

Ancient Manuscripts Research in Universitas Padjadjaran

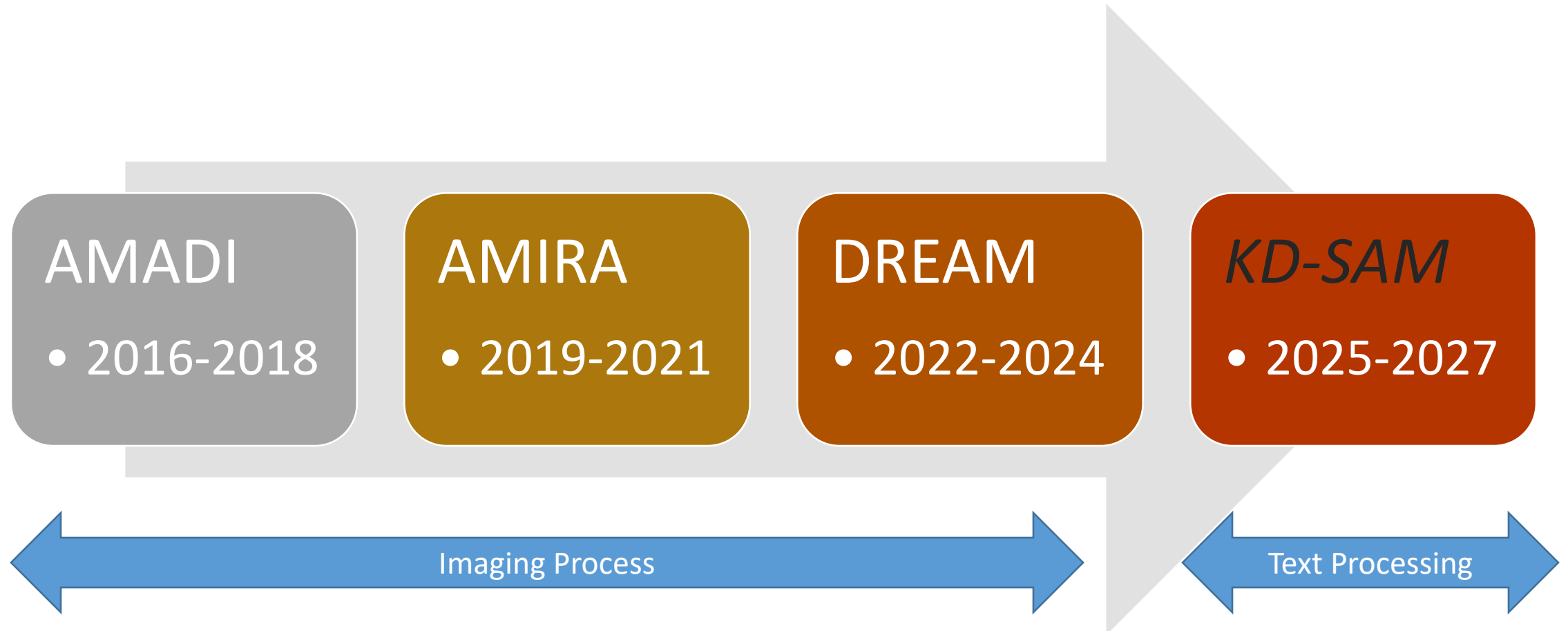
**RISE SMA 2022
30 September 2022
Leiden University**

From West Java for Indonesia to the World through SDGs

www.unpad.ac.id



Research Phases



AMADI (2015-2018)

Ancient Manuscript Digitization and Indexation

International Collaboration Research

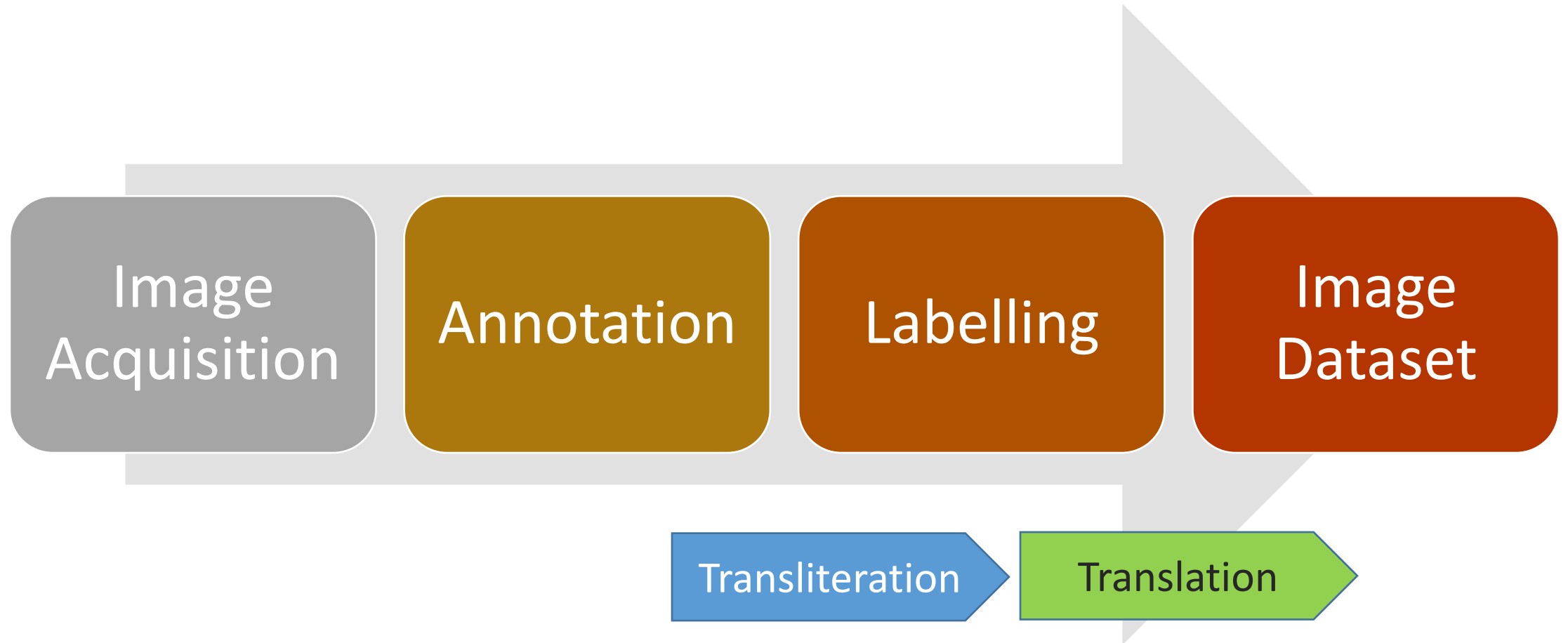
University of La Rochelle France under STIC ASIE program, supporting with French Ministry of Foreign Affairs and International Development (MAEDI), CNRS, INSERM, and IRD

Indonesian Ministry of Education under Foreign Research and Collaboration (KLN) Scheme

Computer Science Department with Faculty of Cultural Science Universitas Padjadjaran

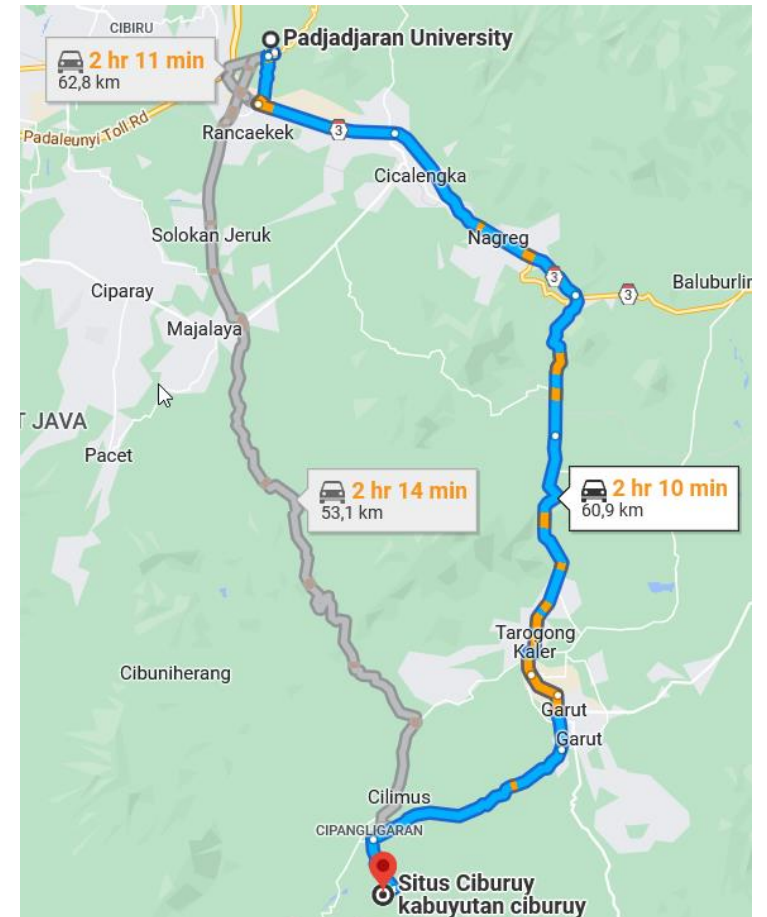


AMADI





Corpus Location





Digitization



3



Binarization Process





Annotation by Philologist

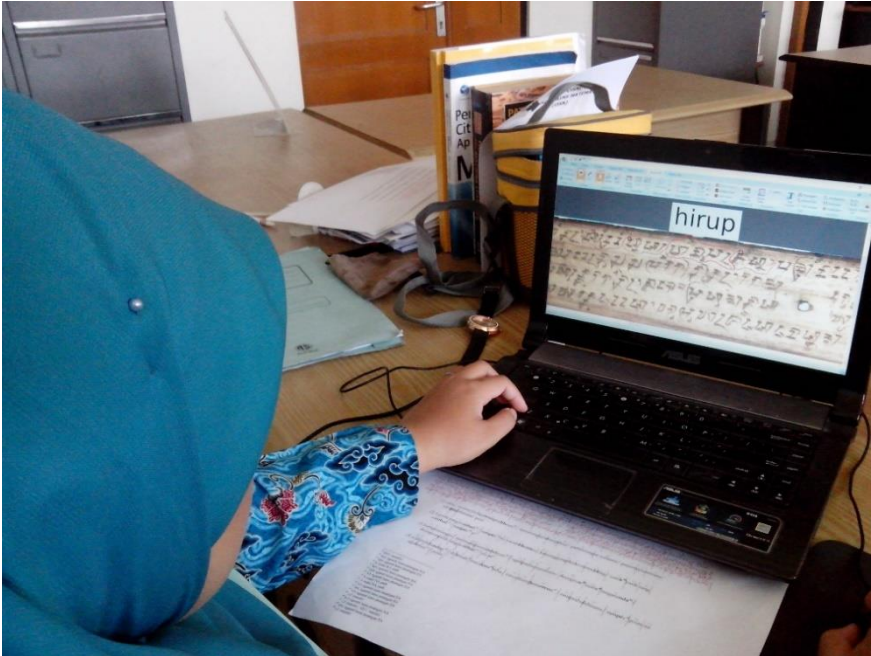
1. 3306.tiff

Handwritten text in red ink with vertical lines indicating syllable boundaries. The text is written in a cursive style and includes several lines of characters, some of which are underlined or marked with asterisks.

- (1) ngeu¹*nm*abujanggalawa/ saur²taandeuwisita/ aki³ayakitadinya/ anak⁴1/sahadyayunan/ sa-*rwa*²deungeunsiutun/
némbal³aki ha-
- (2) may¹canggong/ samapungeura/nu/calik⁴/ ébohnahulunpupulih/ basaaingngasuh**bonycah**⁵/
horéngannuturkeun⁶kacai/ mantaré⁷aingmilanga-
- (3) n/ nasanghyangwatangageung/ siutunhateudingsir/ carékkéng⁸karahsakini/ tostangayunapussageung⁹/
dikawihhan¹⁰babalik¹¹kan¹¹/ tuluy/
- (4) metubonycah¹²lanang/ sa-*rwa*¹³deungeunsiutun/ saur¹⁴taandeuwisita/ kamayangankitaséda/
metukeun*nm*¹⁴kasaktyin/ kamayanganaing



Data Annotation



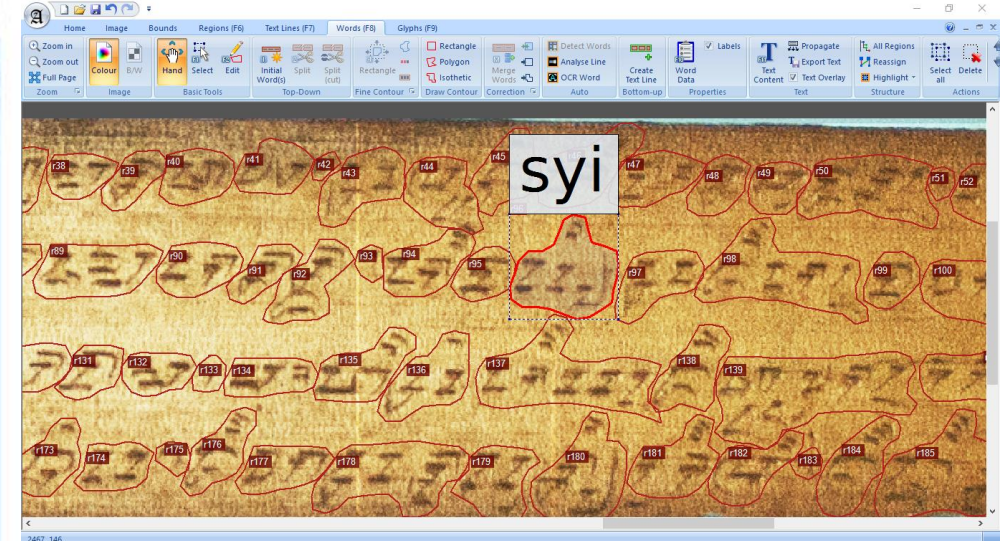
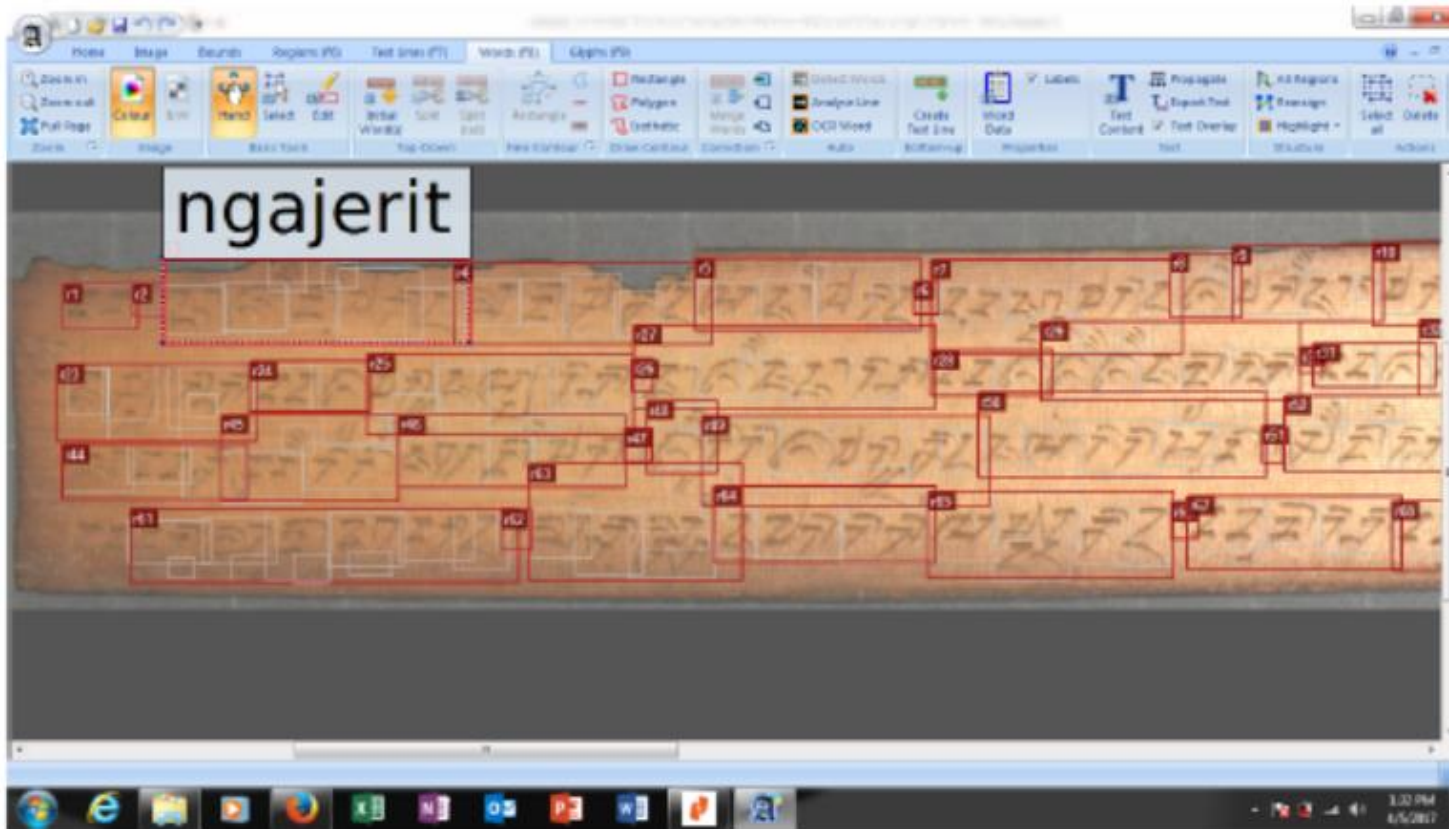
The segmentation and annotation in word level using Aletheia

```
New Text Document - Notepad
File Edit Format View Help
<?xml version="1.0" encoding="UTF-8"?>
<PcGts xmlns="http://schema.primaresearch.org/PAGE/gts/pagecontent/2013-07-15"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" \
xsi:schemaLocation="http://schema.primaresearch.org/PAGE/gts/pagecontent/2013-07-15
http://schema.primaresearch.org/PAGE/gts/pagecontent/2013-07-15/pagecontent.xsd">
  <Metadata>
    <Creator></Creator>
    <Created>2015-10-04T21:40:38</Created>
    <LastChange>2015-10-08T07:50:07</LastChange></Metadata>
    <Page imageFilename="3306_new.tif" imageWidth="3025" imageHeight="326">
    <TextRegion id="tempReg357564684568544579089">
    <Coords points="0,0 1,0 1,1 0,1"/>
    <TextLine id="tempLine357564684568544579089">
    <Coords points="0,0 1,0 1,1 0,1"/>
    <Word id="r1">
    <Coords points="101,9 100,19 94,28 88,42 77,44 67,48 57,48 53,58 55,68 60,77 69,83 80,83 90,81 102,82 111,87 121,89
130,82 140,77 149,72 159,71 169,69 175,60 180,50 181,40 184,30 184,19 177,11 167,7 157,6 102,6"/>
    <TextEquiv>
    <Unicode>ngeunna</Unicode></TextEquiv></Word>
```

The result of segmentation and annotation process in xml format

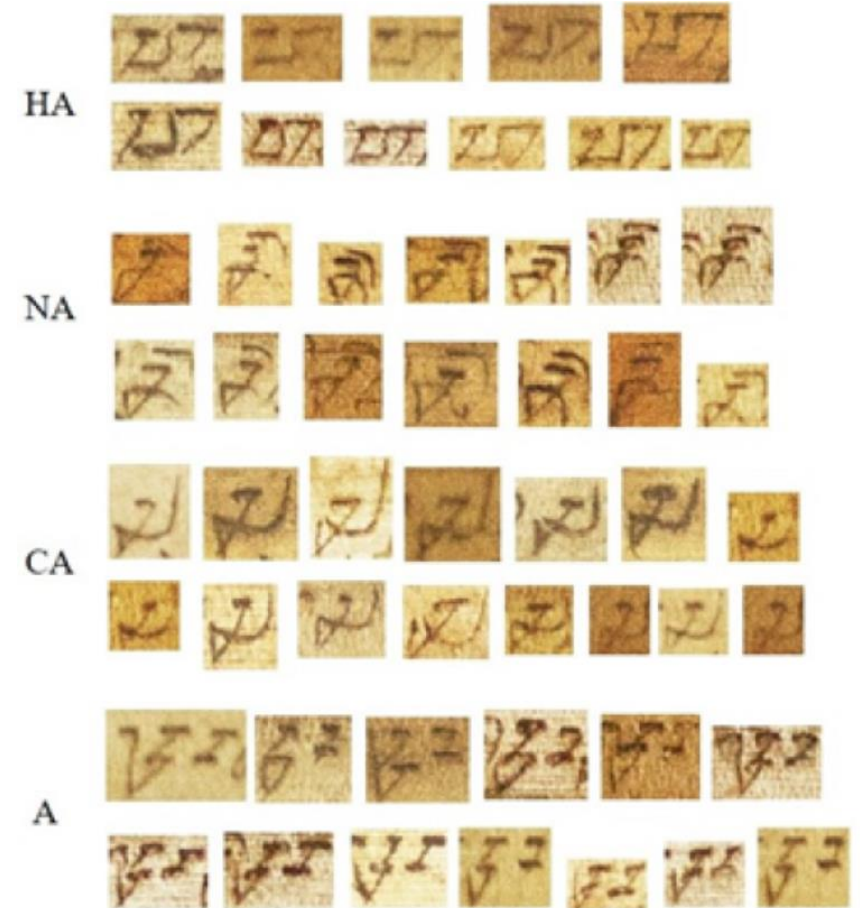


Result of Data Annotation (Word and Syllable Level)





Sample Image of Words and Syllable

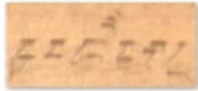




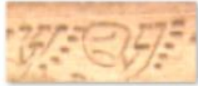
Variation in Words and Syllables



abuing_CB-3-22-90-40_r320_1964-7-2136-88.png



adeg_CB-3-22-90-23_r386_2435-83-2630-170.png



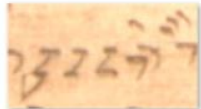
adeg-adeg_CB-3-22-90-9_r354_1550-33-1741-113.png



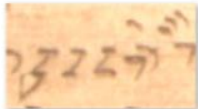
adeg-adeg_CB-3-22-90-22_r36_1403-109-1612-171.png



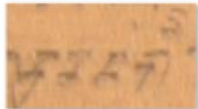
adiing_CB-3-22-90-8_r338_2721-6-2898-97.png



ageung_CB-3-22-90-17_r348_1006-177-1171-267.png



ageung_CB-3-22-90-17_r348_1006-177-1171-268.png



ageung_CB-3-22-90-34_r385_412-158-553-236.png



ageung_CB-3-22-90-35_r398_982-150-1110-220.png



ageungna_CB-3-22-90-35_r353_1586-5-1749-83.png



NA_90.png



NA_91.png



NA_92.png



NA_93.png



NA_94.png



NA_95.png



NA_102.png



NA_103.png



NA_104.png



NA_105.png



NA_106.png



NA_107.png



NA_114.png



NA_115.png



NA_116.png



NA_117.png



NA_118.png



NA_119.png



NA_126.png



NA_127.png



NA_128.png



NA_129.png



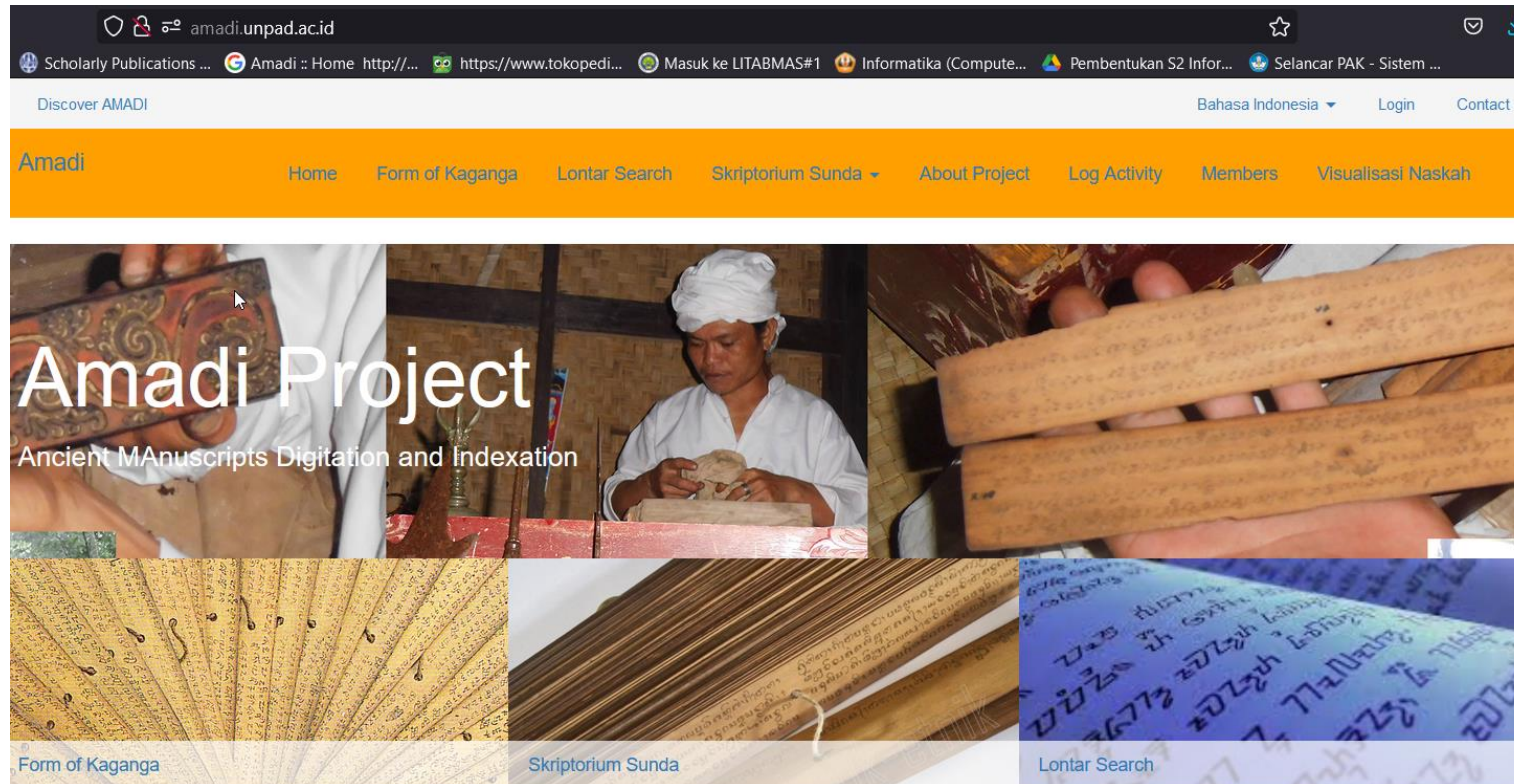
NA_130.png



NA_131.png



<https://amadi.unpad.ac.id>



Form of Kaganga

Aksara swara adalah aksara yang secara silabis memiliki harkat bunyi vokal yang dalam sistem aksara Sunda

Skriptorium Sunda











Skriptorium adalah tempat di mana naskah-naskah manuskrip disalin oleh para juru tulis. Skriptorium

About Project

Dalam rangka melestarikan pengetahuan dalam naskah kuno terutama dalam Lontar bahasa Sunda, studi tentang

- The Handwritten Sundanese Palm Leaf Manuscript Dataset from 15th Century, 2017, 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)



1_19_006A		katuduHhan nya mana hatina mNA,...
1_19_006B		jati na beneR laNGgeNG na temen...
1_19_007A		n tan pa katuduH nu Ageus pun...
1_19_007B		na, luput tan hana, nyamana puhun...
1_19_008A		lan ja léga, saNGkanna...
1_19_008B		muji saNg hYaNG sadi tutuR jati,...
1_19_009A		di ngarannan panutan keuna dituduH...
1_19_009B		pa ngaranna, INGAH? Inya maNGna...
1_19_010A		jati niskala, pun ## ꦩ...
1_19_010B		kadi kalinga niNG puhun puhun...



Number of corpus

Dashboard

[Home](#) / [Dashboard v1](#)

1456
Data 

558
Transliteration 

2
Translate 

522
BOX 1 
[More info](#) 

598
BOX 2 
[More info](#) 

332
BOX 3 
[More info](#) 

AMIRA (2019-2021)

Ancient Manuscript Image Retrieval and Analysis

National Research

Ministry of Research and Technology under University Excellent Research Scheme

Faculty of Cultural Science Unpad and Computer Science Department

Ethnoinformatics Research and Study Center Unpad



AMIRA

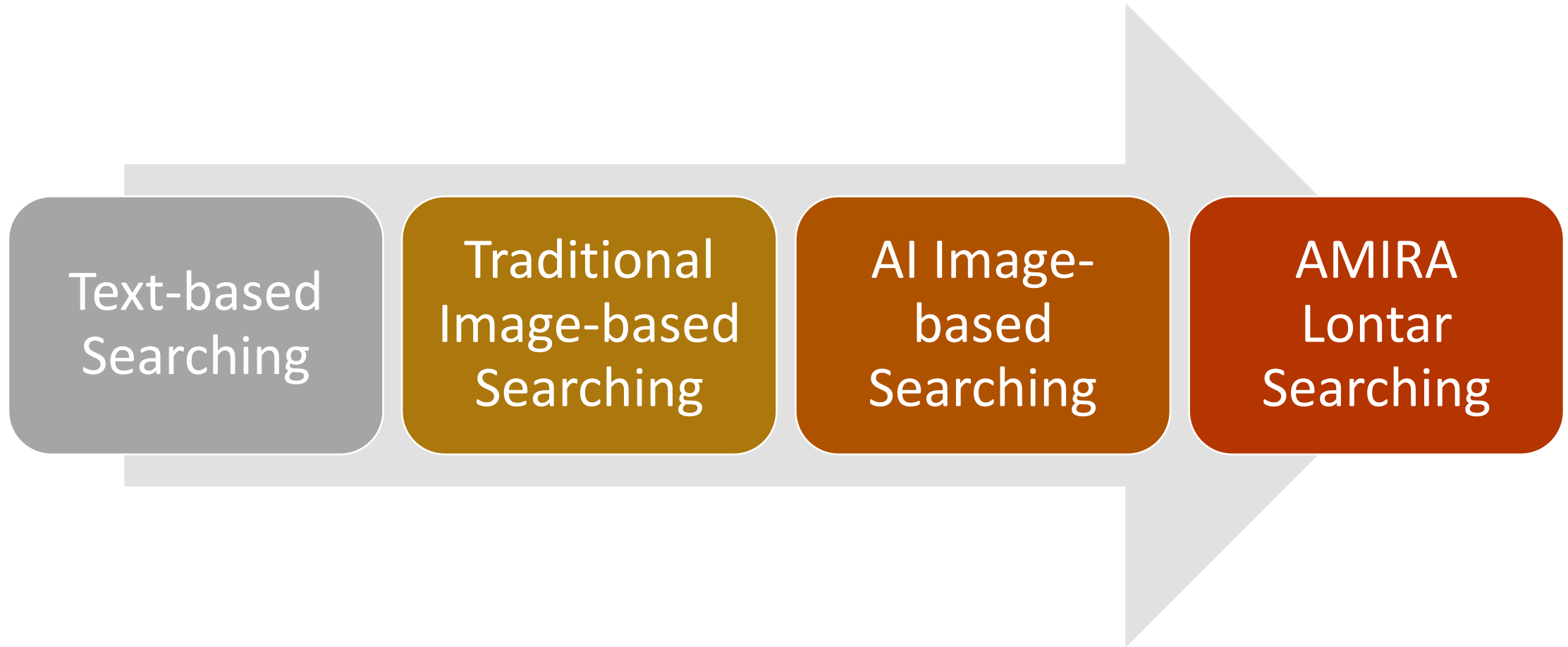




Image Search By Text

Lontar Search

localhost:8000

Lontar

Keyword
raden bibisana

Bayesian Network
 Divergent From Randomness
 PL2F

Document: 3349
Method: DFR:PL2F Probability Value: 7.2756331808674

Document: 3313
Method: DFR:PL2F Probability Value: 4.1520618343339

Search the web and Windows

10:58 AM
04-Aug-16

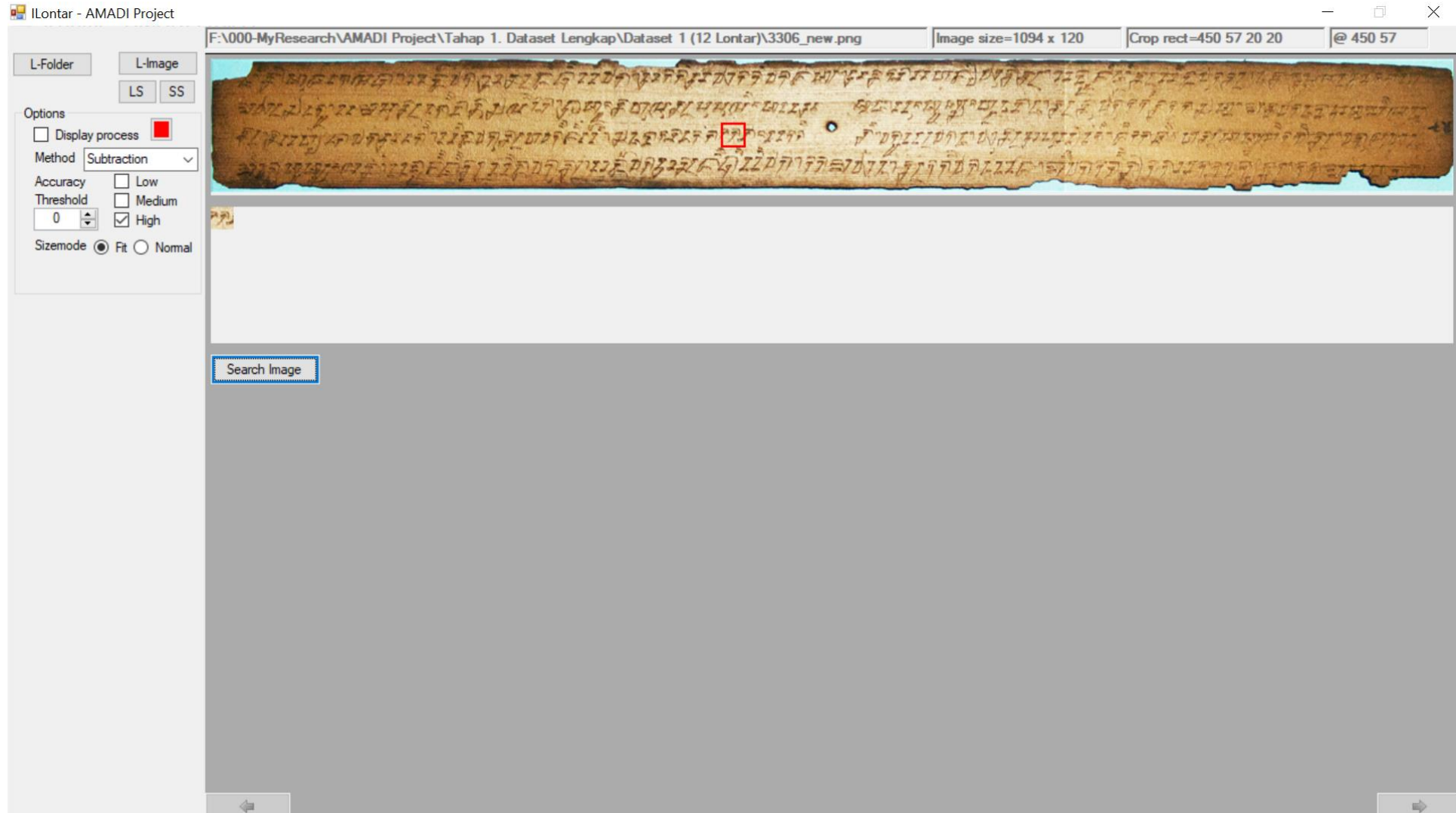
Sundanese ancient manuscripts search engine using probability approach

October 2017

[Journal of Physics Conference Series](#)
893(1):012064



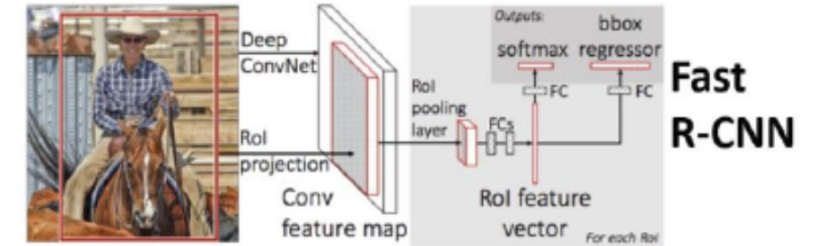
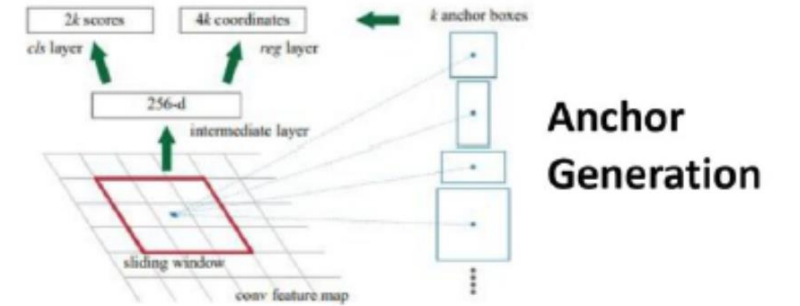
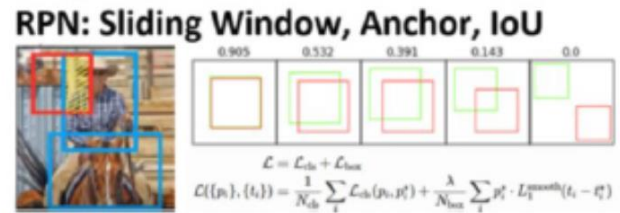
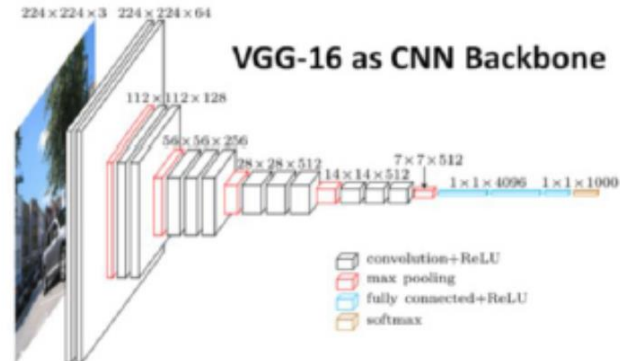
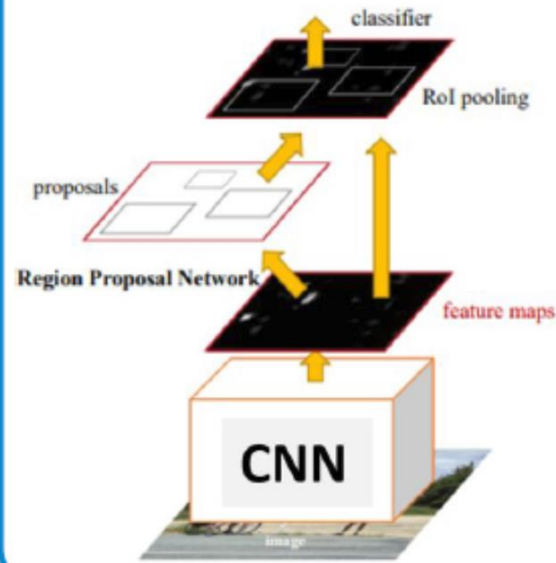
Traditional Image-based Search



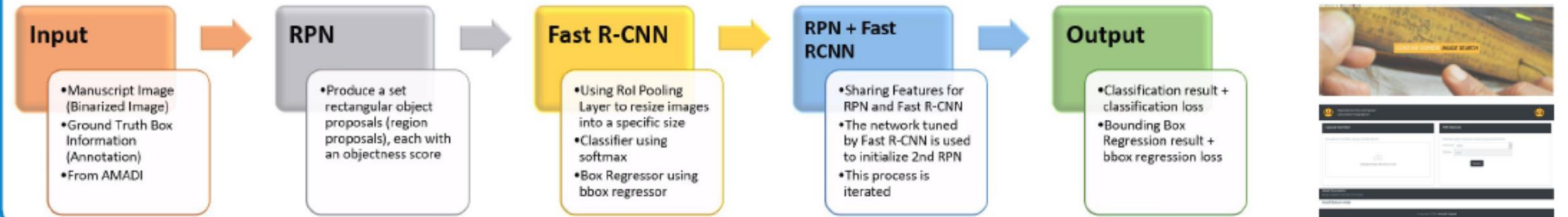


FASTER R-CNN

METHOD : FASTER REGION CONVOLUTIONAL NEURAL NETWORK (FASTER R-CNN)



EXPERIMENT





AMIRA

Ancient Manuscript Image Retrieval and Analysis

login



LONTAR SUNDA IMAGE SEARCH

Copyright 2019: Amira Unpad



Threshold : 100
Metode Seleksi : democracy



Hasil Gambar RA

Pada lontar 3311

Hasil Gambar RA

Pada lontar 3312

Hasil Gambar RA

Pada lontar 3313

Hasil Gambar RA

Pada lontar 3316

Hasil Gambar RA

Pada lontar 3322

Hasil Gambar RA

Pada lontar 3325

Hasil Gambar RA

Pada lontar 3349

Hasil Gambar RA

Pada lontar 3372

3308.png (3025 x 326)



DREAM (2022-2024)

Digital Reconstruction and Enhancement of Ancient Manuscripts

National Research

Ministry of Education, Culture, Research and Technology under Excellent Research Scheme

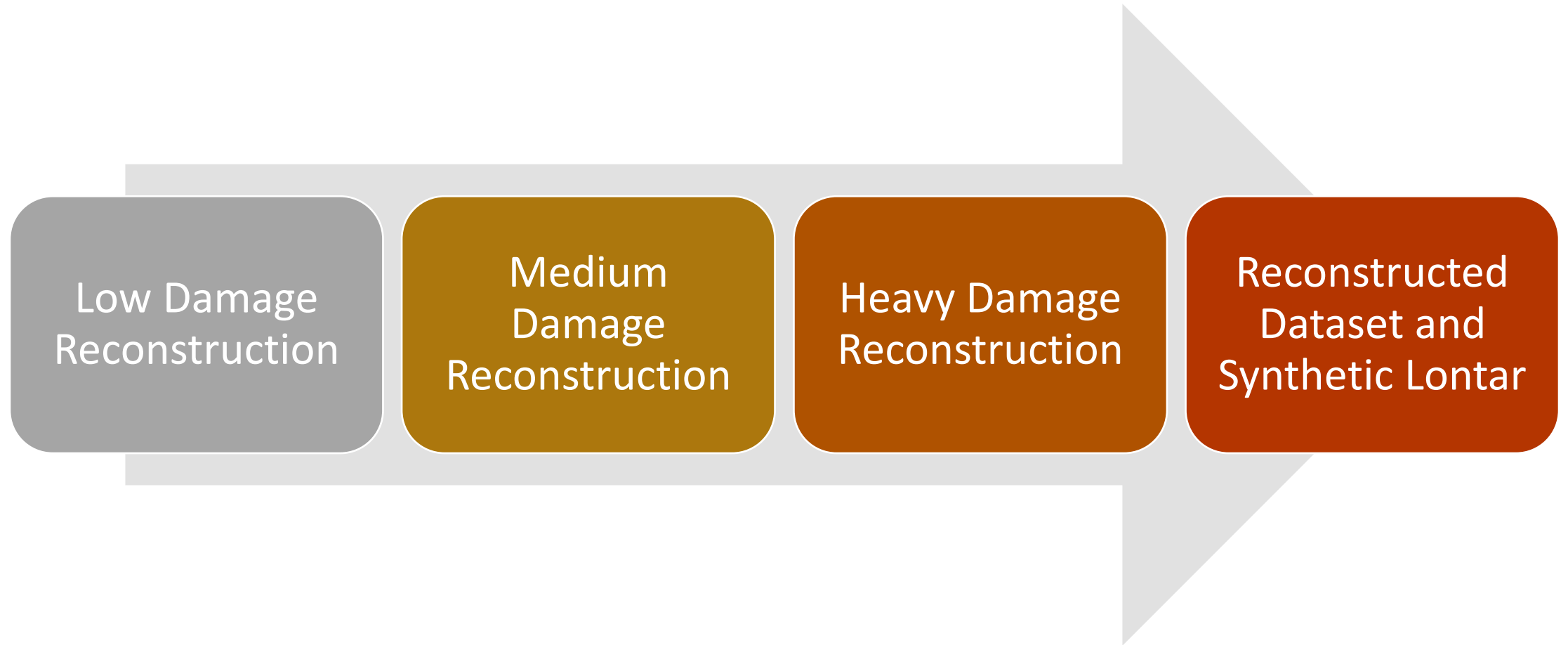
Faculty of Cultural Science Unpad and Computer Science Department

Ethnoinformatics Research and Studi Center Unpad

Research Center for AI and Big Data (AIDA)



DREAM Steps





Low Damage Lontar



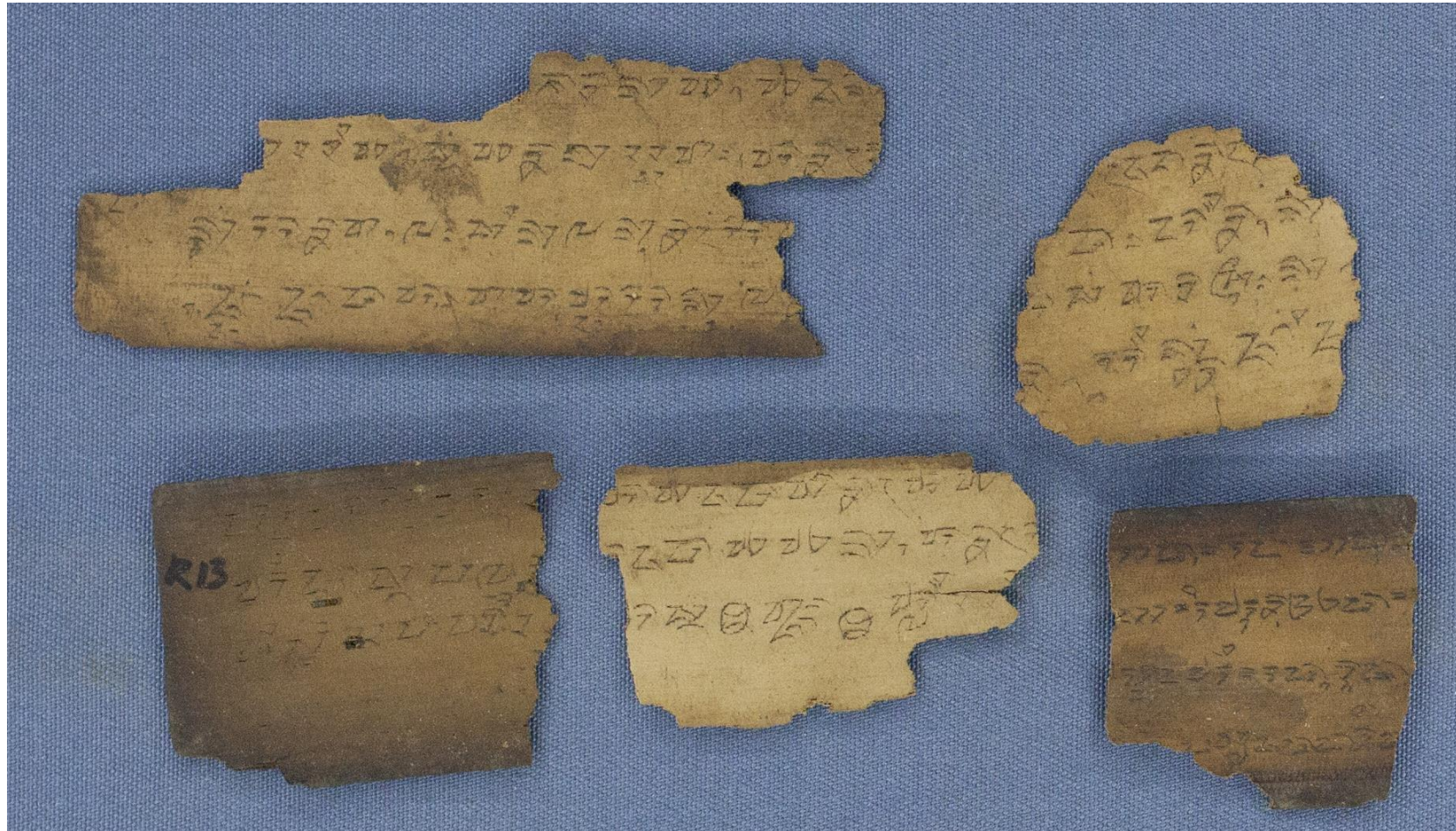


Medium Damage Lontar



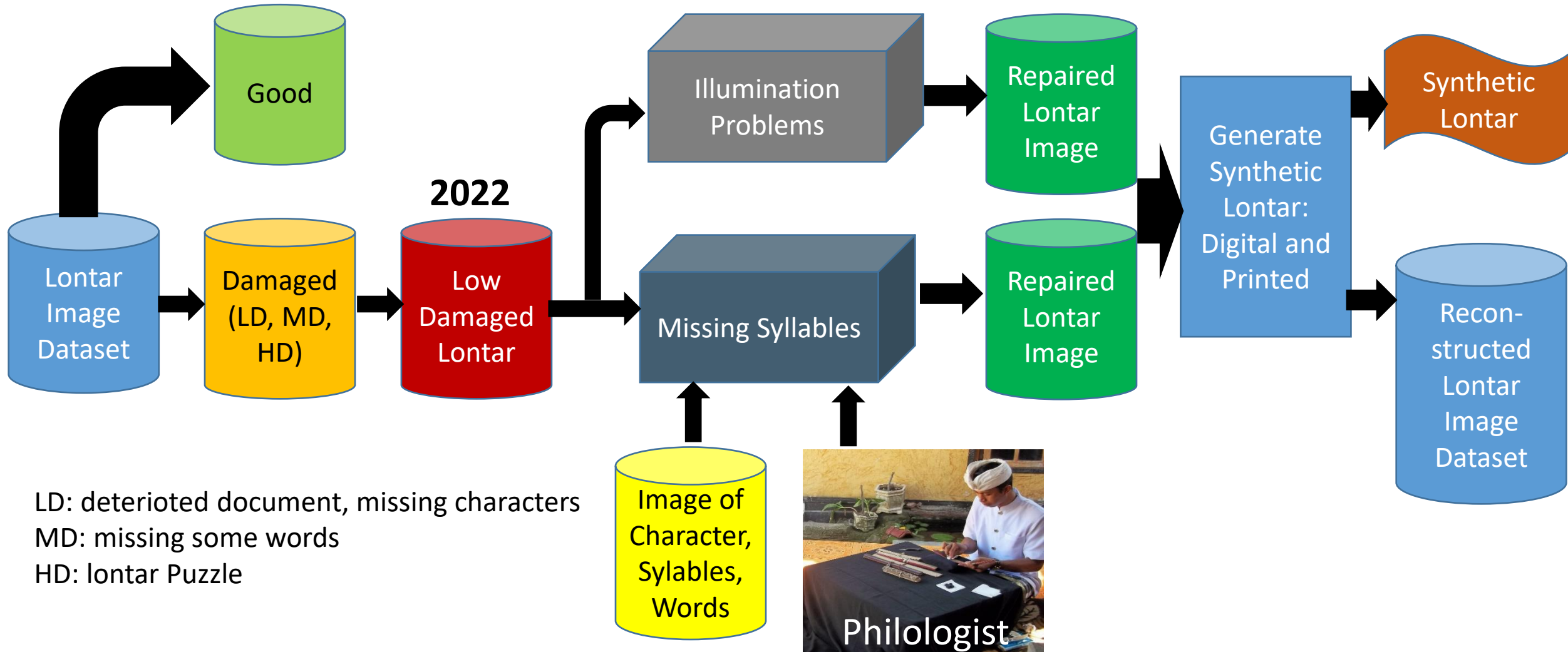


Heavy Damage Lontar



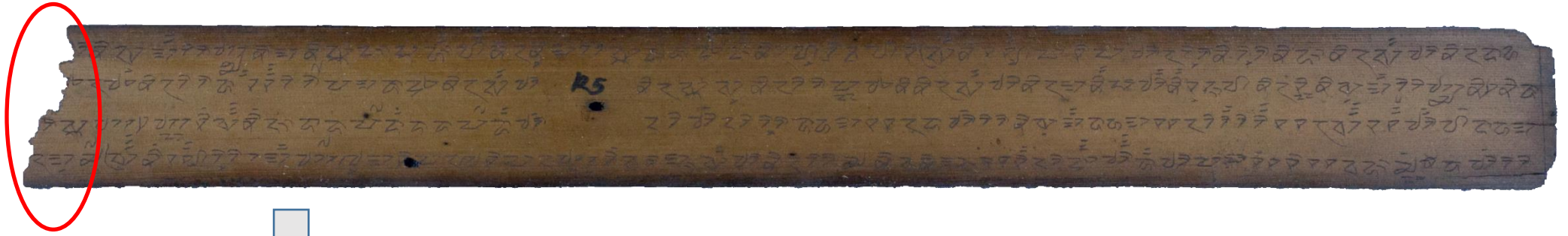


Reconstruction Pipeline

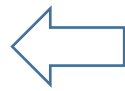
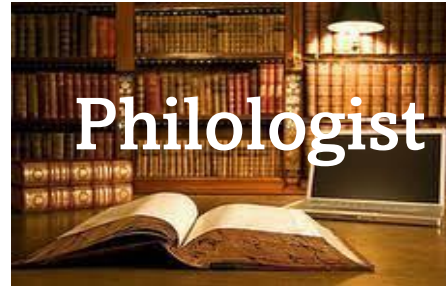




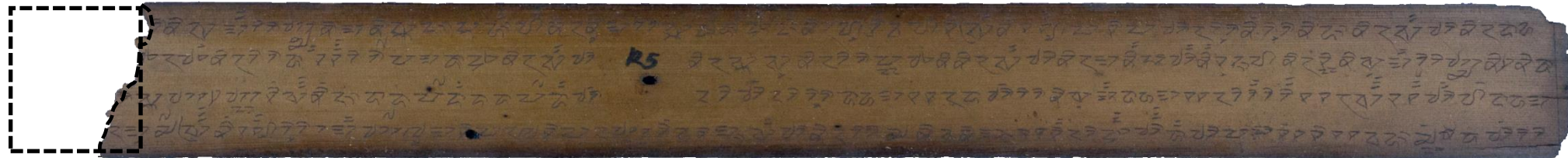
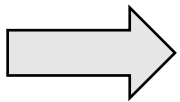
DREAM: Synthetic Lontar Generation



Missing characters identification

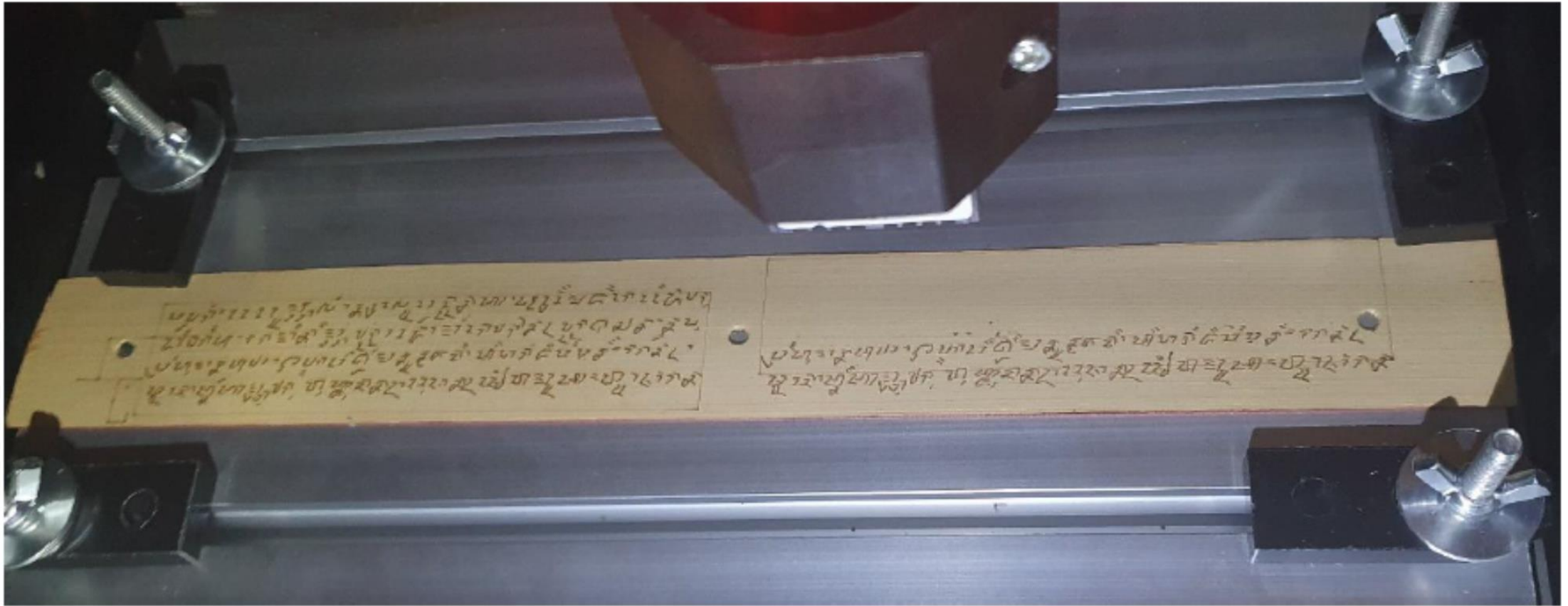


Generate synthetic background and missing characters/words





Synthetic Lontar Generation





Final Results of this Research

- Full fledged lontar manuscripts
- Although it is not original, but the information in the lontar is re-generated based on the prediction of experts
- Ready for the knowledge discovery process, the next research phase.

KD-SAM (2025-2027)

Knowledge Discovery of Sundanese Ancient Manuscripts

International Research (planning)

From transliteration to interpretation

Knowledge through Interpretation



Transliteration

1. 3306.tiff

Handwritten text in red ink, likely a form of Javanese script, with vertical lines separating words. The text is arranged in four lines, corresponding to the four items in the transliteration list below.

- (1) ngeu^{nm}abujanggalawa/ saurta^{andeu}wisita/ aki^{ya}kitadinya/ anak¹sahadyayunan/ sa-rwa²deungeunsiutun/
némbal³aki ha-
- (2) maycanggong/ samapungeura/ nu/ calik⁴/ ébohnahulunpupulih/ basaaingngasuhbonycah⁵/
horéngannuturkeun⁶kacai/ mantaré⁷aingmilanga-
- (3) n/ nasanghyangwatangageung/ siutunhateudingsir/ carékkéng⁸karahsakinini/ tostangayunapussageung⁹/
dikawihhan¹⁰babalik^{kan}¹¹/ tuluy/
- (4) metubonycah¹²lanang/ sa-rwa¹³deungeunsiutun/ saurta^{andeu}wisita/ kamayangankitaséda/
metukeun^{nm}¹⁴kasaktyin/ kamayanganaing



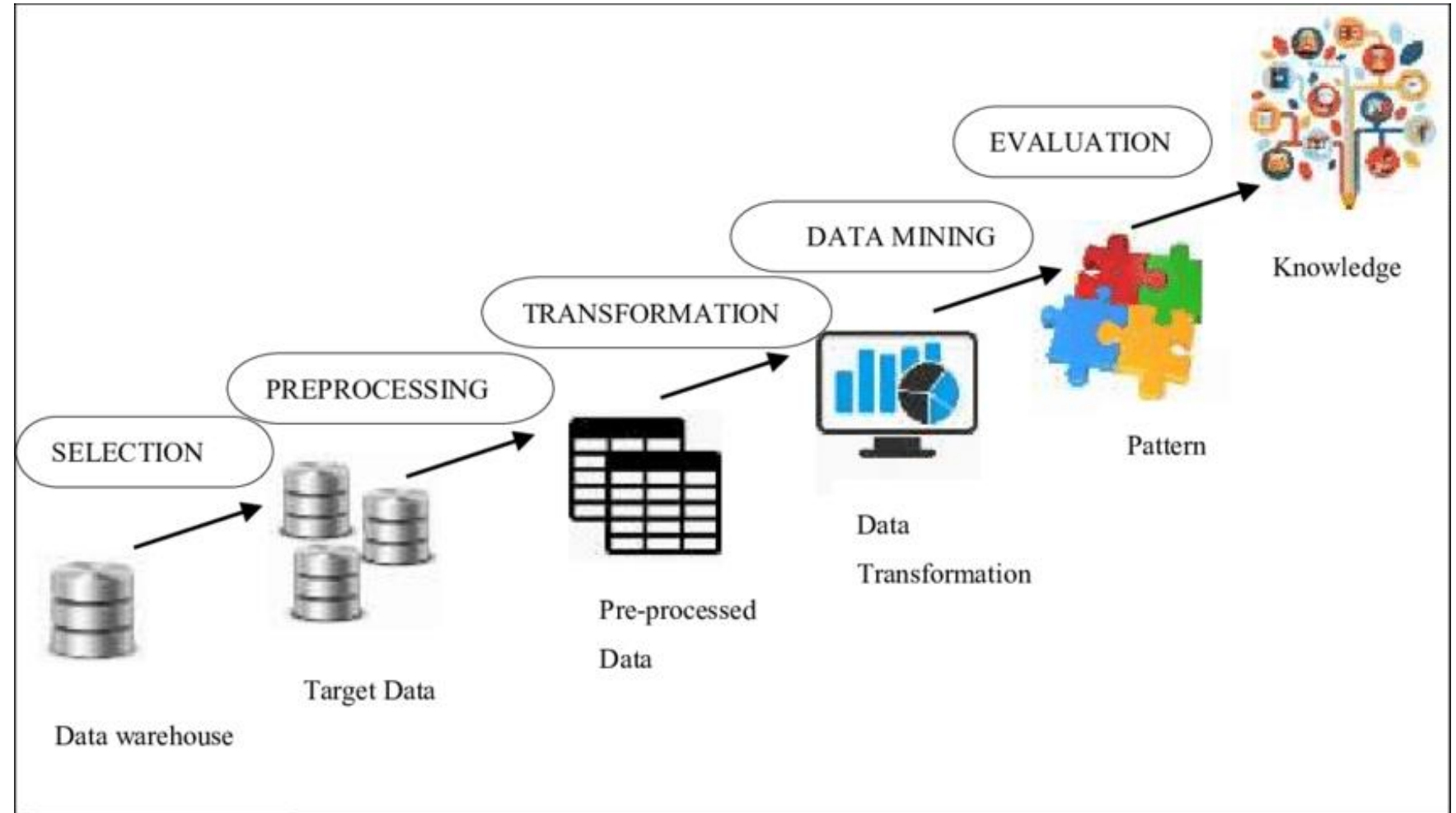
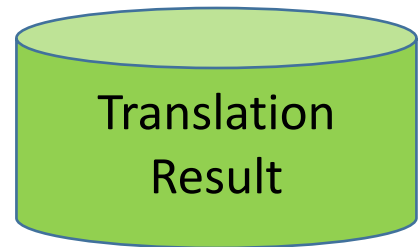
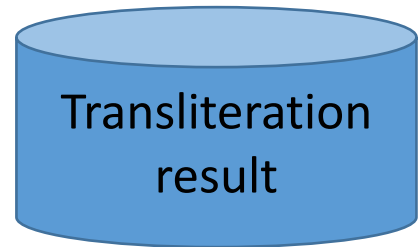
Translation

1.3306.tiff

- (1) ngeun¹ bujanggalawa/ saur taan deui sita/ aki aya kita dinya/ anak¹ saha dyayunan/ sa-rwa² deungeun si utun/
bersama bujanggalawa/ kata Tuan Dewi Sita/ Kadek ada kamur di situ/ ana siapa di ~~itu~~ buaian/ sama dengan si buyung/
némbal³ aki ha-
menjawab katek ha-
- (2) may canggong/ samapun geura nu calik⁴/ éboh na hulun pupulih/ basa aing ngasuh bonycah⁵/ horéngan
may canggong/ Silakan duduk dulu/ ~~kamur~~ akan saya beritahu ketika saya mengasuh anak/ ternyata
nuturkeun⁶ ka cai/ mantaré⁷ aing milanga-
pergi ke sungai/ sementara aku ~~beritahu~~ membaca
- (3) n/ na sanghyang watang ageung/ si utun hateu dingsir/ carékkéng⁸ karah sakini/ tosta ngayun apussageung⁹/
Sanghyang watang ageung/ si buyung tidak kecair/ jadi aku katabakan/ setelah mengayun kitab suci/
dikawihhan¹⁰ babalik¹¹ kan/ tuluy
dinyanyikan pantun/ lalu
- (4) metu bonycah¹² lanang/ sa-rwa¹³ deungeun si utun/ saur taan deui sita/ kamayangan kita séda/ metukeun¹⁴
keluar anak laki-laki/ sama dengan si buyung/ kata tuan Dewi Sita/ benarlal eykahu sakti/ mengeluarkan
kasaktyin/ kamayangan aing
kesaktian/ lenar = aku



Workflow





Any Idea?



Research of Ancient Manuscripts in Universitas Padjadjaran

AMADI

- 2016-2018

AMIRA

- 2019-2021

DREAM

- 2022-2024

KD-SAM

- 2025-2027

Research Collaboration
Joint Supervision for PhD
Others

Ancient Manuscripts Research in Universitas Padjadjaran

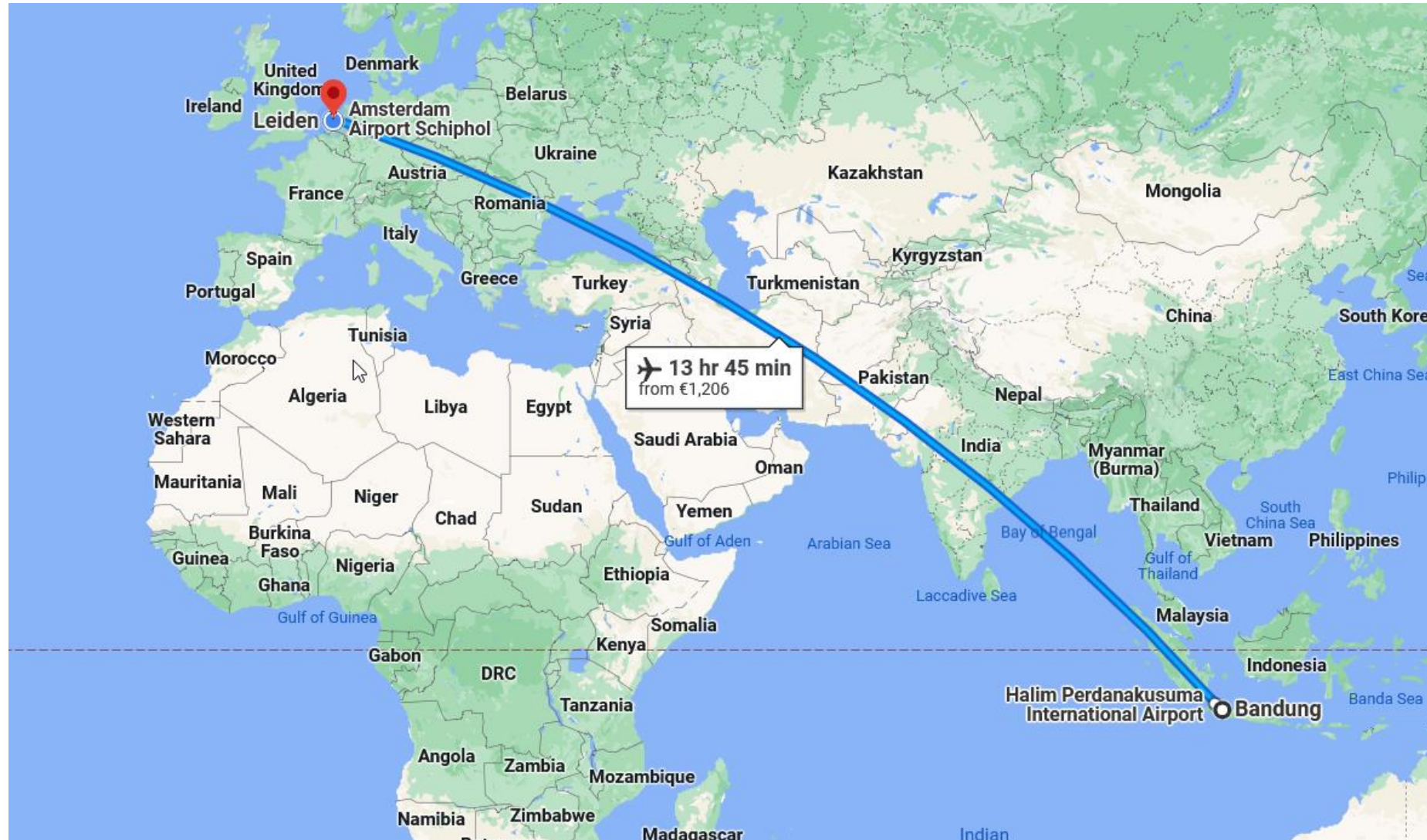
**RISE SMA 2022
30 September 2022
Leiden University**

From West Java for Indonesia to the World through SDGs

www.unpad.ac.id



Bandung (Indonesia) – Leiden (Netherland)





UNIVERSITAS
PADJADJARAN

Main Campus
at Jatinangor



UNIVERSITAS
PADJADJARAN

SITES

CAMPUS OF
JATINANGOR

1.781.780 m²



CAMPUS OF
PANGANDARAN

330.000 m²



CAMPUS OF
BANDUNG



CAMPUS OF
ARJASARI

2.603.743 m²



CAMPUS OF
GARUT

9.603 m²

JATINANGOR CAMPUS







LECTURER EXCHANGE

Part of RISE-SMA Collaboration 2022

