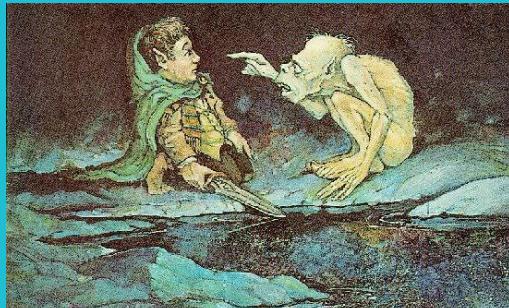


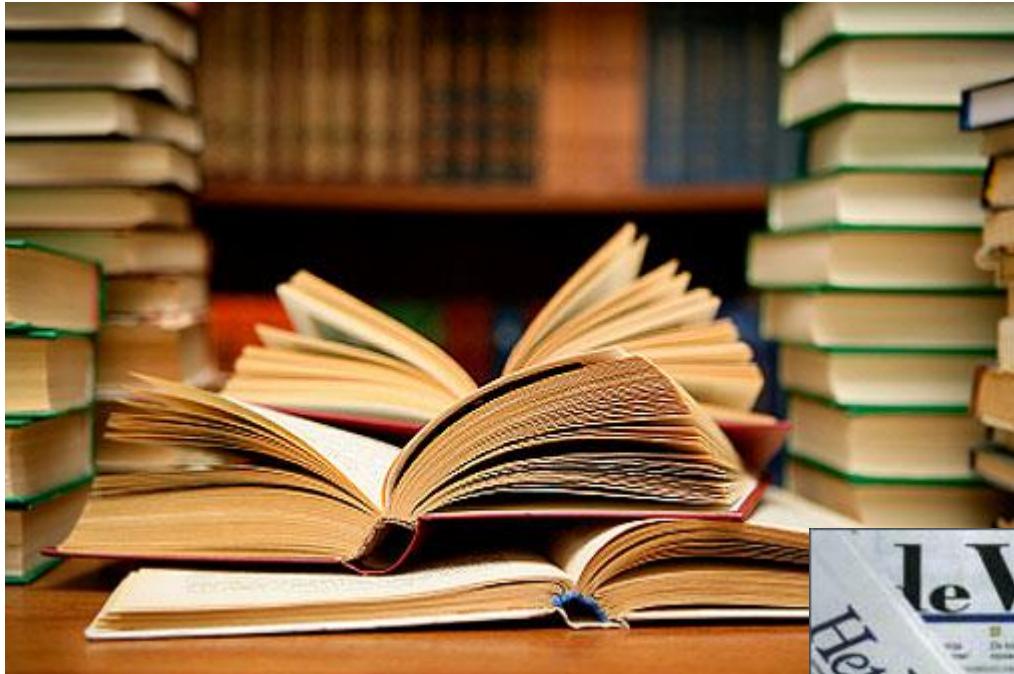
/instituut voor
de Nederlandse
taal/

Riddles in the Dark. Corpus building



Katrien Depuydt
Senior researcher/linguist
katrien.depuydt@ivdnt.org

Excerpt, Copy, Order



Geleerd by = Socie
Een gehoor kind dat
alleen wat dyn onder
hem leggen. Die dient dan
is geleerd, is niet geleerd.

Geleerd
Die dient dan dat die
alleen wat dyn onder
hem leggen.

Geles Korbyk
= gelaas q.v.

Fein

het geleerd van den bauwijn ja

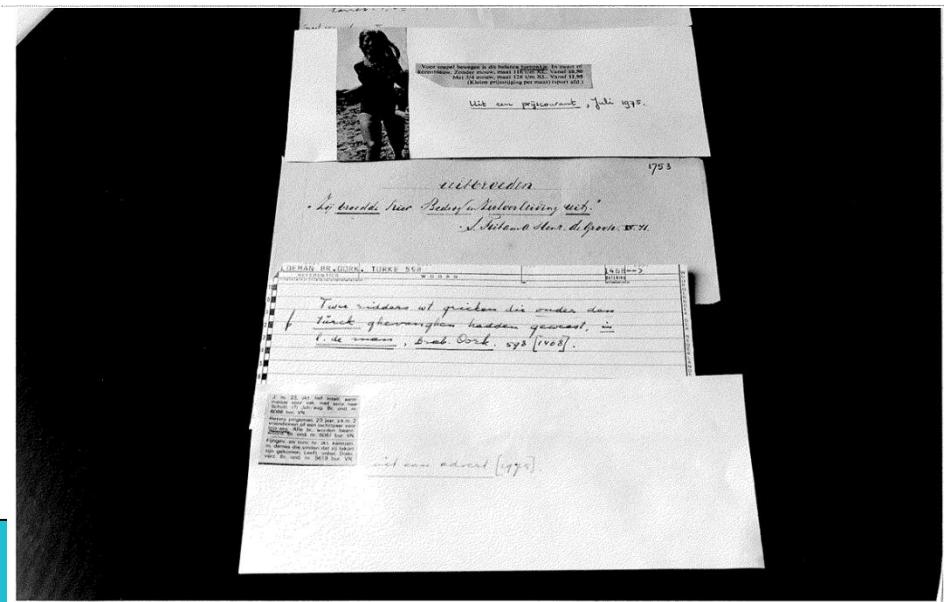
275

~~winkel 127~~ ~~Lijstspeldach~~ ~~1100~~
~~winkel 200~~ ~~1105~~ ~~I~~
 van een te ~~winkel~~ ~~108~~ llen
 worden vo ~~wordt gelaten~~ ~~132~~ lering moet
 vorderen en ~~en de nodige~~ ~~aan de ge-~~
 wordt gelaten ~~winkel~~ ~~een voldoen~~
 en de nodige ~~een~~ ~~I.C. is aan~~
 een ~~en~~
 m ~~winkel 1, 200~~ ~~1524~~
 v ~~winkel 306~~ ~~1527~~
 m ~~winkel 1, 10~~ ~~Liesy. Bubbel 1532~~
 v ~~winkel: W. J. Ap. 26 F~~ ~~Liesy. Bubbel 1532~~
 g ~~2 1/2 v. winkel~~
 d ~~1545~~
 g ~~2 1/2 v. winkel~~
 d ~~1546~~
 g ~~winkel Kiel ve Berchtesgaden~~
 d ~~de winkel~~ ~~1550~~ ~~7a.~~
 g ~~uit den winkel~~ ~~2266~~
 d ~~des maas.~~ ~~1296~~
 g ~~opp 6 winkel cornholte~~
 d ~~1561~~
 g ~~open winkel houten,~~ ~~1564~~
 d ~~Mannise. Ryens. 210 a.~~
 g ~~winkel, Mannise,~~
 d ~~Ryens. 1866.~~

X	winkel	1600	Hartelijnen 62	65
H land hoe De zon kel	X winkel st	173		
	X winkel	Uitst. Com. 1868	657	
H tige ware rees sprai vra ne sch dienst den v zighet Raad deze dat si de pos heeft zichter overig achten vrijste gestel	winkel, h. Univ. n, 128	+ 1600	st lu n- ze	
Den veren Staat dat n de D gers de D (recreat. zaalmoeders ins bro zen tel hei als gez leidit Goos vrijst	winkel (- oven)	1600	Verv. I, q/a	
	Winkel			1,117 a
	Winkel			2,461 b
	winkel (- opvalley)			2,951 a
	Winkel (- opporden)			2,951 b
	Winkel (- opzetten)			2,951 a
	winkel	1600	Attischer Ged.	237
X	winkel	1600	L. Textiel	61 - 524
X	winkel	1600		3,52
X	winkel	1600	Tandendri	10,131
ze ge tu	winkel 297 a	1600		



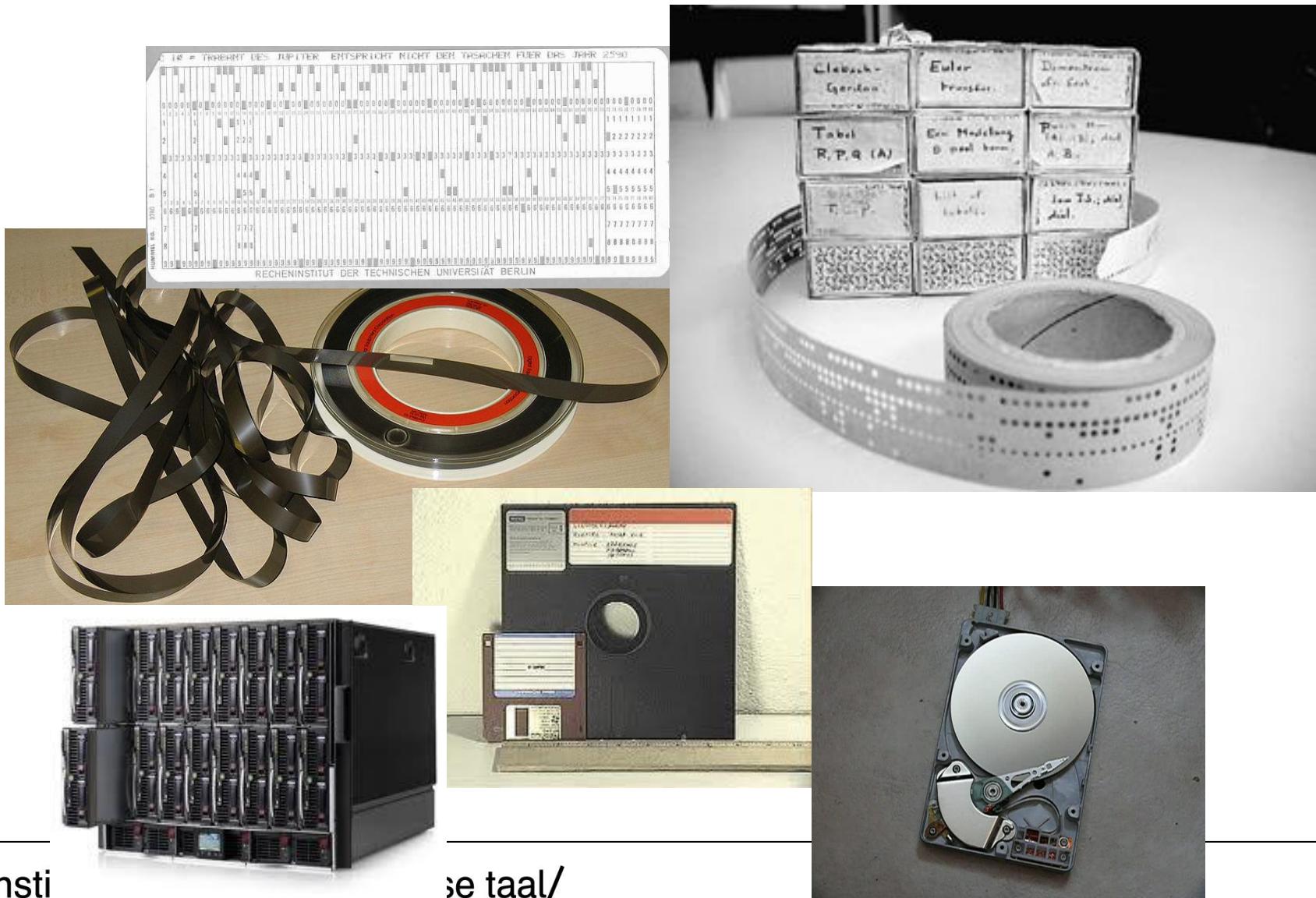
	1	8	15	22	29
2	9	16	23	30	
3	10	17	24	31	
4	11	18	25		
5	12	19	26		

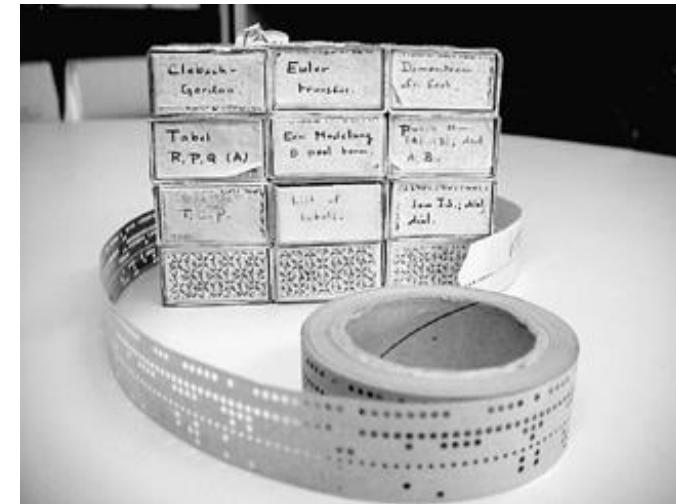
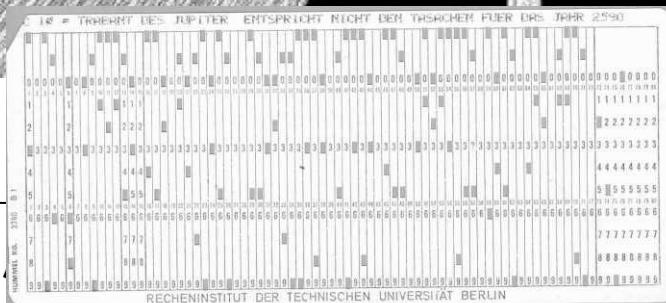


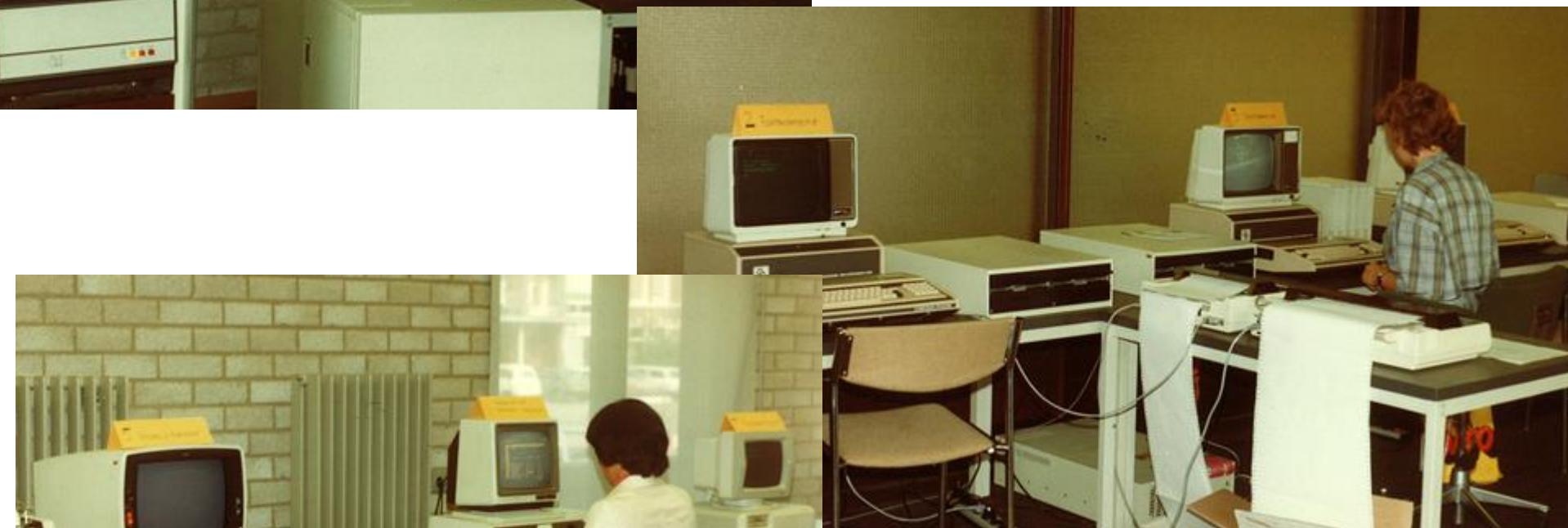


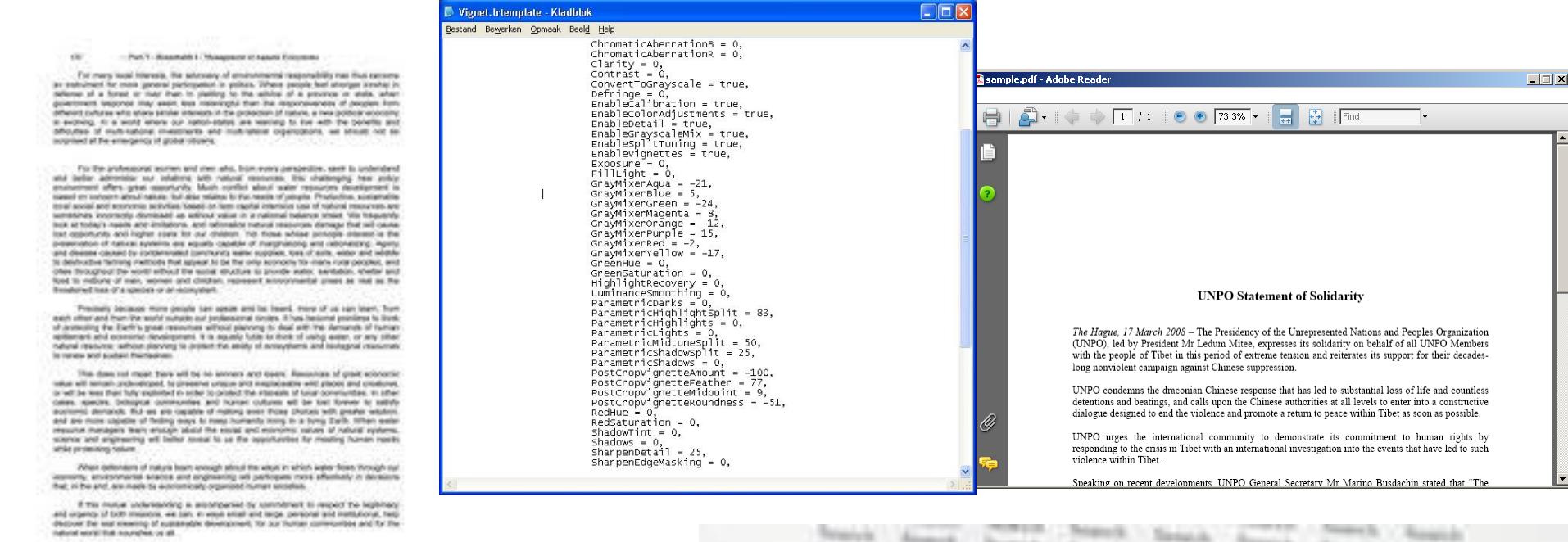
/instituut voor de Nederlandse taal/

Digitisation, conversion









```
C:\WINDOWS\system32\inetres\MetaBase.xml - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address C:\WINDOWS\system32\inetres\MetaBase.xml

<?xml version="1.0"?>
<configuration xmlns="urn:microsoft-catalog:XML_Metabase_V54_0">
  <MBProperty>
    <MB_Global_Location>=1</MB_Global_Location>
    <MB_SessionKey>49634462f000000a400000d00000003f7973b43bc7fb742906b1b83daedd740ace81afbe6ab95908d70aa33c69150e</MB_SessionKey>
    <MB_XMLSchemaTimeStamp>7b1b57751e32c301</MB_XMLSchemaTimeStamp>
    <MB_ChangeNumber>1031</MB_ChangeNumber>
    <MB_HistoryMajorVersionNumber>49</MB_HistoryMajorVersionNumber>
    <MB_SessionTimeStamp>db295a751e32c301</MB_SessionTimeStamp>
  </MBProperty>
  <IIS_ROOT_Location>/</IIS_ROOT_Location>
  <AdminACL>49634462f000000a400000d00000003f7973b43bc7fb742906b1b83daedd740ace81afbe6ab95908d70aa33c69150e</AdminACL>
  <IIS_Computer_Location>/LM</IIS_Computer_Location>
  <IIS_ConfigObject_Location>/LM/IISADMIN/</IIS_ConfigObject_Location>
  <IIS_ConfigObject_Location>/LM/IISADMIN/EXTENSIONS/</IIS_ConfigObject_Location>
  <IIS_ConfigObject_Location>/LM/IISADMIN/EXTENSIONS/DCOMCLSIDS/</IIS_ConfigObject_Location>
  <Custom Name="MD_IISADMIN_EXTENSIONS" ID="1028" Value="{1738644-F196-11D0-9953-00C04FD919C1}" Type="MULTISZ" UserType="IIS_MD_UT_SERVER" Attributes="NO_ATTRIBUTES" />
  <IISConfigObject>
    <IISConfigObject Location="/LM/IISADMIN/PROPERTYREGISTRATION">
      <Custom Name="MD_METADATA_ID_REGISTRATION" ID="1030" Value="0-65535;Microsoft Reserved 65536-524288;Microsoft IIS Admin Objects Reserved" Type="MULTISZ" UserType="IIS_MD_UT_SERVER" Attributes="INHERIT" />
    </IISConfigObject>
    <IISLogModules Location="/LM/Logging">
      <AdminACL>49634462a0000005800000400000003f7973b43bc7fb74032906b16c3daadd740ace81afbe6ab95908b90a63c69150e</AdminACL>
      <IISCustomLogModule Location="/LM/Logging/Custom Logging" LogCustomPropertyServicesString="W3SPVC MSFTPSVC SMTPSPVC NNTPSPVC" />
      <IISCustomLogModule Location="/LM/Logging/Custom Logging/Custom Logging" LogCustomPropertyData-Type="6" LogCustomPropertyName="Date" LogCustomPropertyHeader="date" LogCustomPropertyID="4013" LogCustomPropertyMask="1" LogCustomPropertyName="Date" LogCustomPropertyNodeID="1" />
      <IISCustomLogModule Location="/LM/Logging/Custom Logging/Extended Properties" LogCustomPropertyData-Type="6" LogCustomPropertyName="Extended Properties" LogCustomPropertyNodeID="3" />
      <IISCustomLogModule Location="/LM/Logging/Custom Logging/Extended Properties/Bytes Received" LogCustomPropertyData-Type="3" LogCustomPropertyName="cs-bytes" LogCustomPropertyMask="B192" LogCustomPropertyNodeID="17" />
      <IISCustomLogModule Location="/LM/Logging/Custom Logging/Extended Properties/Bytes Sent" LogCustomPropertyData-Type="3" LogCustomPropertyName="sc-bytes" LogCustomPropertyMask="4096" LogCustomPropertyNodeID="16" />
      <IISCustomLogModule Location="/LM/Logging/Custom Logging/Extended Properties/Client IP Address" LogCustomPropertyName="Bytes Sent" LogCustomPropertyHeader="LM/IISLogName/Custom Logging/Client IP Address" LogCustomPropertyID="10" />
    </IISLogModules>
  </IISConfigObject>
</configuration>
```

/Instituut voor de Nederlandse taal



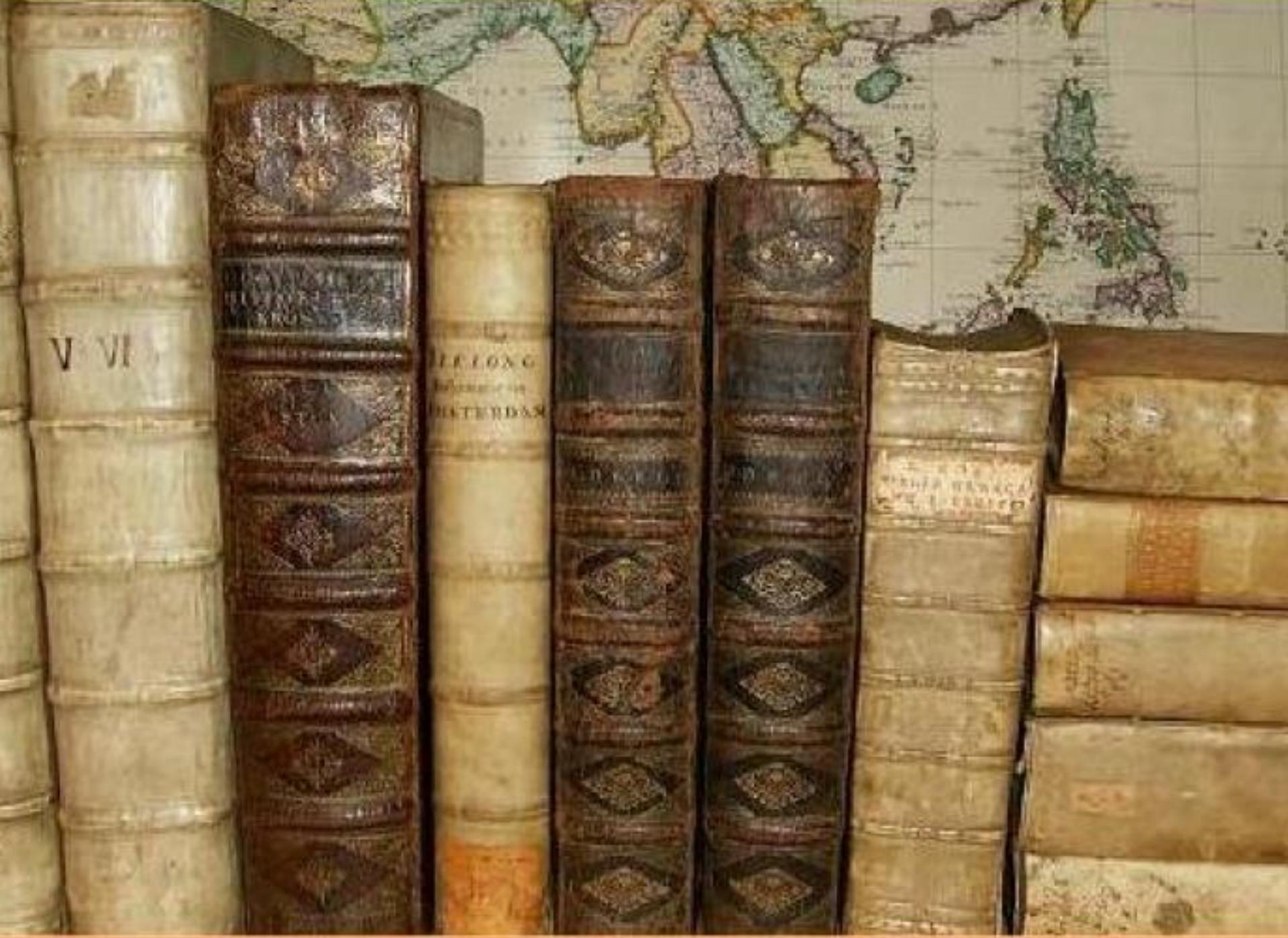
Steps

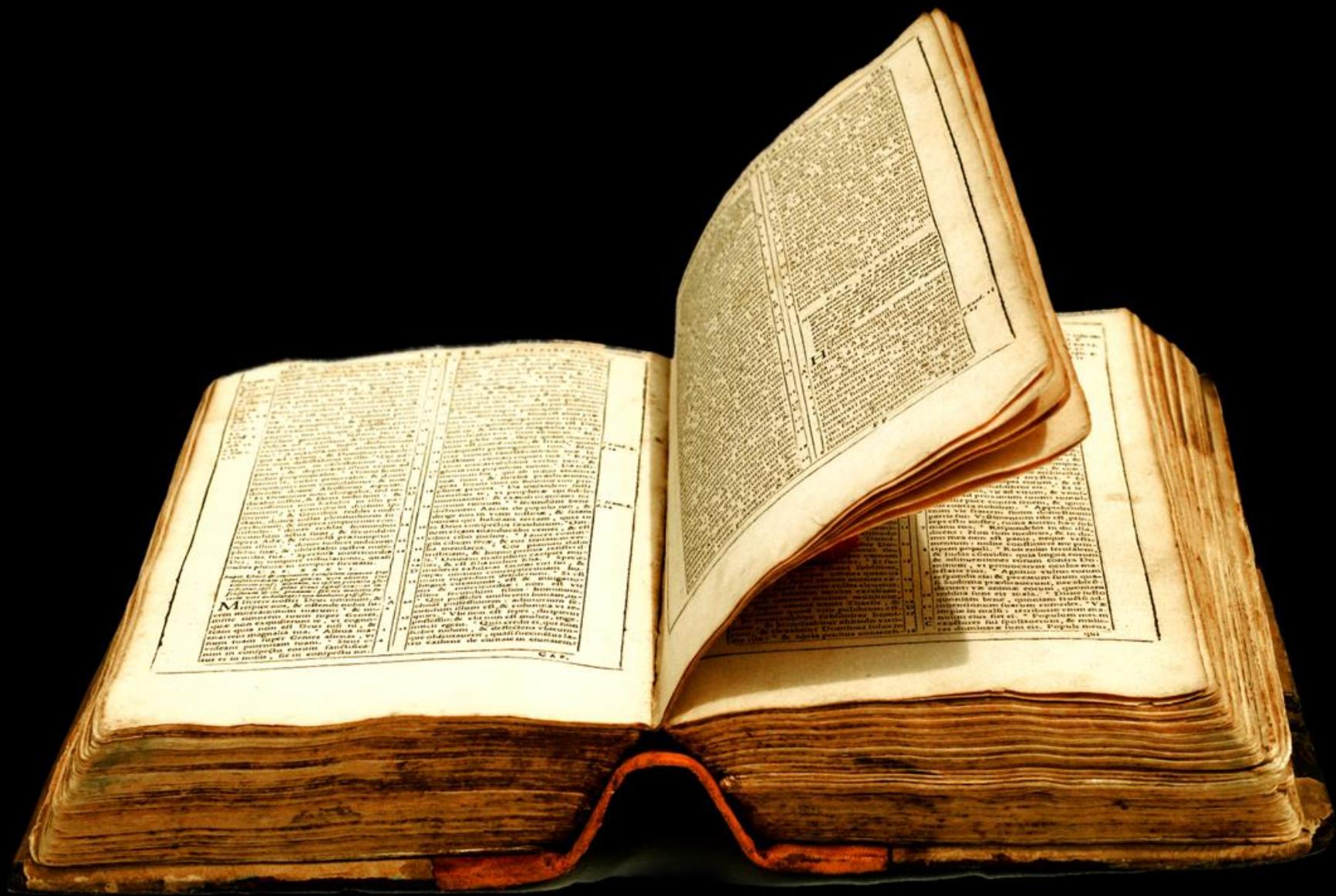
1. Corpus design
2. Corpus building strategy + quality criteria
3. Corpus acquisition [IPR]
4. Processing
 - analysis
 - conversion + metadata

[linguist – SE]



To Lukas Pruijs, 13 januari 1661 by Elisabeth Philipse Amelingh (1661)	... heeft ----- 4 gl - 0 Aen	boeken	en pampier en pennen en ...	boek	NOU
To Lukas Pruijs, 6 juni 1664 by Elisabeth Philipse Amelingh (1664)	... lijeue man stelt alles te	boeck	waet ghij uijt geft dat ...	boek	NOU
To Simon Been, 9 december 1664 by Jacob Been (1664)	... alzoo hij niet op de	bouck	en stondt en hij heeft ...	boek	NOU
To J.H. Martens, 11 november 1780 by J.D. Piest (1780)	... Schrijfpampier + + a 6 Sr t	boek.	50 ditto ongsneeden ditto a ...	boek	NOU
To Gabriel Lixraaven, 10 januari 1781 by Simon Quirijn Lixraaven (1781)	... voorne missive, dat UEDs mijn	boeken	bij berntrop laat verkoopen, als ...	boek	NOU
	... verbeeld hebbe, dat wat geleerde	boeken	aan gaat, schouten altoos geprafereert ...	boek	NOU
To David Wendorp, 16 december 1780 by M. Reuvenhagen (1780)	... beijde, accuratesse goede order der	boecken,	Voorsigtighejd, die prijse ik ...	boek	NOU
To Magdalena Wendorp-Bagge, 16 december 1780 by M. Reuvenhagen (1780)	... rollen matten MR Eenige	boeken	Maar also die twee Scheepen ...	boek	NOU
	... te Senden, wat de	boeken	aan betrefft Zo Gelieft maar ...	boek	NOU
To Christiaan Frederik Coller, december 1780 by JvC Jur Luijt (1780)	... laaten wagten tot ik mijn	boeken	heeft, want weet anders niet ...	boek	NOU
	... houw nu alle de	boeken	bonte die sijn ...	boek	NOU
To Christina Bakker-Smits, 26 september 1780 by Carsten Bakker (1780)	... gantsche regel die in myn	Boek	niet te vinden is, en ...	boek	NOU
To Maartje Nannings, 12 februari 1665 by Paulus Pietersz. Schuitemaker (1665)	... al vrij hoch bij hente	bock	soo het scijnt en men ...	boek	NOU
To Michiel Heusch jr., 19 oktober 1664 by Michiel Heusch sr. (1664)	... reeck. van hem wt v	boeken	Is heden versujmt wij sijn ...	boek	NOU
To Eldert Jansz., 6 november 1672 by Pieter Jansz. (1) (1672)	... maer ick behoefde wel een	boeck	pampier te hebben daerom sal ...	boek	NOU
To Willem Lukasz., november 1672 by Meintje Meinderts (1672)	... dan ick behoefde wel een	boeck	Pampier daerom sal ick opbreecken ...	boek	NOU
To Johan Godfried Meeler, 5 januari 1781 by Geertruyda Westman-Erasmus (1781)	... van concept en wilde syn	boeken	absolut ter secretarij versorgen Eydelyk ...	boek	NOU
	... 't Eijschen, van onvertueren van	boeken	ligten van copien Extracten of ...	boek	NOU





General principles corpus processing

- What is a text?
- What is the format?
- Each token in the text has the correct metadata
- Metadatascheme for all corpora
 - core metadataset
 - date text vs. date witness
 - collection specific metadata

Corpusprocessing for Nederlab

- 2013 – 30 June 2018
- Create a research environment for historians, linguists, literary scholars...
- Diachronic corpus of existing digital collections (ca. 500 – present)
- User-friendly and tool-enriched web interface for scholars
- Uniform text format + metadata
- Enrichment of data by team and by scholarly users
- Focus on data quality by including an editorial staff
- <http://www.nederlab.nl/onderzoeksportaal>
10 billion words currently; aimed at incorporating min. 20 different collections

Corpus processing

- Acquisition
- Analysis
- Conversion of text format > Folia
- Conversion of available metadata > relational database
- Curation of metadata (persons)

Huygens ING

- OCR (RGP 450 volumes)



- Text editions in TEI XML (GT)

Heinsius letters

- Daniel Heinsius (or Heins) (9 June 1580 – 25 February 1655) was one of the most famous scholars of the Dutch Renaissance.
- From Flanders to the Netherlands (Spanish)
- Prof of Latin and Greek in Leiden and librarian
- Publications: scientific and literary (poems, emblemata)



Letters

DE BRIEFWISSELING VAN
ANTHONIE HEINSIUS
1702—1720

DEEL 1
19 MAART — 31 DECEMBER 1702

Over 20.000 letters

Published in 19 volumes

Digitised and made available by Huygens ING

Bewerkt door
Dr. A.J. VEENENDAAL jr.

's-Gravenhage
verkrijgbaar bij MARTINUS NIJHOFF
1976

http://resources.huygens.knaw.nl/retroboeken/heinsius/#page=0&accessor=toc_1&view=homePane

Challenges

- OCR quality
- Parsing structure (format!)
- Editorial matter vs original text
- Metadata

OCR quality + format

Hreda:

PHTLIPPUB VnoBtxoa / Breda

I. D. PKTKHS PI,AS< HAART predican tot) Etten

Grn..iKi \n II.u:irnr- (Jilse.

Nevens d' heer PAIJLUS SNELEEN, schepen d e r stadt Hreda en ouder-
lingh aldaer.

U y t welcke aengekome broederen tot diroctores van de aenstaeiele
verhandelinge verkoren sijn, t o t :

Praeses: D . DAVID A M I A .

Assessor: D. DIONYSIÜS KEL< I .

Scriba: D. DANIEL VAN PEKNE.

De tijt van ordinaire vergaderinge is volgens gewoonte gestelt des
morgens van acht t o t elf u y r e n , en des naemiddaghs van drie tot
ses u y r e n ; behalven des Saturdaeghs van seven t o t tien uyren, en des
Maendaeghs van vier t o t seven u y r e n ; en yder absent sal met twaalf
-t u w e r s , en te laet komende m e t ses stuvers geboedt worden.

Sijn in dese Synodus verschenen de ordinaris Gedeputeerde des
Bynodi: I). D. JOIIANNRS VOLLENHOVE, S. S. Theol. doctor, THEODORUS
VAN e n LEK, Grunjiunm ANSLAEK en PETRUS DU BOIS, respective predi-
< uiten in 's-Gravenhage, L e y d e n , Delft en Gorichem.

Sijn door D. I). directores Synodi in dese vergaederinge geintroduceert
de Edele Hooghachtbare heeren Commissarisen van de Edele Groot-
Moogende heeren Staeten van Hollandt en West-Frieslandt, de heer
en meester HENDRIC DRUYF, raadordinaris in den Hoogen R a d e , ende
d ' a c h t b a r e heer en meester Puuroo M' B o » , borgenneester der stadt

ependant ie leur suis obligé de la permission qu'elles m'ont accordée pour la faire sortir en payant les droits. Je suis . . .
N. Lillieroot.

Le 3 ḡbre 1702.

1073. van VAN REEDE VAN DE LIER, 3 november [1702].
Eigenh. orig. H.A. 788.

Weledele Gestrenge Heer, Desen sal alleen dienen tot recipisce van Uw E.¹⁸ van den 27 octob.¹ en sullen wij ons nae derselver teneur reguleren. Ick kan niet wel begrijpen wat Uw E. belieft te verstaen met het recommanderen een H.^r grave Malborough, ten waer Uw E. daermede meent het compliment aan mij voor desen gedaen; so ick het heb geraden ben ick daerover aan Uw E. verobligert, maar versoeke dat daervan niet gelieve te tonen de minste kennisse te hebben. Ick desperere daer nogh niet van, overmits mylord thresorer uyt sijnen eygen movenment niet lange geleden, al-hoewel in duystere termen, daervan heeft gesproken; het moet met goede gracie geschieden of ick soude daer van afsien en terwijlen ick togh geresolveert ben mij uyt het gewoel te begeven en op het lant te gaen uytstruwen, scheelt het mij weynigh, maar ick kan niet afsyn mijn beklagh te doen over de missive van Haer E. Gr. Mog. geschreven aen de heeren Staten van Utrecht dat men met recht een Uria's brief magh noemen en waerdoor men mij genoeghsaem vastmaekt aen de bank van justitie aldaer, regelreight tegen mijne intentie ende alle billichet, sijnende een vasal van Hollant en het goet waerover de questie is leenoerigh aen deseelve provincie.² Maer als Haer Ed.G.Mog. de saek te degen beliefden in te sien, als de justitie 'tsij waer, hetsey haer daermede bemoeit is het tot deklyn en veragting van haere souvereine macht uyt welkers hoofde sij die resolutie hebben genomen. Nu ick wil hopen dat alles sal werden geredresseert en versoekte Uw E. gunste daerontrent, waermede blijve . . .
F.b. van Reede.

Londen, den 3 nov.

1074. van SAUNIERE DE L'HERMITAGE, 3 november 1702.
Eigenh. orig. H.A. 792.

Nouvelles uit Londen over de toespraak van de koningin in het Parlement, het mislukken van de expeditie naar Cadiz en het aanwijzen van de schuldigen daaraan, de unie met Schotland en de houding van Nottingham tegenover de Anglicaanse kerk.

1075. van VAN VRIJBERGEN, 3 november 1702.
Eigenh. orig. H.A. 800.

Hooghedele Gestrenge Heer, Desen namiddagh ontfongh men dry Hollantse maelen en daermede de gewenschte tijdinge van het bemagtighen der citadelle van Luyck, dat hier veel vreughe geeft.

Het Lagerhuys heeft heeden nemine contradicente geresolveert Haer Maj.¹ in alles bij te staen tot het volvoeren haerer alliantieën en tot inteuigelinge der ongelimiteerde ambitie van Vranckrijck. Het adres sal morgen door commissarissen worden opgestelt. Mij dunckt dat men alle de reeden van de werelt heeft om te vertrouwen dat dese sessiën met gewenscht succes en spoedigh sullen aflopen.

D'heeren Van de Lier en Van Haeren hebben mij gecommuniceert Haer Ho. Mog. resolutie en ordres van de 27^e o c t o b . behelsende de depeschés voor de heer Schonenbergh ende de mesures bij dit hoff en den Staat omtrent de croon Portugaels te

1073. 1. Deze brief is niet gevonden.
2. Zie over deze zaak hiervóór p. 474.

ependant ie leur suis obligé de la permission qu'elles m'ont accordée pour la faire sortir en payant les droits. Je suis - - N. Lillieroot.

Le 3 9 b r c 1702.

1073. van VAN REEDE VAN DE LIER, 3 november (1 7 0 2).
Eigenh. orig. H.A. 788.

Weledele Gestrenge Heer, Desen sal alleen dienen tot recipisce van Uw E. t s van den 27 o c t o b . 1 en sullen wij ons nae derselver teneur reguleren. Ick kan niet wel begrijpen wat Uw E. belieft te verstaen met het recomandieren een H. r grave Malborough, ten waer Uw E. daermede meent het compliment aan mij voor desen gedaen; so ick het heb geraden ben ick daerover aan Uw E. verobligert, maar versoeke dat daervan niet gelieve te tonen de minste kennisse te hebben. Ick desperere daer nogh niet van, overmits mylord thresorer uyt sijnen eygen movenment niet lange geleden, al-hoewel in duystere termen, daervan heeft gesproken; het moet met goede gracie geschieden of ick soude daer van afsien en terwijlen ick togh geresolveert ben mij uyt het gewoel te begeven en op het lant te gaen uytstruwen, scheelt het mij weynigh, maar ick kan niet afsyn mijn beklagh te doen over de missive van Haer E. Gr. Mog. geschreven aen de heeren Staten van Utrecht dat men met recht een Uria's brief magh noemen en waerdoor men mij genoeghsaem vastmaekt aen de bank van justitie aldaer, regelreight tegen mijne intentie ende alle billichet, sijnende een vasal van Hollant en het goet waerover de questie is leenoerigh aen deseelve provincie. Maer als Haer Ed.G.Mog. de saek te degen beliefden in te sien, als de justitie 'tsij waer, hetsey haer daermede bemoeit is het tot deklyn en veragting van haere souvereine macht uyt welkers hoofde sij die resolutie hebben genomen. Nu ick wil hopen dat alles sal werden geredresseert en versoekte Uw E. gunste daerontrent, waermede blijve - - F.b. van Reede.

Londen, den 3 nov.

1074 . van SAUNIERE DE L'HERMITAGE, 3 november 1 7 0 2 .
Eigenh. orig. H.A. 792.

Nouvelles uit Londen over de toespraak van de koningin in het Parlement, het mislukken van de expeditie naar Cadiz en het aanwijzen van de schuldigen daaraan, de unie met Schotland en de houding van Nottingham tegenover de Anglicaanse kerk.

1075 . van VAN VRIJBERGEN, 3 november 1 7 0 2 .
Eigenh. orig. H.A. 800.

Hooghedele Gestrenge Heer, Desen namiddagh ontfongh men dry Hollantse maelen en daermede de gewenschte tijdinge van het bemagtighen der citadelle van Luyck, dat hier veel vreughe geeft.

Het Lagerhuys heeft heeden nemine contradicente geresolveert Haer Maj. 1 in alles bij te staen tot het volvoeren haerer alliantieën en tot inteuigelinge der ongelimiteerde ambitie van Vranckrijck. Het adres sal morgen door commissarissen worden opgestelt. Mij dunckt dat men alle de reeden van de werelt heeft om te vertrouwen dat dese sessiën met gewenscht succes en spoedigh sullen aflopen.

D'heeren Van de Lier en Van Haeren hebben mij gecommuniceert Haer Ho. Mog. resolutie en ordres van de 27 e o c t o b . behelsende de depeschés voor de heer Schonenbergh ende de mesures bij dit hoff en den Staat omtrent de croon Portugaels te

1073. 1. Deze brief is niet gevonden.
2. Zie over deze zaak hiervóór p. 474.

Re-OCR

- Abbyy FineReader Engine CLI for Linux [OCR4Linux.com] (Finereader 11)
- Parameters // Engine and ABBYY XML and ALTO available
- 120.000 pages for 1000 euro
- 10 seconds per page; OCR with 6 threads

Formats

Plain text

Docx/RTF/HTML/....

PDF

XML

- Alto

- → Abbyy xml

- Why?

- Blocks and paragraphs both explicitly coded
 - Detailed layout information on glyph level
 - Other details (separators, etc)

Abbyy XML

```
▼<par align="Justified" startIndent="1200" lineSpacing="1170">
  ▼<line baseline="1136" l="328" t="1109" r="1448" b="1143">
    ▼<formatting lang="DutchStandard" ff="Liberation Serif" fs="9.5">
      ▼<wordRecVariants>
        ▼<wordRecVariant wordFromDictionary="0" wordNormal="1" wordNumeric="0" wordIdentifier="0" wordPenalty="5" meanStrokeWidth="29">
          ▼<variantText>
            Gerritsz.,
          ▼<charParams l="328" t="1113" r="353" b="1137" charConfidence="100" serifProbability="40">
            ▼<charRecVariants>
              <charRecVariant charConfidence="100" serifProbability="40">G</charRecVariant>
              <charRecVariant charConfidence="100" serifProbability="40">g</charRecVariant>
              <charRecVariant charConfidence="23" serifProbability="0">Ü</charRecVariant>
              <charRecVariant charConfidence="23" serifProbability="0">Ü</charRecVariant>
              <charRecVariant charConfidence="23" serifProbability="0">û</charRecVariant>
              <charRecVariant charConfidence="23" serifProbability="0">ü</charRecVariant>
              <charRecVariant charConfidence="16" serifProbability="52">Q</charRecVariant>
              <charRecVariant charConfidence="16" serifProbability="52">q</charRecVariant>
              <charRecVariant charConfidence="15" serifProbability="0">Ö</charRecVariant>
              <charRecVariant charConfidence="15" serifProbability="0">ö</charRecVariant>
```

ALTO

```
▼<TextBlock ID="Page1_Block4" HEIGHT="1778" WIDTH="1051" VPOS="304" HPOS="169" language="nl">
  ▼<Shape>
    <Polygon POINTS="171,305 1221,305 1221,2080 1116,2080 1116,2041 171,2041 171,305"/>
  </Shape>
  ▼<TextLine HEIGHT="25" WIDTH="135" VPOS="309" HPOS="625">
    <String WC="0.7966666818" CONTENT="Artic." HEIGHT="25" WIDTH="79" VPOS="309" HPOS="625"/>
    <SP WIDTH="15" VPOS="309" HPOS="705"/>
    <String WC="0.6566666961" CONTENT="28." HEIGHT="24" WIDTH="39" VPOS="309" HPOS="721"/>
  </TextLine>
  ▼<TextLine HEIGHT="35" WIDTH="1007" VPOS="371" HPOS="206">
    <String WC="0.4939999878" CONTENT="[Betalinge" HEIGHT="34" WIDTH="140" VPOS="371" HPOS="206"/>
    <SP WIDTH="21" VPOS="381" HPOS="347"/>
    <String WC="1." CONTENT="van" HEIGHT="15" WIDTH="47" VPOS="382" HPOS="369"/>
    <SP WIDTH="23" VPOS="374" HPOS="417"/>
    <String WC="0.4679999948" CONTENT="leger" HEIGHT="31" WIDTH="63" VPOS="374" HPOS="441"/>
    <SP WIDTH="1" VPOS="373" HPOS="505"/>
    <String WC="0.4099999964" CONTENT="lasten.]" HEIGHT="34" WIDTH="95" VPOS="372" HPOS="505"/>
    <SP WIDTH="26" VPOS="372" HPOS="601"/>
    <String WC="1." CONTENT="-" HEIGHT="2" WIDTH="30" VPOS="387" HPOS="628"/>
    <SP WIDTH="25" VPOS="373" HPOS="659"/>
    <String WC="0.375" CONTENT="Ad" HEIGHT="25" WIDTH="40" VPOS="373" HPOS="685"/>
    <SP WIDTH="22" VPOS="375" HPOS="726"/>
    <String WC="0.5575000048" CONTENT="art." HEIGHT="23" WIDTH="46" VPOS="375" HPOS="749"/>
    <SP WIDTH="17" VPOS="375" HPOS="796"/>
    <String WC="0.4300000072" CONTENT="24," HEIGHT="29" WIDTH="44" VPOS="374" HPOS="814"/>
    <SP WIDTH="17" VPOS="382" HPOS="859"/>
```

Automatic conversion to TEI

```
</p>
▼<p facs="#zone.00577" rend="line-height:9.4pt" xml:lang="nl">
  ▼<hi rend="font-size:8.5; font-family:Liberation Serif" xml:lang="nl">
    <w facs="#zone.00572" xml:id="w.516">Amst.,</w>
    <w facs="#zone.00573" xml:id="w.517">den</w>
  ▼<w facs="#zone.00574" xml:id="w.518">
    25
    <hi rend="font-size:8.5; font-family:Liberation Serif; vertical-align:super" xml:lang="nl">e</hi>
    </w>
  </hi>
  ▼<hi rend="font-size:8.5; font-family:Liberation Serif" xml:lang="nl">
    <w facs="#zone.00575" lrx="1363" lry="1741" type="p" ulx="267" uly="1676" xml:id="zone.00570"/>
    <w facs="#zone.00576" lrx="675" lry="1777" type="lb" ulx="268" uly="1749" xml:id="zone.00571">
      ▼<zone lrx="364" lry="1777" type="w" ulx="268" uly="1752" xml:id="zone.00572"/>
      <zone lrx="422" lry="1773" type="w" ulx="365" uly="1752" xml:id="zone.00573"/>
      <zone lrx="477" lry="1773" type="w" ulx="423" uly="1749" xml:id="zone.00574"/>
      <zone lrx="603" lry="1773" type="w" ulx="478" uly="1752" xml:id="zone.00575"/>
      <zone lrx="675" lry="1774" type="w" ulx="604" uly="1752" xml:id="zone.00576"/>
    </zone>
    <zone lrx="675" lry="1777" type="p" ulx="268" uly="1749" xml:id="zone.00577"/>
  </zone>
  ——— lrx="1363" lry="1741" type="p" ulx="267" uly="1676" xml:id="zone.00570" ——— lrx="675" lry="1777" type="p" ulx="268" uly="1749" xml:id="zone.00577">
```

Structure recognition: notes

```
▼<div facs="#zone.00647">
  ▼<note facs="#zone.00656" rend="line-height:9.4pt; margin-left:2300" xml:lang="nl">
    ▼<hi rend="font-size:8.5; font-family:Liberation Serif" xml:lang="nl">
      <w facs="#zone.00649" xml:id="w.580">4.</w>
      <w facs="#zone.00650" xml:id="w.581">Zie</w>
      <w facs="#zone.00651" xml:id="w.582">hiervóór</w>
      <w facs="#zone.00652" xml:id="w.583">p.</w>
      <w facs="#zone.00653" xml:id="w.584">480</w>
      <w facs="#zone.00654" xml:id="w.585">en</w>
      <w facs="#zone.00655" xml:id="w.586">496.</w>
    </hi>
    <lb facs="#zone.00648"/>
  </note>
</div>
<div facs="#zone.00657" type="separator"/>
</body>
```

-
1073. 1. Deze brief is niet gevonden.
2. Zie over deze zaak hiervóór p. 474.

Mog. resolutie en ordres van de 27e o c t o b . behelsende de depeschés voor de heer Schonenbergh ende de mesures bij dit hoff en den Staet omtrent de croon Portugael te

1073. 1. Deze brief is niet gevonden.
2. Zie over deze zaak hiervóór p. 474.

Structure recognition: letter segmentation

Task:

- find start and end of letter
- remove editorial matter

Method: approximate matching of (corrected)
table of contents + typographical structure

Would not have been possible with less structured
formatting information

cependant je leus suis obligé de la permission qu'elles m'ont accordée pour la faire sortir en payant les droits. Je suis —
N. Lillieroot.

Le 3 ḡbre 1702.

van Reede van de Lier, 3 11 1702.

1073. van VAN REEDE VAN DE LIER, 3 november [1702].
Eigenh. orig. H.A. 788.

Weledele Gestrenge Heer, Desen sal alleen dienen tot recipisce van Uw E^{ts} van den 27 octob¹ en sullen wij ons nae derselver teneur reguleren. Ick kan niet wel begrijpen wat Uw E. belieft te verstaen met het recommanderen aen H^t grāve Malborough, ten waere Uw E. daermēde meent het compliment aen mij voor desen gedaen; so ick het heb geraden ben ick daerover aan Uw E. verobligēert, maer versoeke dat daervan niet gelieve te tonen de minste kennisse te hebben. Ick desperere daer nogh niet van, overmits mylord thresorer uyt sijn eygen mouvement niet lange geleden, alhoewel in dyustere termen, daervan heeft gesproken; het moet met goede gracie geschieden of ick soude daer van afsien en terwijlen ick togh geresolveert ben mij uyt het gevoel te begeven en op het lant te gaen uytrusten, scheelt het mij weynigh, maar ick kan niet afsyn mijn beklagh te doen over de missive van Haer E. Gr. Mog. geschreven aen de heeren Staten van Utrecht dat men met reight een Urias brief magh noemen en waerdoor men mij genoeghsaem vastmaekt aen de bank van justitie aldaer, regelrecht tegen mijne intentie ende alle billicheyt, sijnde een vasal van Hollant en het goet waerover de questie is leenroerigh aen deselve provintie.² Maer als Haer Ed.G.Mog. de saek te degen beliefdēn in te sien, als de justitie 'tsij waer, hetsey haer daermēde bemoeit is het tot deklyn en veragting van haere souvereine maght uyt welkers hoofde sij die resolutie hebben genomen. Nu ick wil hopen dat alles sal werden geredresseert en versoeke Uw E. gunste daerontrent, waermēde blijve — F.b. van Reede.

Londen, den 3 nov.

van Sauniere de Hermitage, 3 11 1702.

1074. van SAUNIERE DE L'HERMITAGE, 3 november 1702.
Eigenh. orig. H.A. 792.

Nouvelles uit Londen over de toespraak van de koningin in het Parlement, het mislukken van de expeditie naar Cadiz en het aanwijzen van de schuldigen daaraan, de unie met Schotland en de houding van Nottingham tegenover de Anglicaanse kerk.

van Vrijbergen, 3 11 1702.

1075. van VAN VRIJBERGEN, 3 november 1702.
Eigenh. orig. H.A. 800.

Hooghedele Gestrenge Heer, Desen namiddagh ontfongh men dry Hollantse maelen en daermēde de gewenschte tijdinge van het bemagtighen der citadelle van Luycx, dat hier veel vreugde heeft.

Het Lagerhuyz heeft heeden nemine contradicente geresolveert Haer Maj^s in alles bij te staen tot het volvoeren haer alliantie en tot inteuigelinghe der ongelimiteerde ambitie van Vranckrijck. Het adres sal morgen door commissarissen worden opgesteld. Mij denkt dat men alle de reden van de werelt heeft om te vertrouwen dat dese sessien met gewencht succes en spoedigh sullen afloopen.

D'heeren Van de Lier en Van Haeren hebben mij gecommuniceert Haer Ho. Mog. resolute en ordtes van de 27^e octob. behelsende de depesches voor de heer Schonenbergh ende de mesures bij dit hoff en den Staat omtrent de croon Portugal te

1073. 1. Deze brief is niet gevonden.

2. Zie over deze zaak hiervoor p. 374.

▼<div2 type="letter">

▼<head>

1073. van VAN REEDE VAN DE LIER, 3 november [1702].

<lb>

Eigenh. orig. H.A. 788.

</head>

▼<p>Weledele Gestrenge Heer, Desen sal alleen dienen tot recipisce van Uw E.

<hi rend="superscript">ts</hi>

<lb>

van den 27 octob.

<hi rend="superscript">1</hi>

en sullen wij ons nae derselver teneur reguleren. Ick kan niet wel

<lb>

begrijpen wat Uw E. belieft te verstaen met het recommanderen aen H.

<hi rend="superscript">r</hi>

grave Malbo-

<lb>

rough, ten waere Uw E. daermēde meent het compliment aen mij voor desen gedaen;

<lb>

so ick het heb geraden ben ick daerover aan Uw E. verobligēert, maer versoeke dat

<lb>

daervan niet gelieve te tonen de minste kennisse te hebben. Ick desperere daer nogh

<lb>

niet van, overmits mylord thresorer uyt sijn eygen mouvement niet lange geleden, al-

<lb>

hoewel in dyustere termen, daervan heeft gesproken; het moet met goede gracie geschie-

<lb>

den of ick soude daer van afsien en terwijlen ick togh geresolveert ben mij uyt het

<lb>

gewoel te begeven en op het lant te gaen uytrusten, scheelt het mij weynigh, maar ick

<lb>

kan niet afsyn mijn beklagh te doen over de missive van Haer E. Gr. Mog. geschreven

<lb>

aen de heeren Staten van Utrecht dat men met reight een Urias brief magh noemen en

<lb>

waerdoor men mij genoeghsaem vastmaekt aen de bank van justitie aldaer, regelrecht

<lb>

tegen mijne intentie ende alle billicheyt, sijnde een vasal van Hollant en het goet waer-

<lb>

over de questie is leenroerigh aen deselve provintie.

<hi rend="superscript">2</hi>

Maer als Haer Ed.G.Mog. de saek

<lb>

te degen beliefdēn in te sien, als de justitie 'tsij waer, hetsey haer daermēde bemoeit is

<lb>

het tot deklyn en veragting van haere souvereine maght uyt welkers hoofde sij die

<lb>

resolutie hebben genomen. Nu ick wil hopen dat alles sal werden geredresseert en ver-

<lb>

soeke Uw E. gunste daerontrent, waermēde blijve — F.b. van Reede.

</p>

<p>Londen, den 3 nov.</p>

</div2>

Metadata

- The collection of Heinsius' letters is written by over 200 different authors
- Identification of authors + biographical information

Corpus Hedendaags Nederlands

chn.inl.nl

- Integration: 5, 27, 38 miljoen woordencorpora, Parolecorpus, monitorkrantencorpus (internal vs. external)
 - format
 - metadata scheme
- Surinam, Dutch Caribbean
- Corpus workflow
- Corpus design future

Corpus workflow

Workflow tool:

Steps:

1. FTP (NRC/Standaard)
2. Conversion to TEI (utf8-pid; fixup-tei)
3. Remove duplicates (content/source-id)
4. Linguistic annotation
5. Indexing with BlackLab (incremental)
6. Errors
7. Maandelijkse export naar Neoloog
8. *Automatic update*

Build your own corpus

- Born digital
- Digitisation
 - IMPACT CoC (digitisation.eu)/ advice on digitisation
 - PICCL pipeline (CLARIAH)
- Conversion:
 - conversion tools
 - Open Convert
 - (Openconvert.clarin.inl.nl)