

Genegeerd materiaal brengt licht in middeleeuwse handschriften

Paleografen zetten hun horloges weer gelijk

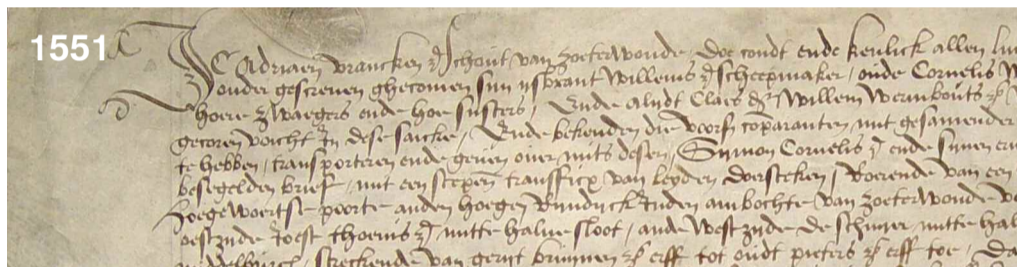
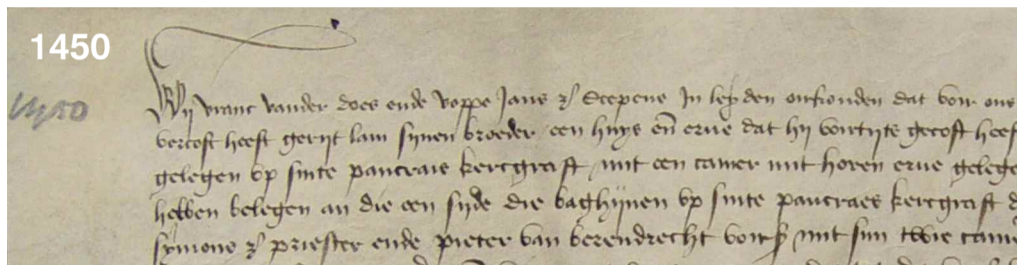
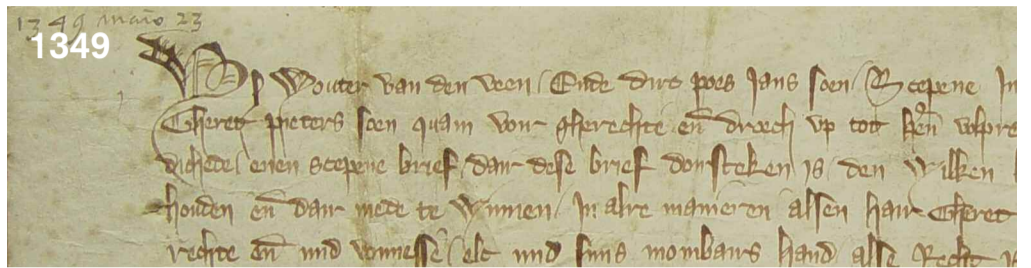
Waar en wanneer werd dat middeleeuwse handschrift geschreven? Een vraag voor de computer. *Jan Burgers en Lambert Schomaker*

Lambert Schomaker

Onderzoekers die zich voor de middeleeuwen interesseren, hebben vaak een probleem. De primaire bronnen die ze bestuderen, handschriften, dragen geen aanduiding van plaats en tijd. Daarmee wordt hun onderzoek bemoeilijkt, want de vraag naar het ontstaan is meestal de eerste die aan een bron wordt gesteld. Gewoonlijk wordt die vraag beantwoord op basis van schriftkenmerken, en daartoe is slechts een handvol gespecialiseerde paleografen in staat. Maar ook die komen lang niet altijd tot eensluidende conclusies, omdat hun expertise vaak meer is gebaseerd op intuïtie dan op objectieve criteria, juist vanwege het gebrek aan voldoende gedateerde manuscripten.

Paleografische meetlat

Door NWO is onlangs subsidie toegekend aan een project om te komen tot een meer objectief, geautomatiseerd procedé van datering en lokalisering van laat-middeleeuws schrift. In het kader van dit project, getiteld *MPS, the Medieval Palaeographic Scale for dating and localizing handwritten manuscripts using digital palaeography*, aangevraagd door de Universiteit van Amsterdam, de Rijksuniversiteit Groningen en het Huygens Instituut voor Nederlandse geschiedenis, zullen een historicus en een informaticus een paleografische ‘meetlat’ construeren met daarin de dateer- en lokaliseerbare elementen in laatmiddeleeuws schrift (1300-1550). Dit zal gebeuren op basis van materiaal dat door paleografen tot nu toe vrijwel volledig is genegeerd, te weten oorkonden en andere archivalische documenten, die meestal precies gedateerd en goed te lokaliseren zijn. Bovendien zijn zij vaak ge-



Typische handschriften uit 1349, 1450 en 1551 bron: Leids Archief, Stadsarchief I

schreven door dezelfde (professionele) scribenten die ook de ongedateerde handschriften vervaardigden, veelal in vergelijkbare schrifttypen.

Doorbraak voor mediëvistiek

Aan de hand van dit corpus, samengesteld in enkele stadsarchieven, zullen de onderzoekers de meetlat vervaardigen. Daarbij maken zij gebruik van de traditionele paleografische expertise, en werken daarnaast aan software die het schrift op een dieper niveau analyseert. Dit nieuwe terrein van de ‘digitale paleografie’ heeft al veelbelovende resultaten opgeleverd: door de Groningse groep van Schomaker is bijvoorbeeld programmatuur ontwikkeld voor de automatische schrijveridentificatie in veertiende-eeuwse oorkonden. Het dateren is een vergelijkbaar probleem, gericht

op de ontdekking van algemene, in de tijd ontwikkelende stijlgroepen. Wanneer de ‘meetlat’ werkt, zou dat een doorbraak betekenen voor het brede veld van de internationale mediëvistiek. Ook niet-specialisten zullen in staat zijn tijd en plaats van vervaardiging van hun bronnen vast te stellen, in ieder geval bij benadering, en de inzet van de computer zal dit tijdrovende aspect van het onderzoek versnellen. Nu bovendien steeds meer afbeeldingen van middeleeuwse handschriften op internet verschijnen, zal het mogelijk zijn de *palaeographic scale* breed in te zetten en al doende te verfijnen. En deze belangrijke stap in de digitale paleografie brengt ons weer dichterbij de heilige graal op dat terrein, namelijk programma’s die de computer in staat stellen handschrift te lezen.

Een mooi tweede leven voor oude data

Afgestoft en opgepoetst: van bureaula naar Datacuratie-service. *Erica Renckens*

Elke wetenschapper heeft wel eens een harde schijf (of misschien zelfs diskette) met oude data. Zelf verzamelde tekstbestanden die ooit, jaren terug, essentieel waren voor eigen onderzoek. Maar de focus van dat onderzoek is in de loop der tijd verschoven en de data komen niet meer van pas. Dat betekent niet dat de data waardeloos zijn. Andere wetenschappers kunnen de data misschien nog goed gebruiken. Maar dan moeten ze wél weten dat de data bestaan en moeten ze er toegang toe krijgen. Dat is het doel van de Datacuratieservice (DCS), gevestigd bij de Radboud Universiteit in Nijmegen, sinds afgelopen januari

officieel van start. “Het bestuur van CLARIN-NL moest regelmatig aanvragen voor standaardisatie afwijzen, terwijl de data in kwestie wel een aanvulling op de CLARIN-infrastructuur zouden zijn”, vertelt Nelleke Oostdijk, projectleider van de Datacuratieservice. “Dankzij de DCS kunnen deze data nu toch worden afgestoft en opgepoetst voor een tweede leven.”

Eén standaard

Momenteel neemt Oostdijk, samen met coördinator Henk van den Heuvel, de Dutch Bilingual Database onder handen. “Het grootste deel van het werk bestaat uit de inventarisatie voorafgaand aan de conversie”, legt Oostdijk uit. Niet alle data liggen op dezelfde locatie: een deel ligt bij het Max Planck Instituut,

een deel bij het Meertens Instituut. “Vervolgens converteren we de data naar één CLARIN-standaard”, aldus Oostdijk. “Want dat werkt het fijnste voor toekomstige gebruikers.”

Licentieconstructie

Mogen alle data zomaar worden hergebruikt? “Nee, en dat is zeker een probleem. Tegenwoordig is het bij de meeste onderzoeksaanvragen een vereiste dat alle IPR (*Intellectual Property Rights*, red.) geregeld zijn, maar vroeger werd daar niet naar gekeken. De wetenschapper verzamelde de data voor eigen gebruik; het regelen van de rechten, zodat andere wetenschappers er ook mee zouden kunnen werken, kostte alleen maar moeite en leverde de onderzoeker zelf geen voordeel op. We werken aan een oplossing. Je

kunt bijvoorbeeld denken aan een licentieconstructie, waarbij de gebruiker alleen toegang tot de data krijgt als hij een licentie ondertekent”, aldus Oostdijk.

De projectleider heeft goede hoop voor de toekomst van DCS. Onderzoekers met oude data kunnen DCS benaderen, maar DCS gaat ook zelf op zoek naar interessante data. Het bestuur van CLARIN-NL beslist vervolgens welke datasets onder handen worden genomen. “Ik ga er vanuit dat, als we hebben laten zien verantwoord om te gaan met IPR-kwesties, de meerwaarde van de Datacuratieservice binnen de geesteswetenschappen breed gedragen zal worden.”

www.ru.nl/letteren/datacuratieservice/

E-DATA & RESEARCH

Jaargang 6 | nummer 4

Nieuwsbrief over data en onderzoek in de alfa- en gamma-wetenschappen.

E-data & Research verschijnt drie keer per jaar en wordt mogelijk gemaakt door: DANS, CentERdata, CBS, CLARIN-NL, Huygens ING, Koninklijke Bibliotheek en de Vereniging voor Geschiedenis en Informatica.

INHOUD

3 Waar sta jij in het netwerk van onderzoekers?

3 Multicultureel panel kleurt kennis over onze cultuur

4 Laten zien hoe goed je database of dataset al is

5 Schuyt over integriteit onderzoeksdata



6 Schatten van de Bodleian Library onder de loep

6 E-science binnenkort in hogere versnelling

7 16,5 miljoen Hollanders in acht variabelen

8 Biografisch portaal wordt semantisch netwerk

8 Ewoud Sanders kan sneller scannen



Scan deze QR code met een smartphone om de website van E-data te bezoeken.
www.edata.nl