

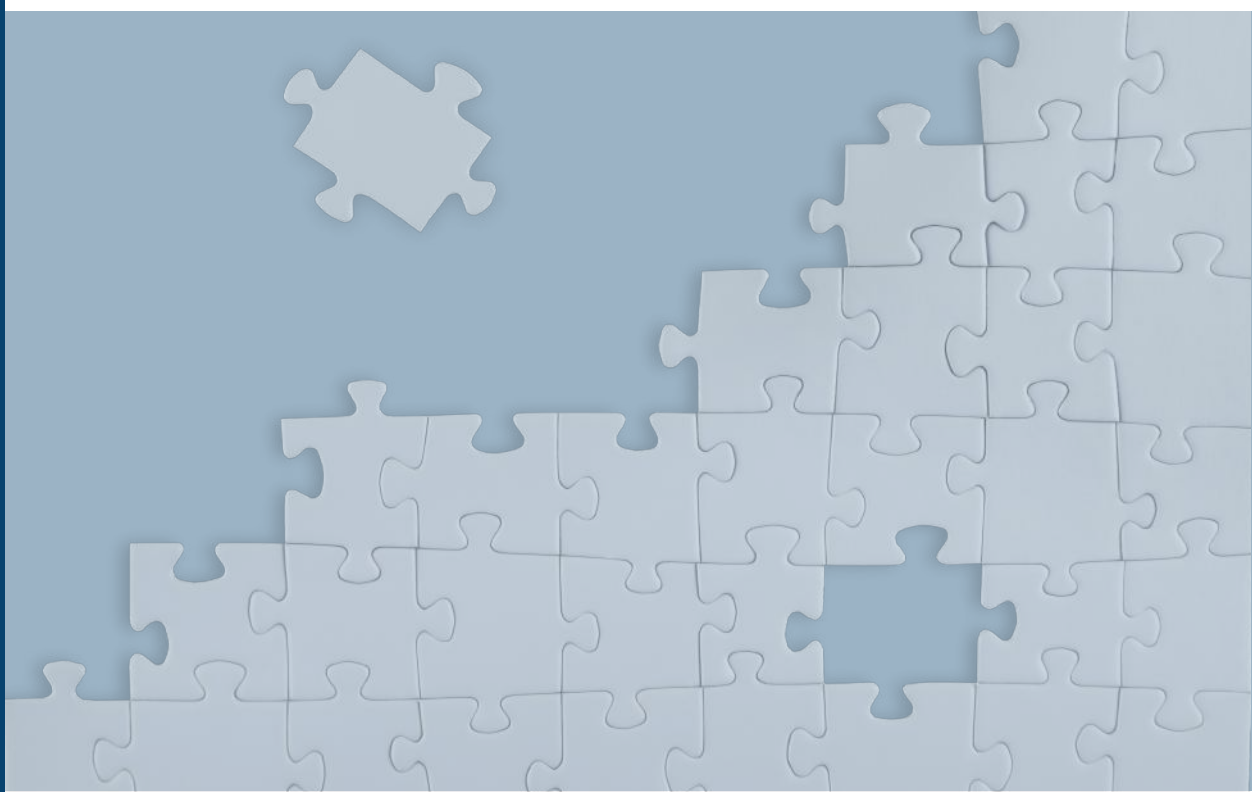
Kalba kaip socialinių ir kultūrinių duomenų šaltinis

Parama moksliniams tyrimams, kuriuose
naudojami kalbos duomenys

CLARIN-LT
strategija 2025-2029



ĮVADAS	3
CLARIN ERIC STRATEGIJA	4
CLARIN-LT NARYSTĖ, ISTORIJA IR VEIKLOS	5
CLARIN-LT STRATEGIJA	6
1 KRYPTIS: vartotojai	8
2 KRYPTIS: teikiamos naudos	10
3 KRYPTIS: veiklų gairės	12



Įvadas

CLARIN-LT strategija 2025-2029 m. suformuota remiantis CLARIN ERIC strategija 2024-2026 metams, atsižvelgiant į Lietuvos narystės infrastruktūroje poreikius ir vietinį kontekstą. Strategijos tikslų įgyvendinimas padės CLARIN-LT prisidėti prie socialinių ir humanitarinių mokslų plėtos, suteikiant prieigą vartotojams prie skaitmeninių lietuvių ir kitų kalbų išteklių ir technologijų per bendraeuropinę platformą.

Taip pat ši strategija dera su projekto „Narystės tarptautinėje mokslinių tyrimų infrastruktūroje CLARIN ERIC (Bendrųjų kalbos išteklių ir technologijų infrastruktūra) plano“ (sut. Nr. VS-15) tikslais ir prisideda prie atvirojo mokslo, tarptautinio ir tarpinstitucinio bendradarbiavimo vystymo Lietuvoje.



CLARIN ERIC strategija

CLARIN vizija: visi skaitmeniniai kalbų ištekliai ir įrankiai yra prieinami per vieną prisijungimą internetinėje platformoje, kuria siekiama padėti humanitarinių ir socialinių mokslų bei kitų su kalbomis susijusių disciplinų tyrėjams.

CLARIN misija: sukurti ir palaikyti infrastruktūrą, kuri padėtų dalintis, naudoti ir išlaikyti kalbos duomenis ir įrankius humanitarinių ir socialinių mokslų bei kitų su kalba susijusių disciplinų moksliniams tyrimams.

CLARIN strateginė kryptis: kalba – tai mokslo ir visuomenės žinių atspindys, žmonių bendravimo ir informavimo priemonė, vienas iš pagrindinių asmenų, bendruomenių, kultūrų ir tautų tapatybės aspektų, žmogaus pažinimo ir kūrybinės raiškos priemonė. Be to, kalbinė medžiaga sudaro didelę istorinių dokumentų, kurie laikomi kultūros paveldu, dalį. Suteikdama prieigą prie kalbinės medžiagos, CLARIN palengvina sudėtingų reiškinių lyginamuosius tyrimus ir sudaro sąlygas kurti duomenimis grįstą analitiką bei kompiuterinį šių reiškinių modeliavimą. CLARIN teikiami ištekliai padeda mokytis ir kurti didžiuosius kalbos modelius – pagrindinę dirbtinio intelekto (DI) technologiją, taip pat kurti metodologines heterogeninių duomenų analizės sistemas.

Pagrindinės CLARIN mokslinės sritys: lingvistika, kalbų studijos, kalbos technologijos, dirbtinis intelektas, literatūrologija, istorija, etninės studijos, žurnalistika ir žiniasklaidos studijos, komunikacijos studijos, etnografija ir antropologija, migracijos studijos, politikos studijos, kultūros studijos, sociologija ir psichologija.

2024-2026 m. CLARIN ERIC strategiją sudaro trys kartinės kryptys ir kiekvienos krypties tikslinės sritys:

Vartotojai: akademinė ir neakademinė bendruomenė.

Teikiamos naudos: duomenų kokybė, suderinamumas, prieinamumas ir vartojimas, bendradarbiavimas ir keitimasis žiniomis.

Veiklų gairės: tvarumas ir poveikis, centrinis CLARIN biuras, nacionaliniai centrai, bendradarbiavimas už CLARIN ribų.

Kiekvienai šiai kryptčiai įgyvendinti keliama tikslai ir juos pasiekti padėsiančios strategijos. Strategijų įgyvendinimas padės vystyti infrastruktūros matomumą ir naudojimą, poveikį ir tvarumą bendroje mokslinių tyrimų infrastruktūrų aplinkoje.



CLARIN-LT narystė, istorija ir veiklos

CLARIN, Bendrųjų kalbos išteklių ir technologijų infrastruktūra (angl. Common Language Resources and Technology Infrastructure) yra Europos mokslinių tyrimų infrastruktūrų konsorciumas (ERIC), siekiantis Europos ir kitų pasaulio šalių socialinių ir humanitarinių mokslų tyrėjams užtikrinti atvirą ir patikimą skaitmeninių kalbos išteklių ir jų analizės įrankių sklaidą ir prieigą vienoje platformoje, bendrus kalbos išteklių kūrimo standartus ir ilgalaikį duomenų saugojimą. Šiuo metu CLARIN vienija 24 šalių, o CLARIN platformoje sukaupta daugiau nei milijonas įvairių kalbų rašytinių, sakytinių, multimodalių šaltinių ir jų apdorojimo įrankių.

CLARIN ERIC siekia prisidėti prie socialinių ir humanitarinių mokslų plėtros ir kurti bendrą platformą, užtikrinančią kalbos išteklių ir jų analizės įrankių sklaidą, naudojimą ir tvarumą. Narystė šioje infrastruktūroje teikia įvairios naudos tyrėjams, mokslo įstaigoms ir šalims narėms:

- Suteikiama prieiga prie bendros platformos (angl. Virtual Language Observatory). Tyrėjai gali ne tik naudotis sukauptais duomenimis, bet ir saugoti sukurtus kalbos šaltinius nacionalinių konsorciimų išteklių saugyklose.
- Skaitmeniniai ištekliai ir įrankiai turi aukščiausius kokybės standartus ir yra žymimi CLARIN ženklų (angl. CLARIN compliance seal).
- CLARIN siūlo moderniausius kalbų analizės įrankius, padedančius tyrėjams anotuoti ir analizuoti duomenis.
- Tyrėjams suteikiama įvairiapusė pagalba kuriant kalbos išteklius, užtikrinant duomenų patikimumą, ilgalaikiškumą, autorines teises ir parenkant tinkamiausius duomenų analizės įrankius.
- Mokslininkai ir visi besidomintys kalbų tyrimais gali kreiptis į sertifikuotus CLARIN žinių centrus, kuriuose įvairių mokslo sričių ekspertai teikia konsultacijas, nukreipia į atitinkamus centrus, rengia įvairius mokymus.
- Kasmet organizuojami moksliniai renginiai, metinė konferencija ir seminarai skatina tyrėjų tarptautiškumą ir atveria įvairias galimybes naujiems projektams.
- CLARIN ERIC nariai dalyvauja įvairiuose tarptautiniuose projektuose.
- Institucijoms, norinčioms steigti CLARIN centrus, suteikiama įvairiapusė pagalba rengiant dokumentus, finansuojamos stažuotės į kitus centrus.
- Narystė CLARIN ERIC infrastruktūroje prisideda prie nacionalinių kalbų, išteklių ir tyrimų matomumo bei sklaidos tarptautiniame kontekste.
- CLARIN centrai, kuriami pagal bendrus techninius reikalavimus ir duomenų bei aukščiausios kokybės reikalavimus atitinkančius programinės įrangos standartus, skatina nacionalinių tyrimų raidą ir pažangą.

Lietuva, atstovaujama CLARIN-LT konsorciumo, tikrąja CLARIN ERIC nare tapo 2014 m. Šiuo metu CLARIN-LT konsorciumą sudaro šešios institucijos: Vytauto Didžiojo universitetas (koordinuojanti institucija), Vilniaus universitetas, Kauno technologijų universitetas, Mykolo Romerio universitetas, Baltijos pažangių technologijų institutas ir Klaipėdos universitetas. CLARIN-LT centro veikla skatina socialinių ir humanitarinių mokslų plėtrą, suteikdama tyrėjams prieigą prie bendrosios Europinės platformos, saugančios skaitmeninius teksto, garso ir vaizdo išteklius ir jų analizei reikalingus įrankius. Potencialūs CLARIN-LT siūlomų paslaugų vartotojai Lietuvoje ir užsienyje yra humanitarinių ir socialinių mokslo sričių tyrėjai, dėstytojai, studentai, mokslo ir studijų institucijos apskritai.

CLARIN-LT prisideda prie CLARIN ERIC infrastruktūros valdymo siūsdamas Lietuvos valdžios atstovą į kasmetinę Generalinę asamblėją, kuri yra svarbiausias CLARIN ERIC valdymo organas, skiria atstovus Nacionalinių koordinatorių forume ir kituose aktyviuose komitetuose. Lietuvos atstovai dalyvauja arba dalyvavo: Žinių infrastruktūros, Mokytojų tinklo, Teisinių ir etinių reikalų, Vartotojų pritraukimo ir Techninių centrų komitetuose.

Per dešimties metų laikotarpį sukurtas nacionalinis konsorciumas įgijo C tipo centro statusą; sukūrė tokio tipo centrui reikiamą techninę infrastruktūrą; įsteigė lietuvių kalbos konsultacijų centrą; parengė teisinę centro veiklą ir licencijų bazę; kuria ir kelia į CLARIN-LT saugyklą įvairius rašytinės ir sakytinės kalbos išteklius, kurie tampa viešai prieinami Lietuvos ir užsienio tyrėjams; kuria ir plėtoja įvairias kalbos analizės paslaugas ir įrankius; siunčia tyrėjus į tarptautines komandiruotes, organizuoja CLARIN-LT centro veiklą sklaidos seminarus ir virtualius mokymus, viešina informaciją CLARIN-LT internetinėje svetainėje. CLARIN-LT centre kuriami ir kaupiami tarptautinius standartus, autorinių teisių apsaugos įstatymus atitinkantys lietuvių kalbos ištekliai tampa laisvai prieinami tiek Lietuvos, tiek kitų šalių tyrėjams, siūlomos ekspertinės kalbos išteklių kūrimo, paieškos ir analizės konsultacijos, organizuojami skaitmeninių išteklių naudojimo mokymai, pristatomi naujausi duomenų saugykloje esančių lietuvių kalbos tekstynų tyrimai, įvairių sričių mokslininkams suteikiamos galimybės dalyvauti tarptautiniuose seminaruose, projektuose ir konferencijose.

Toliau pristatoma CLARIN-LT strategija, suformuota atsižvelgiant į prioritetinius CLARIN ERIC tikslus, socialinių bei humanitarinių mokslų tyrėjams teikiamą naudą ir indėlį į skaitmeninės humanitarikos plėtrą Lietuvoje.

CLARIN-LT strategija

2025-2029 m. CLARIN-LT strategijos įgyvendinimas suskirstytas į tris kertines kryptis, kurios yra svarbiausios CLARIN-LT veiklai:

1 KRYPTIS: vartotojai

Esami ir nauji vartotojai, turintys akademinį ir neakademinį interesą.

2 KRYPTIS: teikiamos naudos

Techninė ir žinių infrastruktūra, kuriama bendradarbiaujant duomenų ir žinių centrų specialistams.

3 KRYPTIS: veiklų gairės

Organizacinės plėtros, veiksmų kryptų koordinavimo ir bendradarbiavimo su suinteresuotaisiais subjektais programa ir įsitvirtinimas platesniame kontekste.

Kiekviena kryptis yra skirstoma į **tikslines sritis** (*Kas mums svarbu?*). Tikslinės sritys yra skirstomos į **tikslus** (*Ką mes norime pasiekti?*). Toliau tikslai yra skirstomi į **strategijas** (*Kaip mes tai pasieksime?*). Šitas skirstymas ateinantiems metams sudaro detalaus darbo plano ir biudžeto pagrindą ir užtikrina, kad planuojami veiksmai ir jų aprašymai atitiktų žmogiškuosius ir finansinius išteklius. Visos žemiau pateiktos strategijos skirtos padidinti CLARIN-LT vaidmenį ir paslaugų pasiūlos tvarumą mokslinių tyrimų infrastruktūros (MTI) kontekste.

1 KRYPTIS: vartotojai

1 TIKSLINĖ SRITIS: AKADEMINĖ BENDRUOMENĖ

1.1. Tikslas: gerinti informacijos sklaidą esamiems CLARIN-LT akademiniams vartotojams.

- **1.1.1. strategija:** toliau aiškintis pagrindines CLARIN-LT vartotojų grupes ir jų poreikius, susijusius su mokslinių tyrimų duomenimis, (didžiaisiais) kalbos modeliais, kalbos analizės įrankiais, dokumentacija, ekspertiniu konsultavimu ir mokymais.
- **1.1.2. strategija:** tobulinti interneto svetainę ir kitus komunikacijos kanalus, kad jie geriau atspindėtų CLARIN-LT misiją, pasiūlą ir poveikį bei geriau atitiktų įvairių bendruomenių, kurioms CLARIN-LT teikia paslaugas, poreikius.
- **1.1.3. strategija:** parengti planą, kuriuo siekiama užtikrinti partnerystę ir nuoseklų matomumą nacionaliniuose ir tarptautiniuose mokslo bei mokymo renginiuose, svarbiuose CLARIN-LT vartotojams, pristatant paslaugas ir skatinant įsitraukimą bei bendradarbiavimą.

1.2. Tikslas: plėsti CLARIN-LT akademinių vartotojų ratą.

- **1.2.1. strategija:** atlikti akademinės bendruomenės apklausą ir nustatyti prioritetines naujas ir besikuriančias vartotojų bendruomenes šiam strategijos laikotarpiui bei išsiaiškinti jų poreikius per apklausas, tikslines grupes ir interviu.
- **1.2.2. strategija:** peržiūrėti ir pakoreguoti planus, skirtus pasiekti naujus vartotojus naudojant CLARIN-LT komunikacijos priemones.



2 TIKSLINĖ SRITIS: NEAKADEMINĖ BENDRUOMENĖ

2.1. Tikslas: plėtoti sistemingą bendradarbiavimą su neakademinėmis bendruomenėmis, kurios domisi CLARIN veikla ir gali paskatinti technines ar visuomenines inovacijas.

- **2.1.1. strategija:** nustatyti ir plėsti esamus ryšius ir bendradarbiavimą su neakademinėmis bendruomenėmis iš verslo, muziejų, bibliotekų ir archyvų sektoriaus, viešojo sektoriaus, nevyriausybinių organizacijomis, švietėjais, mokslo žurnalistais, politikos formuotojais ir platesne auditorija, besidominčia tokiomis temomis kaip kalbų įvairovė ir daugiakalbystė, skaitmeninė revoliucija ir skaitmeninė kalbų lygybė, kalbų technologijos ir dirbtinis intelektas.
- **2.1.2. strategija:** sudaryti viešinimo planą, kuris leistų užtikrinti sistemingą matomumą svarbiuose visuomeniniuose renginiuose.

2 KRYPTIS: teikiamos naudos

3 TIKSLINĖ SRITIS: DUOMENŲ, ĮRANKIŲ, MODELIŲ IR METADUOMENŲ KOKYBĖ IR SUDERINAMUMAS

3.1. Tikslas: gerinti siūlomų duomenų, modelių ir priemonių aprėptį ir kokybę CLARIN-LT infrastruktūroje.

- **3.1.1. strategija:** įdiegti labiau koordinuotus darbo eigos procesus, siekiant pritraukti daugiau išteklių, taip pat ir jautrius duomenis (pvz., iš sveikatos srities, žodinės istorijos). Tobulinti esamas duomenų deponavimo gaires ir kreiptis į tyrėjus, turinčius įdomius ir svarbius duomenų rinkinius, kurių dar nėra CLARIN-LT infrastruktūroje.
- **3.1.2. strategija:** įtvirtinti CLARIN-LT kaip infrastruktūrą, kurioje daugiausia dėmesio skiriama aukštos kokybės kalbos duomenų rinkiniams. Dalintis kokybės apibrėžtimis (formato ir turinio požiūriu), geriausios praktikos pavyzdžiais ir patirtimi su kitais CLARIN centrais.
- **3.1.3. strategija:** tobulinti esamas citavimo gaires ir praktiką, aktyviai kreipiantis į autorius, redaktorius ir konferencijų organizatorius, kad CLARIN-LT duomenų rinkiniai ir įrankiai būtų tiksliau ir plačiau cituojami mokslinėse publikacijose.

4 TIKSLINĖ SRITIS: DUOMENŲ, ĮRANKIŲ, MODELIŲ IR PASLAUGŲ PRIEINAMUMAS IR VARTOJIMAS

4.1. Tikslas: tobulinti CLARIN-LT paslaugų prieinamumą ir vartojimo patogumą.

- **4.1.1. strategija:** stebėti dabartines ir naujas technologijas, standartus ir teisinius reikalavimus, susijusius su duomenų prieinamumu, ir užtikrinti, kad jų būtų laikomasi.
- **4.1.2. strategija:** rengti ir įgyvendinti duomenų prieinamumo ir tinkamumo naudoti reikalavimus ir politiką CLARIN-LT paslaugoms.

4.2. Tikslas: plėtoti nacionalinių techninių paslaugų prieinamumą, tinkamumą naudoti ir sąveiką, derinant veiksmus su kitais nacionaliniais centrais ir CLARIN ERIC.

- **4.2.1. strategija:** bendradarbiauti su CLARIN ERIC pritaikant gerąsias duomenų prieinamumo, tinkamumo naudoti ir sąveikos praktikas savo centrui, dalintis patirtimi su kitais CLARIN centrais.

5 TIKSLINĖ SRITIS: BENDRADARBIAVIMAS IR KEITIMASIS ŽINIOMIS

5.1. Tikslas: konsoliduoti CLARIN-LT bendruomenę, kad ji galėtų dalintis žiniomis ir patirtimi.

- **5.1.1. strategija:** ieškoti galimybių, kaip pagerinti sąveiką ir mažinti darbo fragmentiškumą tarp CLARIN-LT ir kitų CLARIN centrų.
- **5.1.2. strategija:** peržiūrėti ir pakoreguoti CLARIN metinės konferencijos, kurios organizavime dalyvauja ir CLARIN-LT, modelį, kad jis būtų įkvepiantis bendruomenės keitimosi žiniomis šaltinis.
- **5.1.3. strategija:** tobulinti ir skleisti CLARIN-LT mokslinius, mokslo populiarinimo tekstus, kitus susijusius dokumentus, kad jie geriau atspindėtų CLARIN-LT žinias ir patirtį ne tik vietinei, bet ir tarptautinei auditorijai.
- **5.1.4. strategija:** konsultuoti tyrėjus, renkančius kalbos duomenis, apie duomenų valdymo planų (DMP) šablonus.

5.2. Tikslas: integruoti CLARIN-LT žinių centrą į platesnį kontekstą.

- **5.2.1. strategija:** bendradarbiaujant su CLARIN ERIC, skatinti mokslininkus plačiau naudotis K-centrų patirtimi, pasitelkiant įvairias sklaidos priemones, pavyzdžiui, „Tour de CLARIN“.

5.3. Tikslas: stiprinti tarpvalstybinį bendradarbiavimą pagal kompetencijos sritis.

- **5.3.1. strategija:** nustatyti sisteminės galimybes, skatinančias tarpvalstybinį bendradarbiavimą.
- **5.3.2. strategija:** toliau prisidėti prie tarpvalstybinių ir tarpdalykinių renginių organizavimo.

3 KRYPTIS: veiklų gairės

6 TIKSLINĖ SRITIS: TVARUMAS IR POVEIKIS

6.1. Tikslas: plėsti ir stiprinti narystę.

- **6.1.1. strategija:** sukurti augimo strategiją, skirtą pritraukti naujus narius ir bendradarbiauti su visomis Lietuvos akademinėmis ir neakademinėmis institucijomis, galinčioms prisidėti kalbos duomenimis ar analizės įrankiais.
- **6.1.2. strategija:** parengti planus, parodančius pridėtinę narystės vertę esamiems ir būsimiems konsorciumo nariams.

6.2. Tikslas: stebėti tvarumą ir poveikį.

- **6.2.1. strategija:** kartu su CLARIN rengti socialinio ir ekonominio poveikio vertinimo metodiką ir, remiantis parengtais socioekonominės vertės nustatymo rodikliais, koreguoti veiklas.

6.3. Tikslas: stiprinti valdymą.

- **6.3.1. strategija:** peržiūrėti ir pakoreguoti CLARIN-LT konsorciumo narių bendravimo formą (gyvi, hibridiniai, virtualūs susitikimai) ir ciklą. Skatinti kuo daugiau narių dalyvauti bendruose susitikimuose, kad jie prisidėtų prie sprendimų, susijusių su infrastruktūros plėtra.
- **6.3.2. strategija:** įtraukti kuo daugiau CLARIN-LT konsorciumo narių į nacionalines ir tarptautines CLARIN veiklas.

6.4. Tikslas: stiprinti CLARIN-LT vaidmenį šiuolaikinių kalba grindžiamų mokslinių tyrimų ir naujovių kontekste.

- **6.4.1. strategija:** pagerinti CLARIN-LT paslaugų pasiūlą, atsižvelgiant į naujai atsirandančias technologijas, pavyzdžiui, didžiuosius kalbos modelius (LLM) ir dirbtinį intelektą.
- **6.4.2. strategija:** skatinti tyrėjus naudotis CLARIN ERIC siūlomomis dotacijomis, skirtomis lankyti kitų šalių centrus, rengti naujus projektus ir kt.

7. TIKSLINĖ SRITIS: BENDRADARBIAVIMAS UŽ CLARIN-LT RIBŲ

7.1. Tikslas: didinti nacionalinį indėlį į CLARIN infrastruktūrą.

- **7.1.1. strategija:** siųsti CLARIN-LT delegatus atstovauti nacionaliniams interesams CLARIN ERIC komitetuose.
- **7.1.2. strategija:** didinti indėlį į CLARIN veiklą, informavimo ir mokymo iniciatyvas.

8.1. Tikslas: stiprinti bendradarbiavimą su kitomis Lietuvos socialiniams ir humanitariniams mokslams skirtomis infrastruktūromis.

- **8.1.1. strategija:** vykdyti glaudų bendradarbiavimą su kitomis Lietuvos infrastruktūromis, skirtomis socialiniams ir humanitariniams mokslams (pvz., LiDA), taip užtikrinant matomumą, reprezentavimą ir dalyvavimą EOSC, ESFRI ir įvairiose iniciatyvose ar projektuose.
- **8.1.2. strategija:** kartu su CLARIN stiprinti CLARIN-LT, kaip sudedamosios dalies, poziciją EOSC ekosistemoje. Tai, be kitų privalumų, galėtų reikšmingai išplėsti CLARIN vartotojų ratą.
- **8.1.3. strategija:** stebėti kalbos technologijų aljanso ALT EDIC plėtrą ir rengti atitinkamus veiksmų planus.

8.2. Tikslas: stiprinti bendradarbiavimą tarptautiniu lygmeniu.

- **8.2.1. strategija:** plėtoti CLARIN virtualaus žinių centro „SAFMORIL“, kurio narys yra ir CLARIN-LT, tarptautines veiklas.
- **8.2.2. strategija:** plėtoti galimybes bendradarbiauti tarptautiniuose CLARIN inicijuojamuose ar partneriniuose projektuose.



CLARIN-LT

Daukanto g. 27-204

LT-44249

Kaunas

Lietuva

2025 01 11

