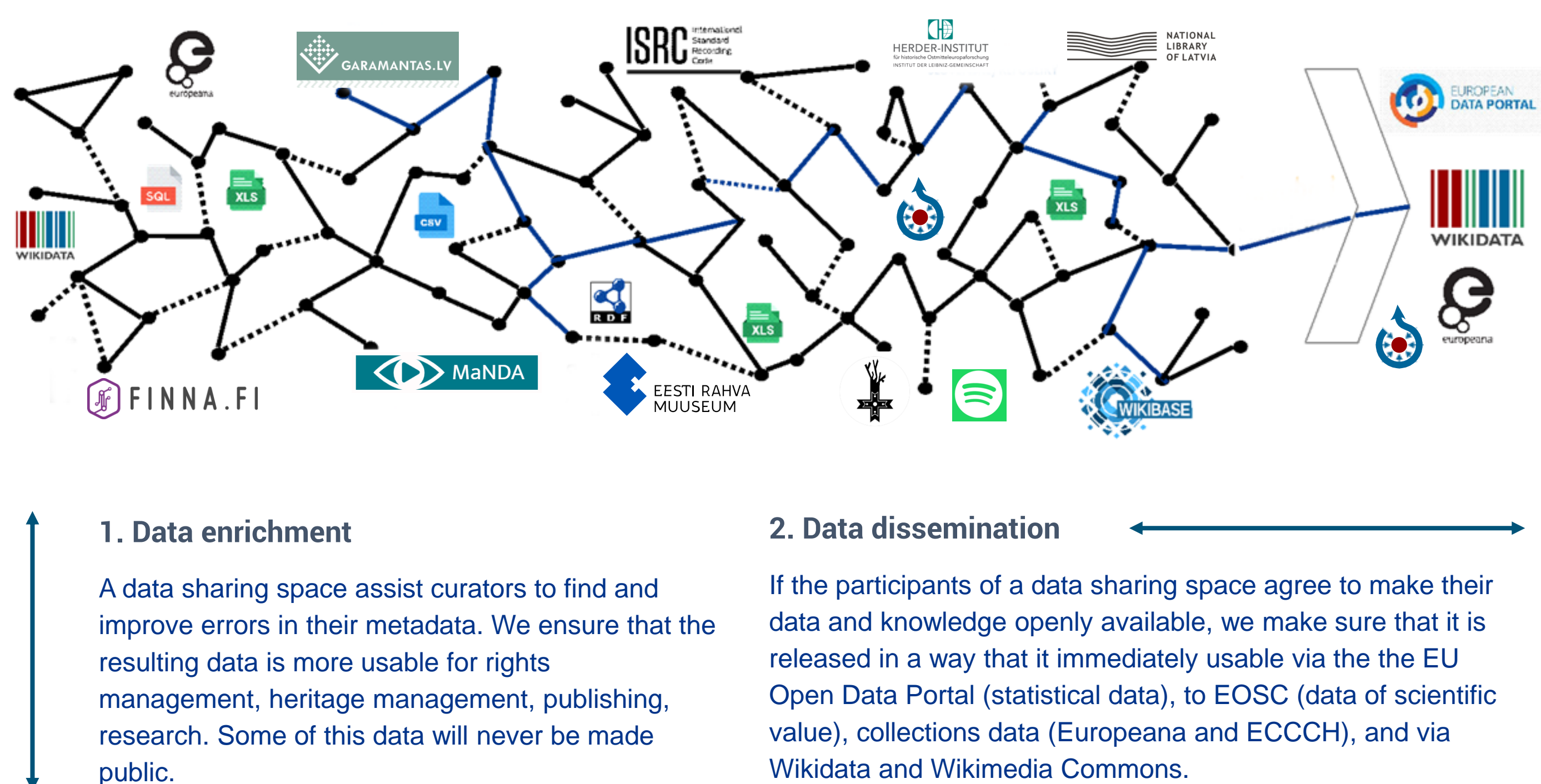
Our demo project is volunteer work for the co-authors without funding, so our data-sharing space is a prototype with few datasets. It is implemented with the Wikibase Suite and its Lexeme extension and aims for compatibility with the main features of libraries, archives, museums, digital rights management and streaming systems, and language technology applications. We placed a few but hopefully highly interesting datasets into the system that you can view using the QR codes in this poster.



We make available the contemporary and folk music collection of Hõimuõied, garments from TextileBase, some original private photographs and other knowledge on www.reprexbase.eu/fu. Make sure to see our presentation on 6 March at 10.30 in the LP05 Session and meet some members of our team: Dániel Antal, Britt-Kathleen Mere, Ieva Pigozne, Bogāta Tímār, Ieva Vīvere. Many thanks to our contributors not present in Tartu, Kata Gábor, Matej Grochal, Anna Mester, and Mihály Nagy, too.

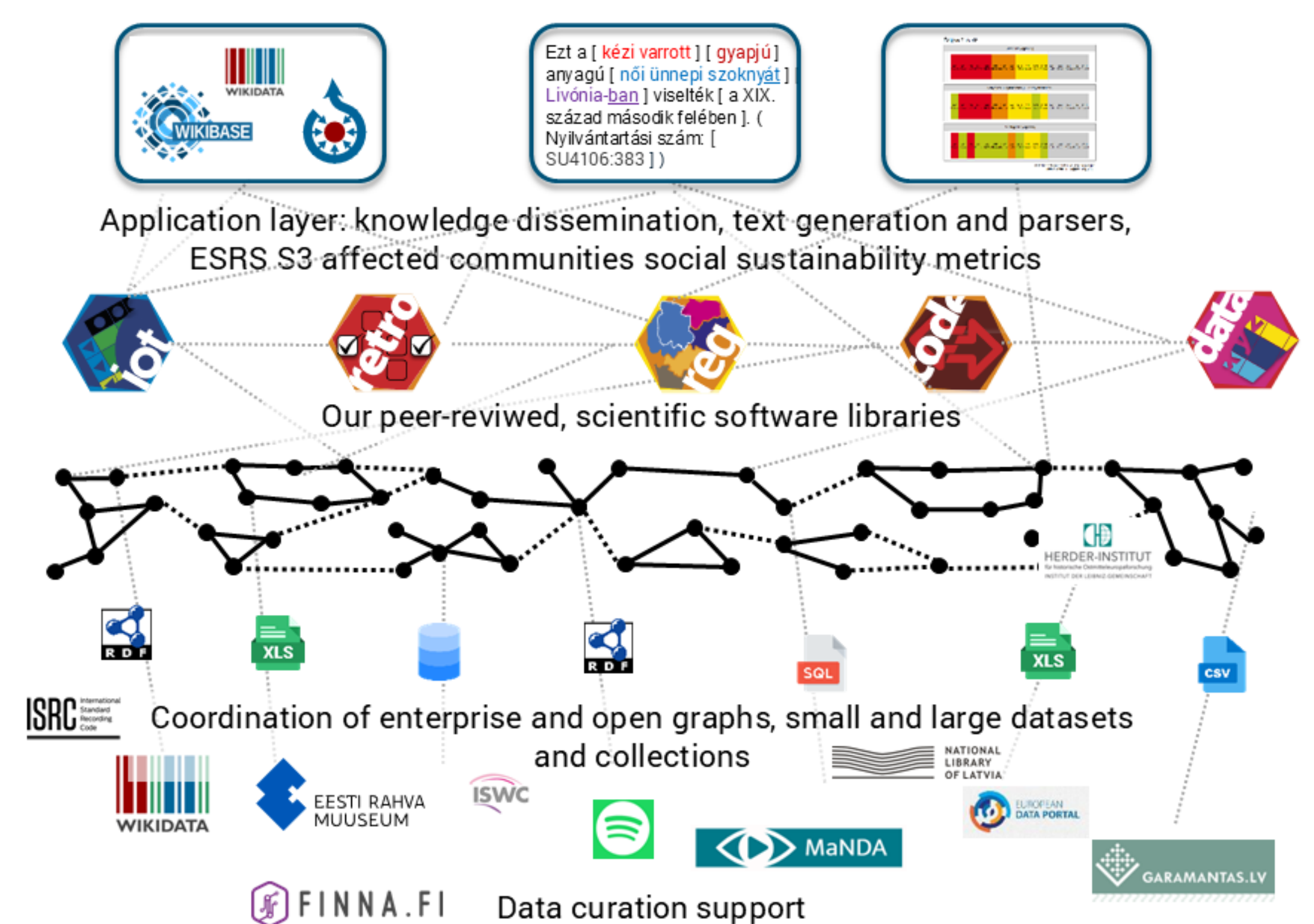
Q1 Which recordings of contemporary musical work and traditional music were released in the Liv and Samoyedic languages in 2012?

A **data (sharing) space** is a system that integrates data whenever needed or permitted. The Statistical Data and Metadata eXchange, European Open Science Cloud, Europeana, the European Collaborative Cloud for Cultural Heritage (ECCCH) will be connected with private systems like Wikidata, Wikimedia Commons, the Spotify or YouTube API. Semantic interoperability means that our system understands public and private cultural APIs.



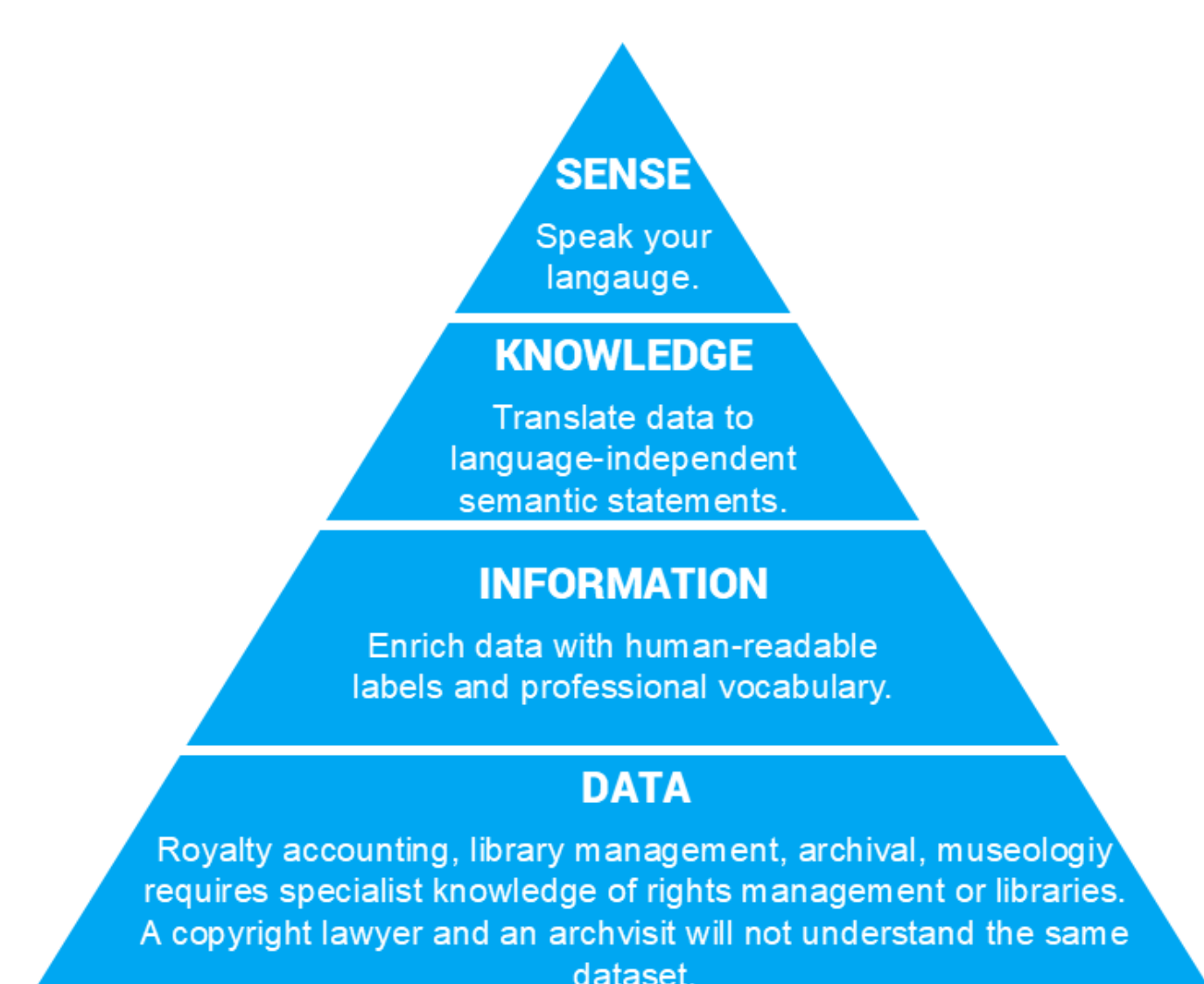
Q2 How can we find new knowledge about the historical or contemporary Khanti-Manysi music tradition?

The organisational interoperability is necessary to create systems that can support application from diverse collections and organisations. We must understand that archivists, rights managers, librarians, museologists, NGOs, private collectors, festival organisers often work with the same data but in a different job or workflow. We do not only need to understand which things belong to the Khanti-Manysi universe, but also how a librarian or a streaming service would handle that part of the cultural universe.



Q3 Which songs refer to dreams in their title or lyrics, regardless of the Finno-Ugric language used?

Starting with a playlist dataset of songs on Spotify, we enrich it with metadata in a way that we can serve first language-independent, then language-specific inferences and queries. In our conference presentation we will explain this by helping a curator who in our 14 language playlist wants to find songs about „imagining events while sleeping” or about a „dream”.



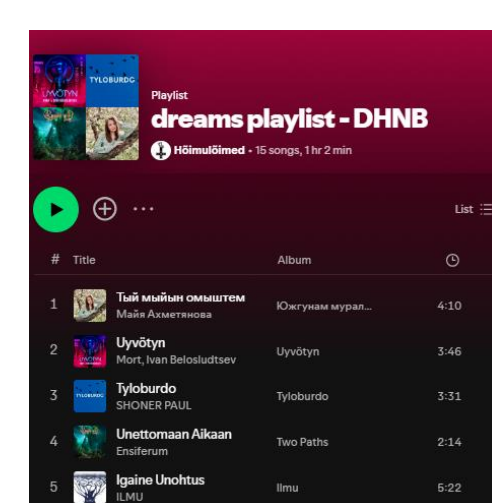
Linguistics AI
Adding lexicographical semantic metadata to our knowledge base (Уивот / Уивотын in Udmurt or álom / álmok in Hungarian allows us to translate the TRUE statement to these languages: "A Тый мыйын омыштем" és az "Elefánt" olyan dalok amik álmokról szólnak.

Transform to knowledge
Converting data and metadata to semantic statements allows humans and AI agents to combine knowledge from different sources. The statement that "Тый мыйын омыштем" and "Elefánt" are songs with dreams in their lyrics is TRUE or FALSE.

Adding labels and metadata
Adding metadata, like ISRC: QZW9M2446460 or song title: Тый мыйын омыштем, or duration of 250005 milliseconds makes the data of a particular sound recording understandable not only for rights managers, but curators of playlists or radio editors, too.

Data needs interpretation
The text QZW9M2446460, Тый мыйын омыштем, or 250005 require unit of measure, definition, text to be understood.

Listen to our Finno-Ugric playlist with songs in Udmurt, Mari, Hungarian, Võro, Karelian, Finnish about **dreaming**.



Q4 What should be the linguistically correct description of a curated list of Samoyedic musical pieces in the Liv language?

Our system generates semantic statements from trustworthy metadata. The semantic statements are then translated to natural language descriptions. We aim for a level of fluency that is suitable for users within the Finno-Ugric communities. This tool can improve the curation and governance of open knowledge projects that cannot recruit many reviewers with a high level of language competency to give an effective oversight for a knowledge base or an AI application.

This [female festive skirt] was [hand-sewn] from [wool] material, and it was in use by [women] on [festive] occasions in the [second half of the 19th century] in [Livonia]. (Inventory number: [SU4106:383])

Ezt a [kézi varrott] [gyapjú] anyagú [női ünnepi szoknyát] [Livónia-ban] viselték [a XIX. század második felében]. (Nyilvántartási szám: [SU4106:383])

label	female festive skirt	női ünnepi szoknya
instance of	skirt	szoknya
cut	traditional cut	hagyományos szabás
fabrication_method	hand-sewn	kézzel varrott
made_from_material	wool	gyapjú
intended_use	female festive wear in Livonia	Ünnepi viselet Livóniában
inventory_number	SU4106:383	SU4106:383



Material provided by The National Museum of Finland, link to this artefact

Please check out our presentation at the LP05 Session, Digital Insights in Cultural Research, on Mar 6, 2025 10:30 AM — 11:00 AM
https://reprex.nl/event/2025-03-07_dreams/

Questions? reprex.nl/contact
Daniel Antal www.linkedin.com/in/antaldaniel

