

Umsókn til Innviðasjóðs 2022

Lýsing á innviðum og fyrirhugaðri notkun

Heiti innviða á vegvísi: Miðstöð stafrænna hugvísinda og lista (MSHL)

Heiti stofnana: Háskóli Íslands – Hugvísindasvið, Háskólinn í Reykjavík, Landsbókasafn Íslands – Háskólabókasafn, Listaháskóli Íslands, Listasafn Íslands, Listasafn Reykjavíkur, Ríkisúttvarpið, Stofnun Árna Magnússonar í íslenskum fræðum, Þjóðminjasafn Íslands, Þjóðskjalasafn Íslands

Leiðbeiningar

Vinsamlegast hafið eftirfarandi atriði í huga við gerð lýsingar

Umsækjandi þarf að svara öllum liðum hér að neðan, að öðrum kosti verður umsókn vísað frá. Þar sem stendur „texti“ skal í staðinn skrifa texta, setja inn myndir eða gera töflur.

Skilyrði er að lýsingu á innviðum og fyrirhugaðri notkun sé skilað á eyðublaði sjóðsins. Ekki er leyfilegt að eiga við uppsetningu eyðublaðsins, taka fyrirsagnir eða spurningar út úr eyðublaðinu. Ef veittar eru rangar eða villandi upplýsingar er umsókn vísað frá.

Engar lengdartakmarkanir eru á skjalinu „Lýsing á innviðum og fyrirhugaðri notkun“, en leitast skal við að hafa lýsingar hnitmiðaðar og forðast óþarfar útskýringar. Nota skal leturgerð og leturstærð eins gefið er í þessu skjali, þ.e. 11 punkta Calibri (body) eða sambærilegt. Ekki er leyfilegt að breyta breidd spássíu. Lýsinguna skal senda inn sem PDF-skjal.

Þetta er eyðublað fyrir umsóknir vegna innviða á vegvísi. Hægt er að fá nýtt eyðublað með færri eða fleiri innviðalýsingum með því að senda fyrirspurn á innvidasjodur@rannis.is.

Frekari upplýsingar og reglur er að finna á vefsíðu Rannís.

Umsóknafrestur er til kl. 15.00 á lokadegi umsóknarfrests sem er mánudagur 1. nóvember.

Samantekt

texti (*samantekt á að vera að hámarki 1/2 bls.*)

Miðstöð stafrænna hugvísinda og lista (MSHL) er samstarf tíu stofnana um innviðakjarna í hugvísindum og listum. Aðilar miðstöðvarinnar vinna samkvæmt tíu ára áætlun að uppbyggingu miðstöðvarinnar og í þessum öðrum áfanga er sótt um styrki fyrir þremur verkefnum:

1. *Sarpur, uppfærsla menningarsögulegs gagnabanka.*
2. *Transkribus – tölvulestur handskrifaðra skjala.*
3. *Sögulegt bæja- og mannanafnatal II.*
4. *Talgreining Ísmús og uppbygging textasafns fyrir eldra talmál.*

Í fyrsta lagi er sótt um styrk til að kaupa og aðlaga nýtt skráningarkerfi fyrir Sarp, menningarsögulega gagnasafnið sem nær yfir safneignir á Íslandi. Skráningarkerfið sem notað er í dag er úrelt, ekki fullnægjandi fyrir skráningu lýsigagna eða birtingu efnis veflægt í nægilega góðri upplausn eða fjölbreyttum formum. Það býður ekki upp á leitir og vinnu með gögn sem fullnægir nútíma rannsónaaðferðum. Til að Sarpur nýtist til rannsókna þarf nýtt skráningarkerfi og í framhaldi nýtt DAMS kerfi (Digital Assets Management System). Í öðru lagi er sótt um styrk fyrir Transkribus, tæki til að lesa handskrifuð skjöl á íslensku frá 17. til 20. aldar. Í þriðja lagi er sótt um styrk fyrir framhaldi sögulega manna- og bæjatalsins, sem var styrkt í vor. Á grunni frumgerðar verður grunnurinn stækkaður til að ná yfir mun lengri tímabil og stærra landsvæði. Eins verða þróaðar aðferðir til að meta gæði tenginga í gagnagrunninum og finna nýjar tengingart. Í fjórða lagi er sótt um styrk fyrir talgreiningu Ísmús, sem er að mörgu leyti sambærilegt við Transkribus. Í þessu tilfalli verða þróaðar aðferðir til að „tölvulesa“ eldri upptökur af töluðu máli, en tækni sem til er í dag til að lesa talað mál á þennan hátt ræður ekki við þessar upptökur.

Efnisyfirlit

	Samantekt	ii
1	Lýsing á innviðum – 1. forgangur.....	1
	1.1. Markmið	1
	1.2. Almenn lýsing á innviðum.....	3
	1.3. Fyrirhuguð notkun innviða	4
2	Lýsing á innviðum - 2. forgangur	6
	2.1. Markmið	6
	2.2. Almenn lýsing á innviðum.....	7
	2.3. Fyrirhuguð notkun innviða	8
3	Lýsing á innviðum – 3. forgangur.....	10
	3.1. Markmið	10
	3.2. Almenn lýsing á innviðum.....	12
	3.3. Fyrirhuguð notkun innviða	16
4	Lýsing á innviðum – 4. forgangur.....	18
	4.1. Markmið	18
	4.2. Almenn lýsing á innviðum.....	18
	4.3. Fyrirhuguð notkun innviða	19
5	Lýsing á innviðum – 5. forgangur.....	23
	5.1. Markmið	23
	5.2. Almenn lýsing á innviðum.....	23
	5.3. Fyrirhuguð notkun innviða	23
6	Skipulag og stjórnun	24
	6.1. Kostnaður og fjármögnun á rekstri innviða	24
7	Heimildaskrá.....	29

1 Lýsing á innviðum – 1. forgangur

Sarpur: Uppfærsla menningarsögulegs gagnabanka

1.1. Markmið

Lýsið markmiðum með kaupum/uppbyggingu á innviðunum.

Markmið umsóknar er að gera menningarsögulega gagnasafnið Sarp að öflugu tæki til rannsókna á íslenskri menningu, list, samfélagi og sögu.

Sarpur, menningarsögulegur gagnabanki er skráningar- og umsýslukerfi sem 60 söfn og aðrar varðveislustofnanir á Íslandi nota til að halda utan um safnkostinn og miðla upplýsingum til fræðimanna og almennings. Kerfið er notað til að halda utan um skráningu listaverka, muna, jarðfundinna gripa, ljósmynda og þjóðháttá, svo fátt eitt sé upptalið. Kerfið varðveitir upplýsingar um gripina og gerir söfnum og öðrum aðilum kleift að sýsla með safnkostinn á einum stað. Sarpur veitir rannsakendum aðgang að gífurlegu magni aðfanga, sem í dag eru tæplega 1,5 milljónir.

Gagnger endurnýjun á Sarpi er nauðsynleg af eftirfarandi ástæðum:

1. Núverandi kerfi Sarps (3. útgáfa) er úr sér gengið og svarar ekki kröfum nútímans um stafræna vinnslu og miðlun gagna. Möguleikar á þróun núverandi kerfis eru mjög takmarkaðir.
2. Í dag miðlar Sarpur gögnum sem eru vistuð á söfnum og stofnunum víða um land, sem skapar hættu vegna varðveislu gagnanna. Með nýju skráningarkerfi verður vistun öll í öruggu umhverfi innan Sarps, sem mun gæta ítrustu varúðarsjónarmiða.
3. Sarpur nýtist ekki vel til rannsókna í dag vegna takmarkaðra leitarmöguleika og alvarlegra takmarkana á birtingu stafræns efnis. Ljósmyndir birtast einungis í lítilli upplausn, erfitt er að birta myndbönd og hljóðupptökur, engir kostir eru til að birta þrívíddarmyndir eða módel.

Áætlun MSHL, í samstarfi við Rekstrarfélag Sarps, gengur út á eftirfarandi aðgerðir:

1. Keypt verður nýtt, alþjóðlegt skráningarkerfi sem uppfyllir alþjóðleg gæðaviðmið (t.d. [Spectrum staðalinn](#)) og verður uppfært reglulega í samræmi við þau.
2. Í kjölfarið verður keypt nýtt umsýslukerfi (DAMS – Digital Assets Management System) sem miðlar upplýsingum úr skráningarkerfinu og nýtist sem tæki til rannsókna. Áætlað er að sækja um styrk fyrir DAMS kerfi 2022.

Nauðsynlegt er að byrja á því að kaupa nýtt skráningarkerfi og færa gögn úr eldra kerfi yfir í það nýja áður en nýtt umsýslukerfi er keypt (möguleikar tenginga úr skráningarkerfi í umsýslukerfi verður vissulega stór þáttur í útboðsgögnum). Um það snýst þessi umsókn.

Umtalsverð forvinna vegna útboðs á Evrópska efnahagssvæðinu hefur þegar farið fram með aðstoð verkfræðistofunnar Framnes ehf. sem hefur lagt mat á kostnað vegna kaupa á nýju skráningarkerfi fyrir Ríkiskaup og Rekstrarfélag Sarps (sjá útdrátt úr skýrslu í viðhengi með umsókn: „Sarpur – útboðsleiðin“). Ríkiskaup birtu [forauglýsingu sl. vor](#), sem sjö aðilar brugðust við. Fimm erlendir og tveir innlendir. Innlendu aðilarnir þóttu ekki góður kostur, en erlendu aðilarnir fimm koma allir til greina. Þeir eru:

1. Axiell Danmark A/S – Axiell Collections
2. Collector Systems LLC – Collector Systems
3. Gallery Systems – TMS Suite
4. Robotron Datenbank-Software GmbH – robotron*Daphne
5. Zetcom GmbH – MuseumPlus

Í september 2021 vann Framnes ehf. nýtt kostnaðarmat, m.a. byggt á nýjum upplýsingum sem komu fram við mat á viðbrögðum við forauglýsingu. Endurnýjað kostnaðarmat er líka í viðhengi (Sarpur - Endurmat á stofnkostnaði vegna kaupa á stöðluðu (erlendu) kerfi“). Þessi umsókn byggist á nýja kostnaðarmatinu.

Nánari lýsing á Sarpí, möguleikum hans og takmörkunum

Sarpur er tvískiptur. Innri vefurinn er vinnuumhverfi safnanna. Þaðan er stýrt hvaða upplýsingum er deilt út á ytri vefinn – <https://sarpur.is/> – þar sem þær eru öllum aðgengilegar. Á innri vef eru skráðar ítarlegar upplýsingar um þau ólíku aðföng sem söfn varðveita. Þannig myndar Sarpur yfirlit yfir safnkostinn í víðu samhengi og þar er safnað upplýsingum sem hægt er að nýta til rannsókna, bæði á safnkostinum sjálfum og á safnastarfinu. Rannsakendur geta fengið aðgang að innri vef, ýmist lesaðgang eða fullan aðgang, eftir því verkefni sem um ræðir.



SARPUR Menningarsögulegt gagnasafn

ADILDARSÖFN | SÝNINGAR | ÞJÓÐHÆTTIR | 1

Oli söfn Oli aðföng Sökðu inn leitarorð. LEITA Um leit

Forsíða > Adildarsöfn > Þjóðminjasafn Íslands > 2001-37 munir

< Aftur í leitarniðurstöður < 29 / 3410 >

2001-37-395

10cm

Landfræðileg staðsetning

Dæmi um birtingu gagna úr nýlegum fornleifagreftri.

Heiti	Dýrabein
Staður	Hofstaðir
Arnað staðarheiti	Hofstaðir
Byggðahéiti	Mývatnsveit
Sveitarfélag 1950	Skútustaðahreppur
Núv. sveitarfélag	Skútustaðahreppur
Sýsla	S-Þingeyjarsýsla

● Nánari upplýsingar

Númer	2001-37-395
Aðalskrá	Jarðfundur
Undirskrá	Almenn munaskrá, Munir, Fundaskrá
Stærð	1,1 x 0,56 x 0,49 cm
Vigt	0,2 g
Efni	Tónn

Með því að kaupa alþjóðlegt viðurkennt kerfi, sem uppfært er reglulega og þróað í samvinnu við söfn um allan heim, verða uppfærslur einfaldari og það mun halda í við þær breytingar sem verða á stafrænni varðveislu og miðlun á næstu árum. Þær útgáfur af Sarpí sem smíðaðar og reknar hafa verið hérlendis á undanförunum 24 árum hafa nýst vel en reynslan hefur sýnt að slík kerfi halda illa í við þróun á alþjóðlegum hugbúnaðarmarkaði. Nýtt alþjóðlega viðurkennt skráningarkerfi mun tryggja einfaldari umsýslu um safnkostinn, auðveldari leit og tölfræðiúrvinnslu, betri samvinnumöguleika og aukna möguleika á stafrænni miðlun gagna. Miðað er við að nýr Sarpur bjóði upp á vefþjónustur þannig

að aðildarsöfnin, og aðrir aðilar, geti miðlað völdum gögnum á sínum vefsvæðum. Þetta ræður núverandi kerfi ekki við.

Nýtt viðurkennt skráningarkerfi mun þannig styrkja og bæta faglegt starf safna, auðvelda aðgengi fræðimanna og almennings að menningararfinum og heimildum um hann óháð staðsetningu.

Núverandi kerfi stenst ekki kröfur samtímans og nýtist takmarkað til rannsókna. Upplýsingatækni hefur fleygt fram, aukin krafa er um nútímalega miðlun til hagaðila og almennings, ekki hvað síst er varðar myndræna framsetningu efnis og samtengingu gagna. Söfn sem og aðrar stofnanir landsins eru að ganga í gegnum stafræna umbreytingu. Markmið með nýju kerfi er að styðja söfnin og samstarfsaðila þeirra á þessari vegferð hvað varðar skráningu, umsýslu, yfirsýn, samstarf og miðlun, auk þess að tryggja öryggi gagnanna.

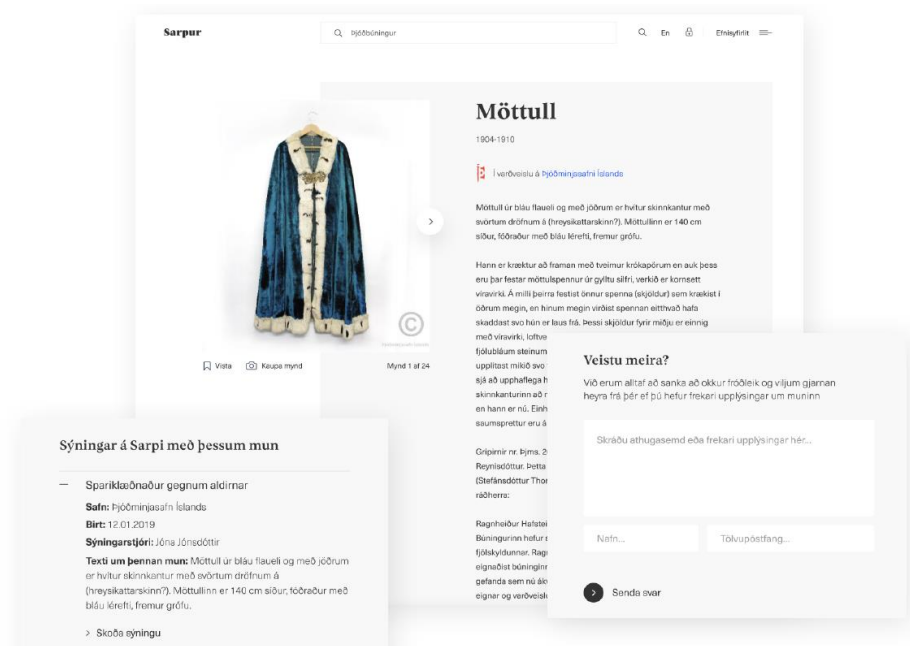
Núverandi kerfi er ekki varðveislustaður fyrir stafræn gögn og nauðsynlegt er að fá nýtt kerfi sem er í senn umsýslutæki og varðveislustaður stafrænna gagna.

1.2. Almenn lýsing á innviðum

Lýsið innviðunum sem sótt er um styrk fyrir og notkunarviði þeirra.

Sarpur hefur verið til sem skráningar- og umsýslukerfi safna frá árinu 1998 og með tilkomu vefsíðunnar sarpur.is árið 2013 opnaðist aðgengi að afrakstri margra ára skráningavinnu í söfnum landsins. Sarpur er lykilverkfæri í safnastarfi, þar sem vönduð skráning skapar undirstöðu fyrir rannsóknir, sýningar og aðra miðlun. Skráning menningararfsins eykur gildi hans og eru það oft upplýsingarnar sem fylgja aðfangi eða safngrip sem gerir það verðmætt og þýðingamikið fyrir sögu okkar og rannsóknir á henni. Í Sarp skrá aðildarsöfnin bæði áþreifanlegan- og óáþreifanlegan menningararf; lýsigögn, ljósmyndir og ýmsar aðrar upplýsingar. Með uppfærðum Sarp verða stórbætti möguleikar til að birta hljóð- og myndbandaupptökur, þrívíddarmyndir (m.a. úr 3D skönnum keyptum með styrk úr Innviðasjóði 2021), módel og fleira. Aðeins hluti þessara gripa er aðgengilegur almenningi á söfnunum því stór hluti safngripanna er geymdur í varðveisluhúsum víðs vegar um landið. Sarpur veitir innsýn inn í safneign safna umfram það sem söfnin hafa möguleika að sýna dagsdaglega og nýr Sarpur mun stórbæta þennan aðgang.

Aðföng



Hluti af kynningu á tillögu að nýju útliti og umgjörð Sarps (með nýju skráningar- og DAMS kerfi). M.a. verður mun auðveldara að kalla eftir gögnum, upplýsingum og hugmyndum frá almenningi. Sjá nánar [hér](#).

Fjöldmargir fornleifafræðingar nota Sarp til að skrá nánast jafnóðum þá gripi sem finnast við uppgröft, en það að skrá jarðfundi beint í Sarp minnkar líkur á mistökum og stuðlar að því að staðlaðar upplýsingar séu skráðar, sem bætir möguleika á síðari samanburði og rannsóknum. Þau sem rannsaka íslenska myndlist nota Sarp og eru verkfæri inni í Sarp til að vinna með þau gögn sem þar finnast. Nýr Sarpur bætir möguleika til skráningar aðfanga „jafn óðum“ og býður upp á ný verkfæri til að nota við rannsóknir.

Á innri vef Sarpis má skrá ítarlegar upplýsingar um þau ólíku aðföng sem söfn varðveita. **Í dag er hægt að gera eftirfarandi í Sarpi:**

- Skrá lýsingar aðfanga, efnisorð, heimildaskrá og eigendasögu.
- Skrá niðurstöður rannsókna safna á aðföngum.
- Vinna og birta forvörsluskýrslur.
- Birta ljósmyndir og ljósmyndir af aðföngum. Ljósmyndirnar eru hins vegar ekki vistaðar í Sarpi og ekki er mögulegt að birta þær í fullri upplausn.
- Skrá og birta miðlunar- og útlánasögu aðfanga.
- Skrá og birta staðsetningu aðfanga, þar sem hún er þekkt (allt frá upprunalegri staðsetningu til seinni staðsetninga og núverandi varðveisluhúsnæðis).

Sarpur er þannig gagnasafn yfir safnkostinn sjálfan, en einnig nýtingu hans í safnastarfinu og meðferð á honum innan og utan safnsins.

Með nýjum Sarpi verður mögulegt að:

- Varðveita stafræn gögn í fullri stærð (og með nýju DAMS kerfi birta þau líka í fullri stærð).
- Bjóða upp á „snertilaus viðskipti“ með stafræn gögn. Með því móti verður hægt að nálgast ljósmyndir og aðrar birtingarmyndir aðfanga án aðkomu starfsmanna safnanna.
- Miðla upplýsingum um aðföng með öðru formi en í dag, t.d. hlóð- og myndbönd, þrívíddarmyndir og þrívíddarmódel.
- Vista stafrænar heimildir (en ekki bara heimildaskrá).
- Bjóða notendum að hlaða inn upplýsingum og gögnum sem teljast ekki vera eign ákveðinna safna. Þetta opnar fyrir möguleika til að búa til heildarskrá yfir verk ákveðinna listamann, bæði verk sem eru varðveitt á söfnum og annars staðar.

Þó núverandi Sarpur geti ýmislegt þá er hann þungur í vöfum og eru ýmsar aðgerðir takmarkaðar í honum, sérstaklega hvað varðar leit og röðun sem eru lykilatriði við rannsóknir á svo umfangsmiklum gögnum. Bætt leit er áhersluatriði í nýjum Sarpi og við þarfagreiningu við undirbúning útboðs var bæði rætt við notendur innan safnasamfélagsins og aðra notendur, bæði fræðimenn og leikmenn í menningarsögu.

1.3. Fyrirhuguð notkun innviða

Lýsið mikilvægi innviðanna fyrir framfarir í rannsóknum á Íslandi. Munu þeir nýtast stórum hópi rannsókna? Munu þeir nýtast á fleiri en einu fagsviði? Eru í gangi verkefni styrkt af samkeppnissjóðum sem myndu nýta innviðina?

Sarpur nýttist nú þegar til rannsókna og hafa fjölmörg verkefni nýtt sér hann til gagnaöflunar. Sem dæmi má nefna eftirfarandi verkefni:

- *Heimsins hnoss* (<https://hh.hi.is>); styrkt af Rannsóknasjóði.
- *Disability before disability* (<https://dbd.hi.is>); styrkt af Rannsóknasjóði.
- *Commodity Entanglement. The Archaeology of the Trade Monopoly* – Gavin Lucas.
- *Fornar rætur Árbæjar* – Borgarsögusafn.
- *Leitin að klaustrunum* – Steinunn Kristjánsdóttir.
- *Muggur – Guðmundur Thorsteinsson* – Kristín Guðnadóttir, Listasafn Íslands.
- *Hamskipti í ferli Gerðar Helgadóttur* – Gerðarsafn – Listasafn Kópavogs.

Hönnuðir og listamenn nýta Sarp nú þegar við rannsóknarvinnu í undirbúningi nýrra verkefna og hann er einnig vinsæll meðal nemenda í leit að heimildum fyrir verkefni og almenningsforvitins um íslenska menningu og listir.

Takmarkanir Sarps eru hins vegar talsverðar, eins og komið hefur fram hér að framan. Þeir nýju möguleikar sem opnast eru fyrst og fremst af tvennu tagi:

- A. Bætt gæðamenning: Umfang Sarps er ákveðinn hvati fyrir söfn til að bæta skráningu lýsigagna, og þar sem Sarpur er notaður í meirihluta íslenskra safna geta söfn lært hvert af öðru. Þegar eru hafnar óformlegar viðræður við Ljósmyndasafn Reykjavíkur, innan Borgarsögusafns og Listasafn Reykjavíkur, um að gerast aðilar að nýjum Sarpi, en þessi söfn hafa rekið sína eigin gagnagrunna byggða á erlendum kerfum. Akureyrarbær er einnig að skoða að nýta Sarp fyrir öll sín söfn en aðeins hluti þeirra er þar nú. Hluti af safnakosti RÚV getur gengið inn í endurnýjaðan Sarp (en það er flóknara mál sem kallar á talsverða forvinnu). Eins skapast möguleikar til að miðla íslenskum kvikmyndarfi og gera aðgengilegi til rannsókna.

Með nýjum Sarpi, sem verður byggður á kerfi sem er í notkun alþjóðlega, stækkar samburðarhringur íslenskra safna og þau geta borið sig betur saman við söfn annarsstaðar. Lögð verður áhersla á að nýr Sarpur bjóði upp á stöðluð gæðaviðmið í skráningu og meðferð safnkosts og hefur sérstaklega verið horft til breska Spectrum kerfisins í því samhengi. Nýr Sarpur með innbyggðu slíku gæðakerfi styður bæði við aukna fagmennsku í safnastarfi og tryggari upplýsingar til þeirra sem nota Sarp til rannsókna.

Sarpur skapar einnig tækifæri fyrir lítil söfn sem ein og sér geta ekki miðlað sínum safnkosti í rafrænum gagnagrunni til þess að eiga aðild að slíkum grunni – sem síðan getur miðlað áfram í veitur eins og til dæmis [Europeana](#). Með nýjum Sarpi sem býður upp á öflugar vefþjónustur þá er hægt að auka slíkt samstarf og auka aðgang rannsækenda að menningarsögulegum gögnum.

- B. Möguleikar til nýrra rannsókna: Með nýju skráningarkerfi, og í kjölfarið nýju DAMS kerfi, opnast rannsækendum aðgengi að fjölbreyttari gögnum en áður og nýir möguleikar til rannsókna verða til. Með bættri leit og auknum möguleikum til að vinna áfram með gögn sem sótt eru í Sarp verður þessi rannsóknavinna ekki bara auðveldari og fjótlegri heldur skapar nýtt kerfi möguleika til nýrra tegunda rannsókna.

Kerfi (skráningar- og DAMS kerfi) sem býður upp á birtingu gagna í fjölbreyttari formum en nú, t.d. þrívíddarformum (bæði myndir og módel), skapar þannig tækifæri sem ekki eru til staðar í dag.

Með tengingu Sarps við Sögulegt manna- og bæjatal (sem verður undirbúin 2022) opnast mikilvægir möguleikar fyrir nákvæmari tengingar aðfanga við einstaklinga og staði. Þannig opnast möguleikar til að tengja aðföng safna við upplýsingar í öðrum gagnabönkum, sem opnar nýja rannsóknarmöguleika og spurningar.

Sarpur verður ekki bara tæki til að afla heimilda, heldur einnig **tæki til rannsókna**.

2 Lýsing á innviðum - 2. forgangur

Transkribus – tölvulestur handskrifaðra skjala

2.1. Markmið

Lýsið markmiðum með kaupum/uppbyggingu á innviðunum.

Markmið verkefnisins er að bylta aðgengi að handskrifuðum heimildum á Íslandi, hvort sem þær eru frá 17. eða 20. öld. Þróun nýrra tæknilausna á undanförunum árum skapar þennan möguleika. Stærstur hluti alls ritmáls fyrir 20. öld er handskrifaður. Á undanförunum árum hefur verið þróaður hugbúnaður sem byggir á gervigreindartækni sem felur í sér að tölvur geta lesið mismunandi handskrift og þar með gert textann lesanlegan og leitarbæran með svipuðum hætti og prentað mál. Fremst í þeim flokki er forritið Transkribus. Markmið þessa verkefnisins er að kaupa aðgang að Transkribus og þróa hugbúnaðinn í þeim tilgangi að búa til íslenskan grunn eða módel (e. module) fyrir handskrifaðan texta frá 17. öld til 20. aldar, og úrval af prentuðum texta með gotnesku lettri. Aukin nýtingarmöguleiki á handskrifuðum heimildum mun auðvelda til muna aðgengi að upplýsingum fyrri alda, og þannig styðja við rannsóknir á ýmsum sviðum.

Vefurinn timarit.is er vel þekktur og gott dæmi um hvernig tækni hefur gert öllum almenningi og fræðimönnum kleift að nýta sér leitarvélar til að leita í bæði prentuðu og vélrituðu efni. Þetta hefur haft í för með sér gríðarlegum vinnusparnaði og hagræði og nýir rannsóknarmöguleikar hafa orðið til. Það verkefni byggði á tölvulestri með OCR tækni. Sú tækni ræður þó ekki við handskrifaðan texta (eða texta prentaða með fátíðum leturgerðum eins og gotnesku lettri).



Dæmi um handskrifaðan texta og tölvutækan texta sem lesinn er af Transkribus, hjá sænska Þjóðskjalasafninu. Hér sést m.a. hvernig handskrifuð lína samsvarar lína úr lestri Transkribus. Þetta er t.d. gagnlegt við þjálfun í lestri handskriftar.

Transkribus mun styðja verulega við uppbyggingu stafrænna gagnabanka yfir íslenskar heimildir, bæta við þá sem fyrir eru og auka gagnsemi gagnabanka yfir skjöl sem til eru í dag en ekki sem leitarbær texti. Má þar nefna sem dæmi: Gagnagrunn um dómabækur sýslumanna, gagnagrunn um sáttanefndabækur, gagnagrunn um manntöl, vef fyrir sóknarmanntöl, gagnagrunn um dánarbú og fleiri sem sjá má á vefnum <http://heimildir.is>. Einnig má nefna gagnagrunninn handrit.is þar sem

eru lýsingar u.þ.b. þrettán þúsundum handritum, þar af hafa um þrjú þúsund handrit verið mynduð og fjöldi mynda er um 710 þúsund. Þessir gagnagrunnar hafa þegar haft umtalsverð áhrif á rannsóknir í sagnfræði hér á landi.

Íslensk skjala- og handritasöfn stefna að því að auka mjög stafræna miðlun skjala með þeim hætti að ljósrita eða skanna þau og setja inn á vefinn. Gallinn er þó að til þess að geta nýtt sér þá til fulls, þarf viðkomandi oftast en ekki að búa yfir skjalalestursþekkingu. Transkribus gæti því orðið til þess að nytsemi þessara fyrrnefndu innviða yrði hámarkuð.

Tæknin sem liggur að baki Transkribus hefur ekki verið notuð hér á landi. Á sama tíma er skjalalestursþekking almennt á undanhaldi og sömuleiðis ákveðin tungumálaþekking sem er oft á tíðum nauðsynleg. Til að tryggja áframhaldandi rannsóknir á fyrri öldum er fjárfesting í innviðum á við Transkribus nauðsynleg. Það verkefni sem hér er sótt um styrk fyrir er sambærilegt öðrum verkefnum á öðrum málsvæðum, en vegna Íslenskunnar og sérkenna hennar þarf að leggjast í ákveðna grunnvinnu svo að nýta megi hugbúnaðinn hér á landi. Verkefnið mun auka gildi þeirra rafrænu innviða sem þegar eru til staðar og stuðla að aukinni dýpt í rannsóknum, m.a. með tengingum sem verða til í gegnum Sögulegt manna- og bæjanafnatal sem er í þróun með stuðningi Innviðasjóðs.

Transkribus mun auðvelda aðgengi að heimildum, auðvelda rannsóknir á rituðum heimildum og gera textaleit mögulega. Það mun opna á aukið aðgengi að skjölum og handritum fyrri alda og notkun heimilda sem hafa lítið verið notaðar í rannsóknum. Þessi innviður er til þess fallinn að vera mjög eflandi fyrir aðgengi allra að heimildum. Fjölmargir háskólar og skjalasöfn í Evrópu eru þegar meðeigendur að Transkribus (sjá lista yfir meðlimi hér: <https://readcoop.eu/members/>).

2.2. Almenn lýsing á innviðum

Lýsið innviðunum sem sótt er um styrk fyrir og notkunarviði þeirra.

Styrkurinn verður notaður til að kaupa notkunarheimildir á rafrænum innviðum og fyrir vinnu við uppsetningu þeirra og aðlögun að íslenskum aðstæðum. Íslenska hefur ákveðin sérkenni sem mörg önnur tungumál hafa ekki. Broddstafir svo sem í, á, é og ó og sérstafi eins og þ og ð. Þetta gerir það að verkum að ekki er fyllilega hægt að byggja á lestri annarra tungumála, þó vissulega geti sú vinna nýst að minnsta kosti að hluta til. Því er nauðsynlegt að byggja upp *Íslenskan lesgrunn* sem verður í framhaldinu aðgengilegur fyrir þá sem vilja nýta sér hann við lestur texta.

Transkribus er alhliða stafrænt verkfæri fyrir stafvæðingu, lestur skriftar með hjálp vélvits, uppskriftar á og leit í sögulegum heimildum – frá hvaða svæði í heiminum sem er og á öllum tungumálum. Tæknin byggir á Pylaia hugbúnaði. Til þess að unnt sé að nota forritið megi notast á íslenskar heimildir þarf að þjálfa forritið. Vinnan við það byggir á vélrænu námi (e. *machine learning*) þar sem byggð er upp vitneskja sem geymd er í þekkingarsafni (þ.e. hin svokölluðu módel). Þetta gerir forritinu kleift að læra að draga ályktanir af framsetningu upplýsinga á handskrifðu máli. Að ná fram þökkalega nákvæmum lestri (yfir 90%, sem er nauðsynlegt til að grunnurinn nýtist vel) krefst þess að settar séu inn í kerfið að minnsta kosti um 50.000 blaðsíður af lesnum texta og í framhaldinu keyrðar inn um 150.000 blaðsíður til samanburðar. Þessi innleiðingarfasir krefst mikillar vinnu við lestur, innslátt og leiðréttingar. Samstarf stofnanna er lykilatriði til að ná að framkvæma slíkt með árangursríkum hætti.

Transkribus varð til í samhengi við tvö Evrópusambands verkefni: TanScriptorium (2013–2015) og READ (Recognition and Enrichment of Archival Documents – 2016–2019). Þróun þess var síðar haldið áfram af Háskólanum í Innsbruck, en frá og með 2019 hefur forritið verið í höndum READ-COOP. Samhliða þessari þróun á hugbúnaðinum hafa hinar ýmsu stofnanir gert tilraunir með hugbúnaðinn. Árið 2016 hóf Þjóðskjalasafn Finnlands tilraun með lestur á 500 blaðsíðum af handskrifuðum texta með góðum árangri, en skjalasafnið hefur nú skannað og lesið tugþúsundir síðna með þeim árangri að nákvæmni lesturs eru á bilinu 93-96% – ekki fjarri því sem tölvulestur á vélrituðum og prentuðum texta (OCR lestur) hefur náð. Reynsla Finna hefur verið góð og er svipuð og annarra aðila sem hafa byggt upp sitt eigin lestrarmódel. Þannig hafa Danir og Svíar tekið í notkun lesvél fyrir sín tungumál. Fjölmargir háskólar og söfn hafa þannig

tekið hugbúnaðinn í þjónustu sína og hefur hann sannað gildi sitt á margvíslegan hátt og auðveldað fræðimönnum og almenningi að nýta sér handskrifað efni.

Áætlað er að nauðsynlegt sé að keyra inn u.þ.b. 15 rithendur á hverri öld til að byggja upp lesgrunninn. Það eru því um 60 ólíkar rithendur sem þarf að lesa og keyra inn í grunn fyrir tímabilið frá upphafi 17. aldar fram á 20. öld. Í upphafi er val á rithöndum mikilvægt og nauðsynlegt að horfa til mismunandi stafgerðar og ritháttar (fljótaskrift o.s.frv.). Leitað verður að rithöndum í skjala- og handritasöfnum og þau valin af kostgæfni.

Verkferlar við gerð íslensks lesgrunns fyrir hverja rithönd er eftirfarandi:

- Skönnun afritun efnis (300 dpi frumritaskönnun og vistun í jpg).
- Lestur hluta þess efnis sem er skannað (t.d. 20 bls.).
- Myndir og texti keyrður inn í Transkribus (20 bls. af lesnum texta + 80 bls. af ólesnum texta). Þessi keyrsla fer fram hjá Transkribus.
- Sá hluti textans sem var ólesinn yfirfarinn og leiðréttur.
- Gögn keyrð aftur í gegn um Transkribus.
- Athugun á gæði lesturs með því markmiði að hann sé yfir 90% réttur.

Þessi verkferill er endurtekinn fyrir hverja þá rithönd sem sett er inn í lesgrunninn. Markmiðið er að í lok verkefnisins verði hægt að lesa allar hendur (ekki einungis þær sem settar hafa verið inn) með yfir 90% réttum lestri. Því er um að ræða mikla vinnu við verkstjórn, lestur og innsetningu efnis auk gæðaprófunar.

Verkefnið verður fyrst og fremst unnið af þremur stofnunum sem eiga aðild að MSHL: Landsbókasafni Íslands – Háskólabókasafni, Stofnun Árna Magnússonar í íslenskum fræðum og Þjóðskjalasafni Íslands. Hver þessara stofnana varðveitir handritaðar heimildir frá mörgum tímum og rithöndum. Fimm ólíkar rithendur frá hverri öld verða valdar hjá hverri stofnun til að þjálfa Transkribus. Hjá hverri stofnun er sérfræðipækning í lestri handrita sem er nauðsynlegt að virkja. Við lok tímabilsins verður íslenska módelið prufukeyrt á aðrar handskrifaðar heimildir frá sama tímabili til að kanna hvort nóg sé gert til að kenna Transkribus að lesa handskrifaða texta frá 17. til 20. aldar nægilega vel.

2.3. Fyrirhuguð notkun innviða

Lýsið mikilvægi innviðanna fyrir framfarir í rannsóknum á Íslandi. Munu þeir nýtast stórum hópi rannsakenda? Munu þeir nýtast á fleiri en einu fagsviði? Eru í gangi verkefni styrkt af samkeppnissjóðum sem myndu nýta innviðina?

Stór hluti þess efnis sem varðveitt er frá umræddu tímabili er handskrifað. Á Þjóðskjalasafni Íslands eru, svo dæmi sé tekið, að minnsta kosti 10-15 hillukílómetrar af handskrifuðu efni. Í handrita- og skjalasöfnum víða um land er gríðarlegt magn handritaðs texta, sem er að miklu leyti ekki sérlega aðgengilegur. Í dag eru afar fáir sem geta lesið handskrifaðan texta án yfirlegu og talsverðrar þjálfunar í að lesa þá. Það er mikil vinna og krefst mikillar elju. Nú er svo komið að almenningur á í erfiðleikum með að lesa handskrifaðan texta frá byrjun 20. aldar og raunar örfáir sem geta lesið skrift á 17. og 18. öld. Það er því brýnt að tækni á borð við þá sem Transkribus býður upp á sé innleidd hér á landi. Áætlað er að þegar verkefnið er langt komið verði hægt að innleiða það í kennslu á skjalalestri í samstarfi við Háskóla Íslands, enda býður það upp á algerlega nýja tækni og aðferðir við lestur skjala.

Sérfræðingar á íslenskum söfnum munu leggja sérþekkingu sína inn í þetta verkefni, sem mun nýtast þeim mörgu sem hafa ekki treyst sér til að lesturs. Allir sem þurfa að nýta sér slíka

heimildir við rannsóknir munu spara tíma og í raun fá aðgang að gríðarstórum gagnabanka efnis, sem hefur að miklu leyti verið óaðgengilegum nema með gríðarlegri yfirlegu í söfnum. Sá tími fræðimanna, sem fer í leit í skjala- og handritasöfnum, er afar mikill. Má segja að lesvél sé næsta skref eftir að gögn eru afrituð stafrænt til að gera þau enn aðgengilegri.

Stafrænn lestur handskrifaðra skjala í þessu magni mun opna á nýja möguleika í rannsóknum á sögu og menningu Íslands. Sagnfræðingar erlendis hafa í auknum mæli notast við aðferðir kenndar við gagnagnótt (e. *big data*) – hugtak sem er best skilgreint útfrá möguleikum okkar á að leita að og safna saman upplýsingum í viðamiklum gagnasöfnum til þess að greina mynstur sem eru efnahags- eða félagsögulega þýðingarmikill. Gott dæmi um hvernig gagnagnótt hefur nýst í rannsóknum er [Gender and Work](#) rannsóknarverkfenið við Uppsala háskóla, sem notaði greiningu á viðamiklu safni heimilda til þess að leita af sagnorðum sem vörðuðu nýju ljósi á vinnu kvenna fyrir á öldum. Með Transkribus gætu slíkar rannsóknaraðferðir á íslenskum heimildum orðið fýsilegar með lágmarks fyrirhöfn.

Almennt séð mun gagnagrunnurinn þó koma til með að nýtast fræðimönnum á hinum ýmsu sviðum, til dæmis sagnfræði, þjóðfræði, mannfræði, félagsfræði og raunvísindum, einkum jarðfræði og náttúrufræði. Sérstaklega mun þetta nýtast þeim sem hafa ekki hlotið þá þjálfun í skjalalestri. Transkribus mun einnig nýtast við málvísindarannsóknir og þróun máltækni. Íslenska módelið sem verður búið til verður hluti þeirra tóla sem eru aðgengileg til rannsókna í gegnum [CLARIN-IS](#). Jafnframt mun verkefnið auðvelda starfsfólki í stjórnsýslu, svo sem í málefnum menningarminja og minjavörslu, í réttarkerfinu og víðar. Það að breyta handskrifuðum texta í stafrænan gjörbreytir aðgangi fólks að gögnum. Þau verða leitarbær og þannig er hægt að orðaleita texta sem telur tugþúsundir síðna og flýta þannig fyrir allri notkun hans.

Aðgangur að íslensku Transkribus módelunum verður frjál. Hins vegar tekur Transkribus gjald fyrir lestur hvers skjals, sem einstaklingar og stofnanir þurfa að greiða fyrir. Þau gagnasöfn sem verða til í verkefninu – alls er sótt um styrk til að kaupa 300.000 „kredit“ sem duga fyrir lestri 300.000 handskrifuðum blaðsíðum – verða aðgengileg í opnum aðgangi.

3 Lýsing á innviðum – 3. forgangur

Sögulegt manna- og bæjatal II

3.1. Markmið

Lýsið markmiðum með kaupum/uppbyggingu á innviðunum.

Textinn hér á eftir er að mestu leyti sá sami og var í umsókn sl. vor, fyrir fyrsta hluta þessa verkefnis. Vinnan er í grófum dráttum á áætlun og ekkert alvarlegt hefur komið upp sem setur hana í hættu.

Í ár er sótt um styrk vegna annars áfanga. Hverjum áfanga lýkur með afurð sem getur staðið sem sjálfstæður rannsóknainnviður.

Markmið verkefnisins er að búa til miðlægan þekkingarbrunn með tímasettum upplýsingum um bæja- og staðanöfn annars vegar og mannanöfn hins vegar svo langt sem heimildir ná. Þekkingarbrunnurinn verður kjarnainnviður fyrir rannsóknir sem tengjast íslenskri sögu og menningu með því að tengja saman og uppfæra marga ólíka gagnagrunna sem ýmsar stofnanir reka í dag þannig að úr verður nýr samræmdur rannsóknainnviður opinn bæði rannsakendum og fróðleiksfúsnum almenningi. Þeir grunnar sem tengjast í þekkingarbrunninum eru ólíkir að uppbyggingu, framsetning gagna er mismunandi og ólíkir staðlar notaðir við skráningu sem gerir mjög erfitt og tímafrekt að vinna rannsóknir þvert á þá. Með nýjum miðlægum þekkingargrunni verður mögulegt í einstökum gagnasöfnum að vísa í samræmd auðkenni bæja- og staðanafna og einstaklinga sem gerir samkeyrslu gagna í rannsóknartilgangi mögulega og rannsóknir sem byggjast á þessum gögnum auðveldari.

Sögulegt bæja- og mannanafnatal ásamt verkfærakistu hugbúnaðarlausna verður sjálfstæður rannsóknainnviður hýstur hjá og á ábyrgð MSHL.

Helstu gagnagrunnar sem um ræðir eru jarðaskrár, manntöl og sóknarmannatöl hjá Þjóðskjalasafni, lykilskrár Landsbókasafns og Landskerfis bókasafna, Sagnagrunnur, Ísmús og örnefnaskrá hjá Árnastofnun og staðfangaskrá Þjóðskrár (listi yfir þessa og fleiri grunna er í viðhengi). Tilkoma þekkingarbrunnisins mun ekki aðeins gjörbreyta möguleikum til rannsókna þvert á þessa gagnagrunna heldur einnig möguleikum rannsakennda og annarra áhugasamra til að bæta við gögnum og upplýsingum og möguleikum safna og stofnana við að birta gögn, til dæmis með landfræðilegri miðlun sem sýnir þróun yfir lengri tíma.

Meginhugmyndin er sú að hægt sé að sækja upplýsingar um einstaklinga eða staði í þekkingarbrunninn og, með tímanum, ýmsar upplýsingar sem tengjast einstaklingum og bæjum í gagnagrunnum sem fyrir eru. Þegar brunnurinn verður gerður opinberlega aðgengilegur býðst öllum áhugasömum ekki aðeins að nota upplýsingarnar sem þar eru til rannsókna, heldur nýtist innviðurinn til að safna staðbundinni þekkingu sem kann að hverfa þegar einstaklingar falla frá. Brunnurinn mun þannig styðja við „hópöflun“ upplýsinga (e. crowdsourcing) og „borgaralegar rannsóknir“ (e. citizen science), þ.e. rannsóknir sem leita eftir samstarfi við almenning um upplýsingaöflun, rýni og mótun rannsóknasurninga.

Með samræmdum auðkennum einstaklinga verður unnt að sækja upplýsingar um búsetur þeirra og starfsheiti eftir tímabilum og fjölmargar aðrar sögulegar og menningarlegar upplýsingar. Samræmt auðkenni bæja og staða er ekki síður mikilvægt enda næst með því yfirlit yfir sögu hvers bæjar eða staðar fyrir sig með því að tefla saman heimildum frá ólíkum tímum. Samræmd auðkenni eru bara einn hluti þekkingarbrunnisins. Hinn hlutinn er þróun verkfærakistu með hugbúnaðarlausnum til að vinna með gögnin. Verkfærakistan mun

samanstanda af tilbúnum og sérsníðuðum lausnum, með áherslu á opinn hugbúnað sem rannsakendur geta sjálfir nýtt og þróað áfram eftir sínum þörfum.

Íslenskar stofnanir og söfn hafa unnið þrekvirki á undanförunum áratugum við að koma gögnum í sinni vörslu á stafrænt form. Gögnin sem til eru hjá ólíkum stofnunum ná oft til sömu einstaklinga og sömu staða, en auðkennin (lýsigögnin) eru ólík eftir stofnunum, sem gerir samkeyrslur erfiðar sem og tengingar við sambærileg gagnasöfn erlendis. Samráð milli þeirra um þróun gagnagrunnanna hefur verið takmarkað og í dag er mjög erfitt að leita þvert á tilbúna gagnagrunna. Samræmd skráning einstaklinga og staða sem er tengd við gagnagrunnana býr til nýjan þekkingarbrunn með því að tengja tilbúna grunna saman.

Við uppbyggingu þekkingarbrunnis um staði og fólk verður eftirfarandi gert:

1. Mótuð stefna um lýsigagnasamvinnu stofnana og rannsakenda um með hvaða hætti skal leggja inn upplýsingar í gagnagrunna. Slík lýsigagnasamvinna mun stórauka samstarf, byggja upp möguleika á aukinni miðlun meðal annars með landfræðilegri birtingu og verður til þess að ekki verður þörf á grunnskráningu manna- eða staðarnafna í hvert sinn sem ný stafræn gögn verða til þar sem hægt verður að sækja þær upplýsingar í þann sameiginlega þekkingarbrunn sem hér um ræðir.
2. Kannað hvaða persónuverndarsjónarmið þarf að hafa í huga við samkeyrslu gagna, vinnu með samkeyrð gögn og birtingu þeirra á ytri vef verkefnisins. Verkefnið mun í öllu fylgja lögum og reglum og heilbrigðum viðmiðum um persónuvernd.
3. Tryggt að grunnurinn verði þannig úr garði gerður að ólík tölvukerfi geti sótt upplýsingar þangað og gefin verður út verkfærakista með hugbúnaðarlausnum sem notendur geta nýtt til að tengjast grunninum, aðlaga sín gögn að honum og skilgreint ýmiskonar úrvinnslu.

Þetta verkefni er afar viðamikið og flókið og ekki á færi einstakra stofnana að taka það að sér. MSHL mun styrkja samstarf stofnananna sem að henni standa og stuðla með þessu verkefni að samstarfi við aðrar stofnanir, söfn, fyrirtæki og einstaklinga. Að auki leggur verkefnið nýjan grunn að öflugu rannsóknarumhverfi sem mun nýtast rannsakendum og samfélaginu öllu um ókomna tíð.

Vegna umfangs verkefnisins er nauðsynlegt að skipta því í þrjá áfanga, sem hver og einn tekur yfir fjóra grunnþætti (umsýsla, gögn, tækniþróun og verkefni/afurðir). Sótt er um stuðning til Innviðasjóðs fyrst og fremst vegna grunnþátta B og C í áætluninni hér á eftir. Í ár er sótt um stuðning vegna 21. áfanga (20221–20232).

1. Áfangi (2021-2022)

- A. UMSÝSLA: Ritstjórn mynduð, ferlar skilgreindir, ákvarðanatataka og lýsingagnastefna skilgreind. Leitað verður samstarfs við aðila sem hafa þekkingu á þessu sviði og þeim boðið til samstarfs.
- B. GÖGN: Forvinnsla og tilreiðing gagna; afmörkuð við ákveðið tímabil og svæði.
- C. TÆKNIÞRÓUN: Þarfagreining, tilbúna lausnir kannaðar, frumgerð verkfærakistu þróuð.
- D. TILRAUNAVERKEFNI/AFURÐIR: Í lok fyrsta áfanga verður til frumgerð verkfærakistu sem býður upp á leiðir til að vinna rannsóknaverkefni á grunni tveggja eða fleiri gagnabanka tilbúin í tilraunaútgáfu.

2. Áfangi (2022–2023) sem hér er sótt um styrk til

- A. **UMSÝSLA:** Samstarf myndað og tryggt við helstu alþjóðlegu rannsóknarnet í stafrænum hugvísindum, eins og ERIC rannsóknainnviðina CLARIN og DARIAH ásamt norræna innviðaverkefninu NDHL. Eins verður leitað samstarfs við innlendar og erlendar stofnanir sem hafa unnið með sambærileg gögn, eins og Minnesota Population Center.
- B. **GÖGN:** Innkeyrsla gagna á stærri skala.
- C. **TÆKNIÞRÓUN:** Önnur ítrun frumgerðar verkfærakistu; þróun gervitauganets til að greina gögnin.
- D. **TILRAUNAVERKEFNI/AFURÐIR:** Tilraunaútgáfa af gervitauganeti sem gefur innsýn í gögnin tilbúin til prófunar. Tauganetið mun m.a. hafa það hlutverk að finna gloppur og galla í gögnunum og tengingum gagnabanka.

3. Áfangi (2023–2024)

- A. **UMSÝSLA:** Vefur opnaður; almenn kynning og fræðsla; þjálfun fyrir nemendur og rannsakendur í aðferðafræði stafrænna hugvísinda og vinnu með gögnin.
- B. **GÖGN:** Áhersla á birtingu í opnum aðgangi.
- C. **TÆKNIÞRÓUN:** Lokaítrun hugbúnaðar.
- D. **TILRAUNAVERKEFNI/AFURÐIR:** Vefviðmót sem býður öllum áhugasömum upp á aðstoð við vinnu með gögnin. Sérstaklega verður hugað að viðmóti sem auðveldar fólki að bæta við upplýsingum um gögnin og stinga upp á leiðréttingu.

Staða verkefnis í lok október 2021:

- A. Vinna við verkefnið hófst í september 2021 skv. tilboði frá ad libitum ehf. Vinna við gerð samninga dróst, en í lok október er búið að ganga frá og samþykka nýja verkáætlun sem var unnin í samstarfi við Upplýsingatæknisvið HÍ sem mun hýsa gagnagrunninn og veita aðstoð við þróun hans og ganga frá samningum milli aðila.
- B. Verktaki hefur fengið aðgang að öllum gögnum sem hann þarf á að halda til að þróa grunninn, þ.e. gagnagrunnum hjá Stofnun Árna Magnússonar, Landsbókasafninu og Þjóðskjalasafninu og vinnur að því að prófa mismunandi möguleika til tenginga og greina galla í gagnabönkunum sem þarf að mynda tengingar við.
- C. Undirbúningur að umsókn um aðild MSHL að DARIAH, sem co-operative partner, er í vinnslu og reiknað með að hún verði samþykkt í ársbyrjun 2022.

3.2. Almenn lýsing á innviðum

Lýsið innviðunum sem sótt er um styrk fyrir og notkunarviði þeirra.

Með verkefninu verður til samræmdur gagnagrunnur yfir bæja- og staðarnöfn og yfir einstaklinga og vensl þar á milli auk upplýsinga um landsvæði sem hver bær tilheyrir (þ.e. stjórnsýslueiningar eins og hreppar, prestaköll o.s.frv.). Auk einfaldra vensla, sem tákna búsetu einstaklings á ákveðnum stað, verður söguleg tenging lykilþáttur þar sem hver einstaklingur getur hafa tengst mörgum stöðum á ólíkum tímabilum. Auk þess geta bæir verið staðsettur í ólíkum sóknum eða hreppum eftir tímabilum og því verður nauðsynlegt að gera ráð fyrir slíkum upplýsingum. Ofan á gagnagrunninn verður byggt leitarviðmót fyrir birtingu á gögnunum auk annarra tengimöguleika við önnur kerfi. Við lok verkefnisins verður til landfræðileg vefsja þar sem hægt verður að nálgast upplýsingar um einstaklinga og staði (svipuð, og tengd, öðrum vefsjám sem þegar eru til eða í þróun, eins og vefsja kirkjubóka, sögustaða, örnefna o.fl.) og gögn sem tengjast þeim í gagnabönkum (t.d. úr manntölum, sóknarmanntali, dánarbúum og fleiri heimildum).

Aðgangur að gagnabrunninum verður mögulegur eftir fjórum mismunandi leiðum:

- Í gegnum opið leitarviðmót á vef. Slíkur aðgangur verður miðaður að almenningi og fræðafólki sem þarf að fletta upp í gögnunum til að fá upplýsingar. Vefviðmótið mun skila niðurstöðum á þann hátt að notendur geta séð allar upplýsingar sem eru fyrir hendi um tiltekinn stað eða einstakling og þær heimildir sem liggja að baki þeim upplýsingum. Ef við á birtast tenglar á tengd gagnasöfn, til dæmis tenglar inn á manntalsvef Þjóðskjalasafns, handrit.is, Sarp og fleiri vefi aðildarstofnanna. Fyrirmynd að þess háttar aðgangi er til dæmis vefurinn málið.is sem veitir með samræmdu leitarviðmóti aðgang að mörgum gagnasöfnum Árnastofnunar er varða íslenskt mál.
- Boðið verður upp á niðurhal á tölvulesanlegum gagnapökkum í gegnum vefinn. Þannig verður hægt að hala niður öllum gögnunum eða ákveðnu mengi þeirra. Þessi möguleiki verður væntanlega mest notaður í þágu stærri rannsókna og verða gögnin þá keyrð saman við kerfi þeirra sem þau sækja.
- Annar tölvulesanlegur aðgangur eru svokölluð forritaskil (e. API = Application Programming Interface) þar sem önnur tölvukerfi geta sent fyrirspurnir beint í grunninn og nýtt sér svör þaðan. Með slíkum aðgangi er mikilvægt að innleiða staðla opinna tengdra gagna (linked open data). Þar má ímynda sér til dæmis vefi eins og Sóknarmannatöl Þjóðskjalasafns. Slíkur aðgangur gerir kleift að birta ítarupplýsingar um einstaklinga sem finnast bæði í sóknarmannatölunum og í þeim miðlæga þekkingarbrunni sem hér er fjallað um. Notendur vefsins geta því séð hvaða frekari upplýsingar um ákveðinn einstakling eða stað er að finna í öðrum stafrænum söfnum.
- Loks verður smíðað ritstjórnarviðmót þar sem samstarfsaðilar geta sótt um að fá skilgreindan skrifaðgang að miðlæga þekkingarbrunninum. Hér er t.d. um að ræða starfsfólk héraðsskjalasafna eða aðila að rannsóknarverkefnum sem eru þess eðlis að nýjar upplýsingar um einstaklinga og/eða staði koma fram. Í slíkum verkefnum geta samstarfsaðilar því aukið við þekkingarbrunninn og tengt sín gögn um leið við önnur gögn og þannig verða hin nýju gagnasöfn samræmanleg við gagnasöfn sem fyrir eru. Í tengslum við skrifaðgang verður sett á fót ritstjórn og aðgangsstýringu þannig hagað að ritstjórar þurfa að yfirfara og staðfesta innsetningar ákveðinna notenda eftir því sem við á. Slíkt verður samkomulagsatriði og skoðað í hverju tilfelli eftir því sem notendur fá skrifaðgang.

Þessi skipting er algeng þegar kemur að stórum hagnýtum gagnasöfnum. Sem fyrirmynd má nefna Beygingalýsingu íslensks nútímamáls sem Árnastofnun rekur þar sem fletta má upp á vefnum bin.arnastofnun.is en einnig má hala gögnunum niður í heild sinni eða senda fyrirspurnir vélrænt í gegnum forritaskil.

Langtímamarkmið verkefnisins er að hér byggist upp yfirgripsmikill þekkingarbrunnur sem í fyrstu verður byggður á samræmdum stafrænum gögnum sem margar stofnananna hafa nú þegar undir höndum. Í þessum þekkingarbrunni mun hvert einindi, þ.e. stök lýsigögn um einstakling og bæ, hafa fast, óbreytanlegt auðkenni (sbr. DOI) sem mun tryggja að tengingar við brunninn úr gögnum utan hans munu halda sér. Með því að nýta sér þessi auðkenni einstaklinga og staða geta aðrar stofnanir sem byggja upp gagnasöfn yfir einstaklinga eða staði tryggt samræmi við önnur gagnasöfn í framtíðinni, bæði innanlands og erlendis. Sem dæmi um vandamál sem þessi þekkingarbrunnur leysir af hendi er að henda reiður á hvaða Jónar Jónssynir á 19. öld eða hvaða þverár í Eyjafjarðarsýslu eru í raun sömu einindi í tveim gagnasöfnum.

Mikill hagur er fólgin í því að tengja auðkenni staða og einstaklinga við erlendar nafnmyndaskrár á borð við VIAF (Virtual International Authority File) og fleiri skrár sem þjóna

svipuðum tilgangi og umrætt verkefni nema í víðara samhengi. Við slíka samtengingu opnast tengimöguleikar íslenskra sögulegra gagna við alþjóðlegt samhengi í gegnum opin tengd gögn (linked open data).

Ekki hefur verið gerð tilraun til uppbyggingu slíks innviðar á Íslandi áður og smíði hans fylgja ýmsar áskoranir. Uppbygging þeirra stafrænu gagna sem þegar eru til er ólík og gögnin byggð upp á ólíkum forsendum. Í upphafi verkefnisins verður gerð greining á ástandi gagnanna til að ganga úr skugga um með hvaða hætti er hægt að sækja upplýsingar í þá. Tæknileg þekking á eðli gagnanna og þróunarumhverfinu í heild sinni verður varðveitt í formi verfærakistu, sem með hagkvæmum hætti auðveldar og samræmir alla vinnslu.

Samstarfsaðilar munu einnig skilgreina þær upplýsingar sem skrá á í gagnagrunninn. Um hverja manneskju er sjálfsagt að skrá og kortleggja eins og hægt er:

- fæðingarár,
- dánarár,
- fæðingarstað,
- dánarstað,
- kyn.

Einnig er æskilegt að skrá (eftir því sem við á):

- menntun,
- stöðu og starf (breytilegt eftir tímabilum),
- stutt æviágrip,
- búsetu eftir tímabilum
- giftingarár,
- ár við skilnað,
- ár við makamissi.

Í sumum tilvikum er vitað um fæðingar- og/eða dánarár án dagsetningar og í enn öðrum tilvikum er aðeins hægt að segja með vissu að viðkomandi hafi verið uppi á tiltekinni öld. Þessar upplýsingar verða skráðar í grunninn með eins skipulegum hætti og tiltækar upplýsingar leyfa þannig að hægt sé að flokka upplýsingar eftir aldri þeirra sem komu við sögu og skópu, varðveittu eða sýsluðu með grunnheimildirnar í söfnum stofnananna. Í sögulegu bæja- og mannanafnatali verður einnig hægt að tengja fólk við staði á landinu, eða jafnvel erlendis. Þannig myndast tengingar við fæðingarstað, uppvaxtarstað, búsetu og dánarstað fólks. Í verkefninu verður unnið með nafnmyndir fólks sem finnst í gagnasöfnum stofnananna og manntölum frá 1703 til 1920. Í öðrum áfanga verður gervitauganet þróað til að greina galla og gloppur í þekkingarbrunninum, ásamt því að geta mögulega stungið upp á nýjum tengingum sem farið gætu hafa framhjá mannlegu auga.

Vitað er að nöfn og mörk sókna, hreppa og sýslna hafa breyst verulega í gegnum tíðina (sjá t.d. Björk Ingimundardóttir 2019). Bæir hafa skipt um nafn, flust til eða farið tímabundið eða varanlega í eyði. Afar mikilvægt er að þessi þróun komi fram. Bæir og staðir verða staðsettir með GPS hnítum þannig að hægt sé að staðsetja þá á korti í þeim gagnagrunnum og gáttum sem nota slíka birtingu. Leitað verður leiða til að tengja grunninn við landnúmeraskrá Þjóðskrár Íslands (Fasteignamats ríkisins), sem þegar hefur skapað sér sess og er virk skrá, sem notuð er við leit að gögnum. Sú skrá nær þó aðeins til hluta þeirra bæja- og staðarnafna sem um er að ræða og því nauðsynlegt að vinna rannsóknir og eiga samvinnu við þá sem gerst þekkja til víða um land.

Grunnupplýsingar varðandi bæi eru:

- nafn bæjar eða jarðar og nafnbreytingar,
- tenging við aðra bæi eða staði (t.d. hjáleigur, þurrabúðir eða smábýli),
- sýsla, prófastsdæmi,
- hreppur (miðað við hreppakort frá ákveðnum tíma t.d. 1920) en breytinga getið,
- sókn, prestakall (miðað við sóknarkort frá ákveðnum tíma t.d. 1920) en breytinga getið,
- staðsetning (GPS).

Einnig er æskilegt að skrá:

- stutt lýsing bæjar eða jarðar.

Eftir að manna- og staðanafnatöflur hinna ýmsu gagnagrunna hafa verið sóttar vélrænt og steipt saman til að mynda sögulegt bæja- og mannanafnatal tekur við vinna við að sameina og samræma upplýsingar. Lögð verður áhersla á að vinna sem mest af þessari vinnu með vélrænum hætti. Hins vegar er ljóst að mannhugur verður einnig að koma að verki enda er mikilvægt að tryggja gæði þekkingarbrunnansins.

Aðilar að umsókninni eru lykilstofnanir sem allar geyma mikilvægar upplýsingar og gagnasöfn sem nýtast við uppbyggingu þekkingarbrunnansins, eða búa yfir tæknilegri sérþekkingu á sjálfvirkri vinnslu með gögn. Í fyrsta hluta verkefnisins verður til frumgerð að samvinnuðum gagnagrunni stofnananna þar sem einblínt var á ákveðið hérað og íbúa þess á takmörkuðu tímabili.

Hér á eftir er lýsing sem á sérstaklega við um annan hluta verkefnis, sem hér er sótt um styrk fyrir:

Í öðrum hluta verður frumgerðin útvíkkuð til þess að ná utan um öll fleiri gögn aðildarstofna verkefnisins, þ.e. öll þau gögn sem skipta meginmáli fyrir manna- og bæjatalið. Markmiðið er að nýta reynslu og þekkingu sem varð til í fyrsta hluta, auk þeirra verkfæra sem þróuð voru í tengslum við gerð frumgerðarinnar, til þess að ljúka samþættingu allra gagna úr manntölum frá 1703 til 1920 og leggja grunn að því að sóknarmannatöl sem nú er verið að koma á stafrænt snið geti samþæst snurðulítið jafnóðum og þau eru unnin. Unnið verður að undirbúningi tengingar við upplýsingar í nýju skráningarkerfi Sarps (þó tenging við Sarp þurfi væntanlega að bíða þar til nýtt kerfi er komið í gagnið og upplýsingar fluttar yfir). Að svo miklu leyti sem verkefnið færir yfir á 20. öldina verður hugað að persónuverndarsjónarmiðum.

Annað markmið þessa hluta er að greina og/eða staðfesta gögnin í nýja gagnabankanum (þ.e. upplýsingar um bæi og einstaklinga úr ýmsum gagnabönkum), til dæmis með því að kortleggja mögulegar breytingar á bæjum (t.d. þegar kot og hjálegur eru klofin frá bæjum og þau mögulega sameinuð hvert öðru eða bænum aftur), rekja búsetuferil einstaklinga sem þekkjast og leita uppi og draga ályktanir um einstaklinga sem lítið sem ekkert er vitað um annað en nafn og mögulega aldur, t.d. vinnuhjú og niðurstættinga. Ýmislegt í þessum gögnum er mjög torrekjanlegt nema með mikilli yfirlegu, sbr. vinnuhjú sem eru skráð í manntöl en eru mögulega á sitthvoru landshorni milli manntala.

Tölvutæknin og framfarir í gervigreind mun nýtast í þessari vinnu. Eitt af því sem t.d. er mikilvægt að gera sér grein fyrir er hvort ólíkar tilvísanir í fólk og bæi úr mismunandi heimildum séu að vísa til sama „hlutar“. Ofangreint vandamál, þ.e. að tengja saman ólíkar lýsingar af

hlutum, einstaklingum eða atburðum og ákvarða hvort verið sé að vísa til sama hlutar, einstaklings, eða atburðar er ekki einstætt fyrir þetta verkefni og skýtur víðar upp kollu. Í því sambandi má nefna vefsíður sem gera verðsamanburð á sömu/samskonar vörum milli ólíkra verslana, þar sem oft getur reynt erfitt að átta sig á hvort verið sé að vísa til sömu vörunnar. Vandamálið, sem á ensku kallast „product matching“, felst í stórum dráttum í því að búa til tölvugerð líkön sem taka sem inntak lýsingar á margskonar formi (t.d. texta og myndir) og úttakið segir til um hversu líklegt sé að tvær ólíkar lýsingar séu að vísa til sama hlutar. Þetta vandamál hefur því þegar verið töluvert rannsakað og hafa m.a. vélrænar lærdómsaðferðir sem byggja á klösun (e. clustering) og (djúp)tauganetum (e. deep neural-networks) verið notuð með góðum árangri.

Í þessum verkþætti komum við aðlaga framangreindar aðferðar að Sögulega manna- og bæjatalinu eins og þörf þykir. Áskoranir við það eru fyrst og fremst að tvennu tagi:

- Annarsvegar að gögnin séu bæði fjölbreyttari að gerð (t.d. landfræðileg) og breytilegri yfir tíma (t.d. fyrirkomulag bæja og hjálega breytist) en slík aðferð ræður að óbreyttu vel við.
- Hins vegar að gagnamagnið sem verið er að vinna með er minna, sem er að vissu leyti jákvætt, en sem hefur einnig auka áskoranir fyrir vélrænar lærdómsaðferðir.

Því er líklega nauðsynlegt að aðlaga núverandi reikniadaferðir töluvert til að geta beitt þeim á þetta verkefni á sem skilvirkastan hátt.

Gert er ráð fyrir því að meistaranemi í tölvunarfræði við Háskólann í Reykjavík vinni þennan hluta verkefnis sem 60 ECTS rannsóknartengt meistaraverkefni og að það taki 12 mánuði.

3.3. Fyrirhuguð notkun innviða

Lýsið mikilvægi innviðanna fyrir framfarir í rannsóknum á Íslandi. Munu þeir nýtast stórum hópi rannsakenda? Munu þeir nýtast á fleiri en einu fagsviði? Eru í gangi verkefni styrkt af samkeppnissjóðum sem myndu nýta innviðina?

Tæknilegar aðferðir, innleiðing verkefnisins og sú þekking sem þar verður til mun ekki eingöngu nýtast umsækjendunum heldur verður verkefnið leiðandi hér á landi í að tengja saman stór gagnasöfn og gera þau aðgengileg stofnunum og fræðimönnum í opnum aðgangi. Þekkingarbrunnurinn verður þannig uppbyggður að allir geta sótt þangað upplýsingar og forrit úr verkfærakistu til greiningar og rannsókna. Með slíkum opnum gagnasöfnum verður til stöðkerfi til að vinna margvíslegar rannsóknir, það hvetur til þverfaglegra rannsókna og nálgana og opnar nýjar rannsóknarspurningar og sjónarhorn.

Tilkoma nýs þekkingarbrunnisins mun þannig gjörbreyta möguleikum til rannsókna í hugvísindum og öðrum fræðasviðum sem vinna með söguleg gögn sem tengjast einstaklingum og stöðum. Hópur rannsakenda sem þekkingarbrunnurinn mun hafa áhrif á er því mjög stór og nær langt út fyrir raðir hugvísinda, t.d. inn í félagsvísindi, menntavísindi og heilbrigðisvísindi.

Þekkingarbrunnurinn mun ekki aðeins styrkja stafræn hugvísindi heldur líka efla hefðbundnari hugvísindi þar sem hann býður upp á mun betri aðferðir við að leita gagna. Helsta mikilvægi þekkingarbrunnisins eru tengingar sem verða til með samtenginu gagnagrunna í gegnum einstaklinga og staði og möguleikar til að skoða mikið magn tengdra gagna. Hann mun því hafa mikil áhrif á þróun stafrænna hugvísinda á Íslandi og gera þau sambærileg við stöðu mála í nágrannalöndum okkar. Verkefni sem geta nýtt sér gögnin eru til dæmis rannsóknir sem byggjast á samkeyrslu stafrænnar endurgerðar kirkjubóka frá Þjóðskjalasafni og skrár yfir

bréfasafn hjá Landsbókasafni. Ef slík söfn innihalda að miklu leyti samræmd auðkenni staða og einstaklinga verður samkeyrsla gagnanna mun auðveldari og mikill tími sparast þar sem ekki þarf starfsfólk til að samræma gögnin handvirkt. Sama á við um samkeyrslu upplýsinga úr öðrum stafrænum gagnagrunnum.

Möguleikar fyrir rannsóknir, rannsóknþjálfun og kennslu eru því nær óþrjótandi. Þekkingarbrunnurinn mun ekki aðeins nýtast í rannsóknum heldur býður hann líka upp á margvíslega notkunarmöguleika fyrir ferðaþjónustuaðila, skipulagsyfirvöld og hin fjölmörgu menningarsetur sem finnast víða um land auk byggðasafna og héraðsskjalasafna. Verkefnið samræmist því vel stefnu ríkis og sveitarfélaga um upplýsingasamfélagið.

4 Lýsing á innviðum – 4. forgangur

Talgreining á Ísmús og uppbygging textasafns fyrir eldra talmál

4.1. Markmið

Lýsið markmiðum með kaupum/uppbyggingu á innviðunum.

Markmiðið með uppbyggingu innviðanna er að:

- búa til textasafn sem verður afrakstur talgreiningar á hljóðupptökum úr þjóðfræðisafni Stofnunar Árna Magnússonar í íslenskum fræðum, og
- þjálfa talgreiningarhugbúnað til að nota við talgreiningar á upptökum með svipaða eiginleika.

Slíkt textasafn er mjög mikilvægt og í raun og veru forsenda þess að hægt sé að rannsaka þjóðfræðaeefnið og viðtölin í þjóðfræðisafninu bæði með aðferðum munnlegrar sögu og þjóðsagnafræði. Það er einnig nauðsynlegt til þess að málfræðirannsóknir á stíl, orðaforða, talanda og framburði heimildarfólksins, ásamt þróun á þessum atriðum, geti átt sér stað. Þjóðfræðisafnið geymir um 2000 klukkustundir af efni sem hljóðritað var á síðustu öld. Elstu upptökurnar eru frá 1903 en langflestar eru hljóðritaðar á sjöunda áratugnum og fram á þann níunda víðs vegar um landið. Efnið inniheldur meðal annars sagnir, ævintýri, kvæði, endurminningar og þjóðháttalýsingar. Stór hluti heimildarfólksins er fæddur á tímabilinu 1880–1910 og það kemur víðs vegar að af landinu. Einnig var talsverðu efni safnað í Norður-Ameríku á slóðum Vestur-Íslendinga. Upptökurnar eru aðgengilegar á vefnum <http://www.ismus.is> sem er samstarfsvefur Stofnunar Árna Magnússonar í íslenskum fræðum (SÁM) og Tónlistarsafns Íslands sem heyrir undir Landsbókasafn Íslands – Háskólabókasafn. Á þeim vef má leita að hljóðupptökum eftir heimildarfólki, stöðluðum efnisorðum sem að hluta lýsa innihaldi upptakanna en þar er einnig að finna útdrætti úr því efni sem farið er með og lýsa innihaldinu í stuttu máli.



Stór hluti af uppbyggingu innviðanna er að þjálfa talgreiningarhugbúnað á dæmum af hljóðupptökum þar til greining á heildarsafninu skilar viðunandi árangri. Í lok verkefnisins verða leitarmöguleikar stórlega bættir þar sem hægt verður að leita mun nánar eftir efni á vefnum ismus.is, en einnig í málheildarviðmóti sem sniðið er að þörfum málfræðinga sem áhuga hafa á að rannsaka efnið. Öll textagögnin verða gefin út með

opnu leyfi og einnig hljóðupptökurnar nema í þeim fáu tilvikum þar sem heimildarfólkið sjálft hefur sett fyrirvara um notkun efnis. Opinn aðgangur er nauðsynlegur til þess að efnið gagnist við margs konar rannsóknir í máltækni, sérstaklega talgreiningu.

4.2. Almenn lýsing á innviðum

Lýsið innviðunum sem sótt er um styrk fyrir og notkunarviði þeirra.

Ólíkt því sem gerðist í mörgum nágrannalöndum Íslands hefur aldrei verið sett á fót hér sérstök þjóðfræðastofnun heldur tóku þjóðminjasafn Íslands og Handritastofnun Íslands (nú SÁM) að

sér það hlutverk. Þjóðminjasafn fór þá leið að safna efni með spurningalistum sem heimildarfólk svarar skriflega en strax var farin sú leið á Árnastofnun að hljóðrita efni frá heimildarmönnum. Skipuleg söfnun á vegum stofnunarinnar hófst snemma á sjöunda áratug síðustu aldar og fljótlega var þjóðsagnafræðingurinn Hallfreður Örn Eiríksson ráðinn til starfans. Í fjörutíu ár ferðaðist hann um allt land og hljóðritaði allt það þjóðfræðaeefni sem fólk var tilbúið að miðla. Hjónin Helga Jóhannsdóttir og Jón Samsonarson söfnuðu einnig á sama hátt á vegum stofnunarinnar og síðan margir fleiri. Á níunda áratugnum var einnig lögð áhersla á að afrita eldra hljóðritað þjóðfræðaeefni sem varðveitt var á öðrum stofnunum, svo sem á Þjóðminjasafni og hjá Ríkisútvarpinu, það elsta hljóðritað á vaxhólka allt frá 1903. Þjóðfræðisafn stofnunarinnar varðveitir þannig ómetanlegar heimildir um kveðskap, söng, sagnaskemmtun, þjóðlíf á fyrri tíð og munnlega sögu í um 2000 klukkustunda hljóðupptökum. Á vefnum Ísmús (<https://ismus.is>) er nú þegar hægt að hlusta á stærstan hluta af hljóðupptökunum (Rósa Þorsteinsdóttir 2013). Þar er hægt að leita eftir stuttum útdrætti sem lýsir að hluta innihaldi þeirra ásamt því að þær eru flokkaðar eftir efnisorðum, formi og flutningsmáta. Með talgreiningu á upptökunum verður hægt að tengja uppskrift af viðtölunum við hljóðskrárnar og þannig bjóða upp á mun betri leitarmöguleika. Snemma árs 2022 er áætlað að ný útgáfa Ísmús líti dagsins ljós þar sem auk hljóðritasafns verður hægt að leita í Sagnagrunni, gagnagrunni yfir prentaðar sagnir í íslenskum þjóðsagnasöfnum sem geymir upplýsingar um yfir 10.000 þjóðsagnir (Terry Gunnell 2010). Einnig verður hægt að leita í sambærilegum gagnagrunni yfir íslensk ævintýri (Aðalheiður Guðmundsdóttir, 2016). Sameining þessara grunna skapar yfirgripsmikinn gagnagrunn yfir þjóðfræðiefni allt frá fyrri hluta 19. aldar fram á þessa öld. Því munu uppskriftir þjóðfræðiefnis sem varðveitt er í hljóðritasafninu og sem talgreinirinn mun skapa auka mjög möguleika á samanburðarrannsóknnum á sögnum, ævintýrum og því hvernig þjóðfræðiefni breytist í flutningi á ólíkum tímum.

Textasafnið sem verður til við talgreiningu á hljóðupptökunum verður aðgengilegt á nýjum vef Ísmús. Fyrirhugað er að útvíkka viðmót Ísmús þannig að hægt verði að leita í textunum og fá niðurstöður. Í viðmótinu verður hægt að hlusta á upptökurnar og samhliða sjá uppskriftina úr þeim. Gögnin sem talgreinirinn skilar frá sér eru þannig úr garði gerð að hvert orð er tímasett og því verður hægt að forrita viðmótið þannig að hvert orð sem heyrst í upptökunni merkist sérstaklega þegar það er talað. Verður þessi vinna unnin hjá SÁM sem hluti af mótframlagi stofnunarinnar.

Textasafnið verður gert aðgengilegt í sérstakri málheild sem aðgengileg verður ásamt öðrum málheildum á vefnum [Risamálheildin](#) þar sem hægt verður að leita í safninu út frá málfræðilegum þörfum. Verður því betur lýst í kafla um fyrirhugaða notkun innviðanna.

4.3. Fyrirhuguð notkun innviða

Lýsið mikilvægi innviðanna fyrir framfarir í rannsóknum á Íslandi. Munu þeir nýtast stórum hópi rannsakenda? Munu þeir nýtast á fleiri en einu fagsviði? Eru í gangi verkefni styrkt af samkeppnissjóðum sem myndu nýta innviðina?

Með því að talgreina hljóðritasafnið, greina efnið málfræðilega og merkja það með lýsigögnum á borð við aldur upptöku, aldur þeirra sem tala, kyn þeirra og hvaðan þeir eru skapast fjölmargir möguleikar til rannsókna, í þjóðfræði, sagnfræði, málfræði og máltækni.

Þeir uppskrifuðu tölvutæku textar sem talgreinirinn mun framleiða upp úr hljóðritasafninu bjóða upp á margs konar nýjar aðferðir við að rannsaka efni safnsins, svo sem samanburð á sögum og kvæðum og margs konar greiningu á munnlegum heimildum sem geta nýst við

rannsóknir á munnlegri sögu. Þeir geta einnig vakið upp nýjar og ferskar rannsóknar-
spurningar á ýmsum sviðum menningarfræða. Með því að vinna með efni hljóðskrána sem
texta fæst mikilvæg tenging við þjóðfræðiefni á textaformi sem finna má í gagnagrunninum
[Sagnagrunnur](#) sem geymir prentaðar íslenskar þjóðsögur og ævintýri. Þannig gefst tækifæri til
að bera saman og rannsaka mismunandi gerðir þjóðsagna og annars munnlegs efnis, svo sem:

- þegar „sama saga“ er sögð á mismunandi hátt af mismunandi fólki. Rannsóknir
hafa sýnt að sagnafólk leggur áherslu á mismunandi þætti sagna í frásögn sinni og
setur þannig sinn persónulega svip á það tilbrigði sögu sem það segir (Rósa
Þorsteinsdóttir, 2011). Rannsóknir á slíkum atriðum gefa ekki aðeins til kynna
persónuleg einkenni heldur geta einnig varpað ljósi á mismunandi samfélag og
ríkjandi viðhorf á mismunandi tímum;
- þegar farið er með bundið mál sem aðallega gengur í munnlegri hefð, svo sem
kvæði, vísur og þulur, á mismunandi hátt. Textasafn af hljóðrituðum þulum mun
gagnast Yelenu Sesselju Helgadóttur sérstaklega vel í áframhaldandi rannsóknnum
hennar á þess háttar textum (Yelena Sesselja Helgadóttir 2020);
- þegar sami sagnaþulur segir sömu sögu eða fer með sömu vísu oftast en einu sinni.
Í hljóðritasafninu er að finna dæmi um slíkt og má nefna sagnakonuna Steinunni
Þorsteinsdóttur (1887–1973) sem segir sama ævintýrið fimm sinnum yfir nokkurra
ára tímabil (Rósa Þorsteinsdóttir, 2011). Uppskriftir hljóðritanna bjóða upp á
möguleika á samanburði á því hvernig efni í munnlegri geymd breytist í
endurteknum flutningi;
- hvernig sögur varðveitast og flytjast milli kynslóða og innan svæðis. Hallfreður Örn
Eiríksson sem safnaði miklu af því efni sem hljóðritasafnið geymir hafði mikinn
áhuga á þessu (Hallfreður Örn Eiríksson, 1971). Honum vannst ekki aldur til að
ljúka rannsókn sinni á því efni sem hann hafði hljóðritað í Suðursveit en þar náði
hann að taka upp sögur þriggja kynslóða. Efnið í slíka rannsókn er því til en mun
auðveldara verður að framkvæma hana þegar hægt verður að fá textana
uppskrifaða á auðveldan og fljótlegan hátt;
- frásagnir einstakra hópa. Sem dæmi um slíka rannsókn má nefna doktorsverkefni
Júlíönu Þóra Magnúsdóttur sem rannsakar sagnir kvenna sem Hallfreður Örn
Eiríksson hljóðritaði (Júlíana Þóra Magnúsdóttir 2018);
- frásagnir Íslendinga í Vesturheimi. Sem fyrr segir má í hljóðritasafninu finna
hljóðritanir með fólki af íslenskum uppruna í Norður-Ameríku sem teknar voru á
árunum 1972–1973 og aftur um 1982. Það efni hefur verið talsvert rannsakað
(Gísli Sigurðsson, 2012) en uppskriftir hljóðritanna opna á áframhaldandi
rannsóknir á tungumáli og frásögnum þessa hóps;
- frásagnir af sögulegum viðburðum og lýsingar á samfélagi sem nýst geta
sagnfræðingum sem rannsaka munnlega sögu. Miðstöð munnlegrar sögu
(munnlegsaga.is) sem nú er hluti af Landsbókasafni Íslands – Háskólabókasafni
varðveitir einnig hljóðritaðar heimildir og mun tæknin sem hér verður til því einnig
geta nýst þeirri stofnun.

Fyrir áður nefndan gagnagrunn yfir íslenskar sagnir í prentuðum þjóðsagnasöfnum,
Sagnagrunn, hefur verið þróuð tenging við [ISEBEL-leitarvélina](#) sem rekin er af Meertens-
stofnuninni í Amsterdam en er samstarfsverkefni fleiri stofnana víða um heim. Verkefnið
miðar að því að bjóða upp á leit í þjóðsögum á mörgum tungumálum og býður það því upp á
aukna möguleika í þjóðsagnarannsóknnum í alþjóðlegu samhengi, hvað er líkt með þjóðsögum
á mismunandi svæðum og hvað er ólíkt. Núverandi leitarvél sækir niðurstöður úr
þjóðfræðigrunnum í Danmörku, Þýskalandi og Hollandi. Þar sem hljóðritasafn SÁM inniheldur
þjóðsögur; sagnir og ævintýri þá verður mögulegt að auka efni ISEBEL-leitarvélarinnar í

sérstakri útgáfu hennar, NordISEBEL, sem fyrirhugað er að þróa sérstaklega með norrænu þjóðfræðiefni frá Skandinavíu og Íslandi (Broadwell og Tangherlini, 2020).

Árið 2020 var samstarfsverkefni Landmælinga Íslands og SÁM gert aðgengilegt á vefnum Nafnið.is (<https://nafnid.is>). Á þeim vef má fletta í skjölum örnefnasafns SÁM og sjá staðsetningar þeirra örnefna sem Landmælingar Íslands hafa hnitsett. Í hljóðritasafninu má finna talsvert af lýsingum á örnefnum og má sjá flöt á því að tengja saman þau textagögn sem innviðurinn mun skapa við örnefnavef SÁM í þeim tilgangi að auka þekkingu á þeim og staðsetningu þeirra.

Talgögnin í þessu gagnasafni eru ólík öðrum íslenskum talmálheildum. Fyrst og fremst að því leyti að þau innihalda eldri upptökur en aðgengilegar eru fyrir, upptökur af eldra fólki, upptökur af flutningi á kveðskap og stundum óvenjulegt orðafar, auk þess sem gæði á upptökunum eru misjöfn. Þessir þættir, ef þeir eru vel skráðir og lýsigögn góð, geta orðið til þess að gögnin verði afar gagnleg við þjálfun á talgreinum sem miða að því að vinna með gögn sem hafa þessi séreinkenni, en ekki síður í rannsóknum, t.a.m. á því hvernig hægt er að greina óþekkt orð í tali, og mæli þeirra sem bera sum orð fram með öðrum hætti en venja er nú.

Textinn sem verður til eftir að allt efnið hefur verið talgreint verður settur upp sem sérstök málheild með sambærilegum hætti og aðrar íslenskar málheildir sem aðgengilegar eru á vefnum malheildir.arnastofnun.is. Flestar íslenskar málheildir innihalda hefðbundið ritmál eða ritmál á einhverjum tilteknum sérsviðum. Málið í þessum upptökum er annars eðlis, talmál fólks sem fætt er á yfir 100 ára tímabili. Þetta efni eykur því á fjölbreytni þeirra rannsókna sem hægt er að vinna með íslensku málheildirnar, en þær hafa t.d. verið notaðar í málfræðirannsóknum, við orðabókagerð, við að búa til mállíkön í máltækni fyrir allt frá talgreinum og talgervlum til þýðingarvéla og nafnaþekkjara.

Með tilkomu gagnasafnsins sem hér er lýst skapast einnig fjölmargir möguleikar til rannsókna á mismunandi sviðum málfræði:

- Bent hefur verið á stílfræðileg einkenni frásagnartexta af ýmsu tagi. Dæmi um slíkt stílbragð er frásagnarumröðun, þar sem persónubeygð sögn er notuð fremst í setningu í fullyrðingarsetningu, t.d. Víkur nú sögunni vestur um haf. Þetta hefur verið rannsakað fyrst og fremst á rituðum textum (t.d. Halldór Ármann Sigurðsson 1994) en örðugra hefur reynt að rannsaka þetta í munnlegum frásögnum. Gagnasafnið sem verður til veitir því möguleika á stílfræðilegum samanburði í rituðum og munnlegum frásögnum.
- Málhafar á upptökunum í Ísmús koma víða að og því má gera ráð fyrir að ýmiss konar staðbundinn framburður komi fyrir sem hægt er að rannsaka. Helstu einkenni íslenskra framburðarmállýskna eru vel skrásett og rannsökuð. Með umrituðum textum verður tiltölulega einfalt að finna viðeigandi hljóðumhverfi þegar tiltekna framburðarmállýskur eru skoðaðar í gögnunum. Til dæmis kemur harðmæli fram á lokhljóðunum p, t, k á milli tveggja sérhljóða (í orðum eins og api, ota, reka). Með því að para saman uppskriftir og upptökur verður einfaldara að rannsaka nánar þekkt tilbrigði í framburði. Þá er ekki síður mikilvægt að gagnasafnið veitir möguleika á nýjum uppgötvunum á sviði staðbundins framburðar. Hér er mikilvægt að hafa í huga að þrátt fyrir að málfræðingar (t.d. Björn Guðfinnsson 1946) hafi kortlagt ýmis einkenni mismunandi framburðarmállýskna höfum við hingað til verið að miklu leyti háð rannsóknum þeirra og því sem þeir komu auga á enda var ekki um það að ræða fyrir um 1980 að til væru upptökur úr íslenskum rannsóknarverkefnum.

- Staðbundinn munur í máli kemur einnig fram í setningafræði og beygingar- og orðmyndunarfræði. Eftirfarandi vísa á að vera lýsandi fyrir mál Svafeldinga, a.m.k. á fyrri hluta síðustu aldar (Ásta Svavarsdóttir og Þóra Björk Hjartardóttir 1996:98): Svartfuglu í sandinum leit / sú ég hugsí hafi verið feit. / Hana drepíð hebbí eg skulað þá / hebbí hún ekki flogíð burt mér frá. Auk hins þekkta staðbundna bö-framburðar (hebbí) er þarna fyrstu persónu orðmyndin hugsí (í stað hugsa) sem er dæmi um staðbundin tilbrigði í beygingarfræði og nokkuð hefur verið ritað um (Stefán Karlsson 1993, Halldóra Kristinsdóttir 2006). Þessu til viðbótar eru setningafræðileg og beygingarfræðileg tilbrigði í notkun háttarsagnarinnar skulu (skulað og orðaraðarleg staða hennar á eftir hjálparsögn sem almennt þekkist ekki í íslensku). Hins vegar hafa heimildir um þetta verið heldur fátæklegar og einkum að finna í orðasöfnum SÁM. Þetta sýnir að staðbundin afbrigði á ýmsum sviðum hafa ekki verið rannsökuð nema að litlu leyti en slíkt verður auðveldara með uppskriftum hljóðritana Ísmús.
- Með umrituðum gögnum verður hægt að rannsaka orðaforða og gera samanburð við ýmis rafræn gagnasöfn, svo sem Risamálheildina (Steinþór Steingrímsson o.fl. 2018), Beygingarlýsingu íslensks nútímamáls (Kristín Bjarnadóttir, Kristín Ingibjörg Hlynsdóttir og Steinþór Steingrímsson 2019) og Íslenska nútímamálsorðabók (Halldóra Jónsdóttir og Þórdís Úlfarsdóttir 2020). Með því að nota orðtökutólið ALEXÍU (Steinunn Rut Friðriksdóttir, Atli Jasonarson, Steinþór Steingrímsson og Einar Freyr Sigurðsson 2021) er tiltölulega einfalt að finna orð í ÍSMÚS sem ekki er að finna í áður nefndum gagnasöfnum. Með skrásetningu lýsigagna verður svo mögulegt að rannsaka betur þann staðbundna orðaforða sem skráður er í orðasöfnum Stofnunar Árna Magnússonar í íslenskum fræðum, svo sem Talmálssafninu (Gunnlaugur Ingólfsson 1988, Ásta Svavarsdóttir og Guðrún Kvaran 1994), og Íslensk-danskri orðabók Sigfúsar Blöndals (1920-1924). Með ÍSMÚS-gögnunum verður einnig mögulegt að skoða betur merkingu slíkra orða í því samhengi sem þau birtast og orðanotkun. Þetta býður t.a.m. upp á samanburð við Íslenskt orðanet (Jón Hilmar Jónsson 2018, Hjalti Daníelsson, Jón Hilmar Jónsson o.fl. 2021).
- Í safninu eru hljóðritanir Vestur-Íslendinga en mjög áhugaverðar rannsóknir hafa verið gerðar á vesturíslensku máli, menningu og sögu eins og sést glögglega í Sigurtungu frá 2018 (Birna Arnbjörnsdóttir, Höskuldur Þráinsson og Úlfar Bragason ritstj.). Sem dæmi hefur aðlögun enskra tökuorða í vesturíslensku verið skoðuð, svo sem í hvaða málfræðilega kyni tökuorð eins og kót (e. coat) eða míting (e. meeting) eru (Vilhjálmur Stefánsson 1903, Gísli Sigurðsson 2001, Birna Arnbjörnsdóttir 2006, Birna Arnbjörnsdóttir og Höskuldur Þráinsson 2018, Ásta Svavarsdóttir 2018). Þá hefur t.d. flámæli í vesturíslensku verið rannsakað (Birna Arnbjörnsdóttir 1987). Með umrituðum gögnunum aukast möguleikar á rannsóknum á þessum sviðum.
- Að lokum má nefna félagsmálfræðilegar rannsóknir en þar sem lýsigögn eru skráð er hægt að rannsaka ýmsar málbreytur með tilliti til félagsbreyta á borð við kyn og aldur. Munur á máli karla og kvenna, sérstaklega framburði, hefur verið skoðaður talsvert, einkum í erlendum rannsóknum (sjá t.d. Labov 1964, 1990) en einnig í íslenskum rannsóknum (sjá t.d. Guðvarð Má Gunnlaugsson 1984). Slíkur samanburður hefur leitt í ljós að konur tala gjarna mál sem er í samræmi við málstaðal eða mál sem nýtur virðingar. Með uppritun á hljóðritunum og upplýsingum um málhafa verður mögulegt að bera saman mál karla og kvenna á ýmsum tímum. Þá hafa svokallaðar breytingar á lífsleiðinni verið rannsakaðar í auknum mæli undanfarin ár og áratugi, þar sem fólk breytir máli sínu á fullorðinsaldri. Má þar nefna samanburð á einhljóðaframburði Vestur-

Skafthellinga í átt til tvíhljóðaframburðar (Guðvarður Már Gunnlaugsson 1984), rannsókn á framburði Elísabetar Englandsdrottningar (Harrington, Palethorpe og Watson 2000), /r/ í frönsku í Montreal (Blondeau og Sankoff 2007), framburðarbreytingar í máli Norðlendinga (Margrét Lára Höskuldsdóttir 2013) og rannsókn á notkun stílfærslu í þingræðum Steingríms J. Sigfússonar (Lilja Björk Stefánsdóttir og Anton Karl Ingason 2018). Slíkar breytingar eru mjög einstaklingsbundnar og þess vegna er fjöldi málhafa venjulega ekki vandamálið heldur það hvort til séu nægilega ríkuleg gögn frá einum og sama málhafanum yfir lengra tímabil. Með gagnasafninu sem hér hefur verið lýst skapast tækifæri til að skoða lífsleiðarbreytingar á 20. öld.

5 Lýsing á innviðum – 5. forgangur

6 Skipulag og stjórnun

6.1. Kostnaður og fjármögnun á rekstri innviða

Gerið grein fyrir kostnaði og fjármögnun á rekstri innviða á styrktímabilinu. Hvernig verður brugðist við ef áætluð fjármögnun stenst ekki? Tilgreinið stöðu fjármögnunar (t.d. hvaðan kemur fjármagnið). Athugið að staðfesting á aðkomu stofnananna að rekstri innviðanna þarf að fylgja umsókn.

Reglur MSHL voru samþykktar í september og stjórn skipuð í október. Sjá reglur og lista yfir meðlimi stjórnar hér: https://www.hi.is/samstarf/midstod_stafraenna_hugvisinda_og_lista

Annar áfangi uppbyggingar MSHL er í fjórum liðum.

Fyrsti forgangur: Sarpur. Uppfærsla menningarsögulegs gagnabanka

Verkefnið er í tvennu lagi:

- A. Kaup og þýðing á nýju skráningakerfi.
- B. Yfirfærsla gagna úr eldra kerfi í nýtt.

Uppfærsla á Sarpi er eitt mikilvægasta verkefnið við uppfærslu gagnabanka til rannsókna undir MSHL. Nýtt kerfi er viðamikil fjárfesting sem kallar á formlegt útboð á Evrópska efnahagssvæðinu. Í dag er búið að kalla eftir áhugayfirlýsingum og að þeim fengnum koma fimm alþjóðleg fyrirtæki til greina. Áætlun um kostnað við kaup á nýju kerfi hljóðar upp á 36.000.000 kr. Mótframlag kemur frá Landskerfi bókasafna (sem rekur Sarp), alls 9.200.000 kr. (án vsk). Hluti kostnaðar við kaupin er þýðing á kerfinu, sem er metin á 3.000.000 kr. Yfirfærsla gagna er mjög viðamikið verkefni. Kostnaður við hana er metinn á 14.000.000 kr. hjá verktaka, plús 4.800.000 hjá verkkaupa. Samtals 18.800.000 kr. Reiknað er með hálfmánaðar vinnu hjá Háskóla Íslands við stjórnun verkefnisins inna MSHL.

Yfirlit yfir kostnað:

• Kaup á nýju kerfi:	33.000.000 kr.
• Þýðing kerfis á íslensku:	1.600.000 kr.
• Uppfærsla á tungumáli – forritun:	1.400.000 kr.
• Kostnaður við útboð:	1.562.400 kr.
• Verkefnisstjórn yfirfærslu:	2.000.000 kr.
• Yfirfærsla gagna:	16.800.000 kr.
• Umsjón:	512.500 kr.
• Kostnaður samtals:	56.874.900 kr.

Mótframlög:

• Landskerfi bókasafna (beint):	9.400.000 kr.
• Landskerfi bókasafna (vinna):	4.800.000 kr.
• Háskóli Íslands (vinna):	512.500 kr.
• Mótframlag samtals:	14.712.500 kr.

Sótt er um styrk úr Innviðasjóði til að fjármagna 75% af kaupum og íslenskun skráningarkerfis, 75% af kostnaði við útboð og 75% af launakostnaði við yfirfærslu gagna. Landskerfi bókasafna greiðir mótframlag af kaupum- og íslenskun kerfis og eins leggur Landskerfi bókasafna fram vinnu við flutning gagna. Reiknað er með um sex mánaða vinnu á 800.000 kr. á mánuði hjá Landskerfum. Öll þessi fjármögnun er trygg.

Rekstrarfélag Sarps mun taka við rekstri og þróun kerfisins að þessum innleiðingarfasa loknum.

Rekstur: Nýtt skráningarkerfi verður rekið af Landskerfi bókasafna og fjármagnað af Rekstrarfélagi Sarps.

Annar forgangur: Transkribus – tölvulestur handskrifaðra skjala

Yfirlit yfir kostnað:

- Tilboð frá ReadCoOp (án VSK): 5.182.100 kr.
- Laun sérfræðings í 50% starfi: 5.052.000 kr.
- Laun sérfræðinga stofnana: 6.150.000 kr.
- Umsjón: 512.500 kr.
- **Kostnaður samtals: 15.798.600 kr.**

Mótframlag frá fjórum stofnunum:

- 25% af tilboði án VSK: 1.300.000 kr.
- Vinna sérfræðinga: 3.587.500 kr.
- **Mótframlag samtals: 4.438.500 kr.**

Tilboðið er fyrir kaupum á aðgangi að Transkribus og 300.000 einingum til skjalalestrar. Sótt er um styrk úr Innviðasjóði fyrir 75% af verði (án VSK) og 75% af vinnu við þróun módelis fyrir íslensku. Miðað er við launakostnað sérfræðinga eins og notað er við umsóknir í Rannsóknasjóð. Öll fjármögnun vinnu og mótframlags vegna kaupa er þegar tryggð.

Rekstur: Transkribus er rekinn af ReadCoOp, sem MSHL verður hluthafi í. Íslensku módelin verða hluti af kerfi Transkribus og rekin af ReadCoOp. Eini kostnaðurinn er árgjald upp á 250 evrur, sem MSHL mun greiða.

Þriðji forgangur: Sögulegt manna- og bæjatal

Yfirlit yfir kostnað:

- Tilboð frá ad libitum ehf. 10.000.000 kr.
- Laun MA nema í 12 mánuði: 5.616.000 kr.
- Laun sérfræðinga stofnana: 4.118.000 kr.
- Umsjón: 512.500 kr.
- **Kostnaður samtals: 20.246.500 kr.**

Mótframlag frá sex stofnunum:

- Vinna sérfræðinga stofnana: 4.118.000 kr.
- Háskólinn í Reykjavík (vinna): 750.000 kr.
- Háskóli Íslands (vinna): 512.500 kr.
- **Mótframlag samtals: 5.380.500 kr.**

Sótt er um styrk úr Innviðasjóði til að fjármagna tilboð incognito ehf. í uppbyggingu og uppfærslu stafrænna innviða. Vinna sérfræðinga verður fjármögnuð af þeim stofnunum sem þeir vinna hjá (Stofnun Árna Magnússonar, Landsbókasafni – Háskólabókasafni, Þjóðskjalasafni Íslands, Háskólanum í Reykjavík og Þjóðminjasafni Íslands). Háskóli Íslands

greiðir laun vegna umsjónar. Sérfræðingarnir vinna þegar hjá stofnununum og er fjármögnun starfa þeirra trygg. Miðað er við launakostnað sérfræðinga eins og notað er við umsóknir í Rannsóknasjóð. Launakostnaður MA nema miðar við útgefnar viðmiðunartölur Háskólans í Reykjavík fyrir 2022, 468.000 kr. á mánuði.

Rekstur: Þegar þekkingarbrunnurinn er tilbúinn kallar hann ekki á mikinn rekstrarkostnað, og sá sem um er að ræða verður fjármagnaður af reglulegu rekstrarfé viðkomandi stofnana skv. samkomulagi við Upplýsingatæknisvið Háskóla Íslands.

Fjórði forgangur: Talgreining Ísmús og uppbygging textasafns fyrir eldra talmál.

Yfirlit yfir kostnað:

- Tilboð frá Tiro ehf. 9.279.600 kr.
- Aðstoðarfólk í 8 mánuði: 4.888.000 kr.
- Laun sérfræðinga stofnana: 5.785.500 kr.
- Umsjón: 512.500 kr.
- **Kostnaður samtals: 20.465.600 kr.**

Mótframlag frá sex stofnunum:

- Vinna sérfræðinga stofnana: 5.035.500 kr.
- Háskólinn í Reykjavík (vinna): 750.000 kr.
- Háskóli Íslands (vinna): 512.500 kr.
- **Mótframlag samtals: 6.298.000 kr.**

Sótt er um styrk úr Innviðasjóði til að fjármagna tilboð Tiro ehf við þróun talgreinis og vegna vinnu aðstoðarfólks við að útbúa prófunar- og þjálfunargögn (átta mánaða vinna). Miðað er við viðmiðunarlaun Háskóla Íslands við ráðningu aðstoðarfólks til rannsókna. Vinna sérfræðinga verður fjármögnuð af Stofnun Árna Magnússonar. Háskóli Íslands greiðir laun vegna umsjónar. Sérfræðingarnir vinna þegar hjá stofnununum og er fjármögnun starfa þeirra trygg. Miðað er við launakostnað sérfræðinga eins og notað er við umsóknir í Rannsóknasjóð.

Rekstur: Þegar nýtt módel fyrir íslensku er tilbúið kallar það ekki á mikinn rekstrarkostnað, og sá sem um er að ræða verður fjármagnaður af reglulegu rekstrarfé viðkomandi stofnana.

Á næstu síðu er GANTT kort sem sýnir þróun og stöðu verkefna MSHL 2021 til 2023. Á síðunni á eftir er listi yfir afurðir. Eins og þar sést er gert ráð fyrir að vinna við sum verkefnum ljúki ekki fyrr en árið 2023.

	2021												2022												2023													
	j	f	m	a	m	j	j	á	s	s	o	n	d	j	f	m	a	m	j	j	á	s	s	o	n	d	j	f	m	a	m	j	j	á	s	s	o	n
A: Stjórnun																																						
Reglur samþykktar																																						
Stjórn skipuð																																						
Samstarfssamningar																																						
Kortlagning innviða																																						
Próun nýrra verkefna																																						
Mat á árangri																																						
B: Mann- og bæjatal																																						
Gengið frá samningum																																						
Fyrsti hluti																																						
Annar hluti																																						
Þriðji hluti																																						
C: 3D skanni																																						
Gengið frá samningum																																						
Kaup																																						
Þjálfun																																						
Rekstur																																						
D: Sarpur, nýtt skráningarkerfi																																						
Gengið frá samningum																																						
Undirbúningur																																						
Útboð																																						
Kaup og þýðing kerfis																																						
Yfirfærsla gagna																																						
E: Transkribus																																						
Gengið frá samningum																																						
Aðild að Read Coop																																						
Próun á íslensku módeli																																						
Ítrun á íslensku módeli																																						
F: ISMÚS																																						
Gengið frá samningum																																						
Próun á módeli																																						
Ítrun																																						
G: Alþjóðlegt samstarf																																						
DARIAH																																						
CLARIN																																						

- A1 Reglur samþykktar af öllum aðilum og staðfestar af háskólaráði HÍ
- A2 Stjórn skipuð
- A3 Samstarfssamningar um styrkt verkefni samþykktir
- A4 Skýrsla um kortlagningu innviða birt
- A5 Endurskoðuð langtímaáætlun um uppbyggingu MSHL samþykkt
- A8 Matsnefnd skilar skýrslu um árangur MSHL
- B1 Samstarfssamningur, samningur um verkaup, samningur við UTS um hýsingu: Lokið
- B2 Fyrstra verkþætti lokið
- B3 Öðrum verkþætti lokið
- B4 Frumgerð afhent
- B5 Þróun tauganets til prufunar á frumgerð skilað
- B6 Fyrsta verkþætti annars hluta lokið
- B7 Öðrum verkþætti annars hluta lokið
- B8 Stækkaðri frumgerð gagnagrunns skilað
- C1 Samstarfssamningur, samningur um vörukaup: Lokið
- C2 3D skannar afhentir
- C3 Námskeið á vegum söluaðila: Lokið
- C4 Rekstraráætlun framkvæmdastjórnar lögð fram til samþykktar
- D1 Samstarfssamningar um styrkt verkefni samþykktir
- D2 Útboð á EEA birt
- D3 Útboðsferli lokið með bindandi ákvörðun
- D4 Kerfi tilbúið til notkunar
- D5 Yfirfærsla gagna úr Sarpi III lokið
- E1 Samstarfssamningar um styrkt verkefni samþykktir
- E2 Aðild að ReadCoOp samþykkt
- E3 Fyrsti hluti íslensks módels prófaður
- E4 Annar hluti íslensks módels prófaður
- E5 Þriðji hluti íslensks módels prófaður
- E6 Fyrsti hluti íslensks módels prófaður
- E7 Lokagerð íslensks módels fyrir 17.-19. aldar hluti af Transkribus
- F1 Samstarfssamningar um styrkt verkefni samþykktir
- F2 Prófunargagnasafn og þjálfunargagnasafn: Leiðrétt
- F3 Gögn afhent verkkaupa
- G1 MSHL cooperative aðili að DARIAH
- G2 Ákvörðun um fyrirkomulag CLARIN liggur fyrir
- G3 Samningar um framtíðarfyrirkomulag CLARIN á Íslandi: Frágengnir

7 Heimildaskrá

Sarpur:

- Wold, T. & Ween, G. 2018. „Digitale visjoner. En kartlegging. Identitet, tilgjengelighet, og digitalt demokrati.“, Nordisk Museologi 2-3, s. 90-106: <https://journals.uio.no/museolog/article/view/6657/5676>
- Forauglýsing, vorið 2021: <https://ted.europa.eu/udl?uri=TED:NOTICE:248633-2021:TEXT:EN:HTML&src=0>
- Spectrum staðallinn: <https://collectionstrust.org.uk/spectrum/>
- Tillögur að nýrri hönnun ytra viðmóts Sarps: <https://www.figma.com/proto/dQizj5igSArFTfBfVgisSw/Sarpur.is?page-id=1%3A2&node-id=345%3A1124&viewport=241%2C48%2C0.06&scaling=scale-down&starting-point-node-id=345%3A1124>

Transkribus:

- van Es, Karin, and Eef Masson. 2018. “Big Data Histories: An Introduction”. TMG Journal for Media History 21 (2): 1–6. DOI: <http://doi.org/10.18146/2213-7653.2018.363>
- Dæmi um rannsókn sem byggist á leit í gagnabanka úr Transkribus: Ågren, Maria. "Making Her Turn Around: The Verb-Oriented Method, the Two-Supporter Model, and the Focus on Practice." Early Modern Women, vol. 13 no. 1, 2018, p. 144-152. Project MUSE, doi:10.1353/emw.2018.0057.

Sögulegt manna- og bæjatal:

Björk Ingimundardóttir. *Prestaköll, sóknir og prófastsdæmi á Íslandi I-II*. Þjóðskjalasafn Íslands. 2019.

Gagnabankar sem tengjast sögulegu bæja- og mannanafnatali:

- Bækur (Landsbókasafn) – <http://baekur.is>
- Dánarbúavefur (Þjóðskjalasafns) – <https://danarbu.skjalasafn.is/>
- Dómabókavefur (Þjóðskjalasafn) – <http://domabaekur.manntal.is/>
- Einkaskjöl (Landsbókasafn) – <http://einkaskjol.is>
- Handrit (Landsbókasafn) – <http://handrit.is>
- Íslandskort (Landsbókasafn) – <http://islandskort.is>
- Ísmús (Árnastofnun / Landsbókasafn) – <http://ismus.is>
- Landamerkjabækur (Þjóðskjalasafn) – <http://jardavefur.skjalasafn.is/landamerkjabaekur/#pg-1>
- Leitir / Gegnir (Landsbókasafn / Landskerfi bókasafna) – <http://leitir.is>
- Manntalsvefur (Þjóðskjalasafn) – www.manntal.is
- Nafnið (Árnastofnun) – <http://nafnid.is>
- Sagnagrunnur (Árnastofnun) – <http://sagnagrunnur.arnastofnun.is>

- Sarpur (Landskerfi bókasafna) – <http://sarpur.is>
- Sóknarmannatalvefur (Þjóðskjalasafn) – <http://salnaregistur.manntal.is/>
- Staðfangaskrá (Þjóðskrá) – <https://skra.is/gogn/grunnoggn-til-nidurhals/stadfangaskra/>
- Tímarit (Landsbókasafn) – <http://timarit.is>
- Túnakort (Þjóðskjalasafn) – <http://skjalaskrar.skjalasafn.is/r/MjU5OTQ2>
- Vefsja kirkjubóka (Þjóðskjalasafn) – <http://vefsja.skjalasafn.is/geoserver/www/vefsja/index.html>
- Vegabréfagrunnur (Þjóðskjalasafn) – <https://heimildir.is/vegabref/>

Talgreining Ísmús og uppbygging textasafns fyrir eldra talmál:

Aðalheiður Guðmundsdóttir. 2016. Ævintýragrunnurinn: Gagnagrunnur yfir íslensk ævintýri. Skíma: Tímarit Sambands móðurmálskennara 39: 28–29.

Ásta Svavarsdóttir og Guðrún Kvaran. 1994. The Computerization of the Spoken Language Archive. K. Hyldegaard-Jensen og V. Hjørnager Pedersen (ritstj.): Symposium on Lexicography VI. Proceedings of the Sixth International Symposium on Lexicography May 7–9, 1992. Lexicographica. Series Maior, Band 57, bls. 127–139. Tübingen: Max Niemeyer Verlag.

Ásta Svavarsdóttir og Þóra Björk Hjartardóttir. 1996. Breytileiki í máli. Erindi um íslenskt mál, bls. 95–109. Reykjavík: Íslenska málfræðifélagið.

Björn Guðfinnsson. 1946. Mállýzkur I. Reykjavík: Ísafoldarprentsmiðja.

Broadwell, Peter M., og Timothy R. Tangherlini. 2020. Geist, geest, geast, spøgelse: Challenges for multilingual search in belief legend archives. Arv: Nordic Yearbook of Folklore 76: 7–28.

Gísli Sigurðsson (ritstj.). 2012. Sögur úr Vesturheimi. Úr söfnunarleiðangri Hallfreðar Arnar Eiríkssonar og Olgu Maríu Franzdóttur um Kanada og Bandaríkin veturinn 1972–73. Reykjavík: Stofnun Árna Magnússonar í íslenskum fræðum.

Gunnlaugur Ingólfsson. 1988. Söfnun Orðabókar Háskólans úr mæltu máli. Orð og tunga 1: 65–72.

Halldór Ármann Sigurðsson. 1994. Um frásagnarumröðun og grundvallarorðaröð í forníslensku. Reykjavík: Málvísindastofnun Háskóla Íslands.

Halldóra Jónsdóttir og Þórdís Úlfarsdóttir. 2019. Íslensk nútímamálsorðabók. Orð og tunga 21:1–26.

Halldóra Kristinsdóttir. 2006. Ég hugsji, voni og ímyndi mér. Um i-endingu ð-sagna í 1. persónu eintölu í nútíð framsöguháttar. BA-ritgerð, Háskóla Íslands, Reykjavík.

Hallfreður Örn Eiríksson. 1971. Um íslensk þjóðfræði. Tímarit Máls og menningar 32/1: 61–69.

Hjalti Daníelsson, Jón Hilmar Jónsson, Þórður Arnar Árnason, Alec Shaw, Einar Freyr Sigurðsson og Steinþór Steingrímsson. 2021. The Icelandic Word Web: A language technology-focused redesign of a lexicosemantic database. Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa), bls. 429–434. Svíþjóð: Linköping University Electronic Press.

Jón Hilmar Jónsson. 2018. Íslenskt orðanet: Tekstbasert kartlegging og presentasjon av leksikalske relasjonar. Nordiske Studier i Leksikografi 14: 1–17.

Júlíana Þóra Magnúsdóttir. 2018. Gender, legend, and the Icelandic Countryside in the Long Nineteenth Century: Re-Engaging the Archives as a Means of Giving Voice to the Women of the Past. Folklore 129/2: 129–147.

Kristín Bjarnadóttir, Kristín Ingibjörg Hlynsdóttir og Steinþór Steingrímsson. 2019. DIM: The Database of Icelandic Morphology. Proceedings of the 22nd Nordic Conference on Computational Linguistics, bls. 146–154, Turku, Finnlandi.

Rósa Þorsteinsdóttir. 2011. Sagan upp á hvern mann: Átta íslenskir sagnamenn og ævintýrin þeirra. Reykjavík: Stofnun Árna Magnússonar í íslenskum fræðum.

Rósa Þorsteinsdóttir. 2013. Ísmús (Íslenskur músík- og menningararfur). An Open-Access Database. The Retrospective Methods Network Newsletter: Limited Sources, Boundless Possibilities: Textual Scholarship and the Challenges of Oral and Written Texts. A special issue of RMN Newsletter ed. by Karina Lukin, Frog and Sakari Katajamäki, 7: 97

101: https://www2.helsinki.fi/sites/default/files/atoms/files/rmn_07_2013.pdf

Stefán Karlsson. 1993. Að hneigja upp á i. Orðaforði heyjaður Guðrúnu Kvaran 21. júlí 1993, bls. 103–106. Reykjavík.

Steinunn Rut Friðriksdóttir, Atli Jasonarson, Steinþór Steingrímsson og Einar Freyr Sigurðsson. 2021. ALEXIA: A Lexicon Acquisition Tool. Proceedings of CLARIN Annual Conference 2021, bls. 64–67. Virtual Edition.

Steinþór Steingrímsson, Sigrún Helgadóttir, Eiríkur Rögnvaldsson, Starkaður Barkarson og Jón Guðnason. 2018. Risamálheild: A Very Large Icelandic Text Corpus. Proceedings of LREC 2018, bls. 4361–4366. Myazaki, Japan.

Terry Gunnell, 2010. Sagnagrunnur: A New Database of Icelandic Legends in Print. Folklore: Electronic Journal of Folklore 45: 151-162: <http://www.folklore.ee/folklore/vol45/gunnell.pdf>

Yelena Sesselja Helgadóttir. 2020. Íslenskar þulur síðari alda. Doktorsritgerð frá Háskóla Íslands. <https://opinvisindi.is/handle/20.500.11815/1939>