

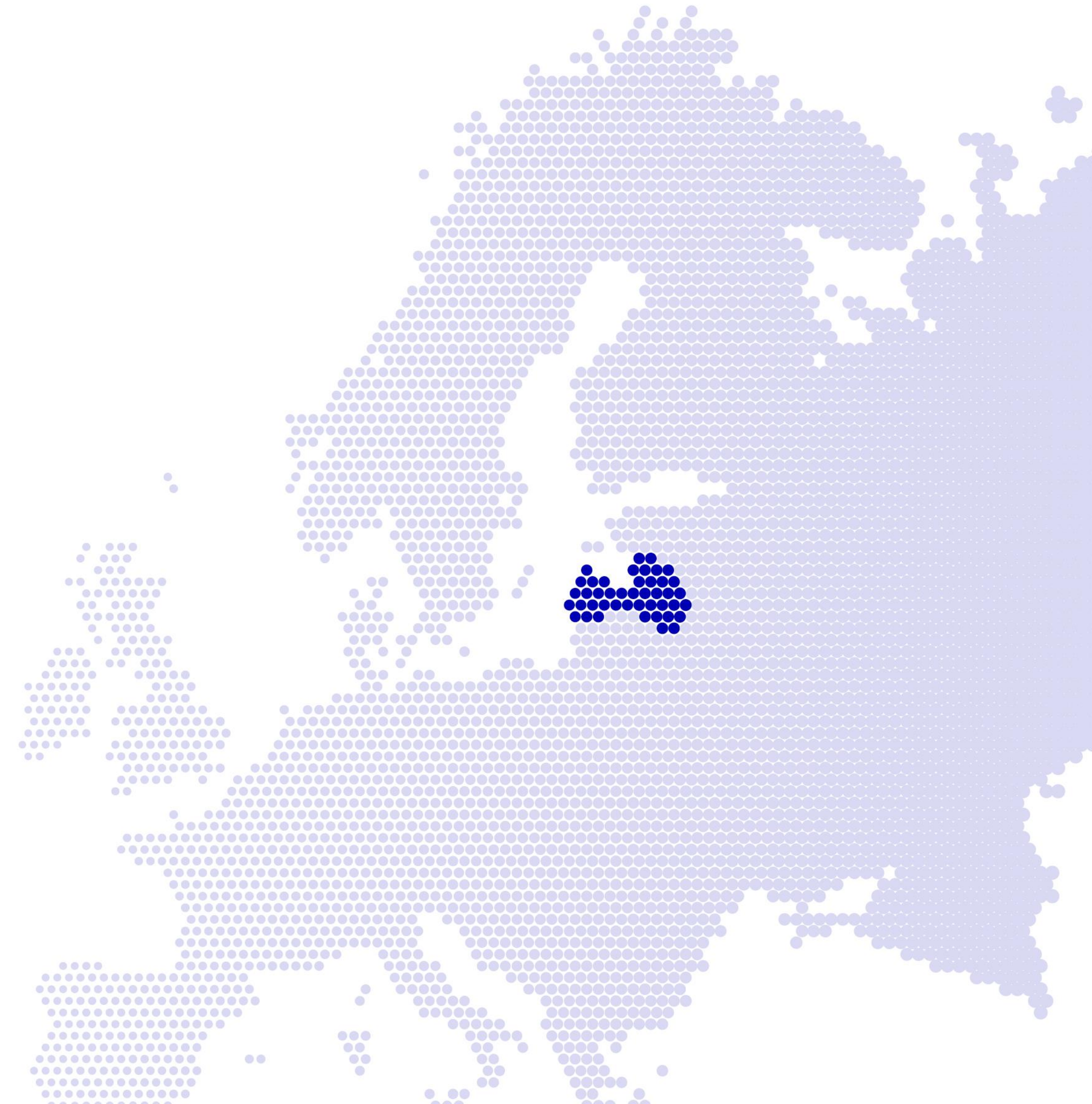


Ministry of
Education and Science
Republic of Latvia

LANGUAGE TECHNOLOGIES IN THE DIGITAL HUMANITIES ECOSYSTEM

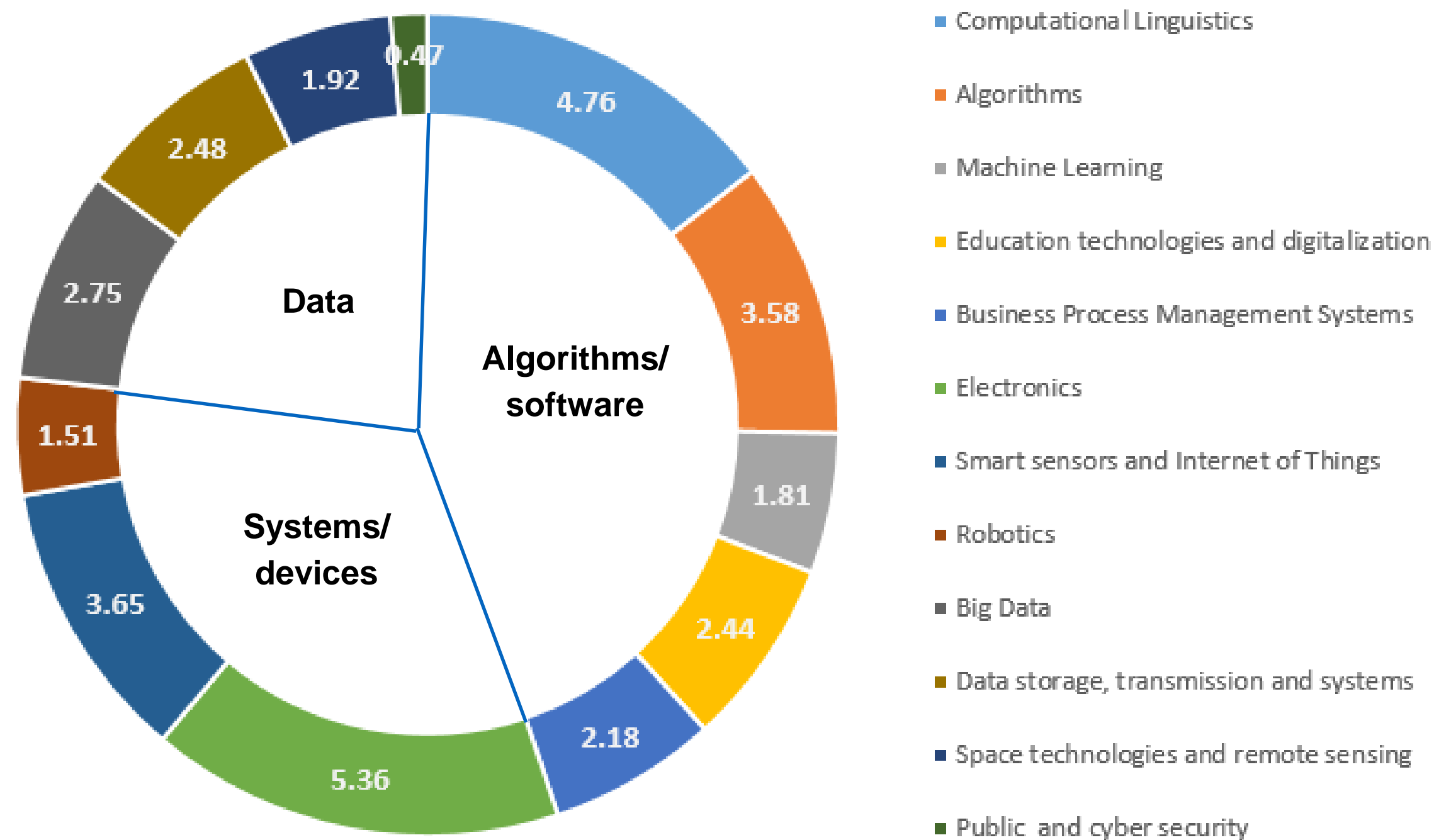
Aleksandrs Mārtiņš Blūms
RIS3 Expert
Aleksandrs.Blums@izm.gov.lv

researchLatvia[★]
Value Through Knowledge

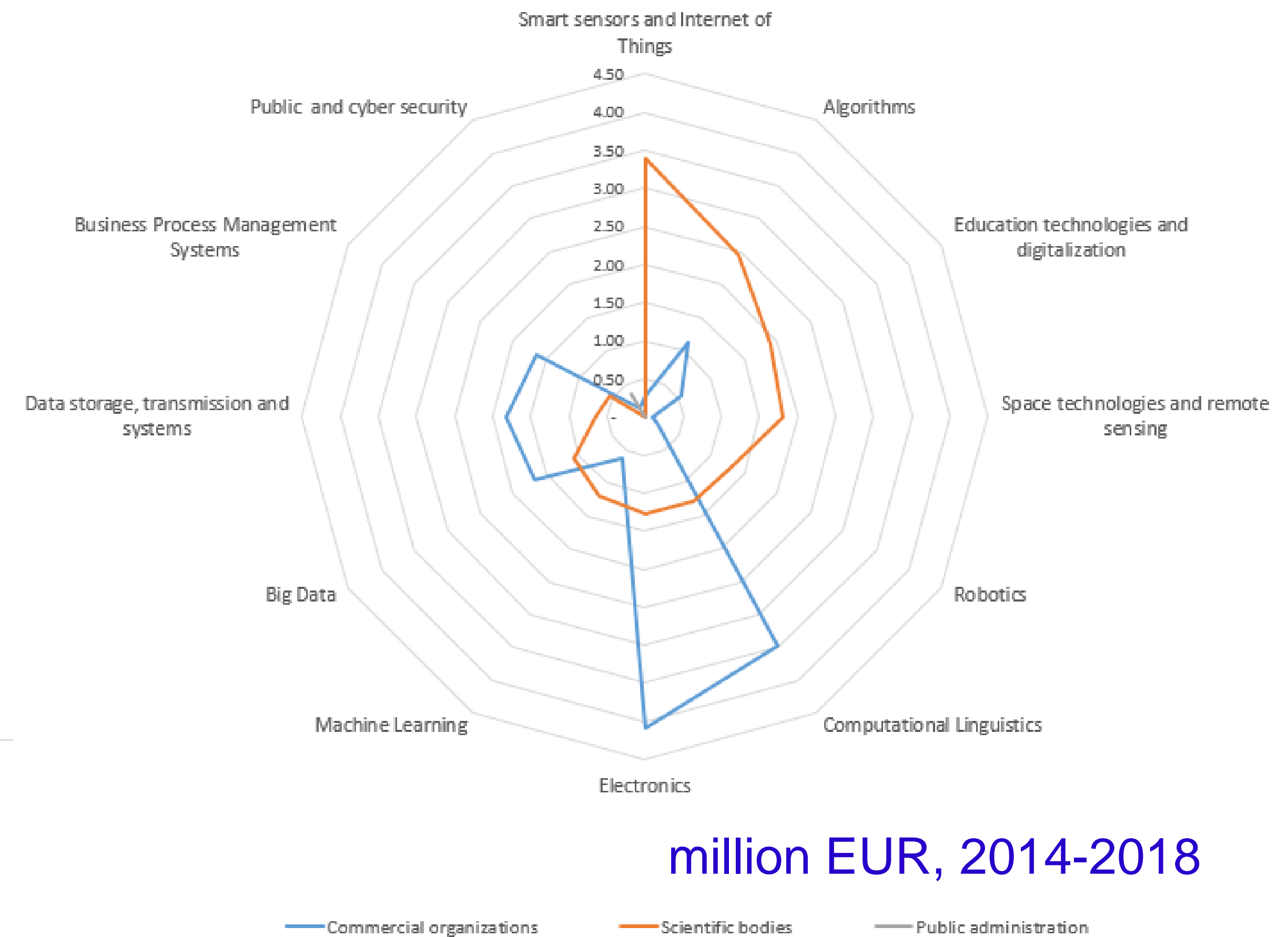


LANGUAGE TECHNOLOGIES IN ICT

Funding for R&D projects in RIS3 ICT thematic areas, scientific and commercial institutions

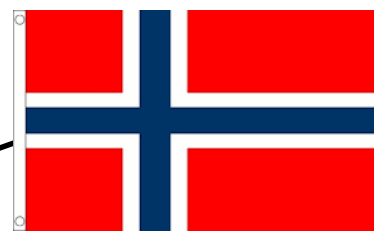


million EUR, 2014-2018



million EUR, 2014-2018

LANGUAGE TECHNOLOGIES IN ICT



Annotated Longitudinal Latvian Children's Speech Corpus

Latvian Speech Recognition Corpus

LVK2018 Balanced corpus of modern Latvian

LVTB Grammar model of Latvian Treebank

UDLV Hybrid Dependency-Constituency Treebank

Latvian Learner Corpus

Corpus of the Saeima with Multilingual Standards



TRL 4-6	Competence Centers (1.2.1.1.)	1,419,752	30%
TRL 3-5	Research results commercialization (1.2.1.2.)	n/a	n/a
TRL 4-8	Horizon 2020	1,375,750	29%
TRL 2-3	Practical oriented research (1.1.1.1.)	1,338,661	28%
TRL 1-2	Post-doctoral research (1.1.1.2.)	133,806	3%
TRL 1-2	Fundamental and applied research program	495,446	10%

STATE LANGUAGE POLICY GUIDELINES 2021-2027

Latvian language research and development



Developing a knowledge base in linguistics and related fields

**Availability of resources for Latvian
language research (incl. in abroad
institutions)**

Societal impacts of research (tangible
benefits for public investments)

Latvian language globally

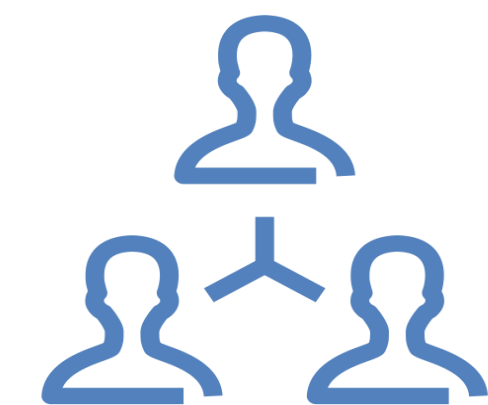


Promoting language, increasing competitiveness

International cooperation in the
field of language policy
Compliance with international
norms (linguistic rights)

National security and legal context,
public involvement in global and
regional context, diaspora, Livonian
language, participation in
international projects

Increasing usage of the language



Environment and resource sustainability

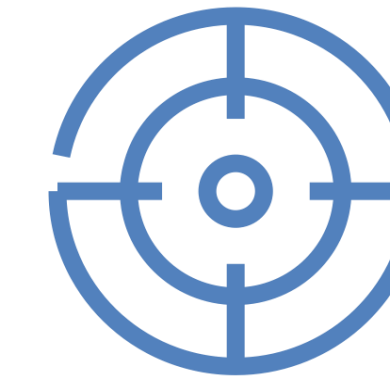
Quality Latvian language
learning opportunities

Developing capacities of
Latvian language educators

STATE LANGUAGE POLICY GUIDELINES 2021-2027

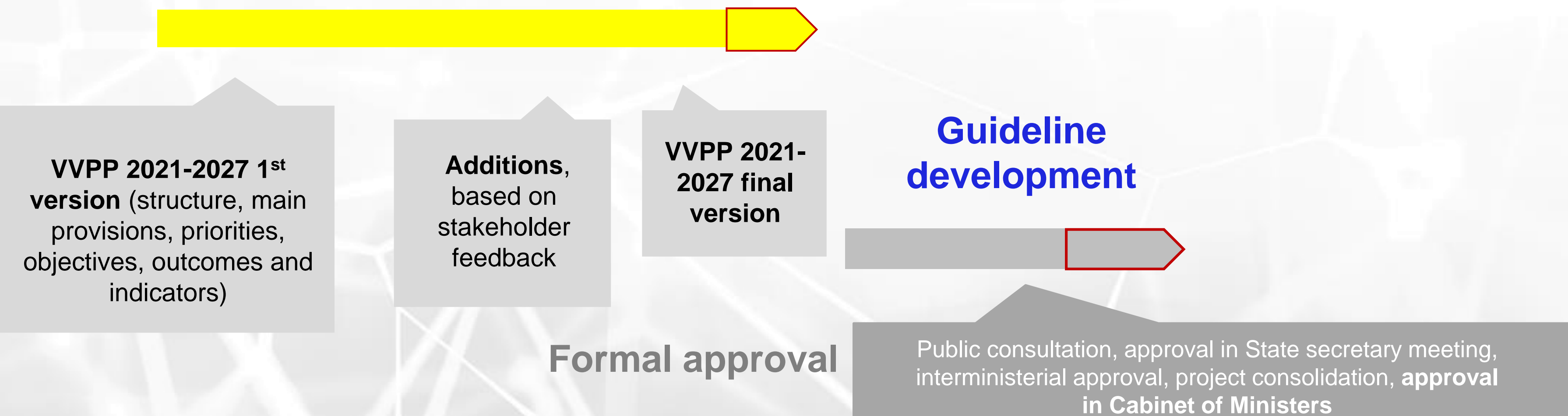
- MoES: VVPP 2021-2027 working groups
- Expert and focus group meetings
- Public consultation
- Online consultation
- Events organized by society and societal partners
- Other initiatives

Involving citizens and experts in drafting the State Language Policy Guidelines 2021-2027



National Development Plan 2027

Strengthen the constitutional provision:
«Trust in Latvia and the Latvian language as the only official language, freedom, equality, solidarity, justice, honesty, work ethic and family as the basis of a cohesive society»



LANGUAGE TECHNOLOGIES IN DIGITAL HUMANITIES

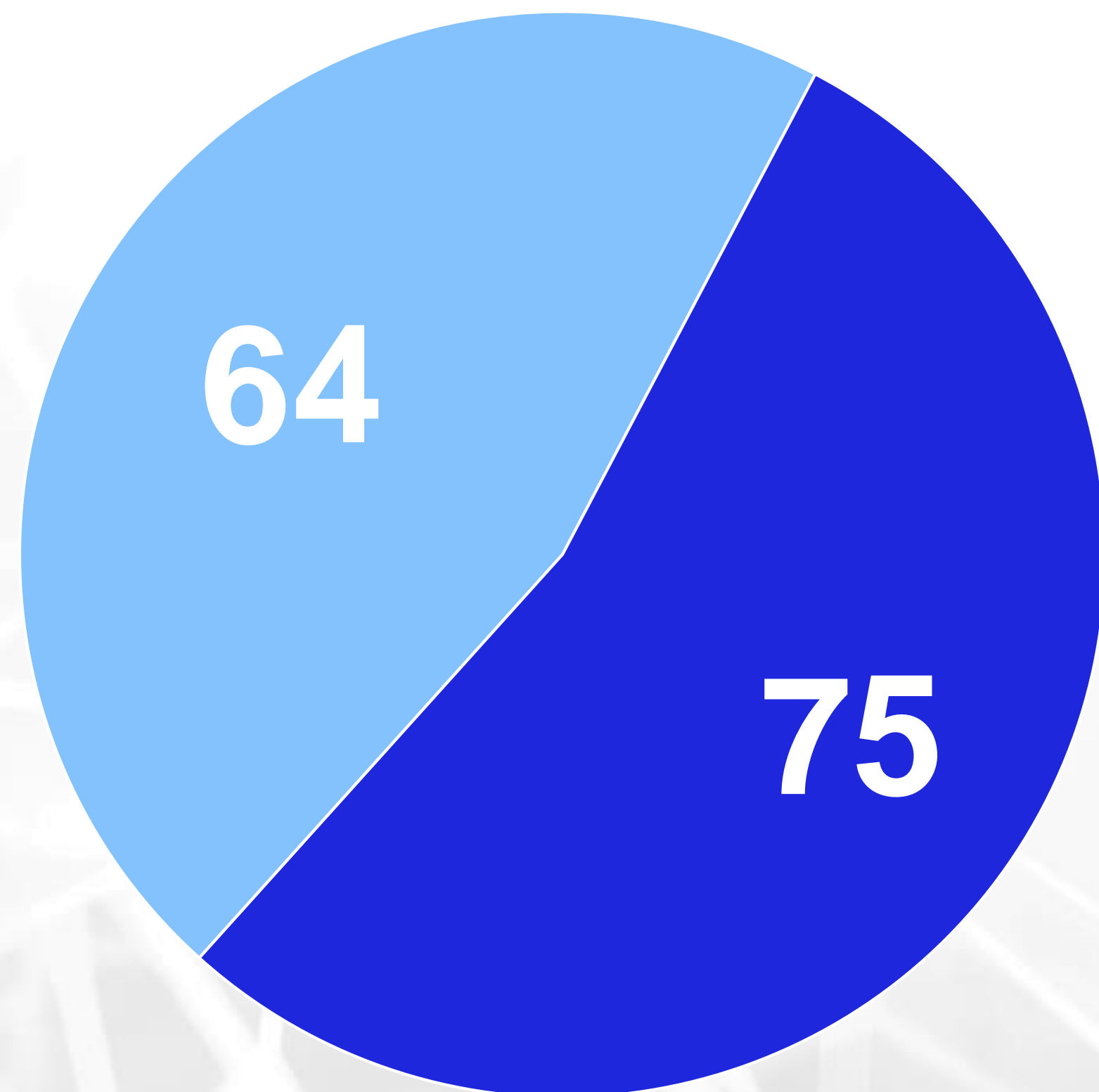
Rīka nosaukums	Resursa tips (Arhīvs)	Turētāji	Apraksts	Mājaslapas saite	Finansē	Finansē	Lietotāji	Redzēju	Ievadīts	Valodu
"Abstract Meaning Representation" (AMR) marķējuma slānis	Valodas korpuss	LUMII	12,7 tūkst. teikumu	https://github.com/LUMII-AMR	ERAF, valsts budžets	2017-2019	Sk. FullStack	Korpusu ne	LU MII	1
"Coreferences" (CoRef) marķējuma slānis	Valodas korpuss	LUMII	23,3 tūkst. saišu	https://github.com/LUMII-CoRef	ERAF, valsts budžets	2017-2019	Sk. FullStack	Korpusu ne	LU MII	1
"FrameNet" marķējuma slānis	Valodas korpuss	LUMII	20,9 tūkst. piemēru	https://github.com/LUMII-FrameNet	ERAF, valsts budžets	2017-2019	Sk. FullStack	Korpusu ne	LU MII	1
"Named Entities" (NE) marķējuma slānis	Valodas korpuss	LUMII	10,6 tūkst. piemēru	https://github.com/LUMII-NE	ERAF, valsts budžets	2017-2019	Sk. FullStack	Korpusu ne	LU MII	1
"PropBank" marķējuma slānis	Valodas korpuss	LUMII	20,9 tūkst. piemēru	https://github.com/LUMII-PropBank	ERAF, valsts budžets	2017-2019	Sk. FullStack	Korpusu ne	LU MII	1
"Universal Dependencies" (UD) marķējuma slānis	Valodas korpuss	LUMII	12,7 tūkst. teikumu	https://github.com/LUMII-UD	ERAF, valsts budžets	2017-2019	Sk. UDLV	Sk. UDLV	LU MII	1
AGROVOC tezaurs	datubāze BIS ALEPH	LLU Fundamentālā bibliotēka							AGROVOC	LLU
AkadTerm	datubāze	LZA	LZA Terminoloģijas izstrādāto un apstiprināto terminu sa	http://termini.lza.lv/term.php			Regulāra uzturēšana	Paredzēts	LZA	1
Austrumu diasporas latgalisko izlokšņu ierakstu datubāze „Reitu d	datubāze	RTA	Brīvpieejas datubāze, kur ievietoti dialektoloģijas ekspedi	https://reiti.rta.lv/index.htm	RTA grants	2018-2019	nav datu	Balstoties	RTA	1
Baltijas valstu lingvistiskā ainava	Datu bāze ar uzrakst	RTA, Reģio	Lingvistiskās ainavas (LA) kā sociolingvistikas pētījumu v	http://ldata.ru.lv/	Projekts „T	2008.-201	nav datu	Lingvistiskā	RTA	1
Citvalodu personvārdu atveide latviešu valodā	lietotne (datubāze)	LVA	Elektroniska daudzvalodu priekšvārdu un uzvārdu vārdnīc	https://www.personvardua	valsts budž	no 2017. g	590 (2019	regulāra pa	LVA	1
CLARIN Latvija	infrastruktūra	LUMII	Vienota valodas resursu un tehnoloģiju infrastruktūra	https://clarin.lv	ERAF, valsts budžets	2008-2021	Pašlaik izve	CLARIN ER	LU MII	1
dalība - projektā -e- pupa, Tava valodas pasaule.	vārdnīca	Latviešu va	Šķirkļu izstrādāšana.	http://epupa.valoda.lv/	Valsts finansējums.				LU HZF	1
Daudzslāņu valodas resursu kopa teksta semantiskai analīzei un	valodas korpuss	LUMII	Marķējuma slāņi: UD, FrameNet, PropBank, NE, CoRef, A	https://github.com/LUMII-	ERAF, valsts budžets	2017-2019	Daudzslāņu	Šī resursa	LU MII	1
Elektroniska lingvokulturoloģiska latviešu valodas vārdnīcas "e-PU	e-vārdnīca	LVA	Vārdnīca paredzēta 7.-12. klases mazākumtautību skolē	http://epupa.valoda.lv/	valsts budž	no 2011.g	654 (2019	Darbs pie v	LVA	1
Interneta agresivitātes indekss	Valodas korpuss, plat	RSU un LU	Rīks paredzēts - Internetā publicēto komentāru sistemāt	http://barometers.korpuss.l	Tiesībsarga	2011 - 201	Specializēt	Vēl 3 - 5 vā	RSU, LU M	1
Latvian NLP Tool Pipeline (NLP-PIPE)	programmriku kopum	LUMII	Mērogojama latviešu valodas automātiskas analīzes rīkk	http://nlp.aialab.lv	ERAF, valsts budžets	2017-2019	NLP-PIPE p	Atvērtās N	LU MII	1
Latviešu literārās valodas vārdnīca (LLVV)	vārdnīca	LUMII, LUL	64 tūkst. šķirkļu	https://tezaurs.lv/llvv/	valsts budž	2006-201	Tēzaurs.lv	Joprojām r	LU MII	1
Latviešu valoda iesācējiem	mācīb līdzeklis	LUMII	Interaktīvs latviešu valodas pašmācības līdzeklis	http://valoda.aialab.lv/latval	valsts budž	1997	Mācīb līdzek	ļu vietnei ir	LU MII	1
Latviešu valoda pamatskolai	mācīb līdzeklis	LUMII	Interaktīvi testi un 250 interaktīvi pašpārbaudes uzdevum	http://valoda.aialab.lv/latval	valsts budž	2003-2004	Mācīb līdzek	ļu vietnei ir	LU MII	1
Latviešu valoda sākumskolai	mācīb līdzeklis	LUMII, LVA	Vairāk nekā 180 interaktīvu pašpārbaudes uzdevumu	http://valoda.aialab.lv/latval	valsts budž	2004	Mācīb līdzek	ļu vietnei ir	LU MII	1
Latviešu valoda vidusskolai	mācīb līdzeklis	LUMII	Latviešu valodas mācību materiāli un interaktīvi pašpārba	http://valoda.aialab.lv/latval	valsts budž	1998-200	Mācīb līdzek	ļu vietnei ir	LU MII	1
Latviešu valodas apguvēju korpuss (LaVA)	valodas korpuss	LUMII	Plānotais apjoms: 1000 tekstvienību	http://lava.korpuss.lv	valsts budž	2018-202	Izstrādes u	Pašlaik kor	LU MII	1
Latviešu valodas pašmācības līdzeklis pieaugušajiem "E-Laipa"	lietotne	LVA	Latviešu valodas pašmācības rīks e-Laipa saturiski ir saist	https://elaipa.lv	valsts budž	no 2017.g	1777 (201	Pašlaik pub	LVA	1
Latviešu valodas rokasgrāmata	lietotne (interaktīva)	LVA	Enciklopēdiska rakstura darbs, elektronisks informatīvs r	http://valodasrokasgramat	valsts budž	no 2018.g	755 (2019	regulāri pa	LVA	1
Latviešu valodas runas korpuss	valodas korpuss	LUMII, LET	100 stundas	http://runa.korpuss.lv	ERAF, privā	2013	Korpora m	Līdz šim nē	LU MII	1
Latviešu valodas runas korpuss (LaRko)	valodas korpuss	LUMII	8 stundas	http://larko.aialab.lv	valsts budž	2014	Korpora or	Korpusu ne	LU MII	1
Latviešu valodas seno tekstu korpuss (Senie)	valodas korpuss	LUMII, LUH	1,1 milj. vārdlietojumu	http://korpuss.lv/id/Senie	valsts budž	2002-2014	Izmanto vē	Nozīmīgs k	LU MII	1
Latviešu valodas sintaktiski marķētais korpuss (LVTB)	valodas korpuss	LUMII	13,6 tūkst. teikumu	http://korpuss.lv/id/LVTB	ERAF, valsts budžets	2010-2014	LVTB izma	Sintaktisk	LU MII	1
Latviešu valodas universālo atkarību korpuss (UDLV)	valodas korpuss	LUMII	13,6 tūkst. teikumu	http://korpuss.lv/id/UDLV	ERAF, valsts budžets	2015-2014	UDLV izma	UDLV tiek j	LU MII	1
Latviešu valodas vēsturiskā vārdnīca (LVVV)	vārdnīca	LULaVI, LU	1200 šķirkļu	https://tezaurs.lv/llvv/	valsts budž	2013-202	Tēzaurs.lv	Seno tekst	LU MII	1
Latviešu valodniecības attīstība: informatīvi izglītojoša elektronisk	lietotne (datubāze)	LVA	Elektronisks informatīvs mācību palīg līdzeklis latviešu vald	http://www.lingvistiskakart	valsts budž	no 2016.g	96 (2019.g	regulāra pa	LVA	1
Latvijas Nacionālais terminoloģijas portāls termini.gov.lv	Terminu datubāze, te	Kultūras ur	Terminoloģijas portāls, kurā apvienoti Latvijas Zinātņu ak	https://termini.gov.lv/	LR ar ERAF līdzfinansē	5476	Katru gadu	VVC, KISC	1	
Latvijas Republikas 5.-12. Saeimas sēžu stenogramu korpuss	valodas korpuss	LUMII, RSU	20,8 milj. vārdlietojumu	http://korpuss.lv/id/Saeima	valsts budž	2013-2014	Specializēt	Korpusu re	LU MII	1
Lībiešu rakstu valodas tekstu krātuve un korpuss	Datubāze un korpuss	LU Lībiešu	Indeksēts rakstu valodas korpuss, ievter publikācijas, rak	http://lingua.livones.net/lv/	Uzturēšana	Dalēji līdz 2	Satura ievi	Resurss ir j	LU Lībiešu i	1
Lībiešu valodas leksikas datubāze	Datubāze un vārdnīca	LU Lībiešu	Leksikas datubāze veidota uz Lībiešu-igauņu-latviešu vār	http://www.livones.net/llv/	-	-	605 (2019	Resurss ir j	LU Lībiešu i	1
Lībiešu valodas morfoloģijas datubāze	Datubāze	LU Lībiešu	Lībiešu valodas vārdu formu datubāze, tiek regulāri papil	http://www.livones.net/llv/	-	-	605 (2019	Resurss ir j	LU Lībiešu i	1
Lībiešu vietvārdu datubāze un karte	Datubāze un ģeotelpi	LU Lībiešu	Lībiešu vietvārdu krātuve, kas piesaistīta ģeotelpiskajai in	http://lingua.livones.net/lv/	FLPP (līdz 2	Līdz 2022.	605 (2019	Resurss ir j	LU Lībiešu i	1
Līdzsvarotais mūsdienu latviešu valodas tekstu korpuss (LVK)	valodas korpuss	LUMII	10 milj. vārdlietojumu	http://korpuss.lv/id/LVK20	ERAF, valsts budžets	2007-201	LVK izman	Līdzsvarot	LU MII	1
Liepājas Universitātes Kurzemes humanitārā institūta dialektoloģij	Arhīvs	Liepājas Ur	Kurzemes reģiona ekspedīciju rezultātā iegūto materiālu	Pieejams lokāli Liepājas Uni	Liepājas Ur	Sākot no 2	Lietotāji ne	Analogo m	LiepU	1
Lietuviešu - latviešu - latgaliešu vārdnīca (LLL)	vārdnīca	RTA, sadar	LLL ir veidota uz „Lietuviešu valodas rakstīto avotu biež	http://hipilatit.ru.lv/dictionary/	Latvijas-Li	2011 - 20	nav datu	Resursu ier	RTA	1
Lietuviešu-latviešu-lietuviešu paralēlais tekstu korpuss (LiLa)	valodas korpuss	LUMII, VDU	8 milj. vārdlietojumu	http://korpuss.lv/id/LiLa	ERAF, valsts budžets	2011-2014	Korpusu iz	Korpusu bī	LU MII	1
Māci un mācies latviešu valodu	digitālā mācību resurs	LVA	LVA vietnes (www.valoda.lv) apakšsadaļā Māci un mācie	http://maciunmacies.valod	valsts budž	no 2013.g	8296 (201	2019./202	LVA	1
Hugo.lv	Lietotne, Platforma	KISC	Automātiskās tulkošanas mobilā lietotne latviešu-angļu-k	hugo.lv	ES finansējums un valsts budžets		Ikvienam iē	KISC	1	
Medicīnas terminu angļu-latviešu-angļu tulkojošā un sarunvārdnīca	Elektroniska mācību v	Ventspils A	Divvalodu tulkojošā un sarunvārdnīca, kas paredzēta mācību nolūkiem	https://tezaurs.lv/mev/	Ventspils A	01.07.-31.	VeA TSF sp	Vārdnīcu ar	Ventspils A	1
Milenbaha-Endzefina latviešu valodas vārdnīca (MEV)	vārdnīca	LUMII	133 tūkst. šķirkļu	https://tezaurs.lv/mev/	privātais fir	1994-2004	Tēzaurs.lv	Nozīmīgs lē	LU MII	1
Mobilās lietotnes prototips. Jaunā botāniskā vārdnīca. Termini latv	Elektroniska vārdnīca	Ventspils A	Daudzvalodu tulkojošā un skaidrojošā vārdnīca (mobilās lietotnes prototips), kurā iel		Ventspils A	01.08.-31.12.2017.	(Nepieciešai	Ventspils A	1
Mobilās lietotnes prototips. Juridisko pamatterminu (saistību tiesī	Elektroniska vārdnīca	Ventspils A	Divvalodu tulkojošā un skaidrojošā vārdnīca (mobilās lietotnes prototips), kas paredz		Ventspils A	15.06.-31.	VeA TSF 2.	Nepieciešai	Ventspils A	1
Morfoloģiski marķēts longitudināls bērnu runas korpuss (LAMB)	valodas korpuss	RPVA, UIT	134 stundas audioierakstu, no tām 34 stundas ortogrāfi	http://runa.lamba.lv	EEZ/NFI	2015-2017		Atlikušo 10	LU MII	1
Mūsdienu latgaliešu tekstu korpuss (MuLa)	valodas korpuss	RTA, sadar	MuLa ir specializēts korpuss — to veido konkrēta tekstu	http://hipilatit.ru.lv/bonito/	Latvijas-Li	2011 - 20	nav datu	Mūsdienu lē	RTA, RTA	1
Mūsdienu latviešu valodas vārdnīca (MLVV)	vārdnīca	LULaVI, LU	46 tūkst. šķirkļu	https://tezaurs.lv/mlvv/	valsts budž	2003-202	Tēzaurs.lv	Ilgtermiņā	LU MII	1
Nacionālās mutvārdu vēstures arhīvs	Krājums	Latvijas Ur	Nacionālās mutvārdu vēstures arhīva aizsākums saistām	dzivesstasts.lv; mutvarduv	LZP granti	???	pētnieku gr	Turpināt dē	LU FSI	1



data.gov.lv/dati/lv/dataset/dh-resursi

LANGUAGE TECHNOLOGIES IN DIGITAL HUMANITIES

Number of reported Digital humanities initiatives in Latvia



■ Language technology initiatives ■ Other initiatives

Resource/tool	Monthly users	Description
Latvijas Valsts arhīva datubāze	54166	Page views
Tēzauris.lv	30000	Unique visitors
Raduraksti	26126	Page views
Literatūras vortāls literatura.lv	24598	Unique visitors
Garamantas.lv	18959	Unique visitors
Akadēmisko resursu interneta meklētājs Primo.lv	11432	
Latgales Dati	10000	Page views
Valsts nozīmes bibliotēku Kopkatalogs	9728	
Redzi, dzirdi Latviju	9208	Page views
Māci un mācies latviešu valodu	8296	
Latvijas Nacionālais terminoloģijas portāls termini.gov.lv	5476	
"Kino skolās"	4441	
Latviešu valodas mācīblīdzekļi	3000	
Literāro tekstu vietne "Ubi Sunt"	2500	
Nacionālā muzeju krājuma kopkatalogs	2000	
Latviešu valodas pašmācības līdzeklis pieaugušajiem "E-Laipa"	1777	
Valodas konsultācijas: elektroniska datubāze	1605	
3td e-GRĀMATU bibliotēka	1417	
DH digitālā koprades platforma iesaisties.lv	1200	
Latviešu valodas rokasgrāmata	755	
Paralēlais latviešu teiku korpuss latviešu un vācu valodā	658	
Elektroniska lingvokulturoloģiska latviešu valodas vārdnīcas "e-PUPA"	654	
Lībiešu valodas leksikas datubāze	605	
Lībiešu valodas morfoloģijas datubāze	605	
Lībiešu vietvārdu datubāze un karte	605	
Citvalodu personvārdu atveide latviešu valodā	590	

CHALLENGES IN DIGITAL HUMANITIES



- ✓ **Nationally unified criteria for** describing and identifying digital materials
- ✓ **Centralizing resources** (stimulating cooperation between institutions; integrating tools in centralized platforms, database compatibility, etc.)
- ✓ **Sustainability** of DH resources (fresh content, more features; ageing resources e.g. websites, funding continuity)
- ✓ **Disseminating information** about existing DH resources
- ✓ **Thematic (topic) differences between fundamental and applied research**

STATE RESEARCH PROGRAMME DIGITAL HUMANITIES

Top-down funding instrument

Three thematic objectives:

- 1) Support existing initiatives and tools
- 2) Reduce existing DH resource fragmentation
- 3) Fulfil specific end-user needs via new or existing tools and resources



2020–2022



Total funding
1 309 380 EUR
(3 years)

Mapping exercise
Jan 24

Stakeholder
consultation
Jan 29

Strategic
council
decision
March

Project start
Q4 2020

Results
2022-23

Aleksandrs Mārtiņš Blūms
RIS3 Expert
Aleksandrs.Blums@izm.gov.lv



researchLatvia

