



TOWARDS A UNIVERSE OF LOCAL TIME MACHINES:

BUILDING AN OPEN ECO-SYSTEM FOR DIGITAL HERITAGE

FUELLED BY COMMON LANGUAGE RESOURCES AND EXISTING INFRASTRUCTURE

BY TOINE PIETERS & IVAR TROOST



# Content



**1. Clarin related research**



**2. Time Machine Europe**



**3. Utrecht Time Machine**



**4. Pilot Living Pasts**

# CLARIN Related Research Projects



- ▶ WAHSP, BILAND, Translantis, Asymenc (2010-2018)  
(newspaper archives, text/sentimentmining approach)
- ▶ Time Capsule (2013-2018)  
(heterogeneous data stores, linked data/RDF approach)
- ▶ DReAM [CLARIAH] (2016-2018)  
(newspaper and broadcast archives, datamining approach)
- ▶ (Utrecht) Time Machine (2018-  
(Towards open infrastructure for digital heritage)

# WAHSP, BILAND, Translantis, Asymenc (2010-2018)



- ▶ Challenges with **sustainability** of project tools
- ▶ Challenges with **IPR** and **digitisation** approaches in an international context
- ▶ Challenges with **stability** and **scalability** due to underestimation of high performance computing for exploratory search and text mining of e.g. the *Delpher corpus* of digitised historical newspapers and journals (National Library of the Netherlands) or the *Europeana newspaper repository*

To what extent can current Clarin  
resources and infrastructure  
help researchers overcome the

- ▶ sustainability
- ▶ IPR and digitisation
- ▶ stability and scalability

▶ ISSUES?



# Time Capsule (2013-2018)

- ▶ Challenges with *metadata standards* and interoperability
- ▶ Challenges with *scalability*
- ▶ Challenges with *visualization of geo-referencing data*
- ▶ Challenges with *sustainability*

<http://www.timecapsule.nu>

To what extent can current CLARIN  
resources and infrastructure  
help researchers overcome the

▶ *metadata standards* and interoperability

▶ ISSUES?



# Zooming in on DReAM

## CLARIAH research pilot Debate Research Across Media

- ▶ In this pilot, heterogeneous datasets (of digitized print and audio-visual media) were made searchable with tools of the CLARIAH Media Suite, combining distant and close reading to do historical public debate analysis.
- ▶ Tools *Collection Inspector*, *Search and Compare*, (Comparitive search), *Workspace* (allowing analysis and annotation of the bookmarked results)



## PUBLIC SPHERE



Radio

Television

Newspapers



CITIZENS

Collection Inspector  
Search and Compare  
Workspace

The screenshot shows the Media Suite web application interface. At the top, there is a dark navigation bar with the following elements from left to right: the logo 'MEDIA SUITE', a 'Data' menu, a 'Tools' menu, a 'Workspace' dropdown menu, and an 'Account' section with the user name 'Mdeh027'. Below the navigation bar, the main content area features a central circular icon with a magnifying glass over a document, surrounded by a 'TV' icon and a 'SEARCH' icon. To the right of this icon is a text block: 'The Media Suite tools offer the core functionalities needed for performing scholarly research tasks with audio-visual media and contextual collections. Tools available in this version of the Media Suite include metadata inspection, exploratory browsing, search, visualization, and analysis (permission support)'. Below this text is a 'VIEW DOCUMENTATION' button. Underneath, there are four tool cards: 'Inspect', 'Explore', 'Search', and 'Compare'. Each card has a 'DETAILS' button and an 'OPEN' button with a right-pointing arrow. At the bottom of the page, there is a dark footer with the 'MEDIA SUITE' logo and 'Version 3.0' on the left, and a navigation menu with 'Tools', 'Data', 'About', 'Documentation', and 'Contact' on the right, along with a 'Login' link.

# DReAM (2016-2018)



- ▶ Challenges with historical bias due to the infrastructure of the Media Suite
- ▶ Challenges with metadata differences
- ▶ Challenges with IPR

To what extent can current CLARIN  
resources and infrastructure  
help researchers overcome the

▶ *Infrastructural methodological bias*

▶ ISSUES?





Time  
Machine

## 2. Time Machine Europe

[HTTPS://WWW.TIMEMACHINE.EU/](https://www.timemachine.eu/)



Time  
Machine

What is the Time Machine  
consortium?



# Time Machine is formed by ...

- ▶ 450+ **consortium** members from 32 countries
- ▶ 195 of Europe's top academic and research institutions
- ▶ Private sector partners from SMEs to international companies: Ubisoft, Flixbus, ICONEM

- Internationally-acclaimed galleries, libraries, archives and museums
- European institution bodies
- Civil society and industry associations



Time  
Machine

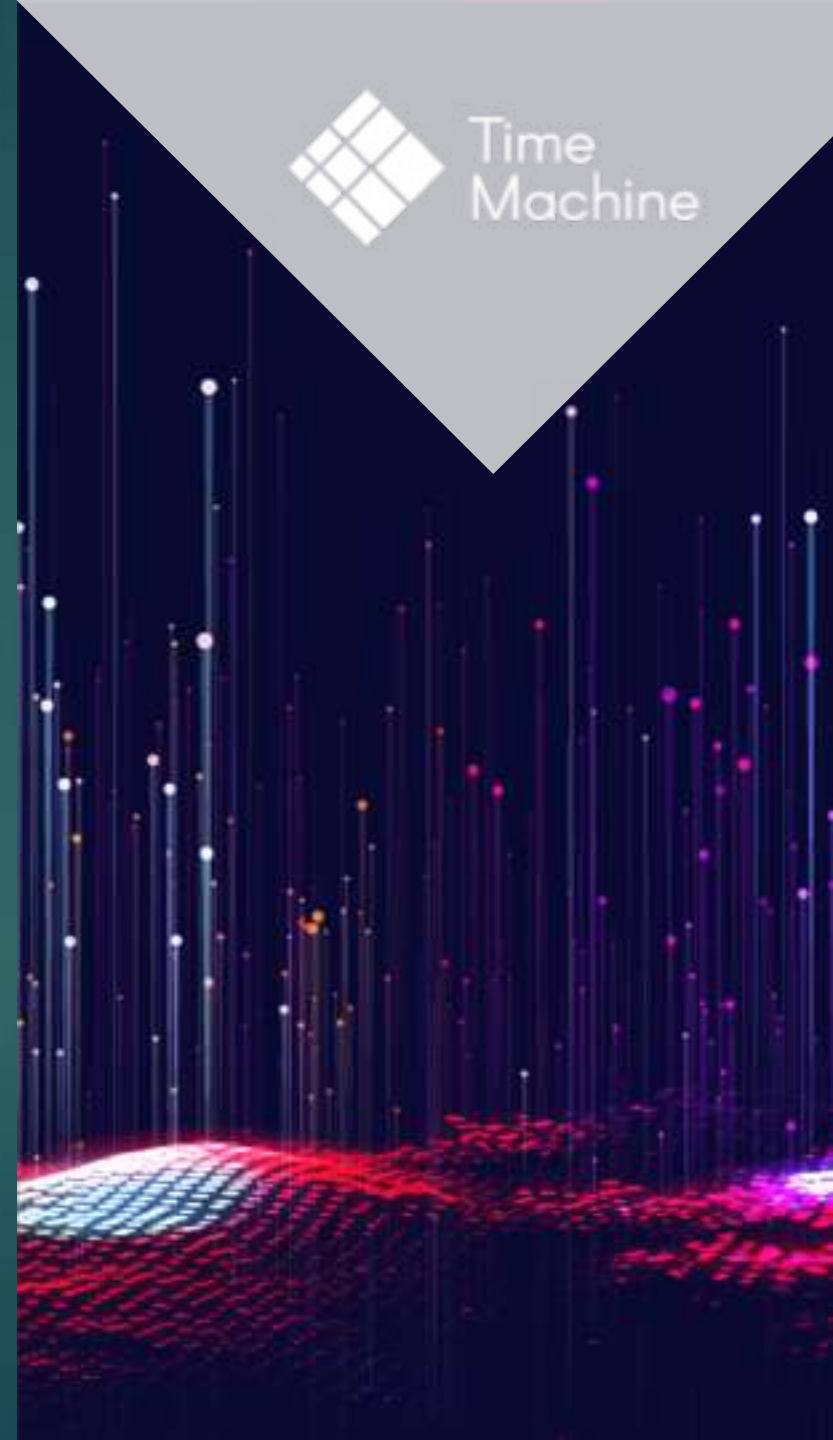


# Time Machine is ...



Time  
Machine

- ▶ An international **collaboration** to bring 5000 years of European history to life
- ▶ Digitising millions of **historical documents**, painting and monuments
- ▶ The **largest** computer simulation ever developed
- ▶ An **open access**, interactive resource





# Time Machine will ...

- ▶ **Revolutionise** education, culture, media, tourism, policymaking and legislation
- ▶ Give historical data **new relevance**
- ▶ **Contextualise history** through advanced AI
- ▶ **Modernise the institutions** that archive Europe's history



# Mass Digitization in a Moment?

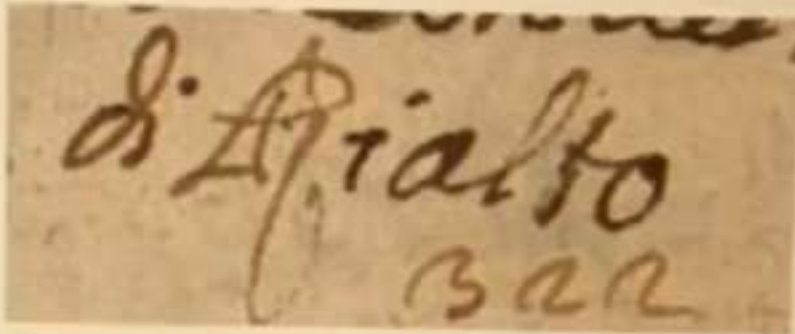


Cultlab3D

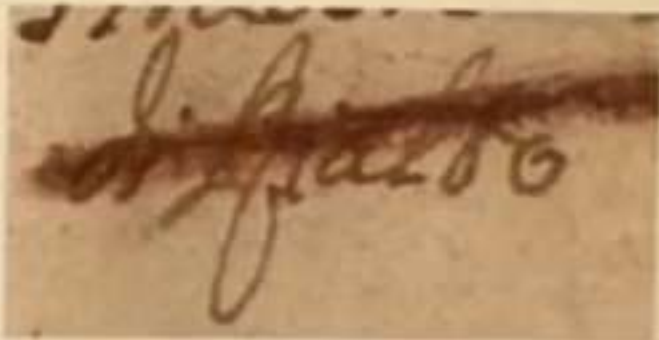


Gaining meaningful knowledge from history?

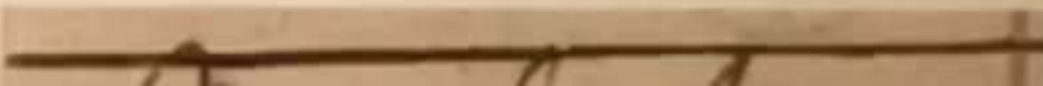
Catastici-434-07 - Page 889



Catastici-434-07 - Page 890



Catastici-434-07 - Page 890



# Time Machine Roadmap



Time  
Machine

12/2016

2/2018

9/2018

3/2019

3/2020

2020



Consortia formation / Hearings by EC

Consultation by EC: 14 Consortia named in three topic areas

Submission of preparatory phase application

Submission of preparatory phase application

Start of preparatory phase to further elaborate plans

Submission of main phase Application (subject of FP9 negotiations)

Intended start of Large Scale Research Initiatives



Time Machine at rank 1 out of 33 consortia





Time  
Machine

1950 1960 1970 1980 1990 2000 2010

2020 2030 2040 2050 2060 2070 2080 2090

How can we align  
Time Machine  
with existing language resources  
and infrastructure?





### 3. Building Utrecht Time Machine as a local time machine

[HTTP://UTRECHTTIMEMACHINE.NL/](http://utrechttimemachine.nl/)

# Questions



- ▶ Technology push or serving public interests?
- ▶ What principles or values matter?



# First step: What is our *principles stack*?

## SSB PRINCIPLES STACK

Harmony & Reflection

Society

Inclusion

Pluralism

Community

Independence

Subjectivity

Interdependence

Technology

Local-first

Upgradeability

Near  
Moderation

Multimodal  
Welcoming

Environment

Sufficiency

Efficiency

Abundance

# What kinds of data should the network hold?



## Primary Data

- Objects
- Texts
- Images
- Audio
- Video
- Etc.



## Secondary Data

- Descriptions of primary data
- Date(s) and location(s) of creation



## Tertiary Data

- Interpretations of primary / secondary data



## Quaternary Data

- Crowd sourcing applications etc.

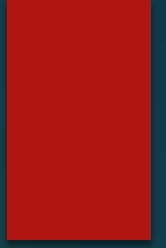
How to align data / encourage shared ontologies?

How to handle fictional data of any kind?

How to handle misinformation / contested information?

*In short: we need shared values and protocols (beyond Triple Stores)*

# Who should have access to the knowledge base?



## ▶ Read

- ▶ Everyone

## ▶ Write

- ▶ Time Machine coordinators?
- ▶ National and Regional Archives?
- ▶ Academics generally?
- ▶ All citizens?

# How can/do Time Machines interoperate?

- What does the network structure look like?
- How do nodes communicate?

# Classical Network Topologies



CENTRALISED



FEDERATED



DISTRIBUTED

# What about the time machines?



CENTRALISED



FEDERATED



DISTRIBUTED



# How is information distributed?

- ▶ Live querying across systems
- ▶ Indices replicated across machines
- ▶ Content replicated across machines  
(more redundancy; higher storage requirements)
- ▶ How to deal with versioning?

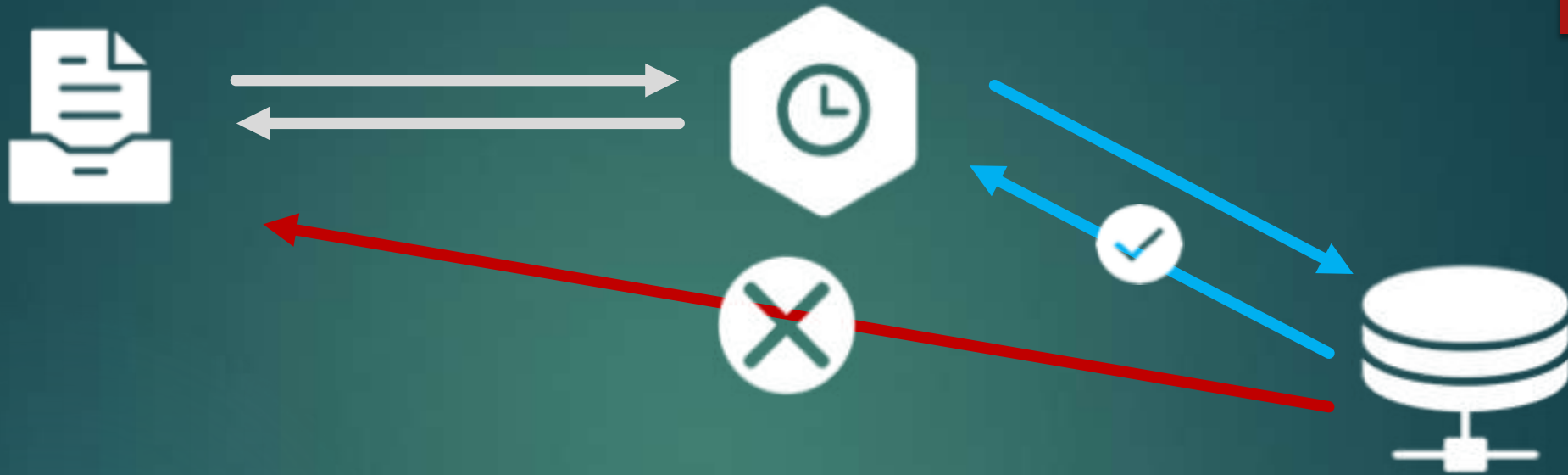


# How is the network expanded?



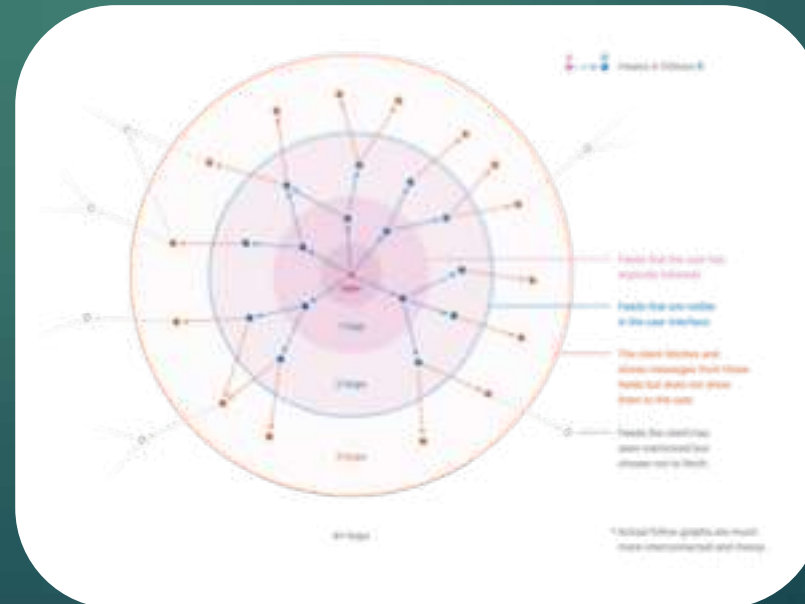
- ▶ Approval of an international council? [centralised]
- ▶ Approval of national or regional councils? [federated]
- ▶ No curation? [decentralised]
- ▶ Trusted web? [decentralised]

# How is the network expanded?

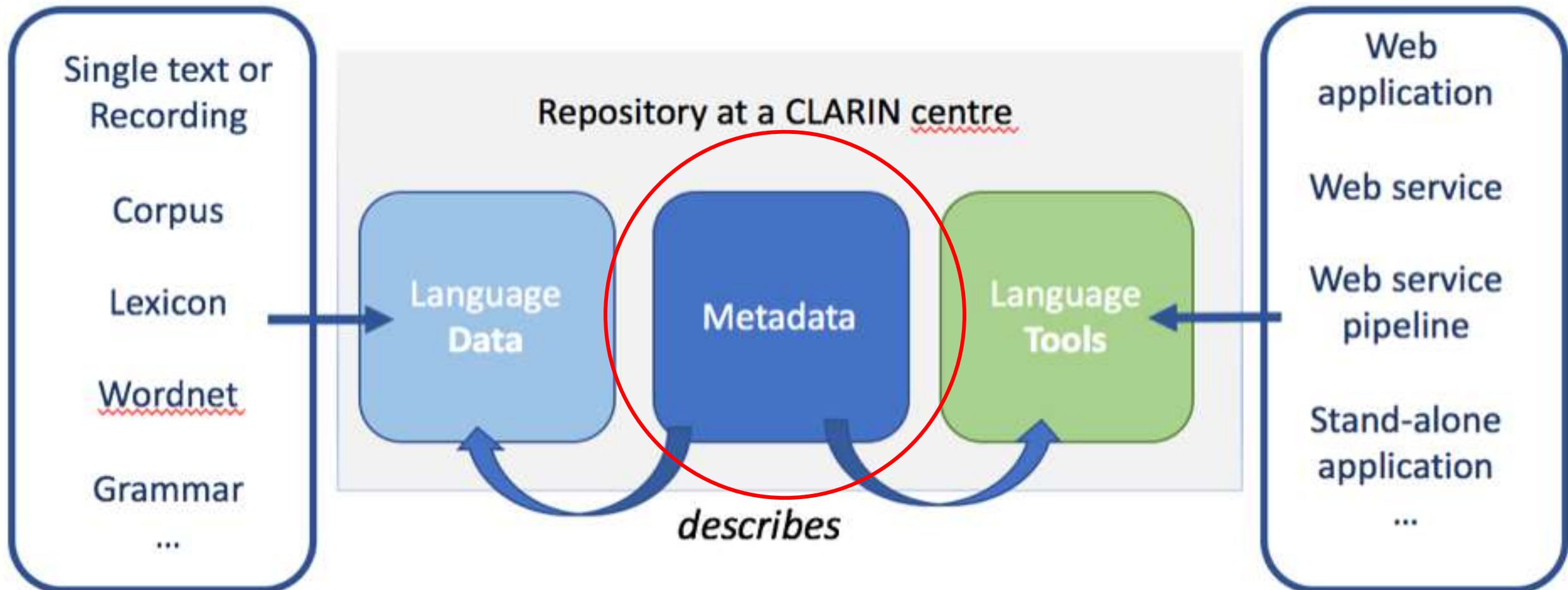


## Tenets:

- Nodes maintain their own truth
- Everyone can host a node
- Nodes added by connected partners are trusted by default, but can be blocked individually

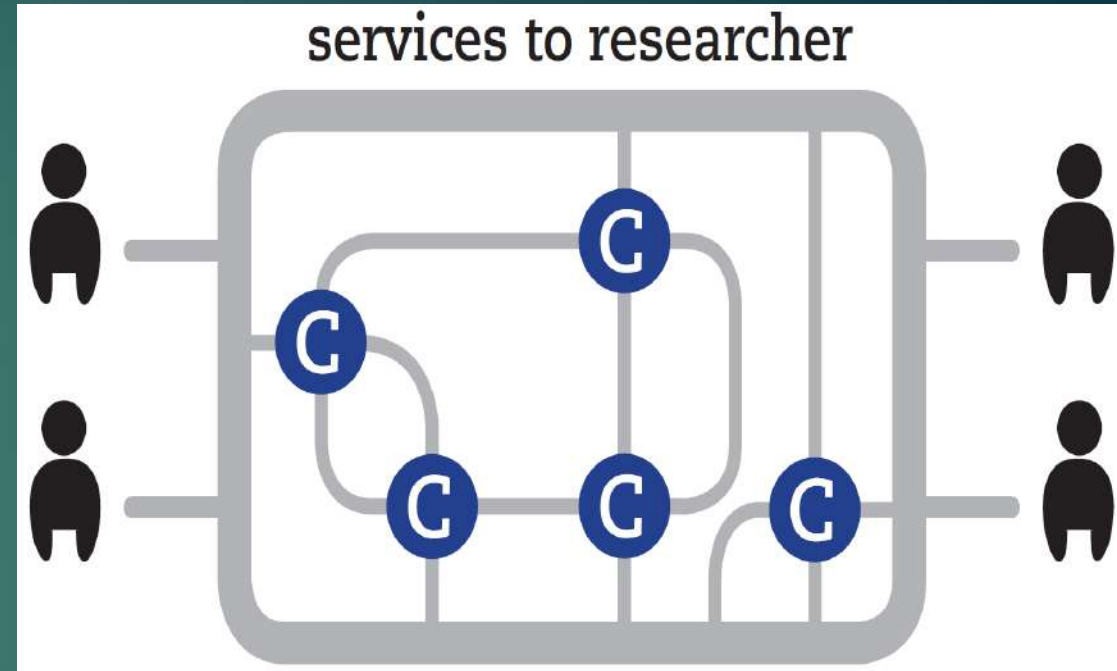


# How can we align with the CLARIN data architecture? *centre repositories*



# CLARIN centres

- ▶ Distributed architecture:  
(http-accessible) files, web applications and web services spread all over Europe
- ▶ Nodes in the network: centres



- ▶ Tools and data from different CLARIN centres are interoperable, so that data collections can be combined and tools from different sources can be chained to perform complex operations to support researchers in their work.

# CLARIN in data types

- ▶ Newspaper archives
- ▶ Literary texts
- ▶ Social Media data
- ▶ Parliamentary records
- ▶ Historical letters
- ▶ Oral History data
- ▶ Disciplinary libraries
- ▶ Institutional archival data
- ▶ Broadcast archives
- ▶ ...

See also the info on the CLARIN Resource Families initiative: <https://www.clarin.eu/resource-families>

# How can we learn/profit from the key strength of CLARIN



- ▶ exchange of metadata
- ▶ exchange formats for the output of analytic tools
- ▶ options for supporting comparative research





Back to the local time machine

## Pilot Living Pasts: Case study 'De Neude'

A testing ground for the Utrecht Time Machine

UTRECHT TIME MACHINE

**LIVING PASTS**

**AUGMENTING URBAN LANDSCAPES AND CULTURAL HERITAGE IN THE DIGITAL AGE**

Advanced Bachelor's level (year 2 or 3) 7.5 ECTS Period 1 – Wed. 13:15–17:00 & Fri. 13:15–17:00

“THE TIME MACHINE PROJECT PUSHES THE BOUNDARIES OF SCIENCE AND TECHNOLOGY TO CREATE A DIGITISATION TOOL FOR GLOBAL HERITAGE.”

Gemeente Utrecht | HE TIONEN UTRECHT ARONDE | Time Machine | Utrecht University

The poster features a background image of a historical street scene in Utrecht, with a yellow tram and people in period clothing. A red banner in the top right corner reads "LATE-REGISTRATION ON AUGUST 19TH & 20TH".







4

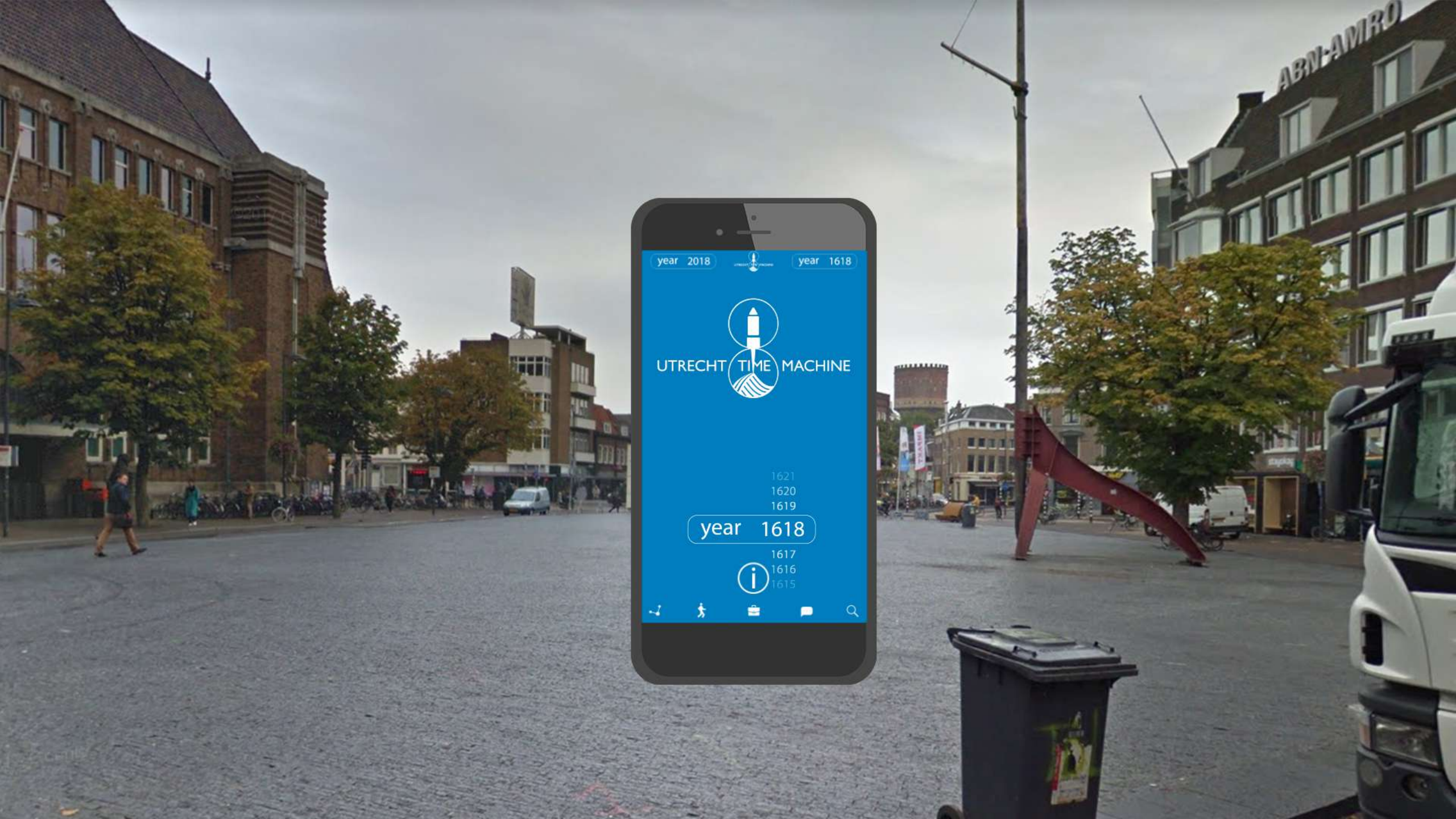
3

**Neude**

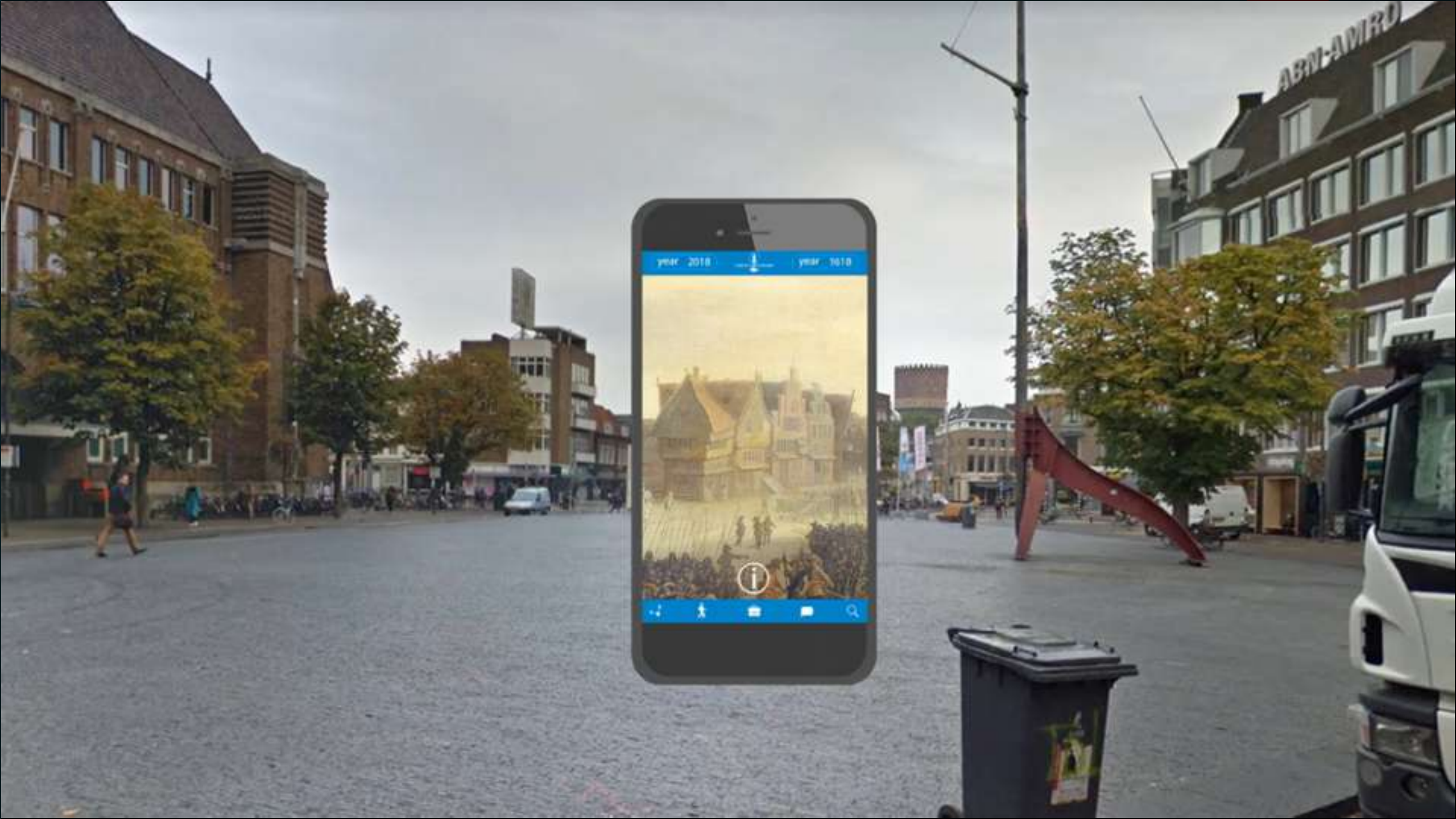












ASV-AMRO



year 2018

year 1618

The Disbanding of the 'Waardgelders'  
(Mercenaries in the Pay of the Town  
Government) by Prince Maurits on the  
Neude, Utrecht, 31 July 1618



painting by Pauwels van Hillegaert,  
oil on canvas, h 98.5cm x w 171.2cm  
collection Rijksmuseum, Amsterdam



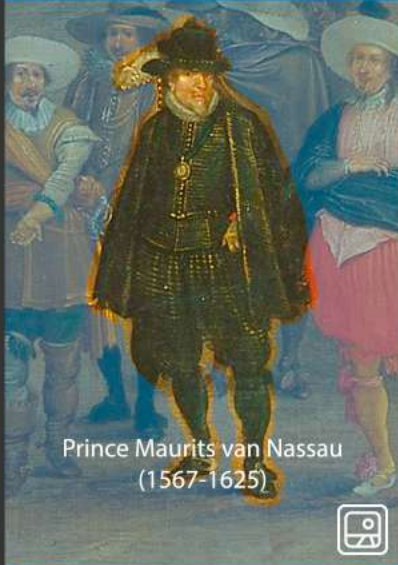




year 2018



year 1618



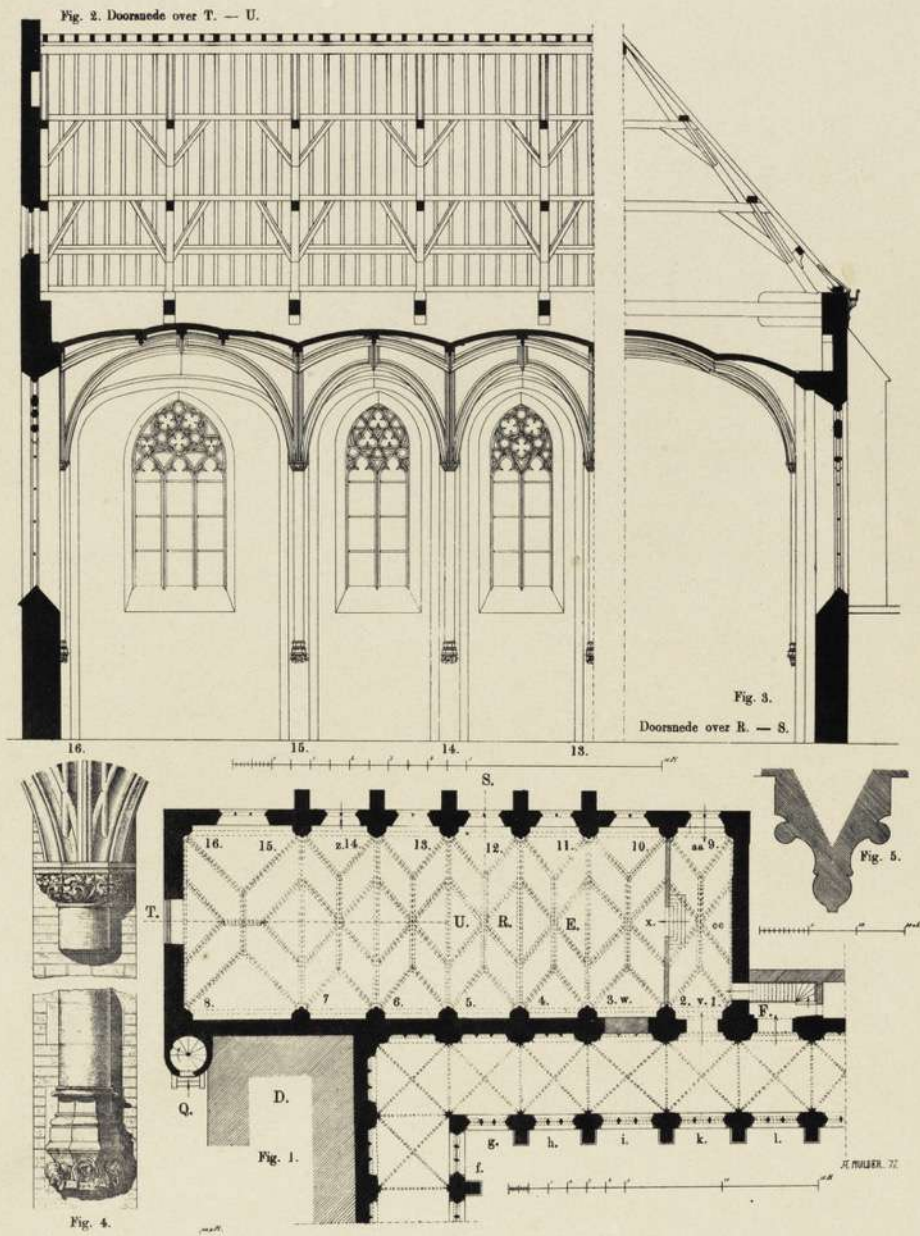
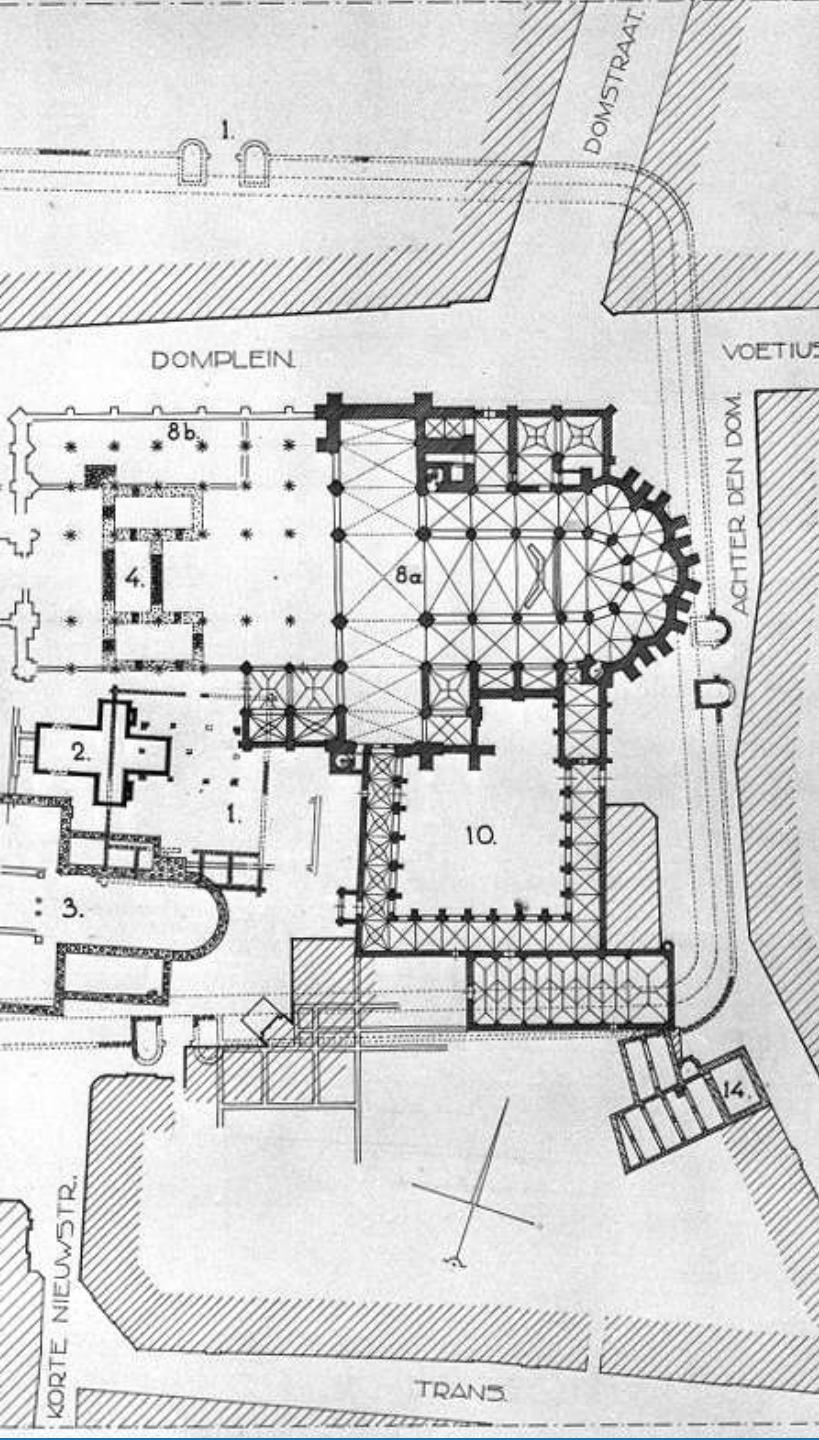
Prince Maurits van Nassau  
(1567-1625)













Inspector

EventSystem  Static

Tag Untagged Layer Default

Transform

Position X 0 Y 0 Z 0

Rotation X 0 Y 0 Z 0

Scale X 1 Y 1 Z 1

Event System (Script)

Script  `<EventSystem>`

First Selected None (Game Object)

Send Navigation Events

Drag Threshold 5

Standalone Input Module (5)

Script  `<StandaloneInputModule>`

Horizontal Axis Horizontal

Vertical Axis Vertical

Submit Button Submit

Cancel Button Cancel

Input Actions Per Sec 10

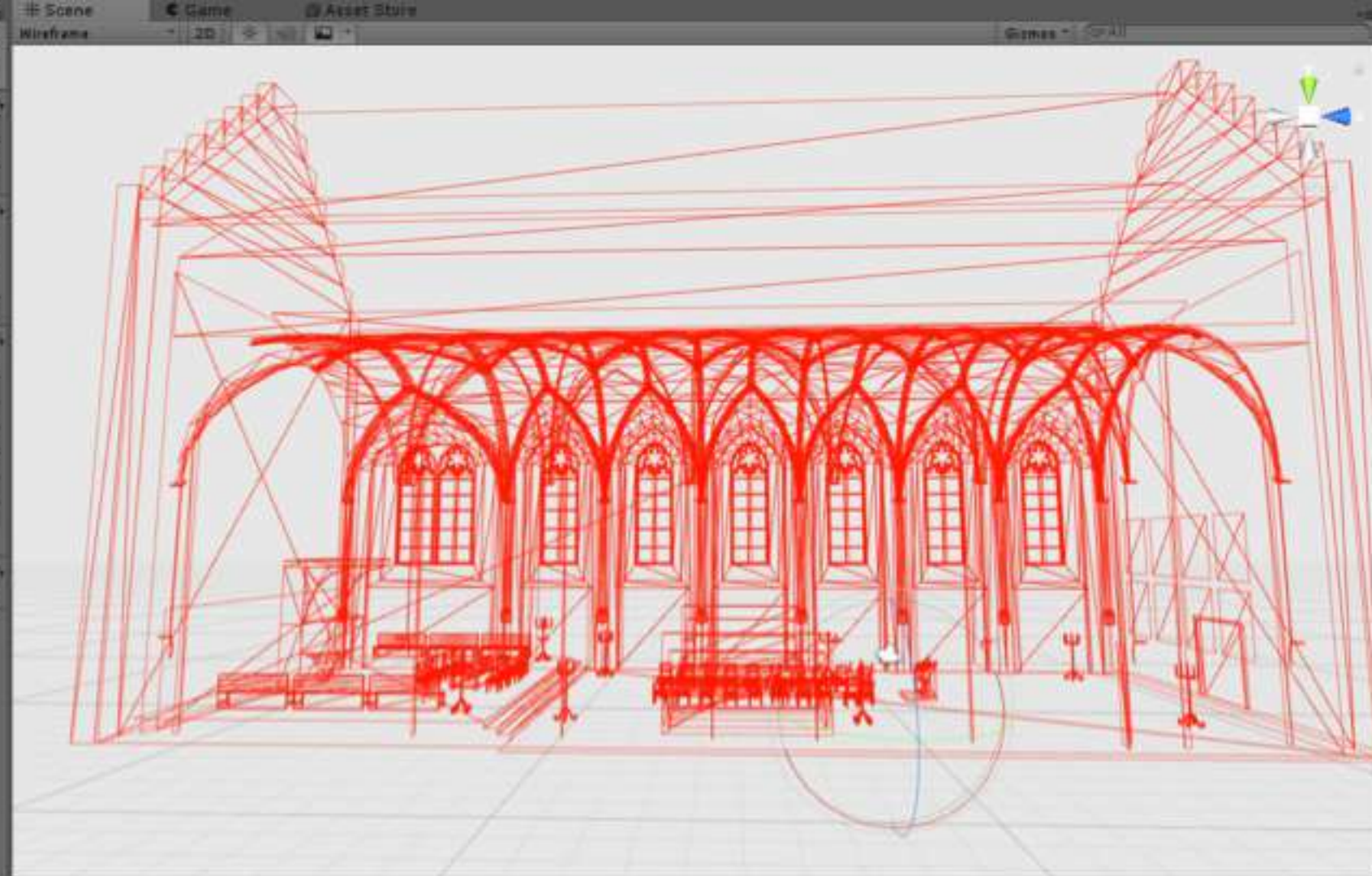
Repeat Delay 0.5

Force Module Active

Base Input (Script)

Script  `<BaseInput>`

Add Component



Lighting

Hierarchy

Create

Aula\*

- backgroundCamera
- CanvasCam
- 3D Canvas
- EventSystem
- Main Camera
  - Aula3D
    - Aula
    - pilaren
    - Table
    - Table (1)
    - CandleStick
    - stoelen
    - CandleCluster
    - Ramen
    - Papieren
    - Ceiling
    - Wapenz
    - Papieren (1)
    - Table (2)
    - Papieren (2)
    - lights
      - SimpleBench (1)
      - SimpleBench (2)
      - SimpleBench (3)
      - SimpleBench (4)
      - SimpleBench (5)
      - BigBench (1)
      - Troon
    - Beek
      - Spreekling
      - BigBench
      - SimpleBench
      - Candlestand
      - Pop
        - Quad
        - Quad (1)
        - Quad (2)

CanvasUI

Console Profiler Frame Debug

Add Profiler

Record Deep Profile Profile Editor Editor Allocation Callstacks Clear on Play Clear Load Save Frame Current

CPU Usage

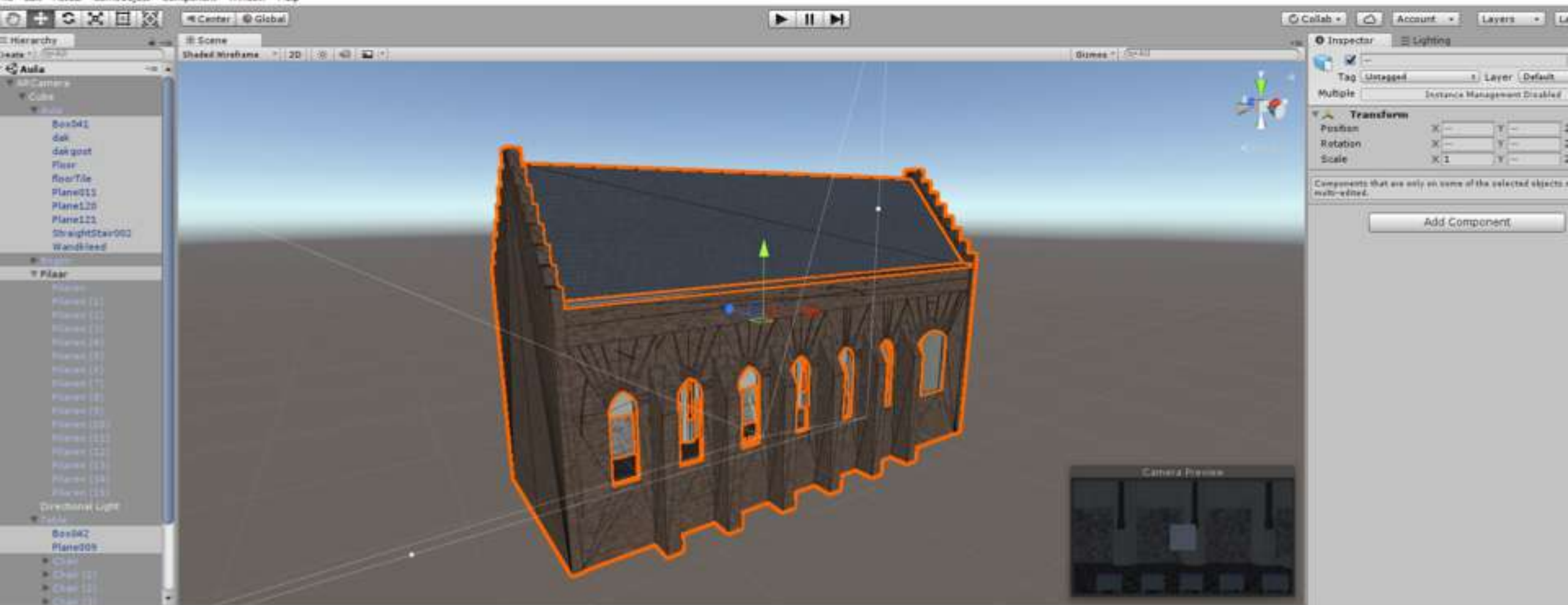
- Rendering
- Scripts
- Physics
- GarbageCollector
- VSync
- Global Illumination
- UI
- Others

1ms (1000FPS)

0.25ms (4000FPS)

0.1ms (10000FPS)

Hierarchy CPU: 5.54ms GPU: 0.00ms No Details



**Project**

- Assets - Vuforia - Prefabs
  - DefaultMdiIndicator
  - DefaultIndicator
  - DefaultPlaneIndicator
  - DefaultIndicator
  - GroundPlaneReference
  - MdiAirReference
  - Axis Y
  - Axis Z
  - Axis X

**Console**

Vuforia initialization successful

**Profiler**

Record | Deep Profile | Profile Editor | Editor | Allocation Callstacks | Clear on Play | Clear | Load | Save | Frame: Current | Current

**Memory**

CPU: 121.13ms GPU: 0.00ms

Overview	Total	Self	Calls	GC Alloc	Time ms	Self ms
EditorOverhead	90.9%	90.6%	2	0 B	110.23	109.85
Initialization.PlayerUpdateTime	8.8%	0.0%	1	0 B	10.47	0.01
Camera.Render	0.1%	0.0%	1	0 B	0.17	0.02
EarlyUpdate.UpdateMainGameV	0.0%	0.0%	1	0 B	0.02	0.00
GPU.RenderTime.RenderContext	0.0%	0.0%	1	0 B	0.01	0.00



experience the past on location  
in augmented reality

Utrecht University Hall



*Fin*



*The Challenge of aligning  
with CLARIN resources and  
infrastructure*

Questions and comments are very welcome!



# Thanks to all UTM Partners!!!

Descartes  
Centre



Utrecht University

HET ILLUSTRATIE  
UTRECHTS  
ARCHIEF!!  
ILLUSTRATIE  
ILLUSTRATIE

CENTRAAL  
MUSEUM  
UTRECHT



Time  
Machine

<https://timemachine.eu/>

<http://utrechttimemachine.nl/>

[t.pieters@uu.nl](mailto:t.pieters@uu.nl)



Utrecht University