

Jezik kot podatki za družbo in kulturo

Infrastruktura podpora za raziskave, ki uporabljajo jezikovne materiale v dobi umetne intelligence

CLARIN.SI
Strategija 2024-2030



OZADJE	2
UVOD	3
DELOVANJE CLARIN.SI	5
STRATEŠKA USMERITEV CLARIN.SI	6
Podpiranje znanstvene odličnosti in inovacij	6
Spodbujanje strateških partnerstev	7
Vzpostavljanje poti do učinkov	8
STRATEŠKI PROGRAM CLARIN.SI	9
1. steber: Naši uporabniki	10
2. steber: Naša ponudba	11
3. steber: Naš prostor	14

Ozadje

V tem dokumentu je predstavljena Strategija infrastrukture CLARIN.SI za obdobje 2024–2030. Delo infrastrukture v tem obdobju bo temeljilo na predstavljeni strategiji. Strategija sledi Strategiji CLARIN ERIC 2024–2026 ([CE-2023-2333](#)), vendar je osredotočena na slovensko vozlišče CLARIN in na njegovo sodelovanje s CLARIN ERIC in z drugimi vozlišči znotraj CLARIN ERIC. Strategija CLARIN.SI je usklajena s slovenskim [Načrtom razvoja raziskovalne infrastrukture 2030](#) (NRRI 2030) in vključuje daljše časovno obdobje kot Strategija CLARIN ERIC. Večina ciljev bo verjetno ostala enaka v dodatnem časovnem obdobju, ki ga pokriva Strategija CLARIN.SI, čeprav bo v kasnejših letih strategije mogoče treba nekatere cilje spremeniti zaradi hitrega razvoja jezikovnih podatkov in tehnologij.



Uvod

Vizija CLARIN.SI: Vsi digitalni jezikovni viri in orodja iz Slovenije in drugod so dostopni prek enovite prijave v namen podpore raziskovalcem s področij humanistike, družboslovja in drugih področij, povezanih z jezikom.

Poslanstvo CLARIN.SI: Vzpostaviti in vzdrževati infrastrukturo za podporo izmenjavi, uporabi in trajnosti jezikovnih podatkov in orodij za raziskave na področju humanistike, družboslovja in drugih področij, povezanih z jezikom.

Strateška usmeritev CLARIN.SI: Jezik je odraz znanstvenega in družbenega znanja, je orodje za človeško komunikacijo in prepričevanje, je eden glavnih členov identitete posameznikov, skupin, kultur in narodov, ter je orodje človeške kognicije in ustvarjalnega izražanja. Poleg tega jezikovno gradivo tvori precejšen del zgodovinskih zapisov, ki veljajo za kulturno dediščino. CLARIN.SI z omogočanjem dostopa do jezikovnega gradiva, v prvi vrsti do slovenskega, pa tudi do gradiva v drugih južnoslovanskih jezikih, omogoča primerjalno raziskovanje kompleksnih pojavov ter razvoj analitike, ki temelji na podatkih, in računalniškega modeliranja teh pojavov. Viri, ki jih CLARIN.SI daje na voljo, spodbujajo tudi učenje in gradnjo velikih jezikovnih modelov, ključne tehnologije umetne inteligence (UI), pa tudi vzpostavitev metodoloških okvirov za analizo heterogenih podatkov.

Ključna znanstvena področja CLARIN.SI: Jezikoslovje, jezikoslovne študije, jezikovne tehnologije, umetna inteligenca, literarne študije, zgodovina, etnične študije, novinarstvo in medijske študije, komunikacijske študije, etnografija in antropologija, migracijske študije, politične študije, kulturne študije, sociologija in psihologija.

Ta dokument opisuje strategijo CLARIN.SI, na podlagi katere bo infrastruktura izpolnjevala svojo zavezo k omogočanju visokokakovostnih raziskav in inovacij, ter ukrepe, s katerimi bi izboljšali dostopnost, trajnost in odpornost infrastrukture za obdobje 2024–2030. Načrtovana strategija CLARIN.SI je usklajena s strategijo CLARIN ERIC 2024–2026 ([CE-2023-2333](#)), s [Programom politike ERA 2022–2024](#) in z [Nacionalnim strateškim načrtom za digitalno desetletje](#) Republike Slovenije. Strategija bo prispevala k načinu izvajanja znanosti v Sloveniji s poudarjanjem sodelovanja, vključevanja vseh in odprtega dostopa.

Infrastruktura CLARIN.SI je bila ustanovljena leta 2015 z vizijo, da bi zagotavljala prvovrstne jezikovne vire za raziskovalce, za katere je razpoložljivost virov kritičnega pomena, pa tudi storitve, ki omogočajo dostop do pomembnih digitalnih jezikovnih podatkov in orodij za obdelavo jezikovnega gradiva iz Slovenije, drugih držav z južnoslovanskimi govorci in od drugod.

Poslanstvo CLARIN.SI temelji na priznavanju izredno pomembne vloge jezika kot vira informacij o družbi in kulturi ter na priznavanju velikega potenciala za primerjalne raziskave kulturnih in družbenih pojavov, tudi v različnih jezikih. Jezik je zaradi svoje večplastnosti, vsestranskosti in globoke povezanosti z vprašanji identitete in izvora sestavni del humanističnih in družboslovnih raziskav. Hkrati pa je jezik zanimiv pojav tudi z vidika informacijske znanosti, podatkovne znanosti, jezikovnih tehnologij in umetne inteligence. Zato je infrastruktura CLARIN.SI ključna za podporo raziskovalcem z zelo širokega spektra področij z različnimi potrebami in veščinami.

V 8 letih obstoja je infrastruktura CLARIN.SI dosegla zrelo operativno fazo. V [Pregledu spremljanja mejnikov ESFRI](#) so ocenili uspešnost evropske infrastrukture CLARIN kot v celoti zadovoljivo. Pregled navaja, da infrastruktura izpolnjuje svoje cilje, zagotavlja pomembne storitve akademski skupnosti na resnično vseevropski, večjezični in multidisciplinarni ravni, pomembno prispeva k razvoju orodij in metod za obdelavo naravnega jezika, ki so široko sprejeti v raziskovalni skupnosti, in zavzema ustrezen položaj, ki zagotavlja, da lahko evropski znanstveniki pozitivno in pomembno prispevajo k razvoju umetne inteligence. V okviru slovenskega prostora je CLARIN.SI vključen v [Načrt razvoja raziskovalne infrastrukture 2030 \(NRRi 2030\)](#), objavljen leta 2022, ki prepoznava pomembno vlogo CLARIN.SI v nacionalni infrastrukturi.

Vendar pa je kljub temu potrebno nameniti CLARIN.SI dodatno pozornost, da se zagotovi dolgoročna tehnična, finančna in organizacijska vzdržnost infrastrukture, vključno z oceno in obvladovanjem tveganja. Za zagotovitev stalne in učinkovite vloge infrastrukture na svojem prostoru je ključno, da ima ambiciozen načrt za spremembe in razvoj, ki so v koraku s hitrimi spremembami na področju delovanja infrastrukture. Poleg tega je pomembno, da načrt upošteva tudi pobude drugih infrastruktur, sektor GLAM (galerije, knjižnice, arhivi in muzeji) in industrijo, katere potrebe po jezikovnih podatkih in jezikovnih primerjalnih analizah so se izrazito povečale.

Delovanje CLARIN.SI

Slovenska raziskovalna infrastruktura CLARIN.SI je vzpostavljena kot dolgoročni infrastrukturni projekt v okviru evropskega konzorcija [CLARIN ERIC](#). CLARIN.SI je organiziran kot [nacionalni konzorcij](#), ki nima narave pravne osebe. V konzorcij se lahko vključijo zavodi, gospodarske družbe, društva in druge pravne osebe iz Slovenije, ki se ukvarjajo s področjem razvoja in uporabe jezikovnih virov, orodij in tehnologij, predvsem za slovenščino in druge južnoslovanske jezike.

Konzorcij CLARIN.SI trenutno združuje [12 članic](#), ki razvijajo ali uporabljajo jezikovne vire in tehnologije v Sloveniji:

- **Univerze:** [Univerza v Ljubljani](#), [Univerza v Mariboru](#), [Univerza v Novi Gorici](#), [Univerza na Primorskem](#)
- **Raziskovalni inštituti:** [ZRC SAZU](#), [Institut »Jožef Stefan«](#), [Inštitut za novejšo zgodovino](#), [Znanstveno-raziskovalno središče Koper](#)
- **Knjižnica:** [Narodna in univerzitetna knjižnica](#)
- **Podjetji,** usmerjeni v razvoj tehnologij za obdelavo naravnega jezika: [Alpineon](#), [Amebis](#)
- **Društvo:** [Slovensko društvo za jezikovne tehnologije](#)

Upravni odbor konzorcija odloča o CLARIN.SI in zastopa člane konzorcija. Vsak član konzorcija ima en glas v Upravnem odboru.

Sedež CLARIN.SI je [Institut »Jožef Stefan«](#) (IJS), ki v okviru konzorcija deluje kot tehnični center. Ta v skladu z določili statuta evropskega konzorcija CLARIN ERIC vzdržuje repozitorij virov in orodij ter druge storitve CLARIN.SI. Tehnični center s spletnimi storitvami CLARIN.SI vzdržujejo tri enote na IJS: [Odsek za tehnologije znanja](#) (IJS E8), [Laboratorij za umetno inteligenco](#) (IJS E3) in [Center za mrežno infrastrukturo](#) (IJS CMI).

CLARIN.SI je vključen v delo **evropskega konzorcija CLARIN ERIC** prek predstavnika slovenske vlade v **Generalni skupščini**, najvišjem upravnem organu CLARIN ERIC, in prek predstavnikov CLARIN.SI v **Forumu nacionalnih koordinatorjev** (NCF), ki delujejo kot glavna povezava med CLARIN ERIC in nacionalnimi vozlišči, v **Stalnem odboru centrov CLARIN za tehnično pomoč** (SCCTC), ki usklajuje dejavnosti centrov za tehnično pomoč, in v Tematskih odborih, ki pomagajo pri izvajanju strategije. Trenutno predstavniki CLARIN.SI sodelujejo v **Odboru CLARIN za pravna in etična vprašanja**, **Odboru CLARIN za standarde** ter **Odboru CLARIN za interakcijo z uporabniki**.

Strateška usmeritev CLARIN.SI

Podpiranje znanstvene odličnosti in inovacij

Jezik prenaša informacije o kulturi in prek njega je mogoče preučevati družbeno, kulturno in znanstveno dinamiko, tako sedanjost kot preteklo. CLARIN.SI z omogočanjem dostopa do jezikovnega gradiva omogoča primerjalno raziskovanje kompleksnih pojavov, kot so človeška komunikacija in kognicija, identiteta posameznikov, skupin, kultur in narodov ter izražanje zamisli in teorij. Poleg tega zagotavlja digitalne vire za razvoj analitike, ki temelji na podatkih, in računalniškega modeliranja teh pojavov. Viri, ki jih CLARIN.SI daje na voljo, spodbujajo tudi učenje in gradnjo velikih jezikovnih modelov, na katerih temelji **umetna inteligenca**, pa tudi vzpostavitev metodoloških okvirov za analizo heterogenih podatkov.



Eden od ključnih temeljev strategije CLARIN.SI je podpiranje podatkovnih registrov in storitev za upravljanje s podatki, ki so v skladu z **načeli FAIR** (najdljivost, dostopnost, interoperabilnost in vnovična uporaba); paradigma interoperabilnosti, ki je danes znana kot **agenda za odprto znanost**, je že od samega začetka ena od prvin CLARIN.SI. Načela FAIR pa bodo v bližnji prihodnosti postala le še pomembnejša. [Lundska deklaracija iz leta 2023 za povečevanje koristi raziskovalnih podatkov](#) namreč poziva k okrepitvi, pospeševanju in povečevanju koristi uveljavljanja načel FAIR in odprtih raziskovalnih podatkov v Evropi, znotraj znanstvenih skupnosti in prek raziskovalnih infrastruktur (RI), da bi povečali splošno uspešnost raziskav in inovacij v evropskem raziskovalnem prostoru (ERP) ter okrepili doseg in vpliv na industrijo in družbo. Poleg tega deklaracija poudarja, da je dostop do visokokakovostnih raziskovalnih podatkov, ki jih je mogoče ponovno uporabiti, ključnega pomena za krepitev in napredek znanja znotraj znanstvenih področij in med njimi ter za omogočanje primerljivih raziskovalnih programov. Še več, deklaracija omenja, kako učinkovito se lahko raziskovalna skupnost spopade z novimi izzivi in krizami, kot sta pandemija covida-19 in ruska agresija v Ukrajini, kar dodatno poudarja pomen odprte znanosti ter tega, da je sposobna podpiranja učinkovitih in hitrih odzivov na prihodnje krize.

Poleg tega lahko načela FAIR in odprti raziskovalni podatki prinesejo večjo korist družbi tudi v nekriznih časih. CLARIN.SI se zavzema za aktivno spodbujanje deljenja in uporabe odprtih podatkov, odprte kode in odprtih standardov, da bi podprli **ponovljive raziskave**. S tem CLARIN.SI prispeva k **trajnostni, interdisciplinarni in odgovorni (ponovni) uporabi raziskovalnih podatkov**, s čimer lajša odkrivanje novega znanja in maksimizira koristnost investicij akademskega sveta, industrije in družbe.

V skladu z inherentno **večjezičnostjo** Evrope in vedno večjo pozornostjo do **jezikovne enakosti** in **digitalne vključenosti** CLARIN.SI teži k temu, da utrdi svojo vlogo infrastrukture, ki podpira nastajajoče raziskovalne programe na področju družboslovja in humanistike ter prispeva k inovacijskemu potencialu naprednih modelov za interakcijo med ljudmi, podatki in tehnologijami za obdelavo podatkov. Ta prizadevanja olajšujeta naslednja dva dejavnika: močna vpetost razvijalcev orodij in zbirk podatkov v lokalni, kulturno specifičen kontekst Slovenije ter zavzemanje za paradigmo interoperabilnosti kot model sodelovanja med vključenimi centri. Zaradi tega je infrastruktura CLARIN.SI v dobrem položaju, tako kar zadeva računalniške vire kot tudi razpoložljivost usposobljenih strokovnjakov, da lahko skrbi za sedanje in prihodnje potrebe raziskovalne in inovacijske skupnosti, tudi na področju umetne inteligence, kar je v skladu s [Tenerifsko deklaracijo o globalni razsežnosti in trajnosti raziskovalnih infrastruktur iz leta 2023](#).

Spodbujanje strateških partnerstev

Osrednji del skupnosti CLARIN.SI predstavljajo **akademski raziskovalci, razvijalci in predavatelji** iz različnih področij, ki delajo z jezikovnimi podatki in ključno prispevajo k vzpostavitvi in delovanju infrastrukture CLARIN.SI z deljenjem virov, tehnologij in znanja. V naslednjih letih se bo sodelovanje z akademskimi partnerji še okrepilo in razširilo.

Čeprav je glavni cilj infrastrukture CLARIN.SI omogočati temeljne raziskave in znanstveno odličnost, imajo dejavnosti znotraj CLARIN.SI tudi pomemben inovacijski potencial. Infrastruktura sodeluje z različnimi deležniki izven akademskega sveta, vključno z **industrijo, državnimi organi in sektorjem GLAM** (galerije, knjižnice, arhivi in muzeji), ki prispevajo kot tudi uporabljajo podatke, orodja in znanje. Dve podjetji in ustanova iz sektorja GLAM (Narodna in univerzitetna knjižnica) so celo formalizirali svoje sodelovanje z infrastrukturo kot člani konzorcija CLARIN.SI. Vedno večji pomen jezikovnih in govornih tehnologij v digitalni preobrazbi držav članic CLARIN ponuja CLARIN.SI zanimive priložnosti za sodelovanje z industrijo, kjer se bo CLARIN.SI opiral na priporočila novoustanovljene delovne skupine CLARIN, zadolžene za pripravo [Akcijskega načrta 2023–2024 za sodelovanje z industrijo](#).

V širšem ekosistemu RI je CLARIN.SI prek CLARIN ERIC dobro umeščen v [grozd Družbene in kulturne inovacije ESFRI](#). Tesno sodeluje s slovenskimi vozlišči RI **DARIAH** in **CESSDA**, tj. [DARIAH-SI](#) in [ADP](#), pri čemer so vse institucije, ki so članice teh RI, tudi članice konzorcija CLARIN.SI, poleg tega pa CLARIN.SI že sodeluje oz. je sodeloval v skupnih projektih z obema RI, kot sta [razvoj parlamentarnih korpusov](#) in projekt [RDA Node Slovenia](#) (RDA vozlišče Slovenija).

Vzpostavljanje poti do učinkov

Glavni cilj CLARIN.SI je prispevati k znanstveni odličnosti in znanstvenemu napredku z zagotavljanjem dostopa do najsodobnejših virov in orodij ter do najboljših praks, usposabljanja in strokovnega znanja. Infrastruktura CLARIN.SI je tudi umeščena v samo središče skupnosti **podatkovne znanosti** in **umetne inteligence**, katerih najbolj vplivne inovacije z izredno hitrostjo prodirajo v večino akademskih področij, kot tudi v naše vsakdanje življenje. Zaradi tega se lahko CLARIN.SI lažje uveljavi v vlogi infrastrukture, ki omogoča raziskave, namenjene reševanju izzivov družbe, kot je na primer razvoj tehnologij za pomoč ljudem z jezikovnimi in govornimi motnjami z uporabo izboljšane obdelave jezikovnih in govornih signalov. Vendar pa je pomembno omeniti, da medtem ko so velepodatki in algoritmi, ki temeljijo na strojnem učenju, izredno pomembni za raziskave, gospodarstvo in nasploh družbo, predstavljajo tudi grožnjo socialni pravičnosti. Za preprečevanje te grožnje morajo raziskovalci in industrija sprejeti smernice za **odgovorne prakse na področju podatkovne znanosti** in za **etično umetno inteligenco**. CLARIN.SI želi prevzeti vlogo pri podpiranju zasnove primerov rabe, ki lahko pripomorejo k boljšemu razumevanju tveganj metod, ki temeljijo na podatkih.

Viri, ki jih ponuja repozitorij CLARIN.SI, in vključenost repozitorija v pregledovalnik VLO (Virtual Language Observatory) infrastrukture CLARIN predstavljajo eno od ključnih komponent [Evropskega jezikovnega podatkovnega prostora](#) (LDS), oziroma za vzpostavitev tega prostora. LDS naj bi bil namenjen zagotavljanju jezikovnih podatkov, pri čemer bi bilo omogočeno finančno vrednotenje ponudbe storitev, ki bi upoštevalo potrebe industrijskega sektorja. LDS je del evropske strategije za podatke, ki stremi k vzpostavitvi enotnega trga za podatke z namenom zagotovitve **konkurenčnosti** Evrope na svetovni ravni in **suverenosti podatkov**.

Zbiranje kvantitativnih podatkov za sledenje družbeno-ekonomskemu vplivu virov infrastrukture in vpliva CLARIN.SI na splošno je zaradi politike odprtega dostopa zelo zahtevno. Medtem ko kvantitativna vzročna veriga od (dostopa do) virov do končnega rezultata (npr. objave ali zasnove strategije) ni jasna, se v kvalitativnem smislu že kaže družbeno-ekonomski vpliv dosedanjih dejavnosti CLARIN.SI. Na primer, Inštitut za slovenski jezik Frana Ramovša redno uporablja konkordančnike CLARIN.SI pri svojih svetovalnih storitvah in pri gradnji slovarjev. Več prevajalskih agencij si pri prevajanju pomaga z vzporednimi korpusi CLARIN.SI, industrijski sektor pa uporablja orodja za označevanje CLASSLA, ki jih je razvil CLARIN.SI, za jezikovno označevanje besedil v slovenščini in drugih južnoslovanskih jezikih.

V obdobju, ki ga pokriva ta strategija, bo infrastruktura CLARIN.SI razvila kvantitativne in kvalitativne kazalnike za znanstveni in družbeno-ekonomski vpliv, ki bodo v skladu s ključno vrednoto infrastrukture – odprtim dostopom. To bo vključevalo tudi oblikovanje ključnih kazalnikov uspešnosti (KPI) za posamezne učinke oziroma ključnih kazalnikov učinka (KII). V skladu s [Povzetkom politike ESFRI glede ocenjevanja učinka RI](#) bo CLARIN.SI za ocenjevanje uporabil prilagojen pristop z mešanimi metodami, vključno z analizo dostopa in bibliometrično analizo, študijami primerov in raziskavo med deležniki.

Strateški program CLARIN.SI

Program izvajanja strategije CLARIN.SI 2024–2030 je organiziran v tri stebre, ki so ključni za delovanje CLARIN.SI:

1. steber: Naši uporabniki

Obstoječi in potencialni novi uporabniki z akademskimi in neakadetskimi interesi.

2. steber: Naša ponudba

Tehnična infrastruktura in infrastruktura znanja z interoperabilnimi podatkovnimi centri sodelujočih središč znanja.

3. steber: Naš prostor

Program za organizacijski razvoj, usklajevanje področij delovanja in sodelovanje z deležniki ter vključevanje v širši kontekst.

Vsak steber je organiziran v **prednostna področja** (Kaj je za nas pomembno?). Prednostna področja so razčlenjena na **cilje** (Kaj želimo doseči?), ki so razdeljeni na posamezne **strategije** (Kako bomo to dosegli?). Strategije bodo podlaga za delovne načrte za vsako od prihodnjih let, da se zagotovi, da so opisi predvidenih ukrepov dovolj natančni in se ujemajo z razpoložljivimi človeškimi in finančnimi viri. Vsem strategijam, predstavljenim spodaj, so skupni trije cilji: **povečanje potenciala za uporabo**, **povečanje potenciala za vpliv** in **povečanje trajnosti** vloge infrastrukture CLARIN.SI in njene ponudbe storitev v prostoru RI.

1. steber : Naši uporabniki

Prednostno področje 1: Baza uporabnikov iz akademske skupnosti

Cilj 1.1: Izboljšati povezovanje z obstoječo bazo uporabnikov CLARIN.SI iz akademske skupnosti.

- **Strategija 1.1.1:** Nadaljevati z analizami ključnih skupin uporabnikov CLARIN.SI in njihovih potreb v zvezi z raziskovalnimi podatki, (velikimi) jezikovnimi modeli, orodji, dokumentacijo, strokovnim znanjem in usposabljanjem.
- **Strategija 1.1.2:** Izboljšati spletno stran in druge komunikacijske kanale, da bodo informacije tako o poslanstvu CLARIN.SI, mreži infrastrukture in ponudbi kot tudi o vplivu CLARIN.SI bolj izpostavljene ter da bo CLARIN.SI prek spletne strani in komunikacijskih kanalov lahko bolje ustregel potrebam svojih uporabnikov iz raznovrstnih skupnosti.
- **Strategija 1.1.3:** Izboljšati osrednje in nacionalne podporne storitve ter ponudbo izobraževanj z novimi programi usposabljanja, delavnicami in spletnimi vadnicami, s posebnim poudarkom na multidisciplinarnih raziskovalnih programih.

Cilj 1.2: Razširiti bazo uporabnikov CLARIN.SI iz akademske skupnosti.

- **Strategija 1.2.1:** Razširiti in izboljšati uporabo spletne analitike, da bi bolje razumeli, kako uporabniki najdejo in uporabljajo ponudbo CLARIN.SI, ter zasnovati načrte za izboljšanje uporabniške izkušnje.
- **Strategija 1.2.2:** Raziskati akademski prostor in opredeliti prednostne nove in nastajajoče skupnosti uporabnikov za obdobje te strategije ter pridobiti informacije o njihovih potrebah z anketami, fokusnimi skupinami in intervjuji.
- **Strategija 1.2.3:** Pregledati in prilagoditi načrte za doseganje novih uporabnikov prek komunikacijskih orodij CLARIN.SI.

Prednostno področje 2: Baza uporabnikov zunaj akademske skupnosti

Cilj 2.1: Razviti sistematično sodelovanje z neakadetskimi skupnostmi, ki se zanimajo za ponudbo CLARIN.SI in lahko pripomorejo k tehnološkim ali družbenim inovacijam.

- **Strategija 2.1.1:** Opredeliti in okrepiti obstoječe odnose in sodelovanje z neakadetskimi raziskovalnimi in razvojnimi skupnostmi iz industrije, s sektorjem GLAM, z javnim sektorjem, nevladnimi organizacijami, s pedagogi, znanstvenimi novinarji, politiki in širšo javnostjo, ki se zanimajo za teme, kot so jezikovna raznolikost in večjezičnost, digitalna revolucija in enakopravnost jezikov v digitalni dobi, jezikovne tehnologije in umetna inteligenca.

2. steber: Naša ponudba

Prednostno področje 3: Kakovost in interoperabilnost podatkov, orodij, modelov in metapodatkov

Cilj 3.1: Izboljšati kakovost metapodatkov, ki jih ponuja CLARIN.SI, in s tem olajšati iskanje in ponovno uporabo virov in orodij.

Strategija 3.1.1: Razviti in izvajati strategije in delovne procese za zagotavljanje kakovosti in nadzorovanje združenih in urejenih metapodatkov prek kombinacije samodejnih postopkov (npr. preverjanje spletnih povezav) in ročnega urejanja ter prek čim boljše izrabe sodelovanja s ponudniki pregledovalnikov virov (kot so VLO ter portala Switchboard in Resource Families infrastrukture CLARIN).

Cilj 3.2: Izboljšati razpon in kakovost podatkov, modelov in orodij, ki jih ponuja infrastruktura CLARIN.SI.

- **Strategija 3.2.1:** Izvesti analizo pokritosti in vrzeli ter tehnično preveriti razpoložljivost, podporo pri enoviti prijavi itd., za vire in orodja, ki bodo služili kot osnova za opredelitev prednostnih nalog za prihodnji razvoj, pa tudi za dejavnosti na področjih usposabljanja in ozaveščanja o ponudbi. Pri tem bo infrastruktura dala poseben poudarek odkrivanju morebitnih novih strateško pomembnih področij uporabe in delovanja.
- **Strategija 3.2.2:** Razviti postopke za zbiranje, zasnovo in dokumentiranje serij preizkusov za orodja, modele in nabore podatkov ter postopke za zagotavljanje in nadzor njihove kakovosti, ki bodo v celoti usklajeni s principi odprte znanosti in načel FAIR.
- **Strategija 3.2.3:** Razviti bolj usklajen pristop, s katerim bi povečali število vnosov jezikovnih virov v repozitorij, po možnosti tudi takšnih z občutljivimi podatki (npr. zdravstveni podatki, ustno izročilo). Izboljšati obstoječe smernice za vnašanje virov in aktivno pristopiti k raziskovalcem, ki imajo v svoji lasti zanimive in relevantne nabore podatkov, ki še niso na voljo prek infrastrukture CLARIN.SI.
- **Strategija 3.2.4:** Uveljaviti CLARIN.SI kot infrastrukturo, ki se osredotoča na visokokakovostne jezikovne podatke. Deliti z drugimi centri CLARIN svoje definicije kakovosti (na podlagi formata in vsebine), najboljše prakse in izkušnje, da bi skupaj dosegali visoke standarde kakovosti. Objaviti zgodbe o uspehu in prakse v zvezi s kakovostjo podatkov.
- **Strategija 3.2.5:** Izpostaviti kakovost in razpon orodij za obdelavo naravnega jezika, ki jih ponuja CLARIN.SI, s tem, da se jih vključi v načrtovano platformo CLARIN »BLARK meets Test Automation« (Minimalen nabor jezikovnih virov in tehnologij (BLARK) in samodejno testiranje), ki bo ponujala informacije o orodjih, ki so na voljo, v kombinaciji s samodejnim testiranjem na standardnih testnih datotekah.

- **Strategija 3.2.6:** Izboljšati obstoječe smernice in prakse citiranja v Sloveniji prek aktivnega spodbujanja avtorjev, urednikov revij in predsednikov konferenc k temu, da izboljšajo citiranje naborov podatkov in orodij CLARIN.SI v znanstvenih objavah.

Cilj 3.3: Izboljšati interoperabilnost metapodatkov, podatkov, modelov in orodij.

- **Strategija 3.3.1:** Sodelovati s CLARIN pri zagotavljanju interoperabilnosti metapodatkov, in sicer na več ravneh: sodelovati z drugimi centri CLARIN (npr. v zvezi z glavnimi priporočili za metapodatke, spodbujanjem uporabe skupnega izrazoslovja in v zvezi s splošno storitvijo CLARIN za FAIR digitalne podatke) ter preučiti, kakšne so možnosti za interoperabilnost repozitorija CLARIN.SI s priljubljenimi repozitoriji, kot so Zenodo, Open Science Framework (OSF), HuggingFace in Github.
- **Strategija 3.3.2:** Raziskati možnosti za zagotovitev bolj enotnega dostopa do orodij in modelov za obdelavo naravnega jezika, ki jih ponuja CLARIN.SI, npr. z zagotavljanjem standardnih programskih vmesnikov. Poudariti dodano vrednost tega v primerjavi z obstoječimi priljubljenimi ogroddji za obdelavo naravnega jezika, kot je SpaCy.
- **Strategija 3.3.3:** Raziskati možnosti za zagotovitev bolj enotnega dostopa do ključnih korpusov, ki jih ponuja CLARIN.SI, v sodelovanju z infrastrukturo CLARIN, npr. tako, da bi jih ponujali prek skupnega konkordančnika, kar bi uporabnikom olajšalo učenje uporabe konkordančnikov in omogočalo primerjalne analize.



Prednostno področje 4: Dostopnost in uporabnost podatkov, modelov, orodij in storitev

Cilj 4.1: Spodbujati visoko raven dostopnosti, uporabnosti in interoperabilnosti med tehničnimi storitvami, ki jih ponuja CLARIN.SI, in storitvami drugih nacionalnih centrov CLARIN.

- **Strategija 4.1.1:** V sodelovanju s CLARIN sprejeti smernice glede najboljših praks, ki bi veljale za vse infrastrukture CLARIN, v zvezi z dostopnostjo, uporabnostjo in interoperabilnostjo podatkov, modelov, orodij in storitev, vključno z olajševanjem uporabe storitev obdelave občutljivih podatkov.

Prednostno področje 5: Sodelovanje in izmenjava znanja

Cilj 5.1: Utrditi povezanost skupnosti CLARIN.SI za boljšo izmenjavo znanja in izkušenj.

- **Strategija 5.1.1:** Opredeliti priložnosti za sistematične pristope in oblikovati prakse za izboljšanje sodelovanja med CLARIN.SI in drugimi centri CLARIN.
- **Strategija 5.1.2:** Pomagati pri pregledu in prilagajanju vsebin konference Jezikovne tehnologije in digitalna humanistika, pri kateri je CLARIN.SI eden od organizatorjev, tako da bo konferenca ostala prostor za bogato in navdihujočo izmenjavo znanja za lokalno in mednarodno skupnost.
- **Strategija 5.1.3:** Izboljšati in v angleščino prevesti zbirko dokumentov CLARIN.SI, kot so pregledni članki in članki o najboljših praksah, da bi bolje zajeli vso znanje in izkušnje skupnosti CLARIN.SI, ter obveščati o teh dokumentih širšo javnost, tudi izven meja Slovenije.
- **Strategija 5.1.4:** Oblikovati najboljše prakse in predloge za načrte ravnanja z raziskovalnimi podatki (NRRP) za raziskovalce, ki zbirajo jezikovne podatke kot nov način izmenjave znanja.

Cilj 5.2: Povečati vključenost središč znanja CLARIN.SI v »širšo sliko«.

- **Strategija 5.2.1:** V sodelovanju s CLARIN promovirati središča znanja CLARIN.SI, da bi raziskovalci več uporabljali njihovo ponudbo strokovnega znanja, in sicer na različne načine, kot so izpostavljanje središč v objavah »Tour de CLARIN« in »CLARIN Impact Stories« (Zgodbe o uspehu CLARIN).

3. steber: Naš prostor

Prednostno področje 6: Trajnost in vpliv

Cilj 6.1: Povečati število članov v konzorciju CLARIN.SI.

- **Strategija 6.1.1:** Oblikovati strategijo rasti za privabljanje novih članov v konzorcij CLARIN.SI, pri čemer je končni cilj vključiti v konzorcij vse relevantne slovenske akademske institucije, zlasti raziskovalne inštitute, pa tudi več industrijskih partnerjev.

Cilj 6.2: Spremljanje trajnosti in vpliva.

- **Strategija 6.2.1:** Oblikovati metodologijo za ocenjevanje družbeno-ekonomskega vpliva v sodelovanju s CLARIN.
- **Strategija 6.2.2:** Opredeliti in obravnavati trenutne in predvidene možnosti razširljivosti in tveganja, povezana z njimi, pa tudi nenehno vrednotenje sestavnih delov tehnične infrastrukture, kot sta avtentikacijska in avtorizacijska infrastruktura (npr. naraščajoče število ponudnikov storitev, ki jih je treba vključiti, vpliv na arhitekturo) in urejanje metapodatkov (npr. podpora, obravnavanje težav, zmogljivost preverjanja povezav).

Cilj 6.3: Kadrovsko in vodstveno okrepiti CLARIN.SI.

- **Strategija 6.3.1:** Proaktivno iskati in zaposlovati podporno osebje za omogočanje neprekinjenega delovanja in posodabljanja tehnične infrastrukture in infrastrukture znanja CLARIN.SI.
- **Strategija 6.3.2:** Spodbuditi člane konzorcija CLARIN.SI k bolj aktivnemu sodelovanju pri vodenju dejavnosti infrastrukture in odločanju o njih.

Cilj 6.4: Okrepiti vlogo CLARIN.SI kot ključne infrastrukture za omogočanje najsodobnejših raziskav in inovacij, ki temeljijo na jeziku.

- **Strategija 6.4.1:** Opredeliti priložnosti za obogatitev ponudbe CLARIN.SI z novimi tehnologijami, kot so govorne tehnologije in viri, veliki jezikovni modeli in umetna inteligenca, predvsem za slovenščino in druge južnoslovanske jezike.
- **Strategija 6.4.2:** Še naprej ponujati manjša nepovratna sredstva CLARIN.SI in izboljšati letni projektni razpis v skladu s spremembami okoliščin.

Prednostno področje 7: Sodelovanje zunaj konzorcija CLARIN.SI

Cilj 7.1: Okrepiti usklajevanje in sodelovanje z drugimi slovenskimi infrastrukturami.

- **Strategija 7.1.1:** Nadaljevati s sodelovanjem, tudi na operativni ravni, s sestrskimi RI v Sloveniji, zlasti z DARIAH-SI in ADP, ter začeti sodelovati z novimi relevantnimi RI, da se čim bolj poveča prepoznavnost, zastopanost in sodelovanje v Evropski komisiji (EC), evropskem oblaku odprte znanosti (EOSC) in ESFRI ter v ustreznih projektih in delovnih telesih.
- **Strategija 7.1.2:** V sodelovanju s CLARIN uveljaviti CLARIN.SI kot vmesno vozlišče v ekosistemu EOSC. S tem bi lahko med drugim pomembno razširili bazo uporabnikov CLARIN.SI in povečali dejavnosti na svojem področju.
- **Strategija 7.1.3:** Spremljati razvoj združenja za jezikovne tehnologije (ALT-EDIC) in pripraviti ustrezne akcijske načrte.

Cilj 7.2: Okrepiti čezmejno sodelovanje.

- **Strategija 7.2.1:** V okviru središča znanja CLARIN.SI CLASSLA opredeliti priložnosti za sistematične pristope in oblikovati prakse za boljšo podporo sodelovanja s partnerji ali mrežami iz drugih držav z govorcji južnoslovanskih jezikov, tj. Hrvaška, Srbija, Bosna in Hercegovina, Črna gora, Severna Makedonija in Bolgarija.
- **Strategija 7.2.2:** V okviru središča znanja CLARIN.SI ELEXIS opredeliti priložnosti za podporo sodelovanja na evropski ravni s središči, ki se ukvarjajo z digitalno leksikografijo.



CLARIN.SI
Institut »Jožef Stefan«
Jamova cesta 39
SI – 1000 Ljubljana
Slovenija

Januar 2024

