Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

UNIVERSITÄT ZU LÜBECK

HEURIST
the iterative database

DATABASING ON DEMAND

HEURIST AS A PLATFORM FOR DATA-DRIVEN STORYTELLING

DR SYLVIA MELZER, UNIVERSITÄT HAMBURG & UNIVERSITÄT ZU LÜBECK

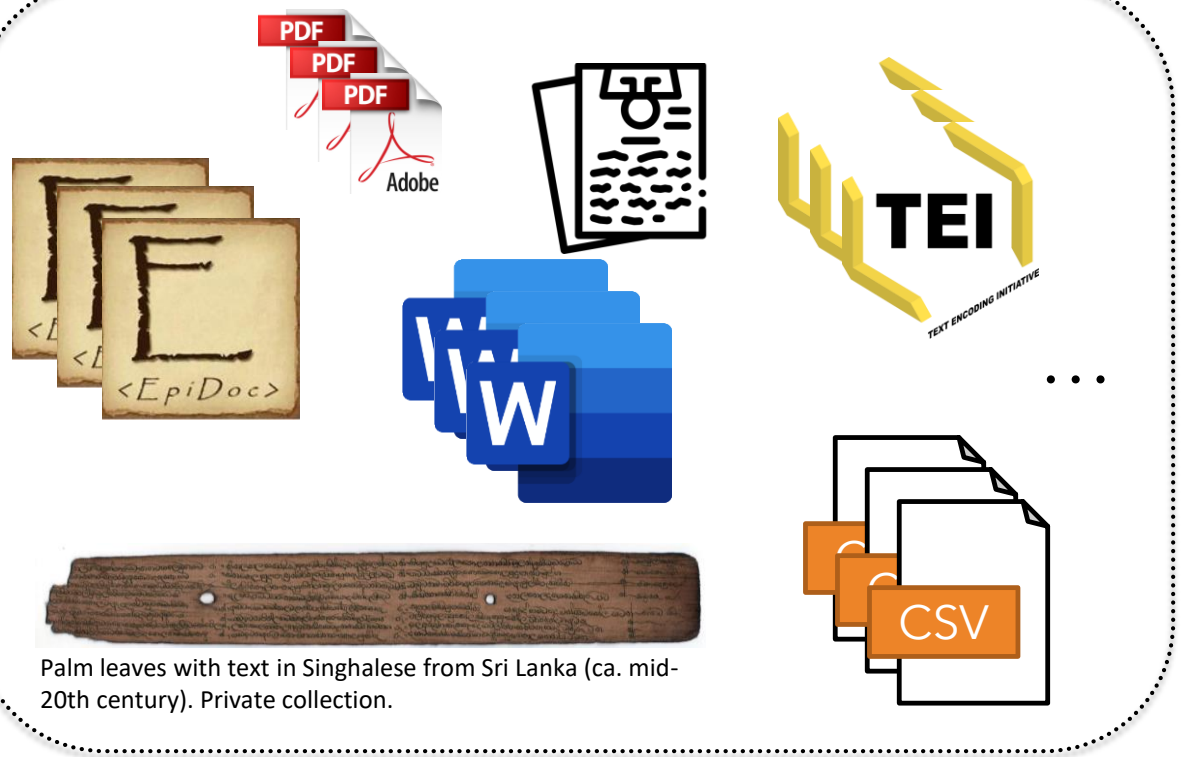
DARIAH ANNUAL EVENT 2022, 1ST JUNE 2022

OUTLINE

1. Motivation
2. Heurist – a database management system for the Humanities
3. Databasing on demand for a large data set
4. Building information systems for publishing research data

MOTIVATION

Research data



Palm leaves with text in Singhalese from Sri Lanka (ca. mid-20th century). Private collection.

- Research data are stored in different formats.
- How to visualize data in a human-legible way?
- How to store large data sets for publishing in short time?

MOTIVATION

```
<ab>
  <lb n="1"/>
  Πλανκίαν Μάγναν
  <lb n="2"/>
  Ἀκυλλίαν θυγατέρα Ἰου
  <lb n="3" break="no"/>
  λίου Σευήρου καὶ
  <expan>
    <abbr>Κλαυ</abbr>
    <ex>δίας</ex>
  </expan>
  <lb n="4"/>
  Ἀκυλλίας, τὴν ἐγ βασιλέ
  <lb n="5" break="no"/>
  ων, ἡρωίδα, ὁ δῆμος ὁ Ἀν
  <lb n="6" break="no"/>
  κυρανῶν τὴν θυγατέρα τῆς
  <lb n="7"/>
  μητροπόλεως.
</ab>
```

machine-readable format

EpiDoc-Stylesheets[1]

human-readable format

1 Πλανκίαν Μάγναν
2 Ἀκυλλίαν θυγατέρα Ἰου-
3 λίου Σευήρου καὶ Κλαυ(δίας)
4 Ἀκυλλίας, τὴν ἐγ βασιλέ-
5 ων, ἡρωίδα, ὁ δῆμος ὁ Ἀν-
6 κυρανῶν τὴν θυγατέρα τῆς
7 μητροπόλεως.

Projects: Research Unit Data Linking

[1]https://sourceforge.net/p/epidoc/wiki/Styleshee

HEURIST – A DATABASE MANAGEMENT SYSTEM FOR THE HUMANITIES

Search
Simple text searches and classification of an object by assignment of several independent terms (faceted searches) to analyze research data.
E.g.: imperial → 1st cent. CE, 2nd cent CE

Extension and customization
Data structures may change during the project work or extensions may be necessary. A tool should support later changes or extensions in a simple way.

Viewing and Publishing
The data should also be made available to other users.

Collaboration
Research is supported when there is an environment for sharing and collaboration with other users through defined working groups and ownership at the dataset level.

Integration
Simple Integration of data in different formats (e.g. CSV) from different sources or projects

DATABASING ON DEMAND FOR A LARGE DATA SET

6570 EpiDoc.xml files

242 TEI.xml files

- Takes a lot of time
- High effort to add data sets to a database
- Integration of additional data sets from other projects, which are also associated with high effort
- Representation of data in required coding (e.g. the Leiden Conventions)

information system

information system

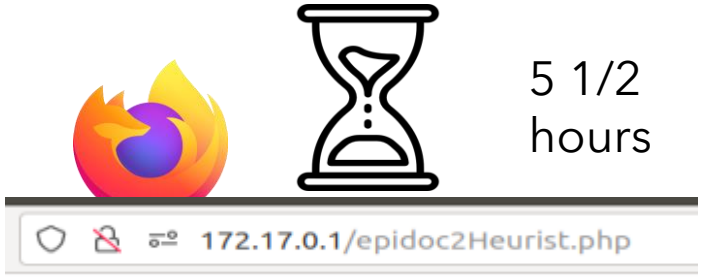
Icons made by Flaticon, Eucalyp from www.flaticon.com

DATABASING ON DEMAND FOR A LARGE DATA SET

6570 EpiDoc.xml files

execute program to generate csv file

output/input

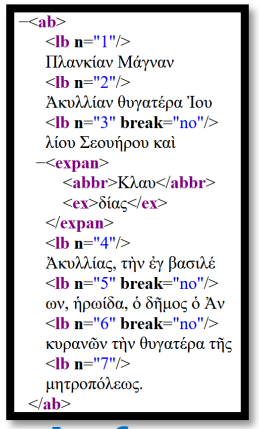


5 1/2 hours

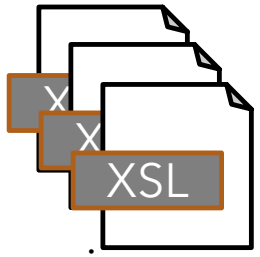
type 172.17.0.1/epidoc2Heurist.php in browser to create a database on demand

database

create website



before



input

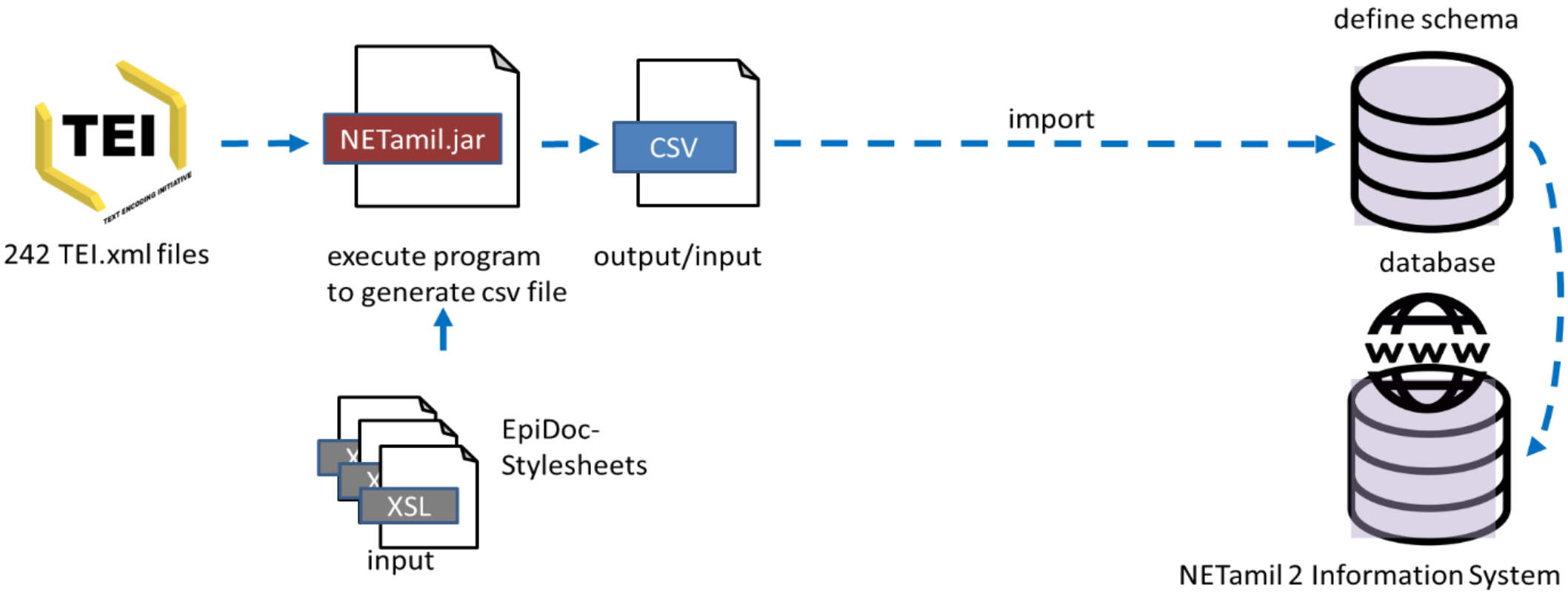
EpiDoc-Stylesheets

5	Πλανκίαν Μάγναν Ἀκυλλίαν θυγατέρα Ἰου-λίου Σευήρου καὶ Κλαυ(δίας) Ἀκυλλίας, τὴν ἐγ βασιλέ-ων, ἡρωίδα, ὁ δῆμος ὁ Ἀν-κυρανῶν τὴν θυγατέρα τῆς μητροπόλεως.

after

information system

DATABASING ON DEMAND FOR A LARGE DATA SET



Storage of data in a short time

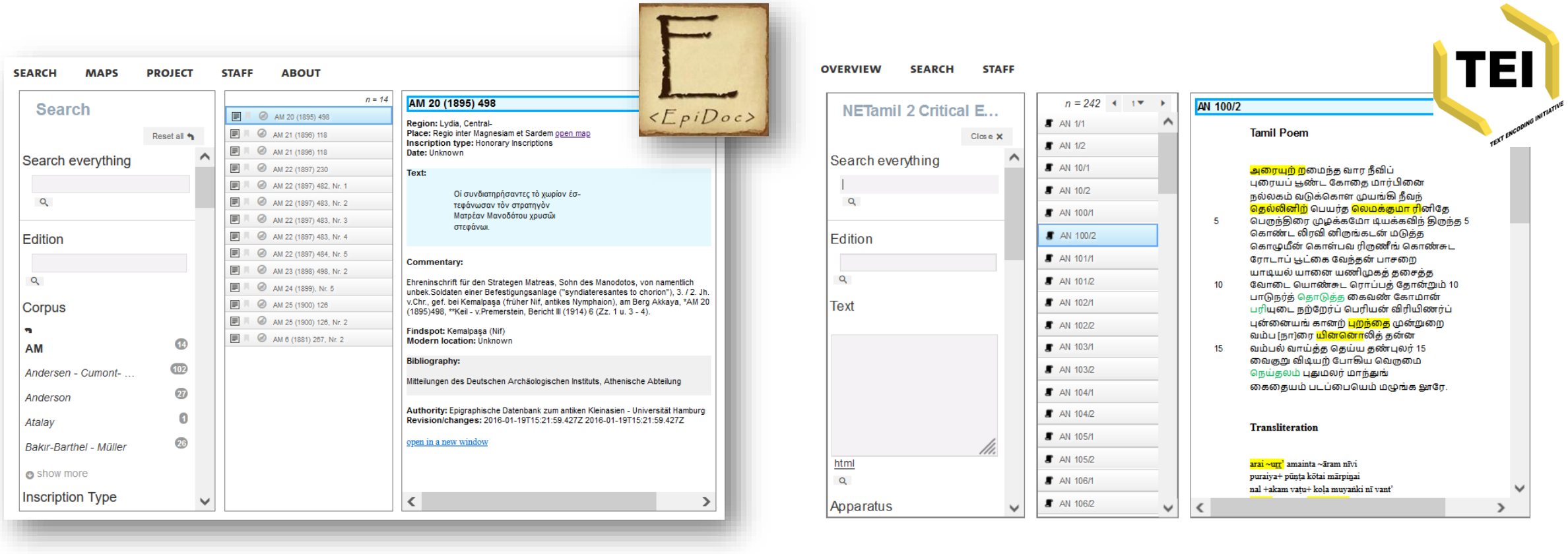
BUILDING INFORMATION SYSTEMS FOR PUBLISHING RESEARCH DATA

EDAK_Template.tpl 4.17 KiB

```
<html>

{*-------------------------------------------------------------*}
{foreach $results as $r} {* Start records loop, do not remove *}
{$r = $heurist->getRecord($r)}
{*-------------------------------------------------------------*}

{if ($r.recTypeID=="70")}{* Description *}

<h1 style="border-color: #00a9ff;
background-color: #e4f8fd;{* #86cde8 *}
border-width: 5px;
border-style: solid;
    color: Black;
     font-family: arial,helvetica,sans-serif;
 font-size: 18px;">{$r.f1181}{*Title*} </h1>

    {if ($r.f731)}{* Image *}
    {wrap var=$r.f731_originalvalue dt="file" width="300" height="auto" auto_play="0" show_artwork="0
     <br/>
   {/if}    {* Image *}
```

- The Heurist web designer for creating project specific layouts
  - Web editor
  - Supported: PHP, HTML

BUILDING INFORMATION SYSTEMS FOR PUBLISHING RESEARCH DATA



Visualization of data in a human-readable format

Thank you for your attention.

Acknowledgment

The research for this contribution was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2176 'Understanding Written Artefacts: Material, Interaction and Transmission in Manuscript Cultures', project no. 390893796.