



SELMA

STOA Workshop on ***Humanities in the digital age: Securing innovation and empowering democracy***

26 April 2023, Brussels, EP

Handling 1000 truths: Innovation in multilingual media monitoring – SELMA project and beyond

Normunds Grūzītis



Institute of Mathematics and Computer Science
University of Latvia



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957017.

Agenda

- SELMA project
 - Multilingual media monitoring
 - Counting diversity
- Beyond SELMA
 - Foreign information manipulation and interference (FIMI)

Empowering
democracy

Defending
democracy

SELMA project

- Partners



- Use cases & applications

- Media monitoring (UC1)
 - Broadcast monitoring & analysis
 - Counting diversity
- News content creation (UC2)
 - Video subtitling & voice-over
- Highly scalable NLP (UC0)
 - SELMA OSS Platform

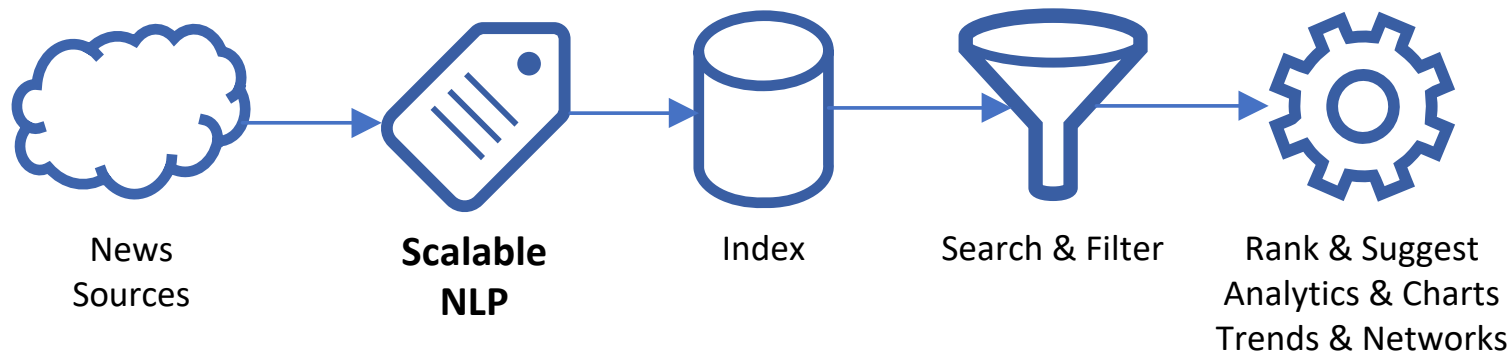
Tech Ready



Media monitoring (MM)

- [It] is the activity of monitoring the output of the print, **online** and **broadcast** media..
- ..to power the **decision-making** process of people and organizations
 - e.g., analysis of current events, identification of [emerging] trends, competitive intelligence, public reputation, brand awareness
- Digital **humanities** and strategic **communication**:
opinions, biases, manipulation, sentiment, connections → **counter-acting**

Multilingual MM: Pipeline



- Text
- Audio/Video
- Ingestion (incl. speech-to-text)
- Multilingual topic detection
- Cross-lingual document clustering
- **Named entity recognition and linking** (NER & NEL)
- Summarization
- Machine translation

global

Summary

Articles

Aggregate

Tiles

Headlines

Analyze

Trending Entities

Entity Network

Saved Articles

Dashboards

Reports

Language English | Spanish | German | Portuguese | French | Dutch | Ukrainian | Russian | Italian | Topics - sport | - arts, culture and entertainment | Places France | Ukraine | Russia | China | European Union | People Emmanuel Macron | Volodymyr Zelenskyy | Vladimir Putin | Organization Kremlin | European Parliament

The Kremlin says Russian President Vladimir Putin has visited headquarters of Russian troops fighting in Ukraine

La Justicia rusa condena a Kara-Murza a 25 años de cárcel, la mayor de la historia a un opositor de Putin

Russische Invasion: Krieg gegen die Ukraine: So ist die Lage

Verhalten sie Putin zur Invasion? Ukraine entlarvt immer mehr Doppel-Agenten – „nur Spitze des Eisbergs“

Ukraine-Krieg: Außenminister Lawrow dankt Brasilien für Unterstützung

Wladimir Putin lobt Manöver im Pazifik: Russlands Flotte in höchster Alarmbereitschaft

NATO believes Ukraine can retake land as allies mull new support

Baerbock nennt ihre China-Reise „zum Teil meh...“

Militärausgaben: Höchster Anstieg in Europa sei...

Josep Borrell acusa Lula da Silva de ter "a..."

"Safe harbours" for blood money. Why are Russia...

Krieg in der Ukraine: "Werden alle auf dem Schlachtfeld vernichten":...

Russian court dismisses Wall Street Journal...

Putin's speech on the anniversary of the Russian Revolution: "We will continue to fight for the interests of the Russian people..."

Macron accused of playing into Xi's hands: with the Last Generation and on €43B chips plan

Russia's tera rede de espionagem russa alinha no Mar do Norte para poder sabotar infraestruturas-chave

politics

Aliados abastecen a Ucrania para que pueda lanzar su contraofensiva

Russia 'chief beneficiary' in deadly Sudan conflict as attempted ceasefire falls apart

G7 foreign ministers show united front as they condemn Russia's war, call out

China affirms ex-Soviet nations' sovereignty after uproar

ЄС має намір ускладнити Росії обхід санкцій,...

EU is still largest importer of...

Росія...
"Als je verwacht dat met de..."

diplomacy

Polish illegal ban on Ukrainian farm products - a hysterical reaction that plays into Putin's hands

EXPLAINE D: How Russia Accidents

economy...

Filters

TOPICS

arts, culture and enter...	36.6k
sport	36.4k
politics	860
unrest, conflicts and war	705
diplomacy	431
defence	385

ORGANIZATION

European Parliament	1.0k
Kremlin	777
NATO	921
Agence France-Presse	615
Armed Forces of Ukraine	474
Telegram (software)	448

PEOPLE

Emmanuel Macron	1.6k
Vladimir Putin	750
Volodymyr Zelenskyy	306
Dmitry Peskov	193
Vladimir Kara-Murza	154
Joe Biden	137

PLACES

France	7.4k
China	6.7k
Ukraine	796
Russia	756
European Union	256
Moscow	557

OTHER

Russo-Ukrainian War	206
2022 Russian invasion ...	189
Easter	80
Euro	63
World War II	57

global

Summary

Articles

Aggregate

Tiles

Headlines

Analyze

Trending Entities

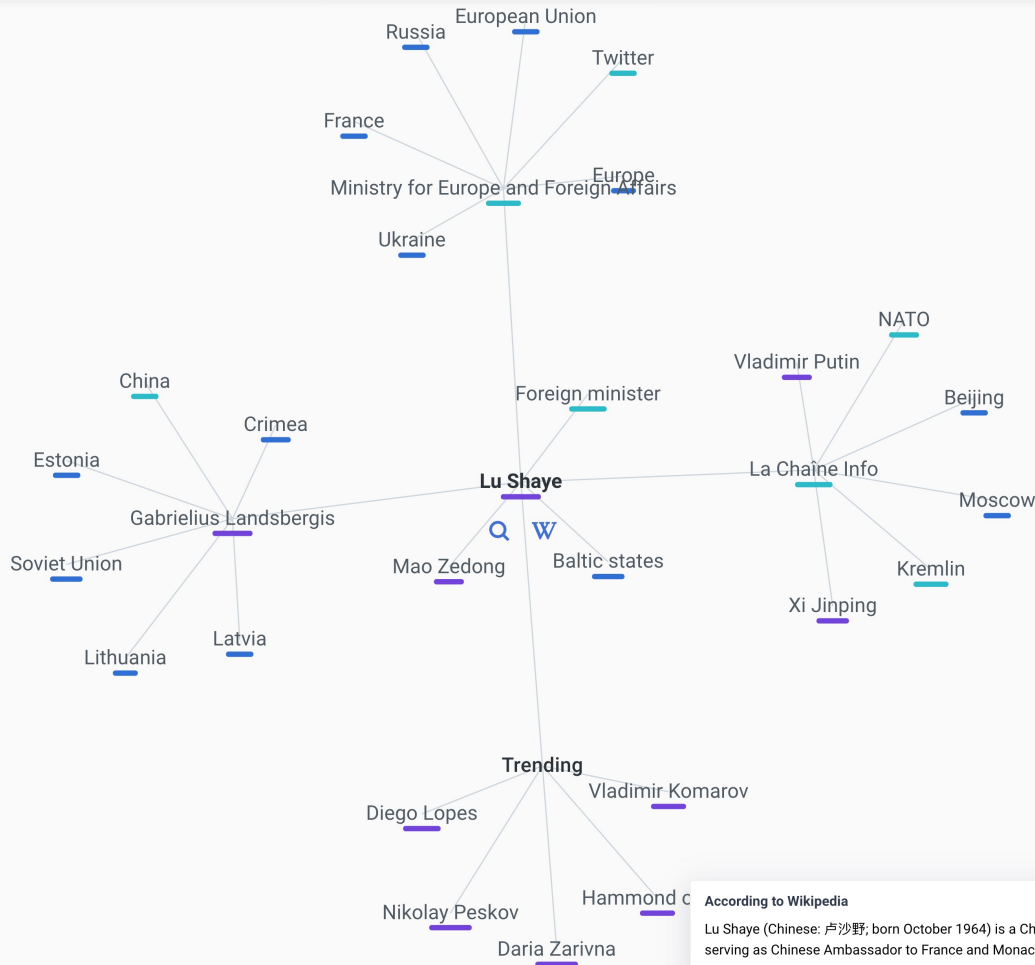
Entity Network

Saved Articles

Dashboards

Reports

Places France | Ukraine | Russia | China | European Union | People Emmanuel Macron | Volodymyr Zelenskyy | Vladimir Putin | Organization Kremlin | European Parliament | NATO



Filters

- ☐ politics 70
- ☐ unrest, conflicts and war 64
- ☐ diplomacy 35
- ☐ defence 34
- ☐ armed Forces 21
- ☐ international relations 21

ORGANIZATION

- ☒ Kremlin 61
- ☒ European Parliament 53
- ☒ NATO 35
- ☐ Telegram (software) 49
- ☐ Agence France-Presse 45
- ☐ Twitter 45

PEOPLE

- ☒ Emmanuel Macron 100
- ☒ Vladimir Putin 74
- ☒ Volodymyr Zelenskyy 20
- ☐ Lu Shaye 25
- ☐ Dmitry Peskov 20
- ☐ Yevgeny Prigozhin 19

PLACES

- ☒ Russia 81
- ☒ Ukraine 77
- ☒ European Union 31
- ☒ China 27
- ☒ France 26
- ☐ United States 49

OTHER

- ☐ Russo-Ukrainian War 24
- ☐ 2022 Russian invasion of... 22
- ☐ Cold War 11
- ☐ Euro 7
- ☐ World War II 7
- ☐ Gross domestic product 7

global

Summary

Articles

Aggregate

Tiles

Headlines

Analyze

Trending Entities

Entity Network

Saved Articles

Dashboards

Reports

Baltics condemn China envoy's stance on ex-Soviet nations

The Nation Kenya web

Today at 12:00 AM



Article

The three Baltic states have strongly condemned comments by China's envoy to France, who appeared to suggest in a recent French television interview that former Soviet republics aren't sovereign nations.

The foreign ministers of Estonia, Latvia and Lithuania in separate announcements late Saturday deemed statements by Lu Shaye, China's ambassador to France, as unacceptable.

In a recent interview with the French news channel LCI, he was asked if he thought that the Crimean Peninsula belongs to Ukraine. Russia annexed Crimea from Ukraine in 2014, a move that most of the world denounced as illegal.

"That depends ... on how one perceives this problem," the envoy told the broadcaster. "There's the history. Crimea was at the beginning Russian, no? It was (Soviet leader Nikita) Khrushchev who gave Crimea to Ukraine in the era of the Soviet Union."

When the channel's presenter noted that according to international law, Crimea is part of Ukraine, the Chinese ambassador drew a parallel to the former Soviet republics — including the three Baltic nations — that broke free after the USSR collapsed in 1991.

"With regards to international law, even these ex-Soviet Union countries,



Soviet Union

1/6



Mark as Unread

Save Article

Topics

politics

diplomacy

unrest, conflicts and war

Entities



Soviet Union



Estonia



Latvia



Lithuania



Lu Shaye



La Chaine Info



Crimea



Ukraine



Russia



World



Gabrielius Landsbergis



Baltic states



Margus Tsahkna



Tallinn

Duplicates

Cluster

• Entity linking: Wikipedia

• Languages:

- PT, FR, EN, ES, DE (prior)
- AR, TR, LV, UK, RU, PL, ... (new)

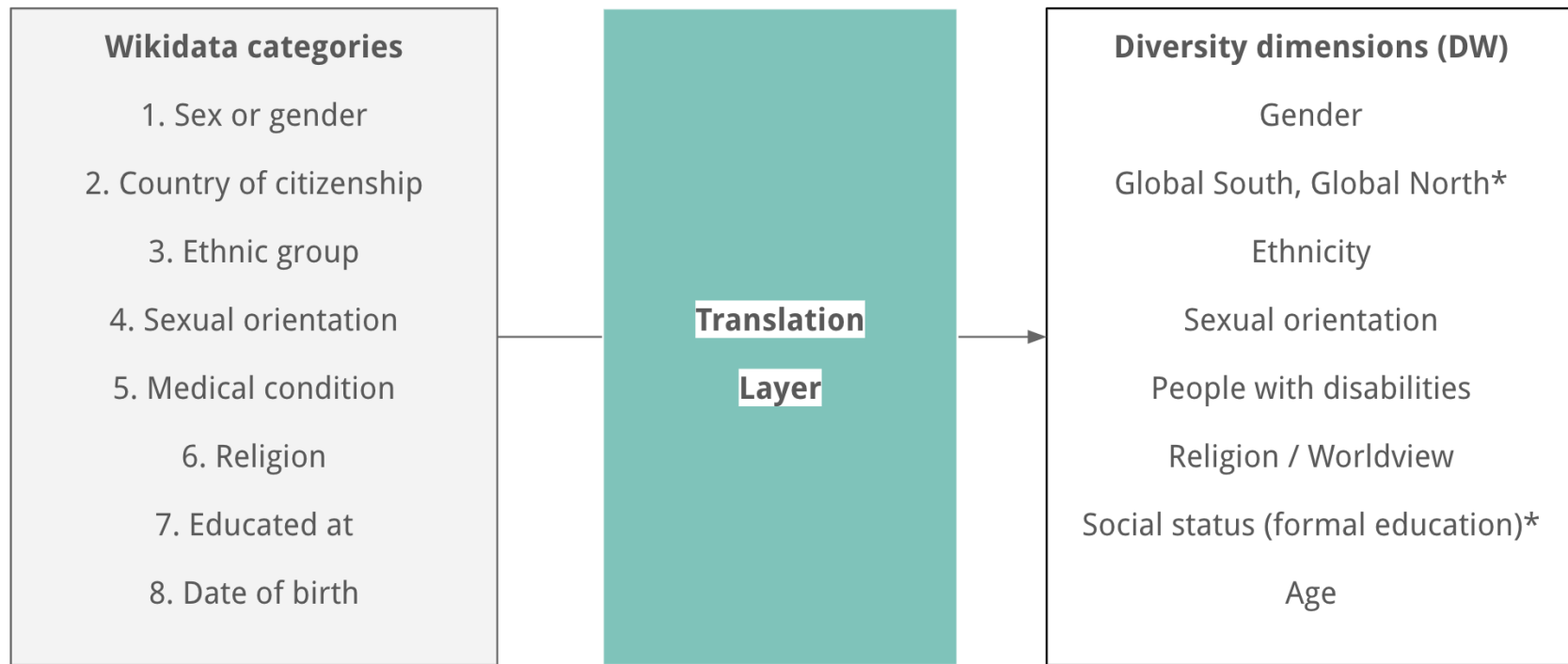
SELMA: Counting diversity

Empowering
democracy

“An editorial team loses **credibility**
if it doesn't represent **diversity** in its many forms:
gender, ethnicity, sexual orientation,
disability and **different world views.**
Credibility is the **most important currency** in journalism.”

Peter Limbourg, Director General Deutsche Welle

Deutsche Welle diversity dimensions



View **DWNEWS** ▾ Date **Today** ▾ Aggregator Topics ▾

Nobelova nagrada za
mirovne aktiviste iz
Belorusije, Ukrajine i Rusije

Europenii, relativ
de acord în fața
crizei

Msichana
Jasiri:
Watumia
mpira wa
magongo...

Burkina : la Cédéao
convaincu par Ibrahim Traoré

Cautious optimism ahead
of Ethiopia-Tigray peace
talks

Spanish Prime Minister
Pedro Sanchez and...

Trump suspected of
holding more classified...



Atomsko oružje: Putin
ucjena koja se shvaća
ozbiljno

Irán dice que la
muerte de
Mahsa Amini...

At least 70
after
Zaporizhzhia

POLITICS

Український Центр
громадянських
свобод став одним і...

مقلاع عملاق #الفضاء
#قمر_صناعي...

EU leaders hold tense talks on...

O que se sabe sobre o teto ao preço do
petróleo russo

Niemcy. 800 milionów
maseczek trafi do
spalarni

Twitter Elon Musk
lawsuit delayed,
allowing more time
to close deal

85 Prozent der

EUA vão
perdoar
condenados...

#México emitió el jueves un aviso
#epidemiológico para todas las
unidades de salud a mantenerse...

Inspeção sueca
reforça suspeitas
de sabotagem em
gasodutos

Policia
mexicana
atribuyó los
ataques a
"Los

#EstadosUnidos: ¿respiro para los
"soñadores" o sentencia de muerte...

SOCIAL ISS

Incêndio florestal atinge
estátuas gigantes da Ilha
de Páscoa

L'Allemagne
inaugure une pla
Rudolf Douala...

ARTS, CULTURE AND ENTERTAIN



HEALTH

Moçambique
Mobilização
contra o assé

EDUCATI

Entity Gender

SENSITIVE

- ☐ Male gender 125
- ☐ Female gender 48

Entity Occupation

SENSITIVE

- ☐ Politician 84
- ☐ Writer 16
- ☐ Journalist 11
- ☐ Football player 10
- ☐ Diplomat 10



Entity Ethnic Group

SENSITIVE

- ☐ Russians 70
- ☐ Ukrainians 54
- ☐ White Americans 39
- ☐ Germans 37
- ☐ Han Chinese 11



Counting diversity: ~~Initial plan~~ current decision

For each entity representing a **person**, we collect the following diversity properties from Wikidata:

date of birth (P569), **binary sex or gender** (P21), ~~country of citizenship (P27), occupation (P106), ethnic group (P172), sexual orientation (P91), religion or worldview (P140), medical condition (P1050), educated at (P69), member of political party (P102),~~ **continent** (P30).

Beyond SELMA: Countering disinformation

Defending
democracy

- Foreign information manipulation and interference (FIMI)
- Tactics, techniques and procedures (TTP)
- Russian FIMI and **Kremlin-related entities** (KRE)
- Media monitoring and analysis w.r.t. KREs and their TTPs
- Construction of a **KRE knowledge graph** – a source of “truth”
- Human-in-the-loop vs. AI-in-the-loop

PiniTree platform for knowledge graphs

- Creation and maintenance of private/shared knowledge bases (KB)

The screenshot displays the PiniTree platform interface, which is used for creating and maintaining knowledge bases (KB). The interface is divided into several sections:

- Story:** The main text area on the left, showing a document with highlighted mentions. The text includes: "Dombrava pieļauj, ka ebreju restitūcijas līdzekļi varot nonākt pie nauda...", "Rīga, 30.sept., LETA. Pieņemot Labas gribas atbildinājuma Latvijas ebreju kopienai likumprojektu, līdzekļi tiks iedoti "naudas atmazgātājiem, personām, kas sadarbojas ar krievu oligarhiem, un tādām personām, kurām reputācija ir tālu no labas", šādu viedokli šodien Saeimas sēdē pauda Nacionālās apvienības deputāts Jānis Dombrava.", "Mārtiņš Bondars (AP) atzīmēja, ka šis ir jautājums ar "garu bārdu". Viņa vērtējumā, lēmuma pieņemšanai būtu arī starptautiska nozīme.", "Dombrava debatēs pauda nostāju, ka ar šo likumprojektu līdzekļus esot paredzēts nodot personām, kuras iekļautas "melnajā sarakstā".", "Politikis kļūstja, ka viņa pārstāvētais politiskais spēks nešķirojot, vai noziegumus paveica nacistiskā Vācija vai Padomju savienība, nosoda abu režīmu noziegumus, kā arī iespēju robežās atbalstot visus upurus, viņu ģimenes un tuviniekus. Pēc viņa vārdiem, pēc neatkarības atjaunošanas visa Latvijas likumdošana esot vērsta uz to, lai veicinātu vēsturiskā taisnīguma atjaunošanu, tostarp atdodot ipašniekiem viņu ipašumus un cenšoties kompensēt radušos zaudējumus."
- Mention:** A blue callout box pointing to the highlighted mentions in the text.
- Related mention:** A blue callout box pointing to the highlighted mentions in the text.
- Selected mention:** A blue callout box pointing to the selected mention in the sidebar.
- Options for linking:** A blue callout box pointing to the options for linking in the sidebar.
- Relations (facts):** A blue callout box pointing to the relations (facts) in the sidebar.
- Mentions in other series:** A blue callout box pointing to the mentions in other series in the sidebar.

The sidebar on the right shows a list of mentions and relations. The top section, titled "Bondars", lists several mentions with their IDs and names: 18b84a15: Aldis Bondars, 652b0b70: Alvis Bondars, ae3f7adc: Andrejs Bondars, a6c6c609: Andris Bondars, d53f0b56: Andris Bondars, 14ac56c7: Anrijs Bondars, 7c1ba6de: Dimitrijs Bondars, f9797b43: Guntis Bondars, 636001a7: Jānis Bondars, 274f65e2: Jāzeps Bondars, 2927a143: Mārtiņš Bondars. Below this, there is a section titled "Active" with a "Select" button and a list of relations (facts) with their IDs and names: b9ae43c1-705c-4... Mārtiņš Bondars, b9ae43c1-705c-4... Mārtiņš Bondars, b9ae43c1-705c-4... Mārtiņš Bondars. At the bottom, there is a section titled "Neighbours:" with a list of mentions and their IDs: 8 Nacio..., 7 Krišjā..., 5 P... and a section titled "Mentions:" with a list of mentions and their IDs: 1 mention in document de845a7a at 2021-0..., 1 mention in document b47282e8 at 2021-0....

TTP analysis with the large language models

- Sample article #1

[https://sputnikglobe.com/20230415/ex-pentagon-officer-\[-\]-1109548524.html](https://sputnikglobe.com/20230415/ex-pentagon-officer-[-]-1109548524.html)

Ex-Pentagon Officer: Russia's Robust Capabilities in Electronic Warfare Second to None

- Hey ChatGPT, are there any FIMI elements or propaganda TTPs used in this article?
 - Appeal to authority
 - Positive portrayal of Russian capabilities
 - Selection of information
 - Reliance on leaked documents
 - Fear appeal
 - Implicit negative portrayal of the US and Ukraine
 - Absence of counterarguments or alternative perspectives

TTP analysis with the large language models

- Sample article #2

<https://www.bbc.com/news/world-europe-65173043>

Nato's border with Russia doubles as Finland joins

- Hey ChatGPT, are there any FIMI elements or propaganda TTPs used in this article?
 - This article **does not** prominently employ propaganda or disinformation TTPs, still:
 - Appeal to authority
 - Framing
 - Emotional language
 - Overall, the article appears to be relatively **balanced** and **informative**, providing **reactions of various parties** involved

Key messages

- Various media spread **different opinions and beliefs**, incl. malicious
- Large language models can handle **many “truths”**
 - Instruction tuning can make them purposely **“biased”** towards a task/truth
- Public and private KBs are used as authoritative sources of **some truth** which, in the case of MM, is “grounded” by **named entity mentions** in the news content, facilitating **fact-checking** but raising **privacy risks**
- AI-aided technology is there to help **empower and defend democracy**
 - Is EU open for it and ready to act while still preserving its values?



Thank you!

Co-authors:

- Guntis Bārzdiņš (IMCS)
- Sebastião Miranda (Priberam)
- Kay Macquarrie (DW)



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957017.