

# 'n Oorsig oor enkele digitiseringsprojekte in Afrikaans: 'n Vertrekpunt vir Afrikaans se digitale toekoms

Roné Wierenga en WAM Carstens

---

Roné Wierenga, Virtuele Instituut vir Afrikaans (VivA) en Navorsingseenheid: Verstaan en Verwerking van Taal in Komplekse Omgewings, Fakulteit Geesteswetenskappe, Noordwes-Universiteit  
WAM Carstens, Navorsingseenheid: Tale en literatuur in die SA konteks, Fakulteit Geesteswetenskappe, Noordwes-Universiteit

---

## *Opsomming*

Grootskaalse digitisering is 'n wêreldwye tendens van die huidige dekade. Ook in Suid-Afrika word al hoe meer digitiseringsprojekte van stapel gestuur. Hierdie artikel het ten doel om te rapporteer oor van die digitiseringsprojekte wat tans deur die Noordwes-Universiteit en die Virtuele Instituut vir Afrikaans (VivA) onderneem word. Hierdie projekte, naamlik die Digitale Bibliografie van die Afrikaanse Taalkunde, Digitale Bibliografie van die Afrikaanse Letterkunde, Historiese Korpus vir Vroeë Afrikaans (1675–1925) en die Argief vir Historiese Afrikaanse Bronne is slegs 'n spreekwoordelike druppel in die emmer met betrekking tot die digitisering van Afrikaans as taal. Dit sluit wel aan by ander bestaande digitiseringsprojekte wat onder meer deur die Wes-Kaapse Argief, universiteitsbiblioteke (soos die Universiteit van Pretoria, Universiteit Stellenbosch en Noordwes-Universiteit se biblioteke), NALN, SADIaR en Nuuseum onderneem is. Die omvang en aard van hierdie projekte is egter nie altyd duidelik of bekend nie.

Dié artikel dien as bekendstelling en ook vorderingsverslag aan persone en instansies wat 'n behoefte daaraan het om saam te werk aan die digitiseringsproses om sodoende gedeelde en afsonderlike uitkomst te bereik. Hierdie artikel het ten doel om die projekte wat hierin uiteengesit word, voor te stel as 'n moontlike vertrekpunt vir die ontwikkeling van soortgelyke digitiseringsprojekte in ander tale. 'n Aantal van die uitdagings wat in hierdie digitiseringsprojekte teëgekom is, word ook kortliks aangebied, omdat hierdie uitdagings ook vir ander tale hindernisse kan wees.

**Trefwoorde:** argief; Argief vir Historiese Afrikaanse Bronne; Digitale Bibliografie van die Afrikaanse Letterkunde (DBAL); Digitale Bibliografie van die Afrikaanse Taalkunde (DBAT); Digitale Biblioteek vir Afrikaans; digitalisering; digitisering; Historiese Korpus vir Vroeë Afrikaans; korpus; Virtuele Instituut vir Afrikaans (VivA)

### *Abstract*

#### **An overview of a few digitisation projects in Afrikaans: The first steps towards a digital future for Afrikaans**

Large-scale digitisation has become a global trend during the current decade. More and more digitisation projects are also being launched in South Africa. Organisations and institutes like the Western Cape Archives, the library services of some South African universities (such as the University of Pretoria, Stellenbosch University, North-West University), NALN, SADiLaR and Nuuseum, among others, are all involved in projects focused on expanding the body of South African texts that are available digitally. However, the extent of these projects, especially regarding the digitisation of Afrikaans texts, is not clear. In fact, the digitisation of Afrikaans sources currently leaves much to be desired. These projects do not necessarily share a common goal, since some of these projects are centred on the preservation of historical documentation or a language form, while others hope to promote research in a particular field or make information more accessible.

This article aims to report on some of the digitisation projects currently being undertaken by North-West University and the Virtual Institute for Afrikaans (VivA). These projects, namely the Digital Bibliography of Afrikaans Linguistics, Digital Bibliography of Afrikaans Literature, Historical Corpus for Early Afrikaans (1675–1925) and the Archive for Historical Afrikaans Sources, are only a proverbial drop in the bucket with regard to the digitisation of Afrikaans as a language.

The terms *digitisation*, *digitalisation*, *digital archive*, *digital library* and *digital corpus* are defined in this article with a focus on discerning *digitisation* (the process by which soft copies are developed from hard copies, usually by means of scanning documents or taking pictures of items) from *digitalisation* (the process by which soft copies are encoded, commonly using optical character recognition [OCR] software). Furthermore, the distinction is made between a *digital archive* or *digital library*, where documents are housed and stored in soft copy, usually as PDFs, and *digital corpora*, where the data from the soft copies are encoded, searchable, statistically analysable, and usually stored in .txt format.

These distinctions and terminological clarifications are presented as a means of understanding how a series of smaller digitisation projects have been utilised as an attempt at large-scale digitisation. The first digitisation project outlined in this article is the Digital Bibliography of Afrikaans Linguistics (Digitale Bibliografie van die Afrikaanse Taalkunde [DBAT]), which was initially based on the Digital Library for Dutch Literature (Digitale Bibliotheek voor de Nederlandse Letteren [DBNL]). Work on the DBAT started in 2004 and the digital database was formally launched in 2008. The DBAT was established to provide national and international linguistic researchers with a database containing the bibliographic information of all the available literature on Afrikaans linguistics as a means of aiding researchers in finding literature for

research. By 2015, a Digital Bibliography of Afrikaans Literature (Digitale Bibliografie van die Afrikaanse Letterkunde [DBAL]) was launched to perform the same function for literary researchers.

Whilst these two databases aided researchers in identifying the available literature on a topic, and in which library these resources are housed, users increasingly expressed the need to gain access to these resources digitally. In order to address this need, DBAT and DBAL initiated digitisation processes, thereby allowing users to gain access to full-text digital copies of resources where possible.

In 2019, the Virtual Institute for Afrikaans (VivA) started working on a digital Historical Corpus for Early Afrikaans (1675–1925). This project aimed to both preserve historical texts written in Early Afrikaans and provide a source of data for linguistic research. Unfortunately, the pandemic of 2020 prevented visits to libraries and archives to acquire and digitise the historical texts. As a result, the decision was made to focus efforts on digitising the texts that were readily available and make these texts accessible online by establishing a Digital Archive for Historical Resources.

This article serves as an introduction and progress report of the digitisation projects that are underway, and invites persons and institutions interested in digitisation to work together towards achieving shared and separate outcomes. This article aims to present the projects detailed herein as a possible starting point for the development of similar digitisation projects in other languages. A number of the challenges encountered in these digitisation projects are also briefly presented, because these challenges can be obstacles for other languages as well. These obstacles include access to texts and resources such as financial support, human capacity, and the necessary technology to perform the digitisation and digitalisation processes. The lack of a collective vision regarding the digitisation of Afrikaans texts, and eventually a Digital Library of Afrikaans, also hampers progress. Challenges like the duplication of work by different institutions as part of their digitisation projects are also outlined, since these challenges translate to the digitisation processes of other South African languages.

**Keywords:** Archive for Historical Afrikaans Resources; Digital Bibliography for Afrikaans Literature (DBAL); Digital Bibliography for Afrikaans Linguistics (DBAT); Digital Library for Afrikaans; digitisation; digitalisation; Historic Corpus for Early Afrikaans; Virtual Institute for Afrikaans (VivA)

## 1. Inleiding

'n Wêreldwye tendens om boeke en bronne in die algemeen te digitiseer, het opnuut vlam gevat in die laaste twee dekades (vergelyk Jamali 2008; Bor Ng en Kucsma 2010; Borgman 2010; De Beer en De Beer 2016; Roux 2020). Coyle (2006:641) stel dat “the Open Content Alliance (OCA), a library initiative formed after Google announced its library book digitization project, has brought library digitization projects into the public eye [...]”

Die digitisering van tekste is egter nie 'n nuwe verskynsel nie; Project Gutenberg digitiseer tekste reeds sedert 1971.<sup>1</sup> Dit is wel betreklik nuut uit 'n Afrikaanse oogpunt. Universiteits- en ook ander biblioteke in Suid-Afrika eksperimenteer die laaste dekade reeds met die digitisering

van boeke (vergelyk Breed, Carstens en Olivier 2016 rakende die DBAT, asook Senekal 2011 en 2012 rakende NALN). Die konstante ontwikkeling van nuwe en verbeterde tegnologie (sien Bor en Kucsma 2011; Borgman 2010; Glorie 2015; De Beer en De Beer 2016; Roux 2020) het tot gevolg dat digitisering op 'n groter skaal kan plaasvind en dat die gedigitiseerde produk van beter gehalte is. Hierdie ontwikkeling het massadigitisering tot gevolg.

Massadigitisering behels egter meer as slegs grootskaalse digitisering. Massadigitisering verwys eerder na digitisering op 'n industriële skaal, dit wil sê na die digitisering van 'n hele argief of biblioteek sonder om tekste oor te sien (Coyle 2006:641). Die doel van massadigitisering is dus nie om versamelings tekste te digitiseer nie, maar eerder om die inhoud van 'n instituut, soos 'n biblioteek of 'n taal in soverre dit haalbaar is, te digitiseer. Nederlands is 'n goeie voorbeeld van 'n taal wat al ver op hierdie pad gevorder het, soos van die versamelings in die Koninklijke Bibliotheek getuig.<sup>2</sup> Massadigitisering is 'n duur en tydrowende proses wat dikwels nie haalbaar is vir kleiner instansies of individue nie en dit is gevolglik 'n proses wat met omsigtigheid aangepak moet word. Die impak op hulpbronne (begrotings en die werkklas van personeel) kan groot wees en dit moet van die begin af deeglik verreken en veral beplan word.

Die teenstuk vir massadigitisering is die grootskaalse digitisering van bepaalde versamelings tekste. Hierdie benadering is die basis vir die bekende Digitale Bibliotheek voor de Nederlandse Letteren (DBNL).

Voor literaire werke is de negendelige reeks *Geschiedenis van de Nederlandse literatuur* het uitgangspunt. De collectievorming richt zich daarmee dus op literair erfgoed uit Nederland en Vlaanderen, aangevuld met secundaire literatuur. (DBNL 2021)

Die DBNL bestaan uit 'n aantal versamelings, insluitend wetenskaplike studies, leksikon en literatuurgeskiedenis, asook taalkunde-versamelings. Sodoende word die digitiseringsproses afgebreek tot kleiner versamelings waarvan die digitisering haalbaar is met betrekking tot die hulpbronne wat die DBNL op 'n gegewe tydstip tot hul beskikking het (of gehad het).

Die digitiseringsprosesse wat in Afrikaans aangepak word, volg oorwegend dieselfde model na as wat vir die DBNL gebruik word.<sup>3</sup> 'n Verskeidenheid digitiseringsprojekte wat ten doel het om Afrikaanse tekste in digitale formaat beskikbaar te stel, is reeds gevestig, soos beskryf in Senekal (2011) rakende pogings om so 'n vroeë proses by die Nasionale Afrikaanse Letterkundige Museum en Navorsingsentrum (NALN) in Bloemfontein aan die gang te kry, asook die pogings sedert 1994 om 'n digitale rekord van taalkundige publikasies in Afrikaans tot stand te bring (sien Breed e.a. 2016).

Vier van die toepaslike projekte word in hierdie artikel uitgelig, naamlik die Digitale Bibliografie van die Afrikaanse Taalkunde (DBAT), die Digitale Bibliografie van die Afrikaanse Letterkunde (DBAL), die ontwikkeling van 'n Historiese Korpus vir Vroeë Afrikaans (1675–1925), en die stigting van 'n Argief vir Historiese Afrikaanse Bronne. Hierdie projekte spesialiseer in die digitisering van versamelings Afrikaanse tekste met die doel om gesamentlik 'n massadigitiseringsproses vir Afrikaans in werking te sit, sodat die taal in soverre dit wel haalbaar is in digitale formaat behou en benut kan word. Die beplande langtermyn doelwit is die uiteindelijke totstandkoming van 'n Digitale Bibliotheek van Afrikaans (DBA).

Dit is belangrik om te meld dat die opteken van bibliografiese inligting oor publikasies in die taal- en letterkunde al 'n ou praktyk is, soos blyk uit die talle bronnegidse wat oor die jare in Afrikaans oor Afrikaanse publikasies verskyn het (sien byvoorbeeld Kettley 1970; Nienaber 1943–1981; Nienaber en Nienaber 1947–1968; Marx 1992; Sandrock 1980). Dit was ook die patroon wat lank in die Nederlandse BNTL nagevolg is. Hierdie werk moet geloof word, omdat dit gedokumenteer het wat op bepaalde tydstippe in die betrokke taal (ook in Afrikaans) gepubliseer is in die verskeidenheid publikasieformate. Die beskikbaarheid van hierdie soort inligting het egter wel 'n beperkte gebruikswaarde, omdat die bron self nie direk vanuit die bibliografie geraadpleeg kon word nie. In so 'n geval word daar dan meestal slegs kennis geneem van 'n bron en word die inhoud van die bron self buite rekening gelaat. Die oorgang na 'n digitale formaat in 'n nuwe tegnologiese era maak derhalwe volledig sin, omdat dit toegang tot die bron self en die ontsluiting van die inhoud daarvan moontlik maak.

## 2. Digitisering in Afrikaans

### 2.1 Terminologiese verklarings

In die literatuur word verskillende opvattinge oor wat *digitisering* en *digitalisering* behels, aangetref (sien Van Huyssteen 2019; Rijswijk 2020). 'n Aantal terme wat verband hou met digitisering en digitalisering kom aan bod in hierdie artikel en word dus hier verklaar, onder meer (a) databasis, (b) digitale biblioteek, (c) digitale argief, (d) digitale korpus, (e) digitale bibliografie en (f) enkodering.

Die term *digitisering* kan gebruik word om te verwys na die implementering van digitale tegnologieë op 'n besigheidsvlak (Rijswijk 2020), na die derde industriële revolusie wat in die 1960's posgevat het en tot gevolg gehad het dat rekenaars algemeen gebruiklik geraak het (Schwab 2017), of na 'n proses waartydens analoog data omgeskakel word in 'n digitale vorm. Vir die doeleindes van hierdie artikel word die term *digitisering* gebruik om te verwys na 'n proses waardeur 'n harde kopie van 'n teks of ander bron omgeskakel word in 'n sagte kopie of digitale kopie (Van Huyssteen 2019). Die digitiseringsproses behels gevolglik die beskikbaarstelling van inligting in 'n digitale formaat.

Digitisering, in hierdie sin, gaan dikwels 'n digitaliseringsproses, dit wil sê die proses waardeur 'n digitale kopie van 'n bron geënkodeer word deur gebruik te maak van optiese karakterherkenningsprogrammatuur (OKH-programmatuur), vooraf. *Digitalisering* is 'n meer gespesialiseerde proses wat afhanklik is van spesiale programmatuur en dit het ten doel om die inhoud van 'n digitale bron verder te ontsluit en bruikbaar te maak vir ander kontekste. Die term *digitalisering* kan egter ook gebruik word om te verwys na sosio-tegniese prosesse wat die gebruik van 'n groot aantal tegnologieë in die alledaagse lewe tot gevolg het, onder meer 'n toenemende afhanklikheid van digitale tegnologieë (Tilson, Lyytinen en Sørensen 2010).

Die terme *digitisering* en *digitalisering* het deurgaans in hierdie artikel betrekking op die ontwikkeling van digitale produkte vanuit gepubliseerde hardekopietekste. Dit is daarom nuttig om aan te dui hoe hierdie digitale produkte van mekaar verskil. 'n *Digitale argief* is 'n virtuele stoorplek vir data wat op 'n eenvoudige wyse gekatalogiseer is. Die doel van 'n digitale argief is bloot om inligting te stoor sodat dit maklik weer onttrek kan word vir gebruik. 'n *Digitale*

*databasis* se inhoud is egter meer gestruktureerd as dié van 'n digitale argief en 'n groter verskeidenheid soekfunksies kan daarin uitgevoer word.

'n *Digitale bibliografie* is 'n databasis wat bestaan uit 'n versameling bibliografiese inligting van publikasies, terwyl 'n *digitale biblioteek* 'n databasis is wat uit elektroniese kopieë van tekste bestaan. 'n Toenemende tendens word egter waargeneem waarin digitale bibliografieë uitgebrei word om ook die elektroniese kopieë van tekste te bevat en nie slegs 'n verwysing na die bibliografiese besonderhede van die teks nie, soos geïllustreer word met die proses wat gevolg is om die DBAT tot stand te bring. Hierdie tendens word gedryf deur die toenemende beskikbaarheid van toegang tot elektroniese dokumente in ooptoegangjoernale of -vaktydskrifte en digitiseringsprojekte by nasionale en universiteitsbiblioteke. Die gevolg is dus dat hierdie digitale bibliografie stelselmatig tot digitale biblioteek ontwikkel.

## **2.2 Die Digitale Bibliografie van die Afrikaanse Taalkunde (DBAT)**

Die DBAT het sy ontstaan te danke aan die DBNL. Die ontstaan en ontwikkelingsproses van die DBAT word volledig in Breed e.a. (2016) uiteengesit; daarom word daar slegs 'n oorsig van die ontstaansproses in hierdie artikel aangebied. Die idee vir die ontwikkeling van die DBAT het vorm gekry toe prof. P.C. Paardekooper (1920–2013), vroeër verbonde aan die KU Leuven se Kortrijk-kampus in België, in 1993 die behoefte vir 'n bibliografie vir Afrikaanse taal- en letterkunde aan prof. Wannie Carstens, verbonde aan die Noordwes-Universiteit (NWU) in Suid-Afrika, uitgespreek het. Carstens het met verloop van tyd danksy die hulp van beperkte finansiële steun uit enkele bronne (primêr die Fonds Neerlandistiek Suider-Afrika asook die PUK-Kanselierstrust) asook die beskikbaarstelling van steun deur die NWU, die idee tot 'n realiteit ontwikkel.

Die eerste fase van die projek (1993–2003) het primêr gefokus op die verkryging van befondsing, die kry van uitsluitel oor die formaat van die beplande versamelings bronne, die besluit oor welke bronne opgeneem moes word en ook hoe dit opgeneem moes word, die vestig van 'n infrastruktuur (personeel en werkruimte) en die identifisering van 'n instansie wat die projek kan huisves. Alhoewel die oorspronklike behoefte was om ooreenstemmend aan die meerdelige Bibliografie van de Nederlandse Taal- en Literatuurwetenschap (BNTL) 'n omvattende Bibliografie vir die Afrikaanse Taal- en Letterkunde (BATL) te ontwikkel, moes die projek afgebaken word tot die Afrikaanse taalkunde om die haalbaarheid daarvan te verseker. Daar is destyds wel onderling ooreengekom dat prof. Dorothea van Zyl van die Universiteit Stellenbosch mettertyd die leiding sou probeer neem met 'n letterkundige ekwivalent van die taalkundeprojek. Aanvanklike pogings in die laat 1990's om die Nasionale Afrikaanse Letterkundige Museum en Navorsingsentrum (NALN) in Bloemfontein te betrek om hierdie literêre taak te onderneem, het ongelukkig weinig konkrete gedigitiseerde resultate tot gevolg gehad, grootliks weens bestuursprobleme en die beskikbaarheid van kundige en voldoende personeel aan die kant van NALN om die taak te verrig.

Die DBAT het vervolgens begin met die ontwikkeling van 'n taalkunde-databasis wat beperk was tot die beperkte versameling taalkundige tekste wat by NALN in Bloemfontein beskikbaar was. Carstens het verder gebruik gemaak van bronne oor aspekte van die Afrikaanse taalkunde wat op die rakke van die universiteitsbiblioteek van die destydse PU vir CHO in Potchefstroom (deesdae Noordwes-Universiteit) beskikbaar was, asook sy eie versameling taalkundige bronne. In hierdie proses is die optekenformaat wel uitgeklaar (op basis van die model van die destydse

BNTL), en daar kon met behulp van assistente (nagraadse studente of kontrakaanstellings) ’n aanvang gemaak word met die versamel van die bronne wat opgeteken moes word.

Die tweede fase (2004–2010) het gepaardgegaan met formele samewerking met die Ferdinand Postma Biblioteek (FPB) van die voormalige PU vir CHO, waar die databasis danksy finansiële steun van die biblioteek op die universiteit se bediener gehuisves kon word (sien Carstens en Van Zyl 2004). Danksy die aanbod om van die biblioteek se bediener gebruik te maak, kon daar wegbeweeg word van die aanvanklike plan om ’n hardekopie-bibliografie soos die BNTL tot stand te bring, en kon daar van die universiteit se digitale omgewing gebruik gemaak word vir die dokumentering van die DBAT. In hierdie fase is daar oorwegend gefokus op die katalogisering van die tersaaklike tekste – hoofsaaklik nagraadse studies (proefskrifte en verhandelinge), akademiese artikels uit Afrikaanse vaktydskrifte, resensies en koerantartikels (veral uit ouer tydskrifte, onder meer *Die Huisgenoot*, *Brandwag*, *Ons Moedertaal*, *Handhaaf*, *Ons eie boek*, *Standpunte* en die *Tydskrif vir Volkskunde en Volkstaal*). Hierdie katalogus is deurgaans bygewerk en uitgebrei soos hulpbronne dit toegelaat het.

In 2010 is die DBAT amptelik aanlyn aan die NWU bekendgestel met die hoop dat dit vir nasionale en internasionale navorsers die geleentheid sal bied om hulself te vergewis van die taalkundebronne wat in Afrikaans beskikbaar is en ook waar en in welke formaat hierdie bronne geleë is (sien Carstens 2010).<sup>4</sup> Uit terugvoer aan die projekteiers het dit spoedig geblyk dat die digitale beskikbaarstelling van inligting (te wete bibliografiese besonderhede, voltekste van bronne) oor taalkundepublikasies in Afrikaans goed byval by gebruikers in Suid-Afrika asook in ander lande gevind het, grotendeels weens die maklike elektroniese beskikbaarheid van Afrikaanse materiaal wat tot op daardie stadium moeilik bekombaar was.

Dit het dus moontlik geword vir taalkundiges in die buiteland om afgesien van die normale verspreidingskanale van akademiese uitsette ook insae te kon kry in die aard en omvang van taalkundige bronne in Afrikaans en so kon daar internasionaal kennis geneem word van die akademiese bydraes (artikels, boeke en nagraadse studies) van Afrikaanse taalkundiges en van afgehandelde en lopende navorsingsprojekte deur nagraadse studente. Die voordeel hiervan is dat dit gehelp het om Afrikaans as deel van die “globale inligtingsnetwerk” (Senekal 2011) te vestig. Hiermee word steeds volgehou.

Tans (Julie 2022) bevat die DBAT (onder beheer van ’n voltijdse personeelid van die NWU se Skool vir Tale asook enkele assistente) meer as 18 000 inskrywings, waarvan 60% in volteks in digitale formaat beskikbaar is. Die DBAT-databasis word weekliks bygewerk en is byna volledig op datum. Al die taalkundige publikasies oor of in Afrikaans wat opgespoor kan word in hardekopie- of e-kopie-formaat word by die databasis ingesluit. Ook publikasies oor aspekte van die Afrikaanse taalkunde wat in ander tale gepubliseer word (soos Engels, Nederlands en Duits) word opgeneem waar dit onder die aandag van die DBAT kom. ’n Vasgestelde lys wat bestaan uit 200 soekstringe, trefwoorde en kategorieë word gebruik om taalkundepublikasies na te spoor op ’n verskeidenheid soekenjins en databasisse, insluitend Google Scholar, JSTOR en ResearchGate. Die lys soekstringe, trefwoorde en kategorieë sluit onder meer in: (a) taalkundige dissiplines soos sintaksis, fonetiek en morfologie; (b) teorieë soos grammatikaliseringsteorie en konstruksiegrammatika; (c) benaderings soos toegepaste taalkunde; (d) navorsingsmetodes soos korpusgebaseerde navorsing of gebruikgebaseerde navorsing; en (e) navorsingsvelde soos Afrikaans in onderwys en onderrig, die wisselwerking tussen Afrikaans en Engels, en die internasionalisering van Afrikaans.

Die prominentste plaaslike vaktydskrifte soos *Tydskrif vir Geesteswetenskappe*, *Tydskrif vir Nederlands en Afrikaans*, *LitNet Akademies*, *Literator*, *SPiL Plus* en *Language Matters*, en buitelandse tydskrifte soos *Internationale Neerlandistiek*, *Journal of Germanic Studies* en talle ander, word deurgaans nagegaan sodra 'n nuwe uitgawe gepubliseer word. Elektroniese kopieë van hierdie vaktydskrifte en joernale word sover moontlik op DBAT beskikbaar gestel. Tog gebeur dit dat die personeel nie al die tersaaklike publikasies in hul soektogte opspoor nie en dat publikasies soms oorgesien word. Om hierdie rede is daar 'n elektroniese vorm op DBAT se webwerf beskikbaar aan die hand waarvan gebruikers publikasies (ook hulle eie) wat nie in die databasis voorkom nie, kan aanmeld sodat dit ingesluit kan word. In hierdie opsig word weer eens gebruike van die DBNL nagevolg.

Die DBAT<sup>5</sup> fokus tans op die digitisering van tekste, met die doel om nie slegs die bibliografiese inligting aan te bied nie, maar ook die digitale kopie van die teks (volteks) vir gebruikers beskikbaar te stel. Hierdie proses is egter arbeidsintensief (wat vanselfsprekend die koste van die projek raak) en tydrowend. Daarom is al die bibliografiese inskrywings waarvoor die nodige kopieregtoestemming ontvang is nog nie in digitale formaat beskikbaar nie, en die digitale kopieë wat wel beskikbaar is nog nie digitaal deursoekbaar nie. Anders as by die DBNL het die DBAT steeds nie by die digitaliseringsproses (dit wil sê die enkodering van digitale formate soos PDF's en JPEG's deur middel van OKH-programmatuur) uitgekom nie. In daardie opsig is die DBAT nog nie so ver gevorderd soos die DBNL nie, alhoewel die DBAT (18 000 opnames) die DBNL (17 500 opnames, soos weergegee in DBNL 2021) in omvang oorskry.

Coyle (2006:641) voer aan die mees ekonomiese en tydeffektiewe wyse om massadigitisering uit te voer, is om die tekste bladsy vir bladsy te skandeer of te fotografeer, en daarna die digitale weergawe deur OKH-programmatuur te laat prosesseer sodat die dokumente elektronies deursoekbaar is. Hierdie prosesse, naamlik die skandering en die digitalisering, kan egter gelyktydig uitgevoer word, maar in DBAT se geval het die beperkte hulpbronne tot gevolg dat die skanderingproses eers grotendeels afgehandel sal moet word voordat die digitaliseringsproses in 'n groter omvang onderneem kan word. Dit is die geval omdat die skandering van tekste deur die huidige personeel en binne die begroting van DBAT uitgevoer kan word, terwyl die digitaliseringsproses nie noodwendig uitvoerbaar is sonder verdere opleiding, samewerking met kundiges / kundige instansies of die aanstelling van derde partye nie. Daar is reeds ondersoek ingestel na die gebruik van bestaande OKH-programmatuur vir Afrikaans en die omvang van die digitaliseringsproses, insluitend uitdagings wat voorsien word (onder meer dat bestaande OKH-programmatuur nie volledig akkuraat is nie en tekste gedeeltelik handmatig nagegaan sal moet word om enige onakkuraathede reg te stel), kostes verbonde aan die proses en die tydskuur van so 'n onderneming is reeds vasgestel.

Die DBAT is dus (veral) wat sekere aspekte, soos die digitisering en volledigheid van die databasis, betref ver gevorderd, maar het nog nie alle vestigingsprobleme (oorwegend beperkte befondsing en die beskikbaarheid van kundige en voldoende personeel om die werk te kan doen) ontgroeï nie.

### **2.3 Die Digitale Bibliografie van die Afrikaanse Letterkunde (DBAL)**

In 2015 is die letterkundige ekwivalent van die DBAT, naamlik die Digitale Bibliografie van die Afrikaanse Letterkunde (DBAL), bekendgestel (sien De Vries en Linde 2021). Die aanvoorwerk vir hierdie projek is reeds vanaf 2012 gedoen deur Hein Viljoen, ook van die



NWU. Weens 'n tekort aan fondse (om projekassistent te kon kry) het die projek stadig gevorder. Aan die begin is daar wel tyd afgestaan aan die ontwikkeling van die kategorieë waarvolgens literêre publikasies ingedeel sou kon word en by nabaat was dit sinvol, omdat die letterkundige produkte makliker in die onderskeie kategorieë, soos literêre genres en strominge, geplaas kon word as wat die geval met die DBAT was.

Tans (Julie 2022) bestaan DBAL uit 17 500 inskrywings, waarvan slegs 50% van die inskrywings verbind is met die digitale weergawe van die teks en onmiddellik op aanvraag aanlyn gebruik kan word. Anders as DBAT is DBAL nog nie op datum nie<sup>6</sup> en is daar letterkundepublikasies, soos uitgawes van vaktydskrifte, joernale en webwerwe, insluitend Versindaba, wat dateer uit die vroeë 2000's wat nog nie by die databasis ingesluit is nie. Dit is hoofsaaklik omdat DBAL sewe jaar ná DBAT gevestig is, maar ook omdat daar meer literêre tekste, dit wil sê primêre bronne soos digbundels, kortverhaalbundels, dramatekste en romans en sekondêre bronne soos akademiese artikels oor literêre werke, as taalkundige tekste in Afrikaans gepubliseer word.

DBAL word tans op dieselfde wyse as die DBAT uitgebrei, in stand gehou en aangepak. 'n Lys van 348 soekstringe, trefwoorde en kategorieë word gebruik om publikasies op te spoor. Hierdie lys bevat onder meer (a) tekssoorte soos kortverhale, poësie en dramatekste; (b) benaderingswyses soos postkolonialisme, kultuurstudie en narratiewe terapie; (c) literêre instansies, instellings en organisasies soos skrywers, uitgewers en leesgroepe; (d) literêre tradisies soos sprokies, mites en rympies; (e) literêre periodes soos die Twintigers; en (f) vertalings, mediasies en vergelykende literatuurstudies.

Minder van die tekste op DBAL is in elektroniese formate beskikbaar, omdat die fokus vir hierdie databasis op volledigheid is en die notering van reeds gepubliseerde tekste steeds voorkeur geniet. Nadat al die reeds gepubliseerde tekste van letterkundige aard (dit wil sê primêre bronne soos romans, kinderverhale, digbundels en dramatekste sowel as sekondêre bronne soos resensies, akademiese artikels, lesersindrukke en navorsing oor die gebruik en beleving van letterkundige tekste in byvoorbeeld die klaskamer) sover moontlik by DBAL ingesluit is, sal die digitisering- en digitaliseringsprosesse wat in DBAT toegepas is, ook die fokuspunt van DBAL word. Soos die werkwyse by DBAT ook is, gebruik DBAL-personeel 'n inventaris van erkende vaktydskrifte en -joernale as vertrekpunt om reeds gepubliseerde tekste te identifiseer en by die databasis in te sluit. Daar word ook dikwels met biblioteke, outeurs en verenigings en organisasies soos NALN saamgewerk om te verseker dat DBAL so volledig moontlik is. NALN se letterkundeversamelings word byvoorbeeld reeds by DBAL ingesluit ten spyte daarvan dat die voltekste nog tot 'n groot mate ontbreek.<sup>7</sup> Soos die geval by DBAT ook is, is daar 'n elektroniese vorm op DBAL se landingsblad beskikbaar waarop gebruikers bronne wat tans nog nie op DBAL ingesluit is nie, kan aandui, sodat hierdie tekste ook by die DBAL-databasis ingesluit kan word.<sup>8</sup>

Naas hierdie prosesse word ook 'n instandhoudingsproses deurgaans volgehou. Hierdie proses behels die sistematiese nagaan van reeds bestaande inskrywings op DBAL om te bevestig dat voltekste bygevoeg word by bestaande inskrywings soos wat die elektroniese kopieë van tekste beskikbaar gestel word, en skakels na tekste of organisasies aktief bly selfs wanneer 'n elektroniese teks verskuif word of 'n webwerf nie meer beskikbaar is nie.

## 2.4 Historiese Korpus vir Vroeë Afrikaans (1675–1925)

Die ontwikkeling van 'n korpus is beduidend meer kompleks as die daarstelling van 'n databasis soos DBAT en DBAL, omdat die gebruik van 'n korpus dit noodsaak dat alle tekste wat by die korpus ingesluit word ten volle geënkodeer en deursoekbaar is.<sup>9</sup> Slegs tekste oor die Afrikaanse taal- en letterkunde word in die DBAT en DBAL onderskeidelik ingesluit. Die opteken van hierdie tekste is in die lig van die huidige stand van die projek nog aan die gang, dit wil sê dit sluit tekste van 'n omvattende aard in, soos artikels in akademiese vaktydskrifte, artikels in tydskrifte wat vroeër populêr was (soos *Huisgenoot* en *Brandwag*), nagraadse studies (verhandelinge, proefskrifte), en boeke oor besondere temas in die taal- of letterkunde. Die digitale beskikbaarheid van al hoe meer tekste dra daartoe by dat ook toepaslike bydraes uit elektroniese publikasies, soos in LitNet se onderskeie afdelings en ander e-nuusinstansies, ingesluit kan word.

Die versameling tekste wat in die Historiese Korpus vir Vroeë Afrikaans (HKVA) versamel word, is eerder gefokus op taaldata, dit wil sê die taal wat as medium vir die publikasie dien en nie die tema waaroor die publikasie handel nie. Soortgelyke korpora is reeds in Afrikaans ontwikkel, soos die Historiese Korpus van Standaardafrikaans (HKS), wat deur prof. Johanita Kirsten, verbonde aan die Noordwes-Universiteit se Vaaldriehoekskampus, ontwikkel is en die Tracing History Trust-korpus (THT-korpus) wat onder leiding van dr. Helena Liebenberg, verbonde aan die Wes-Kaapse Argief, ontwikkel is.

Die HKS is 'n diachroniese korpus wat gestratifiseer is (dit wil sê gebalanseerd saamgestel is) met die doel om taalverandering tussen 1911 en 2010 te kan ondersoek (vergelyk Kirsten 2016 vir 'n volledige beskrywing van die korpus). Die THT-korpus bestaan weer uit 'n versameling dagboekinskrywings uit die tydperk 1675 tot 1779 wat geskryf is deur lede van die Verenigde Oos-Indiese Kompanjie (VOC) wat in die Nederlandse kolonie aan die Kaap gewoon het. 'n Volledige korpusbeskrywing word in Liebenberg (2018) aangebied.

Alhoewel beide korpora nuttig is vir taalkundige navorsing, is daar steeds bepaalde beperkinge aan hierdie korpora. In die geval van die HKS-korpus is slegs 140 jaar se ongestandaardiseerde Afrikaans ingesluit, terwyl die THT-korpus beperk is tot VOC-geskifte en nie taaldata van ander gemeenskappe wat gedurende hierdie tydperk Afrikaans gepraat het insluit nie. Nog te meer is daar geen taaldata uit die 19de eeu in hierdie korpora nie. Hierdie leemte word gedeeltelik opgelos deur die Corpus of Cape Dutch Correspondence, 'n korpus wat bestaan uit 'n versameling briewe uit 1880 tot 1922 (sien Deumert 2004). Hierdie korpus het egter ook bepaalde beperkinge, omdat dit nie ander tekstipes of registers insluit nie.

Die behoefte aan 'n omvattende en gestratifiseerde historiese korpus is verder aangevuur deur die verlies aan historiese tekste in Suid-Afrika. In sommige gevalle word historiese tekste in biblioteke en ander instansies se kelders gestoor, omdat die tekste in onbruik verval het en die rakke in biblioteke oorvol geword het met nuwer publikasies. Hierdie tekste loop derhalwe die risiko om verpulp te word of beskadig te word in die stoorproses. In ander gevalle bedreig vloede en brande, soos in die geval van die brande in die Jagger-biblioteek van die Universiteit van Kaapstad in April 2021 (sien Niémah 2021), die voortbestaan van hierdie tekste. Dit is derhalwe nodig dat die digitiseringsproses in Afrikaans deeglik beplan en gekoördineerd uitgevoer word ten einde moontlike verliese weens die onbekikbaarheid van Afrikaanse bronne te voorkom.

Die omvang en aard van die historiese korpus is gebaseer op 'n hoofstuk in Edith Raidt se 1991-publikasie, *Afrikaans en sy Europese verlede*, waarin sy 'n oorsig gee oor die historiese bronne wat in en oor Afrikaans beskikbaar is. Hierdie hoofstuk dien ook as raamwerk waarvolgens 'n boek oor die aard en omvang van historiese bronne in Afrikaans geskryf sal word.<sup>10</sup>

Die ontwikkeling van die Historiese Korpus vir Vroeë Afrikaans is na aanleiding van die voorafgaande uiteensetting duidelik nog in sy kinderskoene. Die eerste fase van hierdie projek het die inventarisering van die historiese bronne wat beskikbaar is, beide reeds gedigitiseerde en ongedigitiseerde publikasies, behels. Die wêreldwye stilstand wat deur die COVID-19-pandemie (in 2020 en 2021) veroorsaak is, het egter die inventariseringsproses gekniehalter, omdat daar nie fisiese toegang gekry kon word tot biblioteke en argiewe nie. Nietemin het die projek steeds voortgegaan en 'n aantal tekste is wel tydens hierdie periode geskandeer. Daar is egter heelwat inhaalwerk om te doen sodra daar weer vrylik gereis kan word na biblioteke en toegang gekry kan word tot biblioteke wat geslote was vir die publiek tydens 2020 en gedeeltes van 2021.

Die skanderings wat reeds gedoen is, word as 'n proefkorpus beskou, omdat dit gebruik word om verskillende soorte OKH-programmatuur te toets om sodoende te bepaal watter programmatuur die effektiwste sal werk en hoe ekstensief die kwaliteitskontroleproses<sup>11</sup> sal moet wees. Die OKH-proses is meer gekompliseerd in Afrikaans as in Engels of selfs in Nederlands, omdat daar tans geen gespesialiseerde Afrikaanse programmatuur beskikbaar is nie. Om hierdie redes word Nederlandse programmatuur aangepas vir gebruik in Afrikaanse kontekste. Voorts is die OKH-proses nie foutloos nie en dikwels moet tekste getranskribeer word voordat die digitaliseringsproses onderneem kan word. Transkribering neem egter heelwat langer as skandering, met die gevolg dat minder tekste by die korpus ingesluit kan word. Aangesien transkribering aansienlik duurder as die skanderingsproses is, sal die proses ook langer neem omdat die nodige befondsing eers bekom moet word.

Die ontwikkeling van die Historiese Korpus vir Vroeë Afrikaans (1675–1925) het wel tot gevolg gehad dat daar begin is om 'n Argief vir Historiese Afrikaanse Bronne saam te stel as eerste konkrete uitkoms van die groter projek.

### **2.5 Argief vir Historiese Afrikaanse Bronne**

Die ontwikkeling van 'n argief wat historiese Afrikaanse bronne bevat, was nie aanvanklik deel van die beplanning van die Historiese Afrikaanse Korpus-projek nie. Met die skandering van die tekste vir die korpus is daar egter besluit dat dit sinvol is om die skanderings van die tekste aanlyn beskikbaar te stel vir taalnavorsers wat 'n behoefte daaraan het.

Die Argief vir Historiese Afrikaanse Bronne is in Augustus 2021 gevestig en bestaan tot op hede (Julie 2022) uit ongeveer 200 geskandeerde tekste van ouer bronne wat van belang is vir die verklarings van die ontwikkeling van Afrikaans. Hierdie tekste is nie deur 'n kwaliteitskontroleproses nie en is ook nie gekodeer om deursoekbaar te wees nie, maar word slegs in PDF-formaat (sien Senekal 2011 oor die proses om dokumente na PDF-formaat om te skakel) gestoor vir navorsers om handmatig daardeur te kan werk. Dit is nie die ideale manier van doen nie, maar dit word as 'n interimfase beskou waar daar gewerk word met wat tot die medewerkers se beskikking is.

Die digitale kopieë van die historiese tekste word op VivA se webblad gestoor as deel van die Taalportaal. Die argief neem 'n soortgelyke formaat as die Database Geschiedenis Nederlandse Taalkunde (DAGENTA)<sup>12</sup> aan en beoog om dieselfde funksie, naamlik om historiese publikasies toeganklik en bruikbaar te maak vir navorsing, vir Afrikaans te vervul.

Die publikasie van die argief het egter bepaalde uitdagings na vore gebring. Die grootste uitdaging is die kontrolering van die teksinhoud. Aangesien die tekste wat in die argief vervat is, in die era 1675–1925 geproduseer is, word daar in hierdie tekste van woordeskat gebruik gemaak wat in die historiese taalkonteks algemeen was, maar nie in die politieke en sosiale konteks van die hedendaagse Afrikaans gepas is nie. Sommige van hierdie terme is sensitief (soos verwysings na etniese en religieuse groepe) en kan aanstoot gee. Daar moet nog besin word oor die reg van die korpusontwikkelaars om die tekste te sensor en die effek wat die uitsluiting van tekste op die korpus kan hê.

Weens die gebruik van politieke onsensitiewe terme in sommige tekste is daar besluit dat nie al die tekste wat geskander is, beskikbaar gestel sal word vir die breër publiek nie, maar dat navorsers 'n versoek sal kan rig om toegang tot hierdie tekste te verkry, ten spyte daarvan dat hierdie tekste gratis beskikbaar is. Die doel hiervan is om navorsers bewus te maak van hul sosiale verantwoordelikheid wanneer hulle met hierdie tekste werk en om hulle te vermaan om met die nodige versigtigheid en sensitiwiteit met die betrokke tekste om te gaan.

Die reg, of gebrek daaraan, van individue en instansies om die inhoud van digitale bronne soos hierdie argief te sensor is 'n tema waaroor daar al redelik in internasionale kringe gepubliseer is (sien Wilson en Golding 2016; Brilmyer 2018; Chilcott 2019). Wright (2019:331) fokus op die verantwoordelikheid van die argiveerder ten opsigte van historiese taalgebruik wat tans as aanstootlik beskou word. Wright (2019:334) beklemtoon die sosiale verantwoordelikheid van die argiveerder om te verhoed dat etikettering, katalogisering en argiefbeskrywings bydra tot onsensitiewe en aanstootlike sosiopolitieke omstandighede. “Unproblematised reproduction of this language in finding aids and other archival resources implies that the archives and record holders are also accepting of this language” (Wright 2019:335). Daar is dus 'n verantwoordelikheid om verantwoordbaar en met die nodige versigtigheid om te gaan met onsensitiewe taal in 'n digitale argief, sonder om die inhoud van die publikasies wat daarin vervat word – hetsy die taalgebruik of historiese werklikheid – tot niet te maak of aan te pas.

In Australië en die Verenigde Koninkryk is protokole ontwikkel vir die hantering van onsensitiewe taalgebruik in argiewe. Die Aboriginal and Torres Strait Islander Library and Information Resource Network- (ATSILIRN-) protokol is reeds in 2012 in Australië gevestig met die doel om verantwoordbaar om te gaan met argieftekste wat taalgebruik bevat wat kultureel of etnies aanstootlik is. Hierdie protokol behels 'n proses waartydens volledige beskrywings by argiefitems gevoeg is met die doel om die items te kontekstualiseer op 'n manier wat die nodige respek toon waar nodig en ooreenstem met die hedendaagse tydsgees. Onsensitiewe terme in titels van argiefitems, kataloge en ander argiefhulpbronne is behou, maar 'n lys alternatiewe terme wat gebruik kan word vir indeksering en as soekterme is in samewerking met die betrokke gemeenskappe gevestig, sodat diegene wat nie van onsensitiewe terme gebruik wil maak wanneer hulle 'n argief benut nie, nie daarmee gekonfronteer hoef te word nie (Chilcott 2019:364). Hierdie benadering neem veral die besluite en ervarings van die betrokke gemeenskappe in ag en gun hierdie gemeenskappe eienaarskap oor hul geskiedenis. Hierdie protokol kan van onskatbare waarde wees in die Suid-Afrikaanse konteks en kan versoening tussen Afrikaanse gemeenskappe help fasiliteer.

Tans beskik die Argief vir Historiese Afrikaans nog nie oor 'n volledige katalogus, beskrywings van argiefitems of ander argiveringshulpmiddels nie, omdat daar op hierdie stadium gefokus word op die digitisering van inventarisitems en die uitbreiding van die inventaris. Volgende ontwikkelingsfases sal waarskynlik fokus op die toeganklikheid en bruikbaarheid van die digitale argief, en dit sal die verantwoordbare hantering van onsensitiewe taalgebruik en historiese inhoud wat aanstoot kan gee, insluit.

### 3. 'n Digitale biblioteek vir Afrikaans en massadigitisering

Al hierdie digitiseringsprojekte het 'n enkele doel voor oë, naamlik om Afrikaans – en al sy variëteite – te digitiseer en gratis beskikbaar te stel vir navorsing in en oor Afrikaans (sien Senekal 2011, 2012; Glorie 2015; De Beer en De Beer 2016). Hierdie projekte vorm die onderbou vir 'n massadigitiseringsprojek en die ontstaan van 'n potensiële digitale biblioteek vir Afrikaans, waar publikasies in en oor Afrikaans digitaal behou, bewaar en bevorder kan word.

Afrikaans funksioneer egter nie in isolasie nie. Die noue bande tussen Afrikaans en Nederlands het tot gevolg dat sommige van die vroeë Afrikaanse tekste reeds deur Nederlandse instansies gedigitiseer is, of vir Nederlandse instansies van belang is – met die gevolg dat die digitale kopieë dan aan die Nederlandse instansie beskikbaar gestel word. So is daar 'n hele aantal ouer Afrikaanse tekste reeds in die DBNL opgeneem en kan hulle daar geraadpleeg word. Die voordeel is dat van hierdie tekste reeds deursoekbaar is, wat hulle waarde vir taalhistoriese navorsers verhoog.

Verder is Suid-Afrika 'n meertalige land waarin nege amptelike inheemse Suid-Afrikaanse tale naas Afrikaans erken word. In die lig hiervan is daar verskeie tekste (soos taalaanleerdersboeke en tweetalige handleidings) wat vir sowel Afrikaans as ander inheemse Suid-Afrikaanse tale, soos Setswana en Zoeloe, van belang is of kan wees. Hierdie tekste word gekatalogiseer met die oog op gesamentlike digitiseringsprojekte of hulp met die digitisering van publikasies in hierdie tale. Die digitiseringsproses en katalogisering van tekste skep geleenthede vir sinergie tussen Afrikaans en die ander inheemse tale en sodoende kan Afrikaans op die voorgrond geplaas word as 'n taal waarin en waarmee geëksperimenteer kan word. 'n Projekprotokol en -handleiding is derhalwe saamgestel met die doel om die digitiseringsproses wat vir Afrikaans gebruik word, meer toeganklik te maak vir die digitisering van ander Afrikatale vir ingeval daar 'n behoefte aan sou wees.<sup>13</sup>

### 4. Uitdagings rakende die digitisering van Afrikaans

Die voetwerk wat vir die digitisering van Afrikaanse tekste gedoen is, kan vir ander tale (inheemse Suid-Afrikaanse tale in die besonder) 'n hupstoot gee wanneer dit kom by die ontwikkeling van digitiseringsprojekte wat geskoei is op die digitisering van tekste in hierdie tale. Heelwat van die uitdagings wat in die Suid-Afrikaanse konteks teëgekomp is tydens die digitiseringsproses van Afrikaans, kan ook problematies wees by die digitisering van geskrewe tekste in ander Suid-Afrikaanse tale.

Enkele van hierdie uitdagings vir die Suid-Afrikaanse konteks en hoe dit uit die weg geruim is, word hier onder uitgelig.

#### **4.1 Sinergie en die gebrek aan 'n gedeelde visie**

Tydens die inventariseringsfase van die digitiseringsproses van die Historiese Korpus vir Vroeë Afrikaans (1675–1925) is daar grotendeels op die identifisering van relevante bronne en die bekikbaarheid (toeganklikheid) van hierdie bronne gefokus (vgl. Breed e.a. 2016, Trollip 2018, De Vries en Linde 2021). Vervolgens is daar met 'n verskeidenheid instansies in kontak getree om te bepaal watter tekste hierdie instansies tot hul beskikking het en in watter formaat/toestand hierdie bronne is. In die meeste gevalle het instansies ingestem om hul biblioteek-en/of argiefkatalogus beskikbaar te stel. Sommige instansies het bronne wat reeds gedigitiseer is, aangestuur om ingesluit te word in die historiese korpus, terwyl ander hul versamelings oopgestel het om geïnventariseer en gedigitiseer te word.

Tog het sommige instansies hulle teen die projek verset. In die meeste gevalle was hierdie instansies onwillig om hul teksversamelings beskikbaar te stel, omdat hulle hul eie navorsers die alleenreg op hierdie taalkundebronne wil gee. So ook is daar instansies (en individue) wat die projek teengestaan het omdat hulle reeds hul eie digitiseringsprojek geloods het of 'n toekomstige eie digitiseringsprojek in die pyplyn het.

'n Gebrek aan sinergie en 'n gedeelde visie – naamlik die digitisering van Afrikaans in al sy fasette – het noodwendig duplisering en die verdeling van hulpbronne tot gevolg. Die Afrikaanse mark het beperkte hulpbronne, hetsy befondsing, tegnologie, menslike kapitaal of kundigheid, tot sy beskikking en dit behoort volledig en gesamentlik in die belang van die hele taalgemeenskap gebruik te word. Wanneer hulpbronne egter onnodig verdeel word, is die gevolg dat verskeie projekte beperk word deur die hulpbronne wat daaraan toegeken is en dat hierdie hulpbronne boonop aangewend word in fases van die digitiseringsproses wat reeds deur ander projekte met beperkte hulpbronne vervolmaak is. Dit is 'n verkwisting van energie en hulpbronne.

Deur saam te werk en dus ook hulpbronne (soos befondsing, personeel en bewaringskapasiteit) te deel, kan verskeie projekte se uitkomstegelykertyd bereik word op 'n meer tyd- en koste-effektiewe manier. Meer geld, meer tyd om te kan werk en meer mensekrag sal 'n meer gefokuste en gekoördineerde uitset tot gevolg hê. Die totstandkoming van sinergie en die skep en navolg van 'n gedeelde visie oor digitisering in Suid-Afrika in die algemeen en Afrikaans in besonder is daarom van kardinale belang wanneer grootskaalse of massadigitisering beoog word. Om hierdie rede het Viva moeite gedoen om interinstitusionele bande (met byvoorbeeld The South African Centre for Digital Language Resources [SADiLaR]) te bou, en hierop moet verder voortgebou word.

Samewerking deur en met instansies soos SADiLaR, Viva, die Suid-Afrikaanse Akademie vir Wetenskap en Kuns, die Afrikaanse Taalmuseum en -monument, NALN (vgl. Senekal 2011, 2012), universiteitsbiblioteke, Nuuseum, die Wes-Kaapse Argief, die Nasionale Argief, die Nasionale Biblioteek (vergelyk Marx 2021) en andere is daarom nie net gewens nie, maar ook noodsaaklik. Van hierdie instansies is in gereelde kontak en gespreksgeleenthede tussen hierdie partye word dikwels geskeduleer met die doel om samewerking te fasiliteer. Dit kan en behoort wel beter word as wat op hierdie stadium die geval is. Sinergie en 'n gedeelde visie rakende die digitalisering van Afrikaans is derhalwe van uiterste belang.

#### **4.2 Toerusting en tegnologie**

Soos reeds genoem, is die kwaliteit van die skanderings 'n bepalende faktor in die sukses van die digitaliseringsproses. Dit is gevolglik van kardinale belang dat die regte skanderings-toerusting aangeskaf word en dat hierdie toerusting reg gebruik word om die hoogste moontlike kwaliteit skanderings te produseer (vergelyk Carstens 2021). Van die uitdagings wat die kwaliteit van die skanderings kan beïnvloed, is (a) die omgewing waarin die skanderings gedoen word, (b) die wyse waarop die skanderings gedoen word, en (c) uiteraard die gehalte van die apparaat (skandeermasjien) wat gebruik word. Die aard van die apparaat is veral van belang in gevalle waar die kwaliteit van die brondokument swakker is weens faktore soos byvoorbeeld waterskade, ink wat vervaag het of papier wat verkrummel.

Indien die skanderings gedoen word by die instituut of instelling waar die bronne geberg word, moet daar in ag geneem word dat daar spasiebeperkings op die skandeermasjien self kan wees, wat dit moeilik kan maak om groter tekste te skandeer. So ook is dit moontlik dat daar nie gerieflike toegang tot 'n kragpunt is nie, of dat die omgewing stowwerig is. Indien die tekste na 'n sekondêre perseel geneem word om geskandeer te word, moet daar verseker kan word dat die tekste veilig vervoer kan word, dat die sekondêre perseel oor die nodige werkspasie beskik en dat ouer bronne veilig bewaar kan word.

Om die kwaliteit van die skanderings te verseker, is dit noodsaaklik dat die skanderings op 'n stabiele, plat oppervlakte gedoen word sodat onnodige bewegings nie die skanderings negatief kan beïnvloed nie. Des te meer is dit belangrik dat elke bladsy individueel geskandeer word en dat die kwaliteit van 'n teks se skanderings eers getoets word alvorens die volledige teks geskandeer word.

Die programmatuur wat vir die OKH-proses gebruik gaan word, moet ook voor oë gehou word wanneer daar beslis word oor die kwaliteit van die skanderings (vgl. Senekal 2011). Indien die skanderings nie voldoen aan die vereistes van die OKH-programmatuur nie, sal alternatiewe programmatuur bekom moet word, of die tekste sal getranskribeer moet word. Transkribering bring weer probleme op sy eie mee (soos die beskikbaarheid van voldoende kundige persone om die transkribering effektief te kan doen en dit ook te kontroleer vir korrektheid), en dit behoort derhalwe beperk te word.<sup>14</sup>

#### **4.3 Berging en publikasie**

Beter kwaliteit skanderings is dikwels digitaal groter en vereis daarom meer spasie wanneer dit aanlyn geberg word. Daarom is dit noodsaaklik dat genoegsame wolkgebaseerde stoorplek of 'n groot genoeg bediener aangeskaf moet word. In gevalle waar daar nie tydens die skanderingsproses toegang tot vinnige, stabiele internet is nie, sal die skanderings op 'n hardeskyf gestoor moet word voordat dit oorgelaai word op die aangewese virtuele stoorplek.

Alhoewel dit relatief eenvoudig is om die skanderings van een stoorplek na 'n ander oor te skuif, neem hierdie proses onnodige tyd in beslag en daar is altyd die risiko dat inligting oorgesien of verloor kan word. Daarom is dit beter om vroegtydig te besin oor waar die beste stoorplek sal wees en of hierdie stoorplek gepas is vir die formaat waarin die skanderings gepubliseer gaan word (byvoorbeeld as bibliografie, argief of korpus), asook of die plek (ruimte) waar hierdie eindproduk gepubliseer gaan word (byvoorbeeld 'n webwerf of privaat bediener) gemaklik kan koppel met die aangewese stoorplek.

Uit pogings deur Senekal (2011, 2012) om 'n digitaliseringsproses vir NALN van die grond af te kry, kan – en moet – daar geleer word van die aanpak en ook die uitvoer van die groter digitiseringsproses. Senekal wys op die belang van 'n “volledige korpus digitale dokumente”, wat die einddoel van hierdie soort projekte behoort te wees. Hy beklemtoon ook dat daar aanpassings sal moet kom soos 'n mens wyser word oor die digitale insamelingsproses. Om daarom aan die begin tyd te neem om prosesse, standaarde en 'n infrastruktuur te vestig, sal op die lang termyn heelwat probleme en onnodige uitgawes vermy.

## 5. Ten slotte

Die digitisering van Afrikaanse bronne is nog lank nie 'n massadigitiseringsproses nie, maar daar word reeds in die vorm van kleinskaalse digitiseringsprojekte 'n groot verskil gemaak – vergelyk byvoorbeeld Senekal (2011 en 2012) oor die stappe om sodanige prosesse gevestig te kry rakende die NALN-versameling. Verskeie aanlyn hulpbronne is al klaar beskikbaar gestel met die doel om Afrikaanse bronne vir navorsers toeganklik te maak, soos met die DBAT en DBAL, 'n argief vir historiese Afrikaanse bronne, en binnekort 'n Historiese Korpus vir Vroeë Afrikaans, waaroor daar in hierdie artikel verslag gedoen is. Dit kom voor asof sodanige prosesse ook by instansies elders in die land van die grond af begin kom, maar dit illustreer ook weer eens die gebrek aan 'n oorkoepelende plan om digitisering gekoördineerd aan te pak.

Alhoewel verskeie uitdagings teëgekomp is tydens die digitiseringsproses en ook hier bo uitgelig is, kan die lesse uit hierdie proses van nut wees vir ander Suid-Afrikaanse tale om hulle eie digitiseringsprojekte van die grond af te kry. Digitisering behoort weens die koste, tyd en energie wat daaraan gekoppel is eerder 'n gesamentlike (dus instansiegedrewe) projekgeleentheid te wees as 'n individuele een. In hierdie opsig kan die projek rakende die massadigitisering van Afrikaanse bronne as aanleiding tot grootskaalse samewerking (in die besonder ten opsigte van die deel van ervaring en kundigheid met ander tale) tussen rolspelers op hierdie terrein dien.

## Bibliografie

- Borgman, C.L. 2010. The digital future is now: A call to action for the humanities. *Digital Humanities Quarterly*, 3(4):1–30.
- Bor Ng, K. en J. Kucsma (reds.). 2010. *Digitisation in the real world: Lessons learned from small and medium-sized digitization projects*. New York: Metropolitaanse Biblioteekraad van New York.
- Bouma, G., E. Coussé, T. Dijkstra en N. van der Sijs. 2020. The EDGeS Diachronic Bible Corpus. *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*: 5232–9.
- Breed, A., W.A.M. Carstens en J.A.K. Olivier. 2016. Die DBAT: 'n Onbekende digitale taalkundemuseum. *Tydskrif vir Geesteswetenskappe*, 56(2):391–409.



- Brilmyer, G. 2018. Archival assemblages: Applying disability studies' political/relational model to archival description. *Archival Science*, 18:95–118.
- Carstens, A. 2021. Imaging standards and quality management of digital images. Referaat gelewer tydens die UCT Libraries World Digital Preservation Day, Universiteit van Kaapstad. [https://zivahub.uct.ac.za/articles/presentation/Imaging\\_standards\\_and\\_quality\\_management\\_of\\_digital\\_images/16917430](https://zivahub.uct.ac.za/articles/presentation/Imaging_standards_and_quality_management_of_digital_images/16917430) (14 September 2022 geraadpleeg).
- Carstens, W.A.M. 2010. Die DBAT as bemagtigingsinstrument vir die Afrikaanse taalkunde: rol en moontlikhede. Referaat gelewer tydens SAVN-kongres, 27–29 September 2010. [https://collections.nwu.ac.za/dbtw-wpd/textbases/bibliografie-afrikaans/documents-media/dbat\\_lvsa\\_2010.pdf](https://collections.nwu.ac.za/dbtw-wpd/textbases/bibliografie-afrikaans/documents-media/dbat_lvsa_2010.pdf).
- . 2011. Die digitalisering van NALN se knipselversameling: Wannie Carstens lewer kommentaar. LitNet, 15 Junie. [https://argief.litnet.co.za/article.php?news\\_id=109567](https://argief.litnet.co.za/article.php?news_id=109567) (12 Oktober 2021 geraadpleeg).
- Carstens, W.A.M. en D. van Zyl. 2004. Die ontwikkeling van 'n BATL – 'n vorderingsverslag. Referaat gelewer tydens SAVN-kongres, 25 September 2004. [https://collections.nwu.ac.za/dbtw-wpd/textbases/bibliografie-afrikaans/documents-media/batl\\_%20vir\\_afrikaans\\_2004\\_1.pdf](https://collections.nwu.ac.za/dbtw-wpd/textbases/bibliografie-afrikaans/documents-media/batl_%20vir_afrikaans_2004_1.pdf).
- Chilcott, A. 2019. Towards protocols for describing racially offensive language in UK public archives. *Archival Science*, 19:359–76.
- Coussé, E. 2010. Een digitaal compilatiecorpus historisch Nederlands. *Lexikos*, 20(2010):123–42.
- Coyle, K. 2006. Mass digitization of books. *Journal of Academic Librarianship*, 32(6):641–5. <https://www.kcoyle.net/jal-32-6.html> (21 September 2021 geraadpleeg).
- De Beer, C.S. en F. de Beer. 2016. Om digitaal te wees: 'n waarborg of 'n bedreiging vir die toekoms? *Suid-Afrikaanse Tydskrif vir Natuurwetenskap en Tegnologie*, 35(1):137–45.
- Deumert, A. 2004. *Language standardization and language change. The dynamics of Cape Dutch*. Amsterdam/Philadelphia: John Benjamins.
- De Vries, W. en J. Linde. 2021. Lewende, gesaghebbende digitale biblioteek vir Afrikaanse letterkunde. <https://voertaal.nu/lewende-gesaghebbende-digitale-biblioteek-vir-afrikaanse-letterkunde/> (17 Februarie 2021 geraadpleeg).
- Glorie, I. 2015. Computer linguist Gerhard van Huyssteen: “Het gaat goed met een taal waarmee het DIGITAAL goed gaat”. *Maandblad Zuid-Afrika*, 92:240–1.
- Groenewald, R. 2010. Digitiseringskwessies (NALN) 11–12 November 2010. Pretoria: Universiteit van Pretoria.
- Jamali, H.R. 2008. What is not available online is not worth reading? *Webology*, 5(4). <http://www.webology.org/2008/v5n4/a63.html> (21 September 2021 geraadpleeg).

- Kaschula, R. en H. Wolff (reds.). 2020. *The transformative power of language: From postcolonial to knowledge societies in Africa*. Cambridge: Cambridge University Press. Doi:10.1017/9781108671088.014.
- Kettley, G.M. 1970. *Suid-Afrikaanse linguistiese bronnegids*. Saamgestel in opdrag van die Suid-Afrikaanse Akademie vir Wetenskap en Kuns. Johannesburg: Voortrekkerpers.
- Khumalo, L. 2020. Corpora as agency in the intellectualisation of African languages. In Kaschula en Wolff (reds.) 2020.
- Kirsten, J. 2016. Grammatikale verandering in Afrikaans van 1911–2010. PhD-proefskrif, Noordwes-Universiteit (Vanderbijlpark-kampus).
- Liebenberg, H. 2018. Die Wes-Kaapse Argief en die begin van Afrikaans. *Tydskrif vir Geesteswetenskappe*, 58(2):204–36.
- Marx, E. (samest.). 1992. *Bronnegids by die studie van die Afrikaanse letterkunde en taal*. Bloemfontein: Nasionale Afrikaanse Letterkundige Museum en Navorsingsentrum.
- Marx, J. 2021. Bekendes maak planne vir Afrikaanse aanlyn argief. Netwerk24, 17 Mei. <https://www.netwerk24.com/Nuus/Algemeen/bekendes-maak-planne-vir-afrikaanse-aanlyn-argief-20210517> (21 September 2021 geraadpleeg).
- Niémah, D. 2021. Devastation as historical UCT buildings gutted by runaway fire. <http://www.lib.uct.ac.za/news/devastation-historical-uct-buildings-gutted-runaway-fire> (30 September 2021 geraadpleeg).
- Nienaber, P.J. 1943–1981. *Bibliografie van Afrikaanse boeke*. Nege dele. Johannesburg: Voortrekkerpers en Perskor.
- Nienaber, P.J. en G.S. Nienaber. 1947–1968. *Bronnegids by die studie van die Afrikaanse taal en letterkunde*. Talle dele. Bloemfontein: Nasionale Afrikaanse Letterkundige Museum en Navorsingsentrum.
- NWU. 2021a. Digitale Bibliografie van die Afrikaanse Letterkunde. <https://collections.nwu.ac.za/dbtw-wpd/textbases/bibliografie-afrikaans/dbal.html> (14 September 2022 geraadpleeg).
- . 2021b. Digitale Bibliografie van die Afrikaanse Taalkunde. <https://collections.nwu.ac.za/dbtw-wpd/textbases/bibliografie-afrikaans/dbat.html> (14 September 2022 geraadpleeg).
- Raidt, E.H. 1991. *Afrikaans en sy Europese verlede*. 3de uitgawe. Kaapstad: Nasou.
- Rijswijk, K. 2020. *Digital transformation: Ongoing digitisation and digitalisation processes*. Wageningen: Wageningen Academic Publishers.
- Roux, J.C. 2020. African language resources for knowledge societies. In Kaschula en Wolff (reds.) 2020.

SADiLaR (The South African Centre for Digital Language Resources). 2022. Home. <https://www.sadilar.org/index.php/en/> (14 September 2022 geraadpleeg).

Sandrock, U. (samest.). 1980. *Bronnegids by die studie van die Afrikaanse taal en letterkunde, 1900–1970. Akkumulatiewe uitgawe, uit die bronnegidse van P.J. Nienaber, 1900–1968, K.H. Schmidt, 1969, Louisa Coetzee, 1970*. Pretoria: Raad vir Geesteswetenskaplike Navorsing.

Schwab, K. 2017. *The fourth industrial revolution: Currency*. New York: Crown Business.

Senekal, B.A. 2011. Die digitalisering van NALN se knipselversameling: die bemiddeling van 21ste-eeuse navorsing in die Afrikaanse letterkunde. *LitNet Akademies*, 8(2):46–65. [https://www.litnet.co.za/wp-content/uploads/2021/01/LA\\_8\\_2c\\_senekal.pdf](https://www.litnet.co.za/wp-content/uploads/2021/01/LA_8_2c_senekal.pdf).

—. 2012. 'n Inligtingstegnologie-gesentreerde gebruikerskoppelvlak vir navorsingsdoeleindes binne die geesteswetenskappe met spesifieke verwysing na die Afrikaanse letterkunde. *LitNet Akademies*, 9(2):468–99. <https://www.litnet.co.za/assets/pdf/7GWSenekal.pdf>.

Tilson, D., K. Lyytinen en C. Sørensen. 2010. Research commentary – digital infrastructures: The missing IS research agenda. *Information Systems Research*, 21(4):748–59.

Trollip, B. 2018. *Grammatography: Inventorying historical Afrikaans grammars*. Referaat gelewer tydens die 5de Gentse colloquium over het Afrikaans, 4–5 Oktober 2018, Gent.

Van Huyssteen, G.B. 2019. Is ons aan die “digitiseer” of “digitaliseer”? *Beeld*, 4 Januarie. <https://www.netwerk24.com/netwerk24/stemme/menings/is-ons-aan-die-digitiseer-of-digitaliseer-20190103> (28 Augustus 2021 geraadpleeg).

VivA. 2021. Korpusportaal. <https://viva-afrikaans.org/> (30 September 2021 geraadpleeg).

Wilson, J.Z. en F. Golding. 2016. Latent scrutiny: Personal archives as perpetual mementos of official gaze. *Archival Science*, 16:93–109.

Wright, K. 2019. Archival interventions and the language we use. *Archival Science*, 19:331–48.

## Eindnotas

<sup>1</sup> Sien: <https://www.gutenberg.org/>

<sup>2</sup> Sien: <https://www.kb.nl/>

<sup>3</sup> Van die begin af het werk aan die DBAT (wat voorafgegaan is deur die Biblioteek van die Afrikaanse Taal- en Letterkunde [BATL]) geskied in ooreenstemming met die model en struktuur van die aanvanklike BNTL (en later die DBNL). Hiermee word die hulp wat personeel betrokke by die Nederlandse projek (soos Paul Doorenbosch, Ella Kamp,

Karina van Daalen-Oskam en Cees Klapwijk) goedgegunstiglik oor soveel jare verleen het met die konseptualisering, vestiging en instandhouding van DBAT en later DBAL met groot dank en waardering erken.

<sup>4</sup> Sien: [https://collections.nwu.ac.za/dbtw-wpd/textbases/bibliografie-afrikaans/documents-media/dbat\\_persverklaring\\_%2024mei2010.pdf](https://collections.nwu.ac.za/dbtw-wpd/textbases/bibliografie-afrikaans/documents-media/dbat_persverklaring_%2024mei2010.pdf)

<sup>5</sup> Sien: <https://collections.nwu.ac.za/dbtw-wpd/textbases/bibliografie-afrikaans/dbat.html>

<sup>6</sup> Met “op datum” word hier bedoel dat die rekordhouding van alle toepaslike bronne wat enigsins oor die afgebakende veld beskikbaar sou kon wees tot op ’n ooreengekome datum (soos 2021) volledig is.

<sup>7</sup> Vergelyk Senekal (2011, 2012) oor pogings om die literêre versameling van NALN te digitiseer.

<sup>8</sup> Sien: <https://collections.nwu.ac.za/dbtw-wpd/textbases/bibliografie-afrikaans/dbal.html>

<sup>9</sup> ’n Vergelykende korpus Bybelvertalings (Nederlands, Engels, Duits en Sweeds) dien hier as goeie voorbeeld van wat haalbaar is as tekste wel beskikbaar gestel word én elektronies deursoekbaar is. Bouma, Coussé, Dijkstra en Van der Sijs (2020) doen hieroor verslag.

<sup>10</sup> Hierdie projek sal onder die vaandel van die Virtuele Instituut vir Afrikaans (VivA) geskied.

<sup>11</sup> Volgens Carstens (2021) word die impak van gedigitiseerde versamelings grotendeels bepaal deur die proses wat gevolg word om die kwaliteit van die skanderings te verseker. Dit raak beide die apparaat wat gebruik word om die werk te doen én die ooreenstemming van die gedigitiseerde artefak met die oorspronklike bron in terme van inhoud, grootte, kleur en formaat. Dit dra by tot vertroue in die oorspronklikheid van die produk.

<sup>12</sup> Sien: <https://cls.ru.nl/dagenta/>

<sup>13</sup> Khumalo (2020) rapporteer oor projekte wat reeds in hierdie verband aan die gang is.

<sup>14</sup> Die projek om ’n aantal vrywillige Nederlandssprekendes in te span om ’n verskeidenheid ouer Nederlandse tekste te transkribeer, het eers na bykans 15 jaar afgeleef. “Onder de bezielende leiding van Herman Wiltink, Nicoline van der Sijs en Hans Beelen hebben meer dan honderd vrijwilligers sindsdien in totaal meer dan tweehonderd werken getranscribeerd, waaronder liedbundels, taalkundige werken, pamfletten, vroegmoderne medische teksten en psalmberijmingen.” Dit is gedoen met die hulp van die Stichting Vrijwilligersnetwerk Nederlandse Taal (SVNT). Dit is ’n tydsame en tydrowende proses. Vergelyk: [https://www.dbnl.org/nieuws/nieuws.php?l=2022\\_09\\_01](https://www.dbnl.org/nieuws/nieuws.php?l=2022_09_01)