



# GLOBALISE

**Datasprint: Mapping Places in the Indian Ocean World**

# Programme

13:00 Introduction

*outline & aims of today's datasprint*

Lodewijk Petram / Leon van Wissen

*Globalise places*

Ruben Land

*outline of breakout sessions*

Jules Schoonman, Melvin Wevers, Rombert Stapel

13:45 Breakout sessions: start

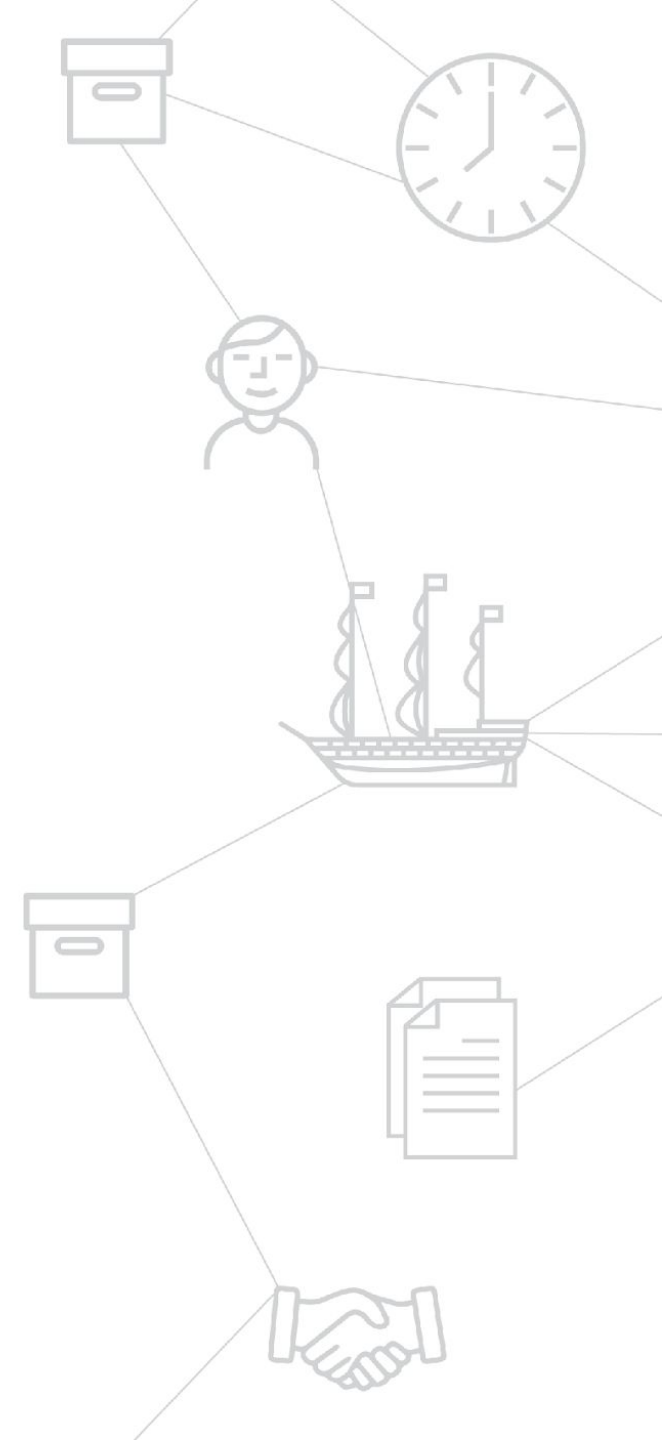
15:00 Break

15:15 Breakout sessions: wrapping up

16:00 Session results and conclusions

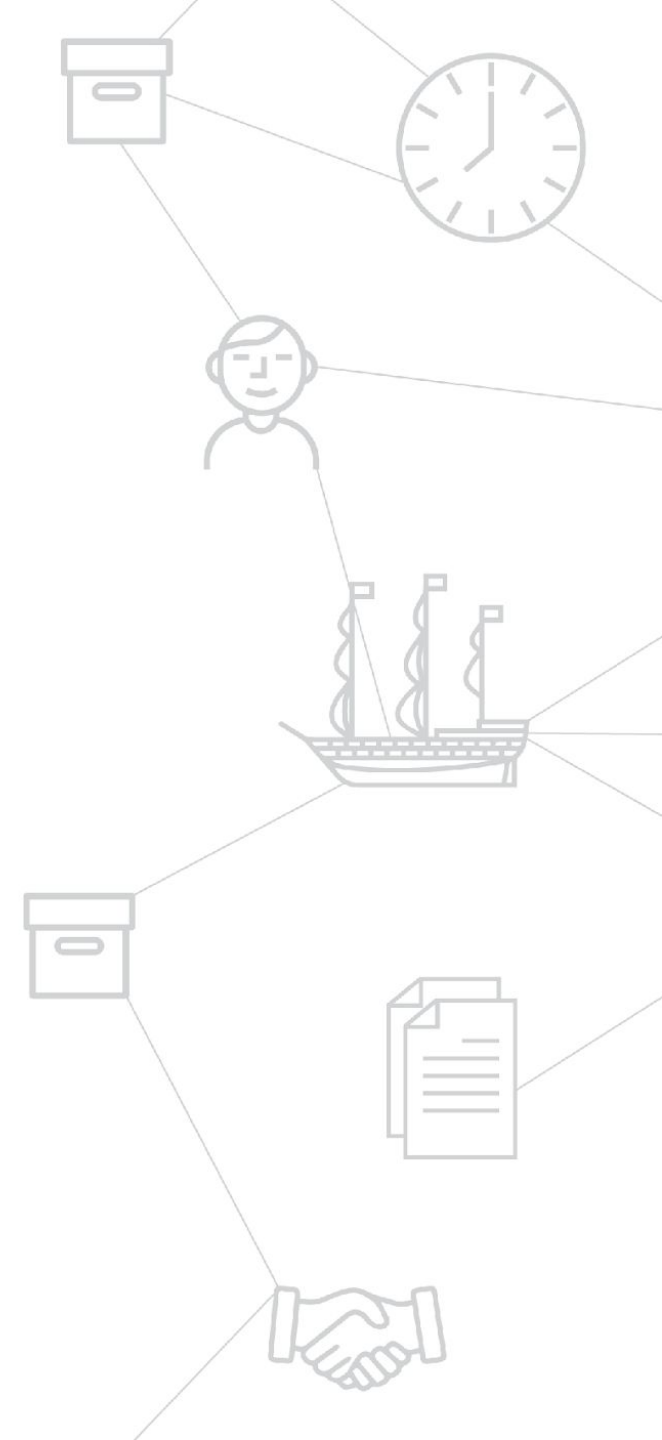
16:30 Reflection

17:00 Drinks



# The GLOBALISE project

- 2022-2026
- Unlocks a unique source...
- ...for large-scale, in-depth and structured analyses
- Makes a wealth of contextual data accessible and re-usable
- Aimed at scholars and citizen scientists





# ‘Overgekomen Brieven en Papieren’ (OBP)

- Key series of VOC archives
- Unique source
  - scope
  - richness
- Lens on world history
- 4.8 mio. scans

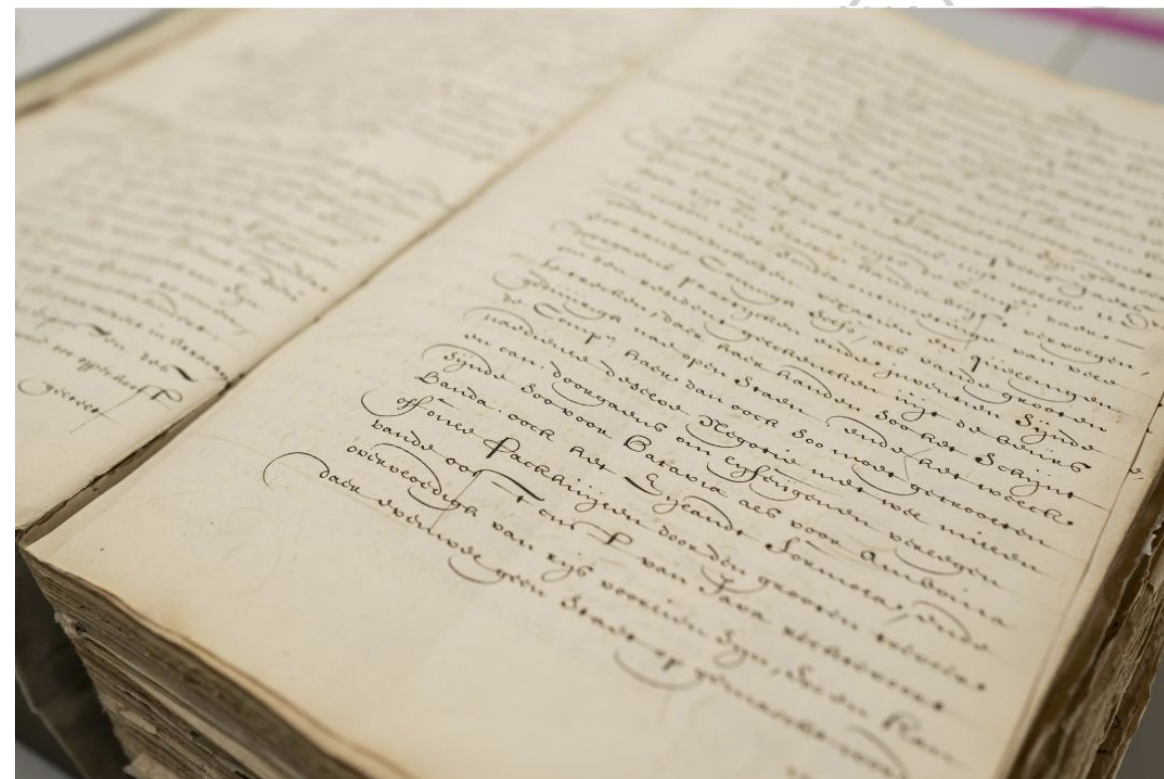


Photo Dave Straatmeyer



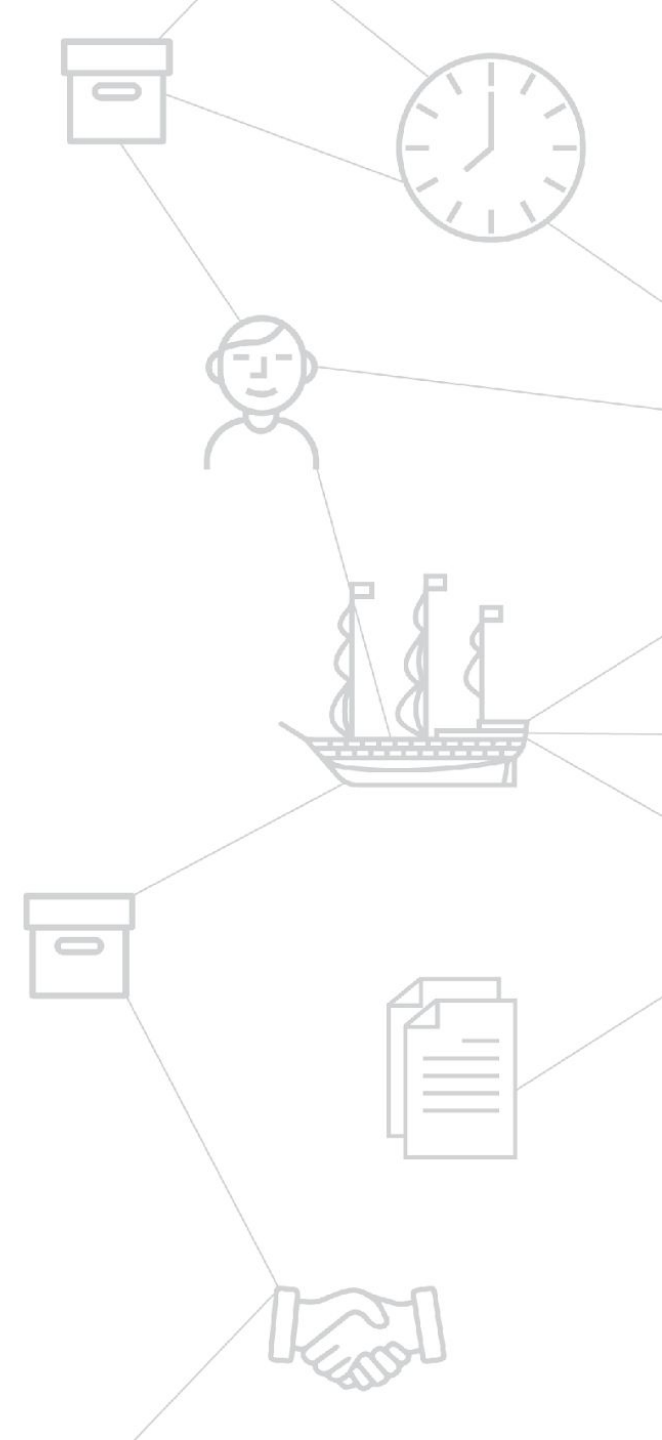


# GLOBALISE

Places all around a large part of the world in the VOC archives

# Content

- Place names data set
- Data set construction
- First version of the OBP HTR
- Examples
- Caveats
- Challenges ahead



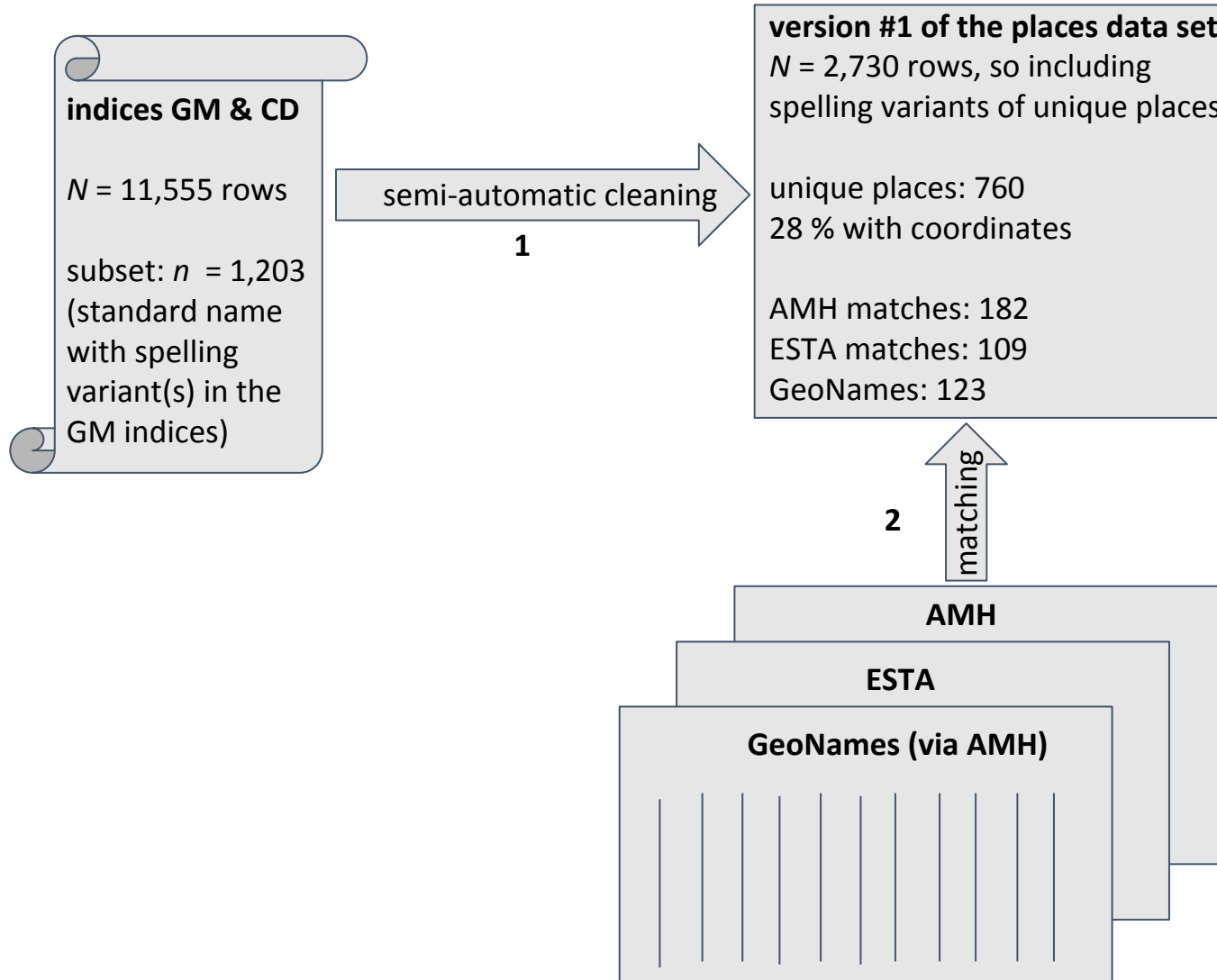
# Place names data set

- First version
- Purpose of the data set:
  - Assisting the annotation process for GLOBALISE
  - Expanding knowledge on (Asian) historical places
- Based on a subset of the indices of:
  - The Generale Missiven, containing the letters from governors-general to the directors of the VOC
  - The Corpus Diplomaticum, containing contracts of the VOC with local rulers
- Links created with other data sets
- Enriched with coordinates and alternative spelling variants in other data sets





# Data set construction





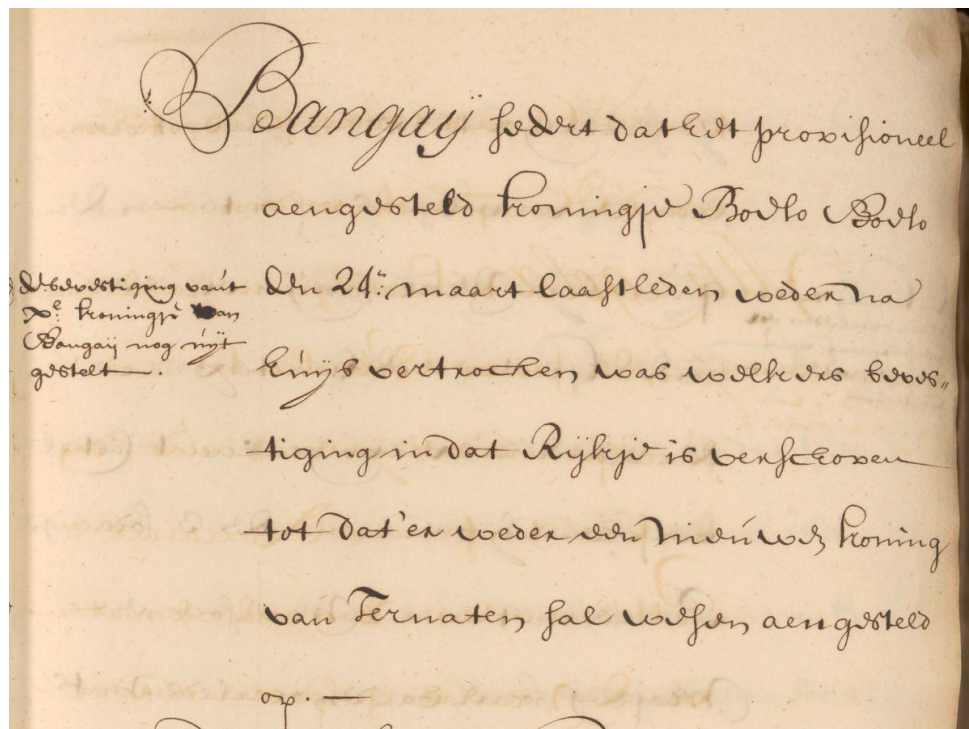
# Example

row_id	unique_name	standard_name	glob_id	source	esta_id	amh_id	latitude	longitude	coord_source
ADD_670	Poelicarro	Poelicarro	GLOB_670	GM/CD					
TEMP_10280	Pollicarro	Poelicarro	GLOB_670	GM/CD					
ADD_1281	Poedecerij	Pondicherry	GLOB_656	ESTA, AMH	ESTA_313	amh_549f	11,9382	79,827976	AMH
ADD_1280	Poedechery	Pondicherry	GLOB_656	ESTA, AMH	ESTA_313	amh_549f	11,9382	79,827976	AMH
TEMP_10888	Pondicheri	Pondicherry	GLOB_656	GM/CD	ESTA_313	amh_549f	11,9382	79,827976	AMH
ADD_656	Pondicherry	Pondicherry	GLOB_656	GM/CD, ESTA, AMH	ESTA_313	amh_549f	11,9382	79,827976	AMH
ADD_1279	Pondichéry	Pondicherry	GLOB_656	ESTA	ESTA_313	amh_549f	11,9382	79,827976	AMH
TEMP_10156	Pondij	Pondicherry	GLOB_656	GM/CD	ESTA_313	amh_549f	11,9382	79,827976	AMH
ADD_1278	Pondycherry	Pondicherry	GLOB_656	ESTA	ESTA_313	amh_549f	11,9382	79,827976	AMH
ADD_1277	Puducherry	Pondicherry	GLOB_656	ESTA	ESTA_313	amh_549f	11,9382	79,827976	AMH
TEMP_09234	Panani	Ponnani	GLOB_366	GM/CD, AMH		amh_601f	10,7769	75,920835	AMH
ADD_1430	Panany	Ponnani	GLOB_366	AMH		amh_601f	10,7769	75,920835	AMH
ADD_1431	Pannanij	Ponnani	GLOB_366	AMH		amh_601f	10,7769	75,920835	AMH
TEMP_10278	Ponanij	Ponnani	GLOB_366	GM/CD		amh_601f	10,7769	75,920835	AMH
ADD_366	Ponnani	Ponnani	GLOB_366	GM/CD, AMH		amh_601f	10,7769	75,920835	AMH
TEMP_10172	Ponecayl	Ponnekail	GLOB_662	GM/CD		amh_504f	8,63222	78,111944	AMH
ADD_1727	Ponnecaijl	Ponnekail	GLOB_662	AMH		amh_504f	8,63222	78,111944	AMH
ADD_1726	Ponnecail	Ponnekail	GLOB_662	AMH		amh_504f	8,63222	78,111944	AMH
ADD_1728	Ponneceijl	Ponnekail	GLOB_662	AMH		amh_504f	8,63222	78,111944	AMH
ADD_662	Ponnekail	Ponnekail	GLOB_662	GM/CD, AMH		amh_504f	8,63222	78,111944	AMH
ADD_1729	Ponnekayl	Ponnekail	GLOB_662	AMH		amh_504f	8,63222	78,111944	AMH
ADD_1725	Punneikayal	Ponnekail	GLOB_662	AMH		amh_504f	8,63222	78,111944	AMH
TEMP_10159	Poimerijn	Ponneri	GLOB_381	GM/CD		amh_528f	9,50139	80,206667	AMH
TEMP_09271	Pomerijn	Ponneri	GLOB_381	GM/CD		amh_528f	9,50139	80,206667	AMH
TEMP_10157	Pommerijn	Ponneri	GLOB_381	GM/CD		amh_528f	9,50139	80,206667	AMH
ADD_1436	Ponnereijn	Ponneri	GLOB_381	AMH		amh_528f	9,50139	80,206667	AMH
ADD_381	Ponneri	Ponneri	GLOB_381	GM/CD		amh_528f	9,50139	80,206667	AMH
TEMP_09260	Ponnerijn	Ponneri	GLOB_381	GM/CD, AMH		amh_528f	9,50139	80,206667	AMH
ADD_1437	Ponnorijn	Ponneri	GLOB_381	AMH		amh_528f	9,50139	80,206667	AMH
ADD_1438	Poonerijn	Ponneri	GLOB_381	AMH		amh_528f	9,50139	80,206667	AMH
ADD_1435	Pooneryn	Ponneri	GLOB_381	AMH		amh_528f	9,50139	80,206667	AMH
ADD_346	Ponorogo	Ponorogo	GLOB_346	GM/CD					
TEMP_07891	Pranaraga	Ponorogo	GLOB_346	GM/CD					
TEMP_10282	Poena	Poona	GLOB_671	GM/CD					
ADD_671	Poona	Poona	GLOB_671	GM/CD					
ADD_878	Parahyangan	Preanger	GLOB_87	ESTA	ESTA_297	amh_539f	-7,2953	107,49566	AMH
ADD_1400	Prangerlanden	Preanger	GLOB_87	AMH	ESTA_297	amh_539f	-7,2953	107,49566	AMH
TEMP_02223	Preanger	Preanger	GLOB_87	GM/CD, ESTA, AMH	ESTA_297	amh_539f	-7,2953	107,49566	AMH
ADD_1398	Preangerlanden	Preanger	GLOB_87	AMH	ESTA_297	amh_539f	-7,2953	107,49566	AMH
ADD_879	Priangan	Preanger	GLOB_87	ESTA	ESTA_297	amh_539f	-7,2953	107,49566	AMH
ADD_880	Prianger	Preanger	GLOB_87	ESTA	ESTA_297	amh_539f	-7,2953	107,49566	AMH
ADD_881	Prianger bovenlanden	Preanger	GLOB_87	ESTA	ESTA_297	amh_539f	-7,2953	107,49566	AMH

# Example

[illegible]

# First version of the OBP HTR



Loghi  
pipeline

'**Bangaij** sedert dat het provisioneel aengesteld koningje Boelo Boelo den 24:' maart laastleden weder na huijs vertroocken was welkers beves,, tiging in dat Rijkje is verschoven tot dat 'er weder een nieuwen koning van **Ternaten** sal wesen aengesteld op. —'

# Data set construction

## indices GM & CD

$N = 11,555$  rows

subset:  $n = 1,203$   
(standard name  
with spelling  
variant(s) in the  
GM indices)

semi-automatic cleaning

1

**version #1 of the places data set**  
 $N = 2,730$  rows, so including  
spelling variants of unique places

unique places: 760  
28 % with coordinates

AMH matches: 182  
ESTA matches: 109  
GeoNames: 123

2

matching

AMH

ESTA

GeoNames (via AMH)

manual cleaning & filtering

3

## OBP HTR output

unigrams with  $n > 1$   
 $N = 41,767,160$  rows

count of unigram  
per year

share of which are  
place names is  
unknown

examples: den,  
batav, compagnie,  
blijkens

## OBP GLOBALISE places data set

unique places: 552  
37% with coordinates

match percentage: 73%

input data

4

App

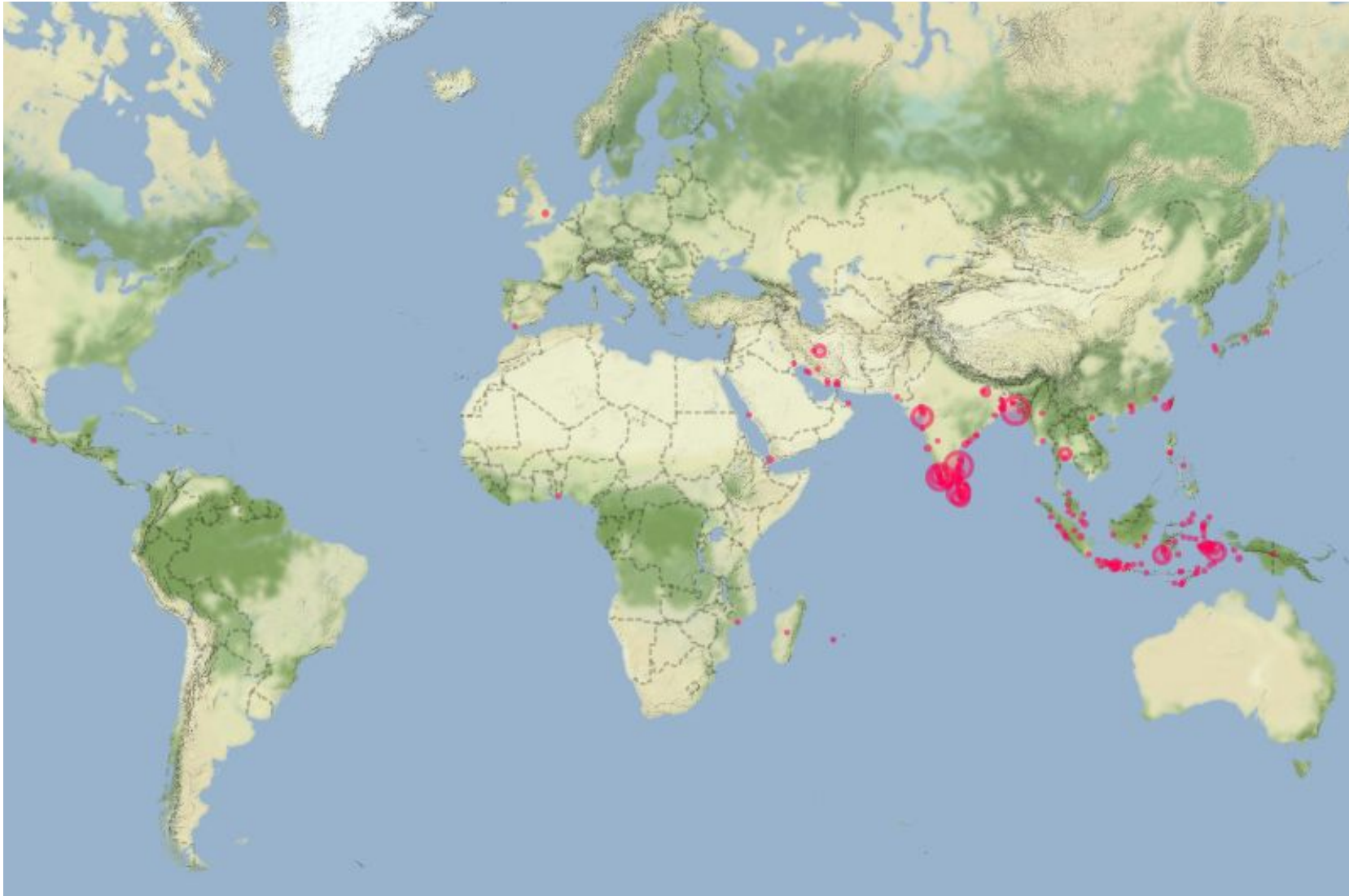
# Application

Interactive GLOBALISE places application:  
[https://globalise.shinyapps.io/mapping\\_places/](https://globalise.shinyapps.io/mapping_places/)



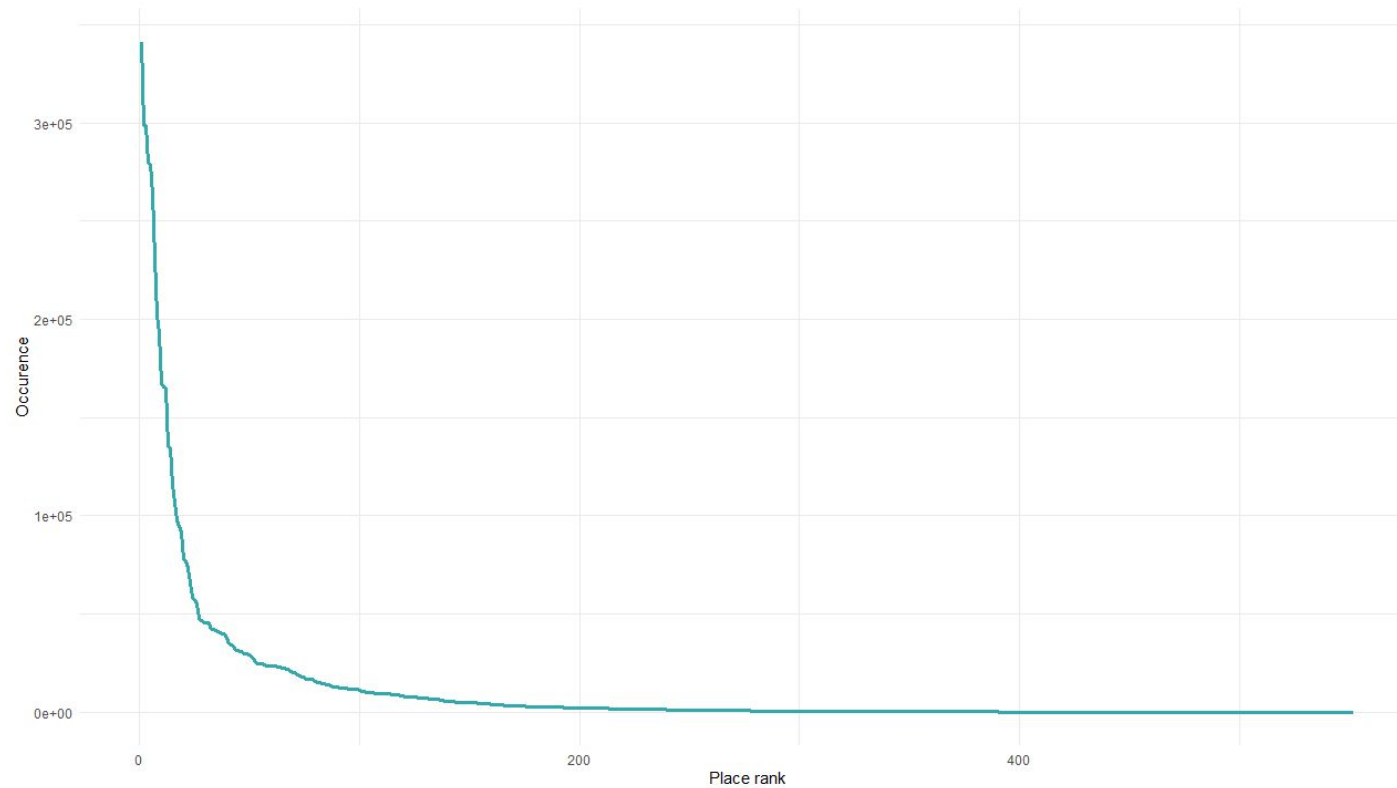


# Spatial distribution



Map source:  
OpenStreetMap  
Period: 1614-1795  
Circle size indicates  
number of occurrences

# Occurrence of places

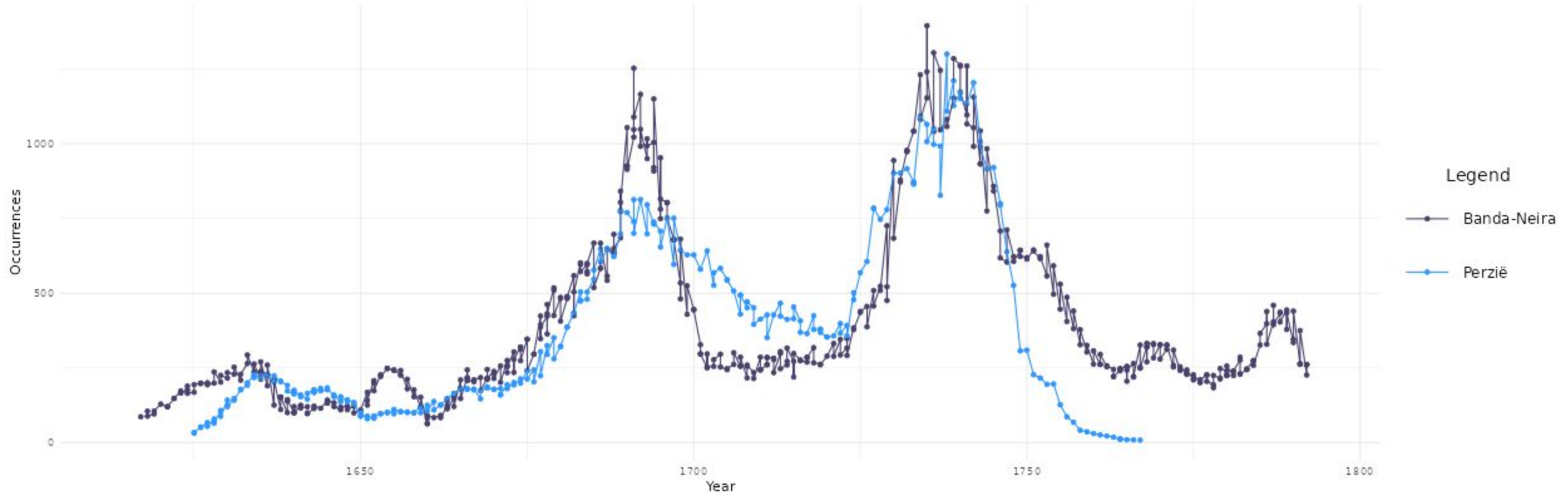


Rank	Standard name	Frequency
1	Ceylon	340,992
2	India	298,872
3	Indië	298,474
4	Coromandel	279,427
5	Bengalen	279,025
6	Cochin	263,718
7	Colombo	231,235
8	Surat	202,517
9	Banda-Neira	192,453
10	Makassar	167,391

The number of occurrences in the OBP of a place name decreases exponentially with decreasing rank

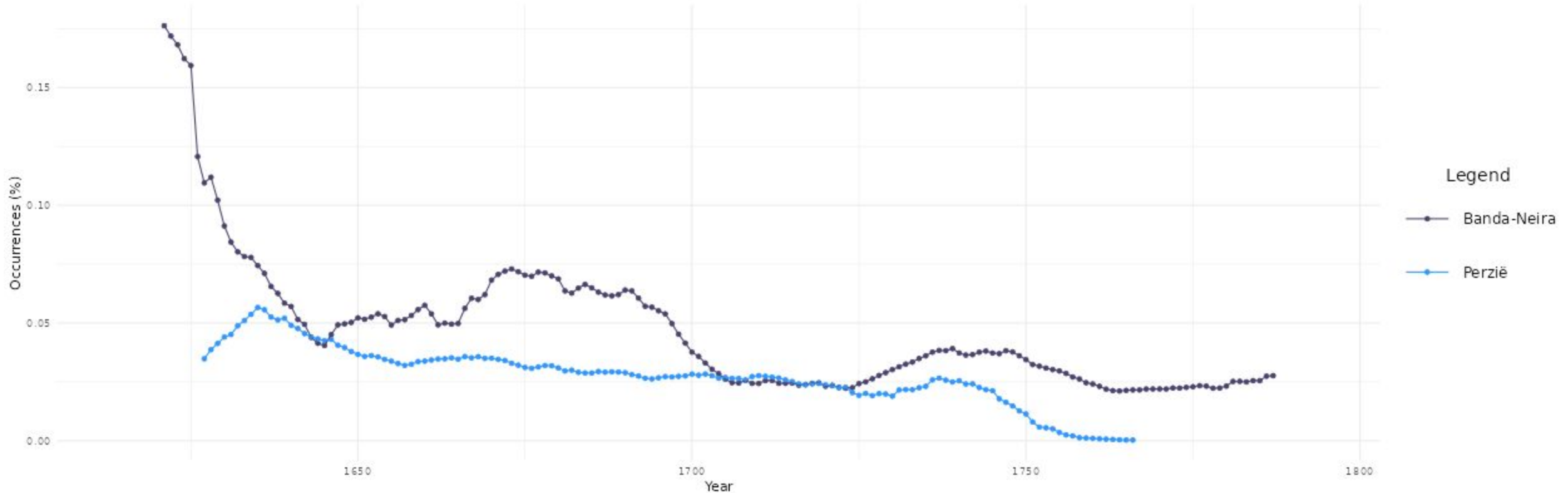


# Occurrences throughout time



Note: Absolute number of occurrences, so without taking into account the total amount of written work per year, based on the moving average with a window of 15 years.

# Occurrences throughout time



Note: Number of mentions of a place relative to the total number of words that appear more than once in an inventory number of the OBP collection of the Dutch National Archives. Based on the moving average with a window of 15 years.

# Caveats

- Based on indices with particular spelling variants
- Current foundation of the data is a subset of the indices for workability
- Only the most well-known places have coordinates
- Coordinates not the perfect way to represent a place



# Challenges ahead

- Adding more places to the data set
- Linking to other data sets
- Finding more coordinates
- Choosing standard names
- Error checking

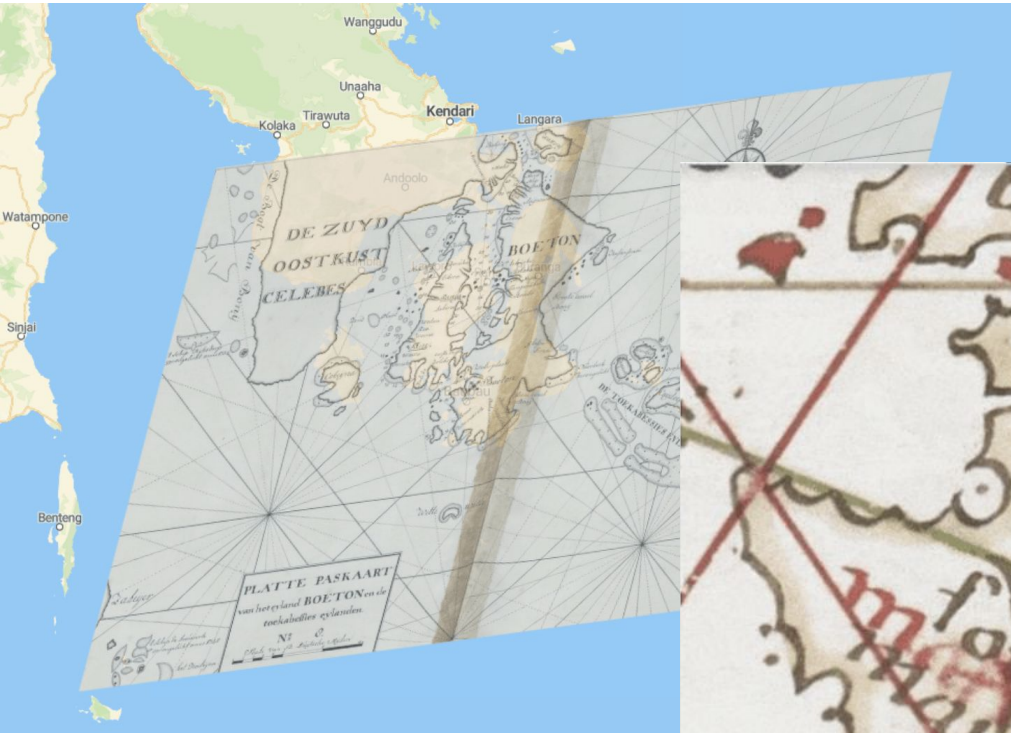


# Thanks for listening!

# Are there any questions?

Interactive **GLOBALISE** places application: [https://globalise.shinyapps.io/mapping\\_places/](https://globalise.shinyapps.io/mapping_places/)

# Datasprint



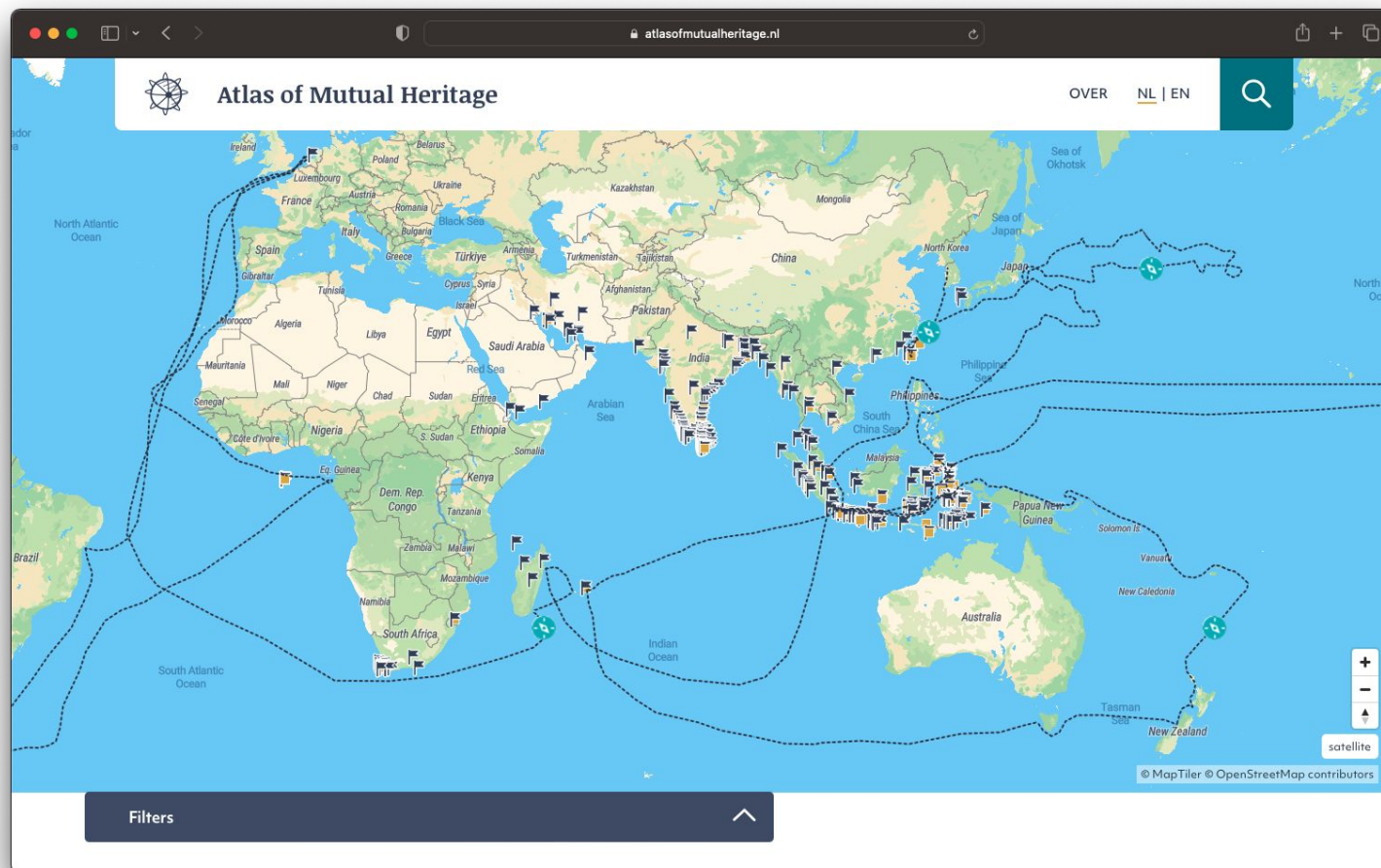
# Breakout sessions

1. Georeferencing early modern maps
2. Data extraction from early modern maps
3. Curating and linking new places data(sets) via World Historical Gazetteer

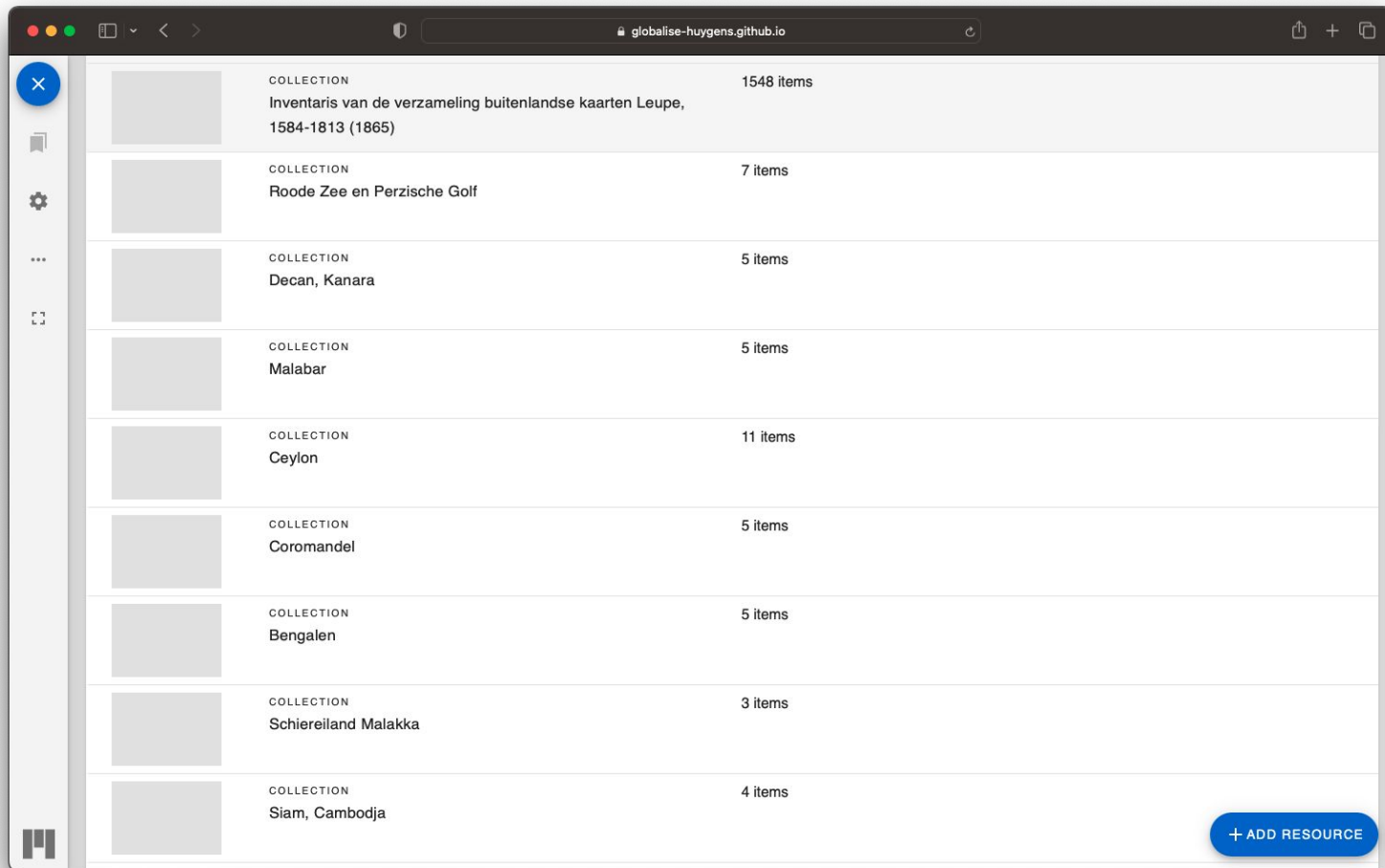




# Atlas of Mutual Heritage / Leupe collection



# Atlas of Mutual Heritage / Leupe collection



COLLECTION	1548 items
Inventaris van de verzameling buitenlandse kaarten Leupe, 1584-1813 (1865)	
COLLECTION	7 items
Roode Zee en Perzische Golf	
COLLECTION	5 items
Decan, Kanara	
COLLECTION	5 items
Malabar	
COLLECTION	11 items
Ceylon	
COLLECTION	5 items
Coromandel	
COLLECTION	5 items
Bengalen	
COLLECTION	3 items
Schiereiland Malakka	
COLLECTION	4 items
Siam, Cambodja	

+ ADD RESOURCE



# Notes & data

<https://github.com/globalise-huygens/datasprint-amh>

