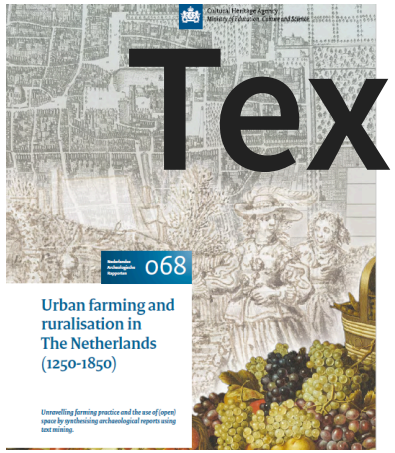
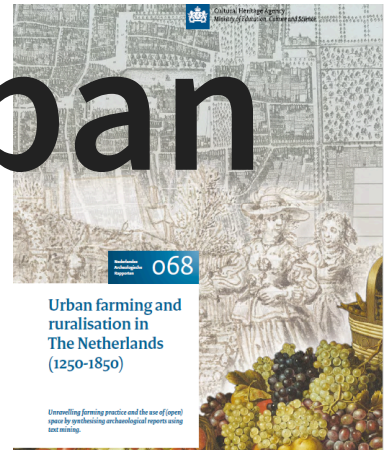


Text Mining for Urban Farming



Gaining new insights by synthesising archaeological and archaeobotanical research with the help of computational methods and conventional analyses

Ronald M. Visser

Anja D. Fischer

Annika L. Blonk-van den Bercken

Heleen van Londen

https://ronaldvisser.github.io/TextMiningUrbanFarming_IWGP2025

SAXION UNIVERSITY OF APPLIED SCIENCES
NMF erfgoedadvies
UNIVERSITY OF AMSTERDAM

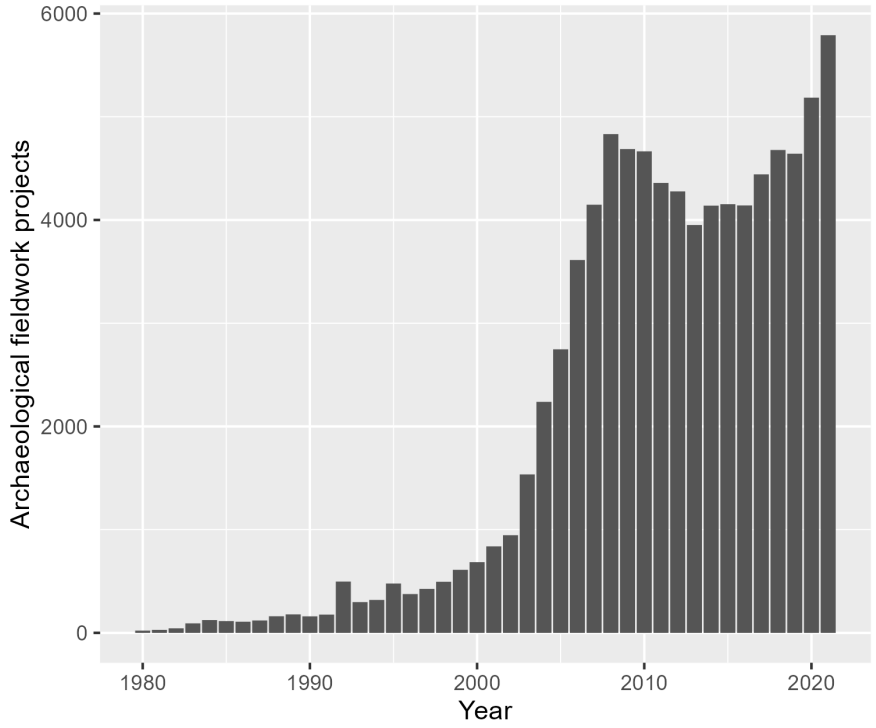
Structure

- Context and problem definition
  - Malta archaeology
  - Urban Farming
  - Archaeological data
- Text mining: a new method
- Results
- Conclusion

Context: Malta-archaeology

Netherlands:

- Decades of development-led archaeology
- Many projects
- Many excavation reports



(Source: Visser 2025, fig. 03.18)

Syntheses > Valletta Harvest

- Cultural Heritage Agency of the Netherlands (RCE)
- Harvesting archaeological reports
- Synthesis of themes
- Dutch overview: https://www.cultureelerfgoed.nl/onderwerpen/bronnen-en-kaarten/overzicht/synthese-archeologische-onderzoeksrapporten

Aim of project

- Overview of urban farming in towns
- Inventory of archaeological information
- Identify broad patterns

Urban farming

“the use of land in, and directly around a town for agricultural activities with the aim of producing food”

- arable farming
- horticulture
- orchards
- livestock farming
- fish farming

Dutch towns

Cities from “Atlas of the Dutch urban landscape: a millennium of spatial development” (Rutte and Abrahamse 2016)

- Towns
- Failed towns

Towns		
Alkmaar	Enkhuizen	Oldenzaal
Amersfoort	Franeker	Purmerend
Amsterdam	Goes	Roermond
Arnhem	Gorinchem	Rotterdam
Bergen op Zoom	Gouda	Schiedam
Bolsward	Groningen	Sneek
Breda	Haarlem	Tiel
Coevorden	Harlingen	Utrecht
Culemborg	Hoorn	Venlo
Delft	Kampen	Vlaardingen
Den Bosch	Leeuwarden	Vlissingen
Den Haag	Leiden	Weesp
Deventer	Maastricht	Wijk bij Duurstede
Doesburg	Medemblik	Woerden
Doetinchem	Middelburg	Zaltbommel
Dordrecht	Monnickendam	Zierikzee
Edam	Naarden	Zutphen
Eindhoven	Nijmegen	Zwolle

Failed towns		
Ameide *	Hardenberg	Ravenstein
Asperen	Hattem	Rijssen
Batenburg	Heukelum	Sluis
Bredevoort	IJsselstein	Staverden *
Buren	Leerdam	Steenwijk
Delden	Megen	Veere
Gennep	Montfoort	Vianen
Goor	Nieuwpoort	Vollenhove
Grave	Ommen	Vreeland *
Hagestein *	Oudewater	Woudrichem

(Source: Fischer et al. 2021, Table 3.1)

Collect archaeological reports

- 84 Towns
- Jacob van Deventer maps
- Buffer 600 metre around centre
- Select excavations from Archis 3 (https://archis.cultureelerfgoed.nl)

Nijmegen with buffer of 600 metre

Selection of data

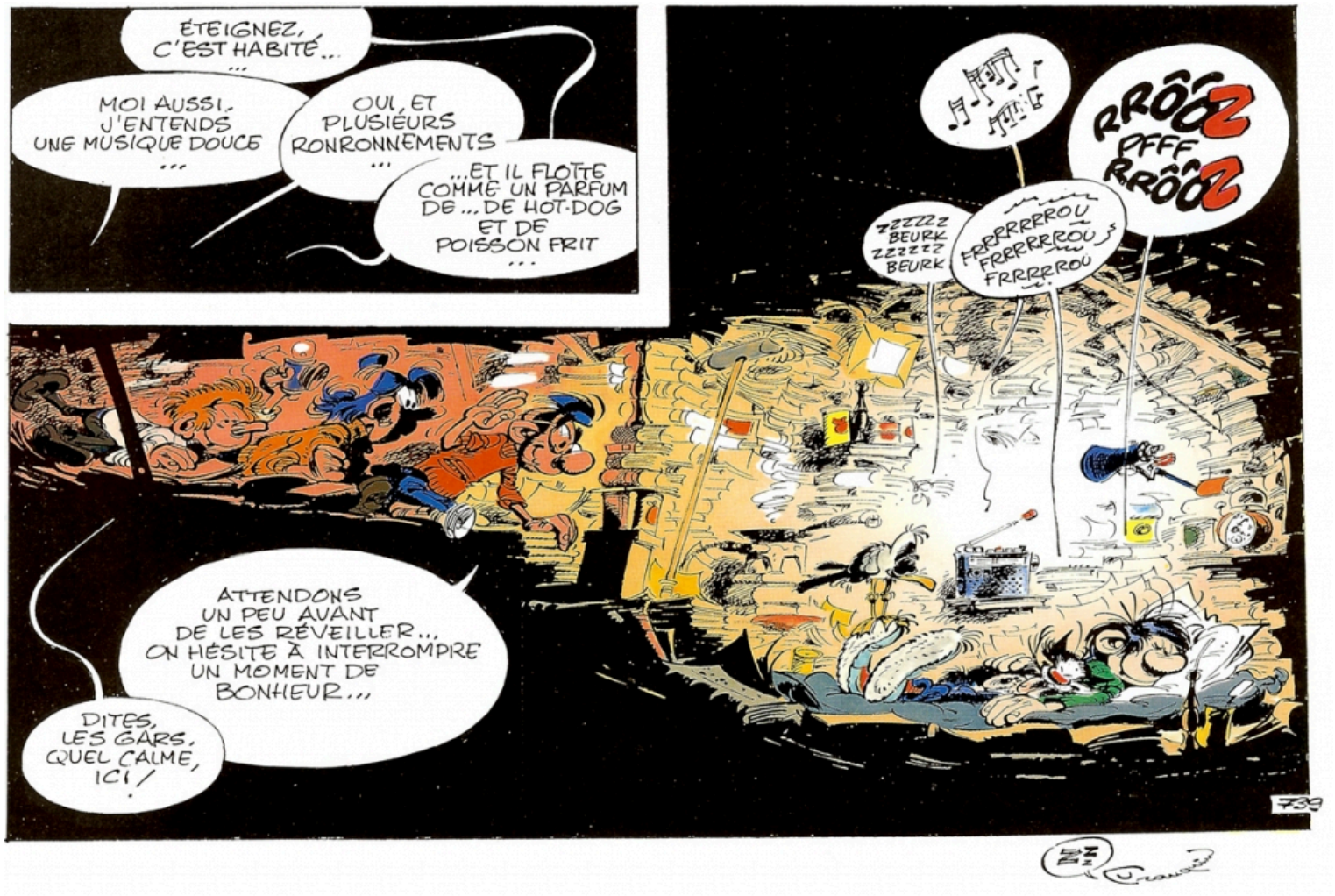
- Excavation between 1997-2017
- 2,278 ARCHIS case identifications
- Download 2,278 pdf-reports from various sources

Table 3.2 Overview of report availability (N=1,380).

	Total number of ARCHIS identification codes	Number of available reports	Number of ARCHIS case-identification codes published under different code	Number of unavailable reports
Towns	2,050	1,245	189	616
Failed towns	228	135	16	77
Total	**2,278**	**1,380**	**205**	**693**

(Source: Fischer et al. 2021, Table 3.2)

Too much data



https://i.redd.it/4yxnn98j7mwz.jpg

Saved by text mining: reading by script

1. Convert PDF to plain text (OCR if needed)
2. Term Document Matrix (count individual words)
3. Create word cloud of document
4. Store TDF in PostgreSQL-database

https://ronaldvisser.github.io/TextMiningUrbanFarming_IWGP2025

Finding the right words: keywords

- animal husbandry, e.g. animals
- arable farming, e.g plant names, tools
- horticulture, e.g. plant names, utensils
- general urban farming, e.g. farming terms
- orchards, e.g. fruits
- NOT: too general or ambivalent terms

(Source: Fischer et al. 2021, fig. 2.22)

Matching keywords with text from documents

- Valuing keywords
  - Important value 4, e..g plant names
  - Less important 1, e.g. arable farming
- Match keywords with text of document (SQL-query)
- Scoring document
  - Sum(Occurence of keyword * value of keyword)

Most common keywords

Table 4.3 Top 20 common keywords, in order of number of occurrences.

Keyword (English)	Occurrence	Score (%)
Flowerpot	80	6.61
Dung	66	5.45
Agrarian	64	5.28
Dung pit	51	4.21
Farmstead	51	4.21
Barn	44	3.63
Layer of arable soil	39	3.22
Garden	36	2.97
Vegetable garden	35	2.89
Orchard	35	2.89
Arable farming	32	2.64
Agrarian	28	2.31
Livestock	23	1.9
Garden wall	23	1.9
Vegetable gardens	22	1.82
Vegetable	22	1.82
Gardens	19	1.57
Fertilization	19	1.57
Cultivated	17	1.4
Arable farming	16	1.32

(Source: Fischer et al. 2021, Table 4.3)

Ranking (scored) reports

Using mean (μ) and standard devation (σ)

A. > μ+2σ (c. 2.5%);

B. > μ+σ and <= μ+2σ (c. 13.5%);

C. > μ-σ and <= μ+σ (c. 68%);

D. > μ-2σ and <= μ-σ (c. 13.5%);

E. <= μ-2σ (c. 2.5%).

Table 3.5 Description of suitability reports per group.

Group	Description
A	reports which are (very) suitable
B	reports which are likely to be suitable
C	reports which may be suitable
D	reports which are unlikely to be suitable
E	reports which are not suitable

(Source: Fischer et al. 2021, Table 3.5)

Ranked reports

- Relevant
  - rank A: 26 reports
  - rank B: 183 reports
- Possibly relevant
  - rank C: 899 reports
- Irrelevant
  - rank D: 152 reports
  - rank E: 48 reports

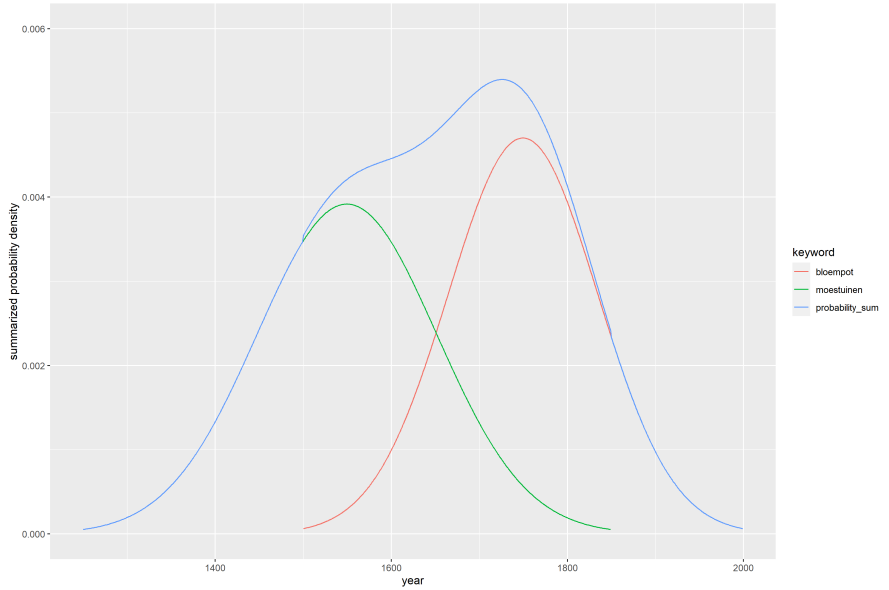
Sample of 100 reports of each rank read to test assumption

Time saving

- Start: 1380 reports (1448 PDF)
- Relevant 265 reports analysed (nearly 20%)
  - Database for structured analysis of keywords
  - Valuing results (relevance)
  - Date
  - Themes

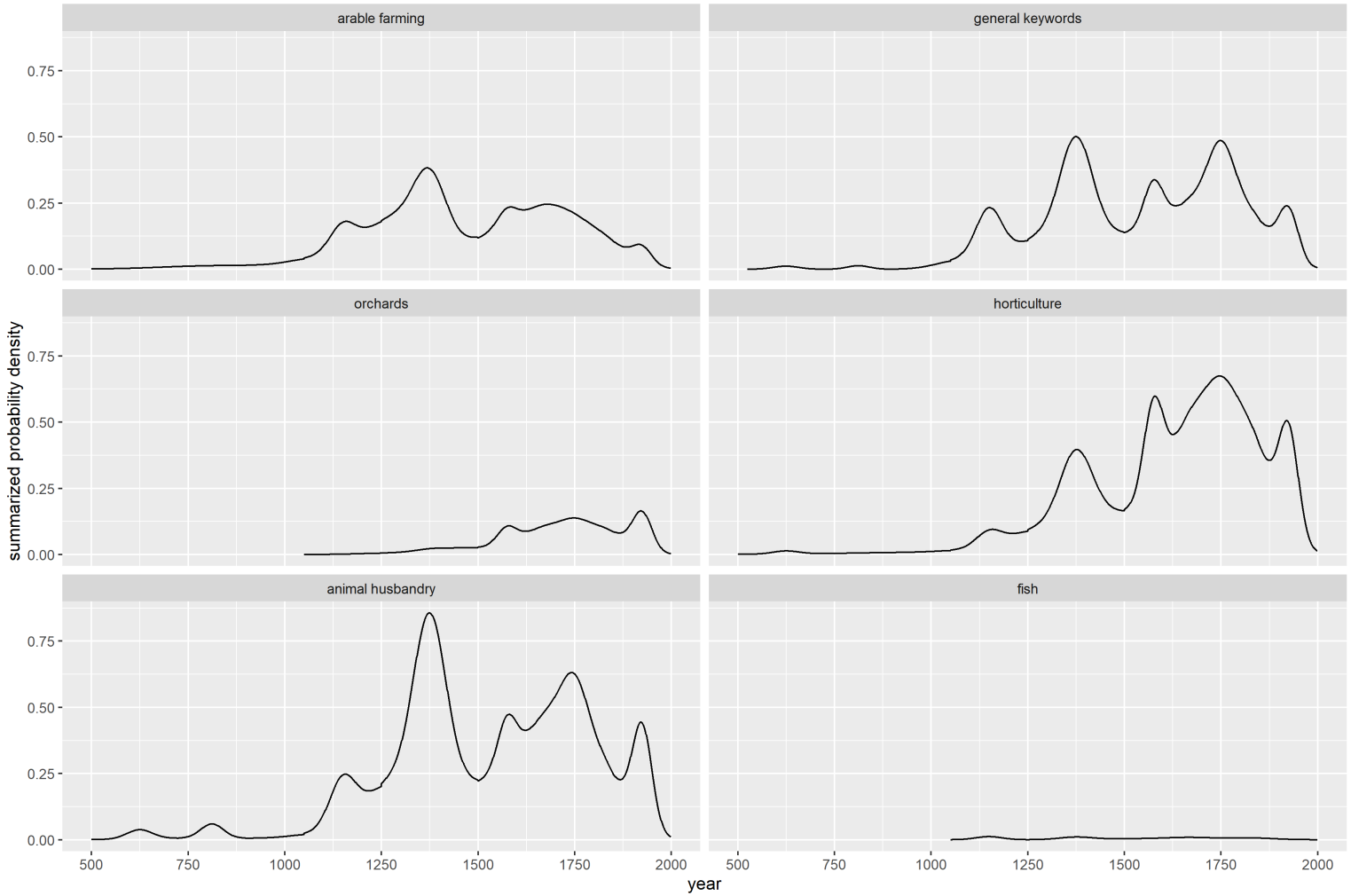
Dating challenges

- Varied dates
  - Periods (Archaeological Basic Register (ABR))
  - Exact dating
  - Date ranges
- Combined with summarized probabilities



(Source: Fischer et al. 2021, fig. 3.8)

Temporal patterns:



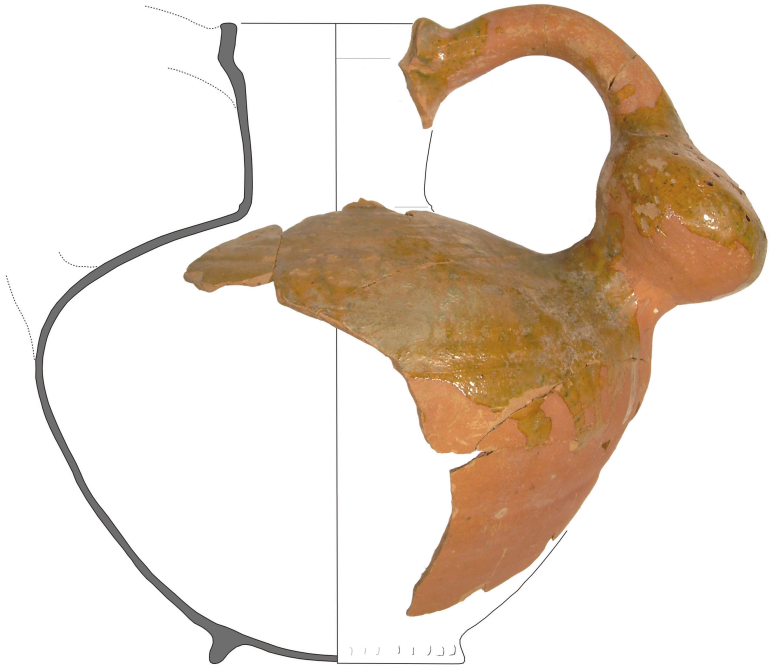
(Source: Fischer et al. 2021, fig. 8.3a)

Example theme: horticulture

(Amersfoort in 1652 by Blaeu in Fischer et al. 2021, fig. 8.41)

Example theme: horticulture

(Vegetable garden plots (Depuydt 2014, 37) in Fischer et al. 2021, fig. 6.16)



(Watering can made of pottery (dating 1500-1525: Vermunt & Van der Kallen 2013, 32) in Fischer et al. 2021, fig. 6.18)

Example:patterns in Groningen

- oldest centre:
  - mixed farming practice
- northern 17th century expansion:
  - horticulture
  - animal husbandry
  - fish farming

(Sites plotted on the Van Deventer map in Fischer et al. 2021, fig. 8.19)

https://ronaldvisser.github.io/TextMiningUrbanFarming_IWGP2025

Conclusion

- Archaeologists produce (too?) many reports
- Text mining saves time (>80% of all reports irrelevant)
- Defintion of keywords is important
  - Rating
  - Valuing
- Analyses of documents still necessary
- Patterns of urban farming
  - common
  - spatial distribution over towns
  - temporal variance

https://ronaldvisser.github.io/TextMiningUrbanFarming_IWGP2025

Questions?

If you want to know more:

This presentation:

https://ronaldvisser.github.io/TextMiningUrbanFarming_IWGP2025

Open Access book, Fischer et al. (2021):

https://www.cultureelerfgoed.nl/publicaties/publicaties/2021/01/01/urban-farming-and-ruralisation-in-the-netherlands

Reproducible code, Visser (2022):

https://github.com/RonaldVisser/Mining_Archaeological_Reports /

https://doi.org/10.5281/zenodo.7157758\

SAXION UNIVERSITY OF APPLIED SCIENCES NMF erfgoedadvies UNIVERSITY OF AMSTERDAM

References

Fischer, A. D., H. van Londen, A. L. Blonk-van den Bercken, R. M. Visser, and J. Renes. 2021. *Urban Farming and Ruralisation in the Netherlands (1250 up Tot the Nineteenth Century), Unravelling Farming Practice and the Use of (Open) Space by Synthesising Archaeological Reports Using Text Mining*. Nederlandse Archeologische Rapporten 68. Amersfoort: Rijksdienst voor het Cultureel Erfgoed. https://www.cultureelerfgoed.nl/publicaties/publicaties/2021/01/01/urban-farming-and-ruralisation-in-the-netherlands.

Rutte, Reinout, and Jaap Evert Abrahamse, eds. 2016. *Atlas of the Dutch urban landscape: a millennium of spatial development*. Bussum: THOTH Publishers.

Visser, Ronald M. 2022. “Text Mining of Archaeological Reports for Urban Farming (Data and Code).” https://doi.org/10.5281/zenodo.7157759.

———. 2025. “Relating Roman Rings. An Interdisciplinary Study Using Archaeology, Data Science and Tree Rings to Understand Timber Provision in the German Provinces of the Roman Empire.” PhD thesis, Amsterdam. https://doi.org/10.5463/thesis.1062.