

Appendix 1. Lijst van onderzoekers die zijn geconsulteerd over de opzet van Nederlab. Alle geconsulteerde onderzoekers hebben aangegeven dat zij van plan zijn Nederlab in hun onderzoek en onderwijs te integreren.

Rijksuniversiteit Groningen (RUG)

prof.dr. K. van Berkel
dr. G. Bouma
dr. S. Corbellini
prof.dr. G.J. Dorleijn
prof.dr. J. Hoeksema
prof.dr.ir. J. Nerbonne
prof.dr. G.J.M. van Noord
prof.dr. B.A.M. Ramakers
prof.dr. L.R.B. Schomaker

Radboud Universiteit Nijmegen (RUN)

dr. S. Aalberse
prof.dr. R.A.M. Aerts
prof.dr. A.P.J. van den Bosch
prof.dr. P. Fikkert
dr. B.J.M. van Halteren
dr. H. van den Heuvel
dr. L.G.M.M. Hustinx
dr. L.E. Jensen
prof.dr. J.H.T. Joosten
prof.dr. A.M.C. van Kemenade
prof.dr. S.A. Levie
Y.F.M. Linders, MA
dr. B.L.J. Los
dr. N.H.J. Oostdijk
prof.dr. J.B. Oosterman
dr. M. Rem
dr. J.G.M.M. Rosendaal
dr. M.P.J. Sanders
dr. J.M. Sanders
prof.dr. P.J.M.C. Schellens
dr. N.H.H. Veldhorst
dr. C.C. De Wulf

Universiteit Tilburg (TU)

prof.dr E.J. Krahmer
dr. M.W.C. Reynaert
drs. T. Roselaar (IVN)
drs. S. Wubben

Technische Universiteit Delft (TUD)

prof.dr. A.P. de Vries (TU Delft/CWI)

Universiteit van Amsterdam (UvA)

prof.dr. J.A. Baggerman
dr. F.R.E. Blom
prof. dr. L.W.M. Bod
dr. Y. van Dijk
dr. J. Don

prof.dr. O.C.M. Fischer

dr.ir. J. Kamps
prof.dr. J.C. Kennedy
dr. W.T.J.M. Kuiper
prof.dr. J.T. Leerssen
dr. M. Marx
dr. H.C.B. Perridon
prof.dr. N.C.F. van Sas
dr. C.F.M. Streng
prof.dr. T.L. Vaessens
prof.dr. W.R.E. Velema
dr. B.M.A. de Vries
prof.dr. F.P. Weerman
dr. H.H. Zeijlstra

Universiteit Leiden (UL)

prof. dr. W. van Anrooij
dr. J. Audring
prof.dr. G.E. Booij
dr. C.L.J.M. Cremers
dr. M.B. Elenbaas
prof.dr. J.L. Goedegebuure
prof.dr. T. van Haften
dr. A.J.E. Harmsen
dr. M.J. Janse
dr. C.H. Reintges
dr. G.J. Rutten
dr. A.J. van Santen
prof.dr. IM. Tiekens-Boon van Ostade
dr. M.A.C. de Vaan
prof.dr. H. te Velde
prof.dr. A. Verhagen
prof.dr. M.J. van der Wal
dr. G. Warnar
dr. T. van der Wouden

Universiteit Twente (UT)

prof.dr. F.M.G. de Jong

Universiteit Utrecht (UU)

prof.dr. W.J. van den Akker
dr. G. Bloothoof
prof.dr. G. Buelens
prof.dr. N.F.M. Corver
dr. J. Evers-Vermeul
dr. G.J. Johannes
dr. J.M. van Koppen
drs. S. Krauwer
dr. J. de Kruijff
prof.dr. W.W. Mijnhardt

prof. dr. A.B.M Naaijkens
prof.dr. J.E.J.M. Odijk
dr. J.L. Salman
dr. E. Stronks
prof. dr. S. Stuurman
prof.dr. W. Zonneveld

VU Amsterdam

dr. L.M. Aroyo
drs. E. Akkerman
dr. E. Hagen
prof.dr A.A. den Hollander
prof.dr. I.B. Leemans
prof.dr. S. Legene
dr. P.H. Moser
dr. J. Noordegraaf
prof.dr. B.J. Peperkamp
prof.dr. W.P.M.S Spooren
prof.dr. G.J. Steen
prof.dr. P.T.J.M. Vossen

Onderzoeksinstellingen

dr. M.C. Le Bailly (Huygens ING)
prof.dr. L.C.J. Barbiere (MI)
dr. J. Beeken (INL)
prof.dr. H.J. Bennis (MI)
dr. P. Boot (Huygens ING)
dr. M. Bouwhuis (SARA)
dr. H. Brand (FA)
prof.dr. H. Brinkman (Huygens ING)
ir. D. Broeder (MPI)
dr. A. Bruyn (INL)
prof.dr. J. Burgers (Huygens ING)
prof.dr. L.M.E.A. Cornips (MI)
dr. K.H. van Dalen-Oskam (Huygens ING)
lic. K. Depuydt (INL)
drs. S. Derks (Huygens ING)
dr. J. de Does (INL)
drs. G. Gerritsen (Huygens ING)
ing. R. Haentjes Dekker (Huygens ING)
dr. B.J.H. Hermans (MI)
dr. A.B.G.M van Kalmthout (Huygens ING)
ir. M. Kemps-Snijders (MI)
drs. T. Kenter (INL)
dr. E. Kloek (Huygens ING)
dr. J. Kruijssen (MI)
dr. F. Landsbergen (INL)
prof.dr. M. van Oostendorp (MI)
dr. H. Piersma (NIOD)
dr. G.J. Postma (MI)
prof.dr. P.Th. van Reenen (MI)
dr. H.W. Roodenburg (MI)
dr. T. Schoonheim (INL)

dr. N. van der Sijs (MI)
prof.dr. A.P. Versloot (FA)
dr. H. Wals (Huygens ING)
drs. J.J. van Zundert (Huygens ING)

Dataleveranciers

dr. H. Balk (IMPACT, KB)
dr. W. van Bergen (UBL)
drs. L. van den Bosch (Nederlandse Taalunie)
dr. C. Cucchiaroni (Nederlandse Taalunie)
drs. P.M. Doorenbosch (KB)
dr. P.K. Doorn (DANS)
dr. I. Haslinger (KB)
C.A. Klapwijk (DBNL)
drs. M. de Niet (DEN)
drs. J. Oomen (VU/Beeld en Geluid)
drs. M. Slabbertje (UBU)
dr. R. van Stipriaan (DBNL)
dr. B. Zeeman (UBA)

Buitenlandse onderzoekers

prof.dr. J. van der Auwera (Antwerp Un.)
drs. J. Beelen (Oldenburg Un.)
prof.dr. L.J. Beheydt (Louvain Un.)
dr. R. de Bies (Anton de Kom Un. Suriname)
prof.dr. J. De Caluwe (Ghent Un.)
dr. E.L. Coussé (Ghent Un.)
prof. dr. W. Daelemans (Antwerp Un.)
prof.dr D. Geeraerts (Louvain Un.)
dr. C.R. Groeneboer (Pusat Kebudayaan Belanda (Erasmus Taalcentrum) Jakarta, Indonesia)
prof.dr. R. Grüttemeier (Oldenburg Un.)
prof.dr. J.M. van der Horst (Louvain Un.)
prof.dr. R.B. Howell (Wisconsin Un.)
prof.dr. M. Hüning (Berlin Un.)
lic. M. Kestemont (Antwerp Un.)
prof.dr. J. van Keymeulen (Ghent Un.)
prof.dr. E. Leijnse (Namur Un.)
prof.dr. A. Marynissen (Cologne Un.)
prof.dr. J. Nuyts (Antwerp Un.)
prof.dr. J. Pekelder (Paris-Sorbonne Un.)
prof.dr. D.J. Prinsloo (Pretoria Un.)
dr. I. Schuurman (Louvain Un.)
prof.dr. R. Severing (Netherlands Antilles Un.)
lic. V. de Tier (Ghent Un.)
prof.dr. J. Tollebeek (Louvain Un.)
dr. F. Van de Velde (Louvain Un.)
prof.dr. W. Vandenbussche (Brussels Un.)
lic. E. Vanhoutte (KANTL, Gent)
dr. R. Vismans (University of Sheffield, UK)
prof.dr. G. de Vogelaer (Münster Un.)
prof.dr. F. Willaert (Antwerp Un.)

Appendix 2. Overzicht van te verwerken data en metadata

Overzicht van te verwerken corpora

historische bronnen	Estimated size 01/01/ 2013 pp x million	Estimated size 31/12/ 2017 pp x million	Quality	Reprocessing?	Linguistic encoding	year I	year II	year III	year IV	year V
DBNL-XML	4	6	++	n	y	100%	100%	100%	100%	100%
DBNL-pdf's scans	7	10	--	n	n	100%	100%	100%	100%	100%
KB-newspapers	5	8	-	n	n	-	20%	100%	100%	100%
KB-EDBO	2	1,9	--	n	n	-	20%	60%	100%	100%
KB-ProQuest	2	10	--	n	n	-	20%	30%	50%	100%
KB-Google	10	30	--	n	n	-	20%	30%	50%	100%
KB-stat gen	2,5	2,5	+/-	n	n	-	100%	100%	100%	100%
KB-metam-tijdschr	1,5	1,5	--	n	n	-	100%	100%	100%	100%
KB-Linschotenver.	0,02	0,02	--	n	n	-	100%	100%	100%	100%
KB-ANP-nieuwsb	1,5	1,5	--	n	n	--	100%	100%	100%	100%
Behoud Ned Bp (KB, UBU/L/A)	1,5	3	--	n	n	-	20%	30%	50%	100%
UGent-Google	30	60	--	n	n	-	20%	30%	50%	100%
MI-CRM 14de eeuw	0	0	++	n	y	-	100%	100%	100%	100%
MI-GTRP/MAND/SAND	0	0	++	n	y	-	100%	100%	100%	100%
HI-ING (xml-proj)	0,05	0,10	++	n	y	100%	100%	100%	100%	100%
HI-ING (historici.nl)	0,2	0,4	-	n	n	100%	100%	100%	100%	100%
INL-27 milj wrdn	0,08	0,08	+/-	n	y	-	100%	100%	100%	100%
INL-5 milj wrdn	0,015	0,015	+/-	n	y	-	100%	100%	100%	100%
INL-38 milj woorden	0,1	0,1	+/-	n	y	-	-	100%	100%	100%
INL-corp Gysseling	0,01	0,01	+	y	y	50%	75%	100%	100%	100%
INL-corp oud-nederl	<0,01	<0,01	-	y	y	50%	75%	100%	100%	100%
INL-Parole	0,06	0,06	++	y	y	50%	100%	100%	100%	100%
NTU, Corp. gespr Ned.	0,03	0,03	0	n	y	-	-	100%	100%	100%
NTU, Corea	<0,01	<0,01	0	n	n	-	-	75%	100%	100%
NTU, Lassy Groot	7	7	0	n	y	-	25%	50%	75%	100%
NTU, Daeso	<0,01	<0,01	0	n	y	-	-	75%	100%	100%
NTU, DPC	0,03	0,03	0	n	y	-	-	100%	100%	100%
NTU, D-Coi	0,15	0,15	+	n	y	-	50%	100%	100%	100%
NTU, SoNar, JASMIN	0,7	1,4	+	n	y	-	25%	50%	75%	100%
RUN -hist taaldata 14de	<0,1	<0,1	+/-	y	y	-	25%	50%	75%	100%
RUN -hist taaldata 15de-17de	0,1	0,2	+/-	y	y	-	25%	50%	75%	100%
Utlb., D-Tuna	<0,01	<0,01	0	n	y	-	-	100%	100%	100%
TU Eindhoven, Corpus	<0,01	<0,01	0	n	y	-	-	100%	100%	100%
UAntw, SumNL	<0,01	<0,01	0	n	y	-	100%	100%	100%	100%
UvA-varia	<0,01	<0,01	0	y/n	0	-	25%	50%	75%	100%
VU-varia	<0,01	<0,01	+	0	0	-	25%	50%	75%	100%
UL-hist brieven	0,02	0,02	+/-	y	y	-	25%	50%	75%	100%
Nat. A. (UL/MI)Sailing Letters	0,1	0,2	+/-	y	y	-	25%	50%	75%	100%
RUG-varia	<0,01	<0,01	0	0	0	-	25%	50%	75%	100%
UU-varia	2,5	2,8	--	n	n	-	25%	50%	75%	100%
RUN- varia	<0,01	<0,01	0	0	0	-	25%	50%	75%	100%
Fryske Akad.: Fr Lang database	0	0	+	n	y	-	50%	100%	100%	100%
It Wumkes	0,2	0,3	--	n	y	-	100%	100%	100%	100%
Beeld&Geluid, transcripties	0	0	0	0	0	0	0	0	0	0
SMLC-Diplomatica Belgica (before 1250)	0	0	0	0	0	0	0	0	0	0
SMLC-Narrative Sources (before 1500)	0	0	0	0	0	0	0	0	0	0

legenda:

column 'quality': ++ (text accurate, encoded in XML/TEI), + (text accurate, encoded in xml), +/- (text accurate, structured but not available in xml), - (text generated with ocr, but with some structural encoding), - (text generated with ocr, with no structural encoding)
0 = to be assessed

overzicht van te verwerken metadata

databases/catalogues/bibliographies for historical sources	Estimated size 01/01/2013 in records	Estimated size 31/12/2017 in records	Quality of metadata	Reprocessing?	year I	year II	year III	year IV	year V
DBNL	100 000	200 000	++	n	100%	100%	100%	100%	100%
KB-newspapers	75 000	125 000	++	n	100%	100%	100%	100%	100%
KB-EDBO	10 000	10 000	++	y	50%	100%	100%	100%	100%
KB-ProQuest	10 000	50 000	++	y	-	40%	60%	80%	100%
KB-Google	160 000	160 000	++	y	-	40%	60%	80%	100%
KB-stat gen	0	0	+/-	y	-	30%	50%	80%	100%
KB-metam-tijdschr	4500	4500	+/-	n	-	100%	100%	100%	100%
KB-Linschotenver.	100	100	+/-	y	-	100%	100%	100%	100%
KB-ANP-nieuwsb	1 500 000	1 500 000	0	n	-	100%	100%	100%	100%
KB-Metam./Beh.Ned Bp	450 000	>450 000	++	y	-	40%	60%	80%	100%
KB-CBKinderboeken	26 000	30 000	++	y	-	40%	60%	80%	100%
UGent-Google	300 000	300 000	++	y	-	40%	60%	80%	100%
MI-CRM 14de eeuw	0	0	++	y	-	100%	100%	100%	100%
MI-gTRP/MAND/SAND	0	0	++	y	-	100%	100%	100%	100%
HI-ING (historici.nl)	1000	2000	++	y	-	100%	100%	100%	100%
INL-27 milj wrdn	0	0							
INL-5 milj wrdn	0	0							
INL-38 milj woorden	0	0							
INL-corp Gysseling	3000	3000	+/-	y	-	100%	100%	100%	100%
INL-corp oud-nederl	0	0							
INL-Parole	0	0							
NTU Corp. gespr Ned.	0	0							
Ned. Taalunie, Corea	0	0							
Ned. Taalunie, Lassy Gr.	0	0							
Ned. Taalunie, Daeso	0	0							
UTilb., D-Tuna	2400	2400	+/-	y	-	100%	100%	100%	100%
Ned. Taalunie, D-Coi	0	0							
Ned. Taalunie, SoNar	0	0							
RUN –hist taaldata 14de	0	0							
RUN –hist taaldata 15de-17de	0	0							
KULeuven et al, DPC	0	0							
TU Eindhoven, Corpus	0	0							
UAntw, SumNL	0	0							
UvA-varia	0	0							
VU-varia	0	0							
UL-hist brieven	0	0							
NA-Sailing Letters	10 000	38 000	+/-	y	-	100%	100%	100%	100%
RUG-varia	0	0							
UU-varia	17 000	17 000	++	n					
RUN- varia	0	0							
Fryske Akad.: Fr Lang database	0	0							
It Wumkes	1000	1500	+/-	y	-	100%	100%	100%	100%
Beeld en Geluid, metadata	0	0							
SMLC-Diplomatica Belgica (before 1250)	0	0							
SMLC-Narrative Sources (before 1500)	0	0							

legenda:

- ++ metadata present to supply metadata necessary for XML container, with sufficient bibliographical metadata to be able to connect digital text to the original printed work.
- +/-: metadata present but insufficient to infer exact source of texts (holds for contributions to journals, for instance)

Appendix 3. Overzicht van de infrastructuur- en toolprogramma's die in deze aanvraag zijn vermeld

CATCH en CATCHPlus: Binnen CATCH ontwikkelen technologen, geesteswetenschappers en de erfgoedsector generieke methoden en technieken (tools) op het snijvlak van geesteswetenschappen en informatica ten dienste van en in interactie met collectiebeherende instellingen.
(<http://www.catchplus.nl/en>)

CLARIAH: Aanvraag als onderdeel van de Nationale Roadmap voor Grote Onderzoeksfaciliteiten (via NWO) met als doel een infrastructuur te ontwerpen voor de hele e-humanities, inclusief audiovisuele bronnen, muziek, beeld, etc. CLARIAH zal voortbouwen op de resultaten van CLARIN. DANS en CLARIN zijn de hoofdaanvragers. (http://www.clarin.nl/system/files/CLARIAH_Brief_Aanvraag_Progress.pdf)

CLARIN-EU: CLARIN (Common Language Resources and Technology Infrastructure) is opgericht om een geïntegreerde, interoperable Europese onderzoeksinfrastructuur van taalbronnen te ontwerpen en de technologieën nodig om deze bronnen te onderzoeken. In 2011 is de opbouwfase van de European Research Infrastructure Consortium (ERIC) gestart, die is gevestigd in Nederland. ERIC wordt gefinancierd door de aangesloten landen en door de Europese Unie. (<http://www.clarin.eu/external/>)

CLARIN-NL: Lopend project om een Common Language Resources and Technology Infrastructure (corpora, datasets, tools, standaardisatie, virtuele werkruimten, duurzame toegang, etc.) in Nederland te realiseren. Speciaal bedoeld voor onderzoek naar taal en tekst, maar ook met gesproken taal wordt rekening gehouden. CLARIN is niet beperkt tot taalkunde, maar is bestemd voor ieder onderzoeksgebied waarin taaltechnologische tools en/of bronnen voor onderzoeksdoeleinden worden ingezet. Clarin loopt in principe tot 2014, en wordt gefinancierd door de Nederlandse regering via NWO. (<http://www.clarin.nl/>)

DANS: Instituut dat is opgezet door KNAW en NWO om Data Archiving and Networked Services te verlenen; DANS bevordert duurzame toegang tot digitale onderzoeksgegevens.
(<http://www.dans.knaw.nl/en>)

IMPACT: IMPACT (Improving Access to Text) is gestart in 2008 met subsidie van de Europese Commissie en wordt gecoördineerd door de Koninklijke Bibliotheek in Nederland. Het doel is om de toegang tot historische teksten significant te verbeteren door innovatie in de ocr-technologie en om de belemmeringen weg te nemen die de massadigitalisering van het Europese culturele erfgoed bemoeilijken. (<http://www.impact-project.eu/>)

NEDERLAB: Huidige NWO-groot aanvraag met als doel een virtueel diachroon digitaal corpus van Nederlandstalige data/teksten te bieden in een onderzoeksomgeving, samen met de tools die nodig zijn om onderzoek op dit corpus te verrichten. Nederlab is met name bedoeld voor het opsporen van veranderingen in de taal en de cultuur in de loop van de tijd, hoewel het corpus uiteraard ook voor andere soorten onderzoek geschikt zal zijn. Nederlab is meer een toepassingsprogramma dan een infrastructuurprogramma, en zal uitgaan van de CLARIN-standaarden.
(http://www.diachronie.nl/index.php?option=com_content&view=article&id=45&Itemid=20)

SARA/BigGrid: SARA is het onafhankelijke nationale supercomputercentrum voor het wetenschappelijk onderzoek en onderwijs in Nederland. Het biedt een geïntegreerde ICT-onderzoeksinfrastructuur en levert diensten voor dataopslag op basis van de BigGrid Infrastructure. (<http://www.sara.nl/>;
<http://www.biggrid.nl/>)

SURF: SURF is de ICT-samenwerkingsorganisatie voor het hoger onderwijs en onderzoek. SURFnet zorgt ervoor dat onderzoekers, docenten en studenten eenvoudig en krachtig samen kunnen werken met behulp van ICT. SURF stelt onder andere standaarden voor autorisatie en authenticatie vast.
(<http://www.surf.nl/>)

TLA: The Language Archive is een onderdeel van het Max Planck Institute for Psycholinguistics (MPI) in Nijmegen. Het hoofddoel is digitale taalbronnen op te slaan en te bewaren, toegankelijk te maken voor onderzoekers en andere geïnteresseerde gebruikers, en nieuwe technologieën ter bevordering van taalonderzoek te ontwikkelen en te integreren. (<http://www.mpi.nl/research/research-projects/the-language-archive>)