



golden
agents

From archive to annotation

The pipeline of scanning, crowdsourcing, transcribing and modelling

Jirsi Reinders, Leon van Wissen

Luik, 13 September 2019

 @goldenagents

Industrial scanning



**3,5
Kilometers**



**731
Notaries**

**Amsterdam
Notarial Archives
1578 - 1915**

**6.419.017
Scans**

**3600
Bookshelves**

**20.000
Scans
Per day**

**30.000
Inventory numbers**



Crowd sourcing with VeleHanden



Alle Amsterdamse Akten

Mijn profiel: Jirsi Reinders | Uitloggen

ALWEIJN DE JAGER 1658-1664 - 5 - Expert - 2976 - (1658-1659) - KLAD00634000445

Handwritten document image with blue annotations. The text includes: "Oeffening de op Augustus 1659", "Commissie van", "Jager Notaris D. J. Leenders van", "nagelende", "Dirck Leenders", "n. J. van", "Claas Carstens", "Pieter Boudeswart", "Jan Rigoo", "Gerrit Heijndricks".

nederlands

Attestatie ⓘ

20-08-1659

Overzicht invoer:

- Naam: Dirck Leendersz ✓
- Naam: Jan Gerritsz ✓
- Naam: Claas Carstensz ✓
- Naam: Pieter Boudeswart ✓
- Naam: Jan Rigoo ✓
- Naam: Gerrit Heijndricks ✓

Naam toevoegen

Locatie toevoegen

Alle van verwersgraft, dienstmeit gooit
vulligheid voor reqts huis daardoor

Afronden Scan heropenen voor invoer Rapporteer deelnemer Opmerkelijk! Probleem

Transcribing with the crowd



🏠 [Crowd leert computer lezen - Intern](#) Mijn profiel: [Jirsi Reinders](#) | [Uitloggen](#)

SIMON VAN SEVENHOVEN 1681-1705 - 3 - Basis Paleografie - 5301 - NOTD02572000148

Aldert Reijtjes van Hinlopen, te Rijpe op't
schip genaemt de Jonge Arents Cab, van 1623
loijt $2\frac{1}{2}$, $2\frac{3}{4}$, & $4\frac{1}{4}$ vort, t...
anderrijde
b. de ladder $2\frac{1}{2}$ onparant } mette, andere —

Text Region 1

1	Aldert Reijtjes van Hinlopen, schipper op't	#
2	schip genaemt	#
3		#
4		#
5		#
6		#

[Afronden](#) [Te moeilijk](#) [Onbruikbaar](#) [Opmerkelijk](#) [Tussentijds bewaren](#) [Bekijk vorige](#) [Bekijk volgende](#)

[Uitgebreide invoerstructuur \(pdf\)](#) [Algemene handleiding \(pdf\)](#)



Crowd leert computer lezen

Mijn profiel: Jirsi Reinders | Uitloggen



HENDRIK SCHAEF 1636-1665 - 3 - 1285 - KLAB0552000012



x² ab Special Characters ? Unclear

Text Region 1		#
1	Gerrit Alberte Ron	#
2	Wsena Pietnt	#
3	O huijden den 10en. Janry. @ 1642. Compareerde	#
4	voor myj Henrick Schaeff Nots. etc. ende den etc.	#
5		#
6	Gerrit Albertsz vojij, grootschipper won	#
7	binnen deser Stadt Amsteldam my Notario bekend	#
8	Ende heeft inder bester forma. stille ende maniere	#

Results

ANTHONY VAN DE VEN 1666-1670	(110)	
CORNELIS STAAL 1749-1753	(200)	
CORNELIS VAN LOON 1700-1743	(623)	
DANIEL VAN DEN BRINK 1734-1785	(614)	
DAVID DES POMMARE 1698-1724	(110)	
DIRK VAN DER GROE 1670-1720	(546)	
HENDRIK SCHAEF 1636-1665	(469)	
HERMANUS VAN HEEL 1745-1771	(500)	
JACOB DE WINTER 1648-1710	(610)	
JACOB JANSZ WESTFRISIUS 1612-1654	(110)	
JAN BARELS DE JONGE 1716-1753	(110)	
JAN FRANSSSEN BRUIJNINGH 1593-1624	(110)	
JAN VERLEIJ 1741-1775	(525)	
PALM MATHIJSZ 1610-1641	(436)	
SIMON VAN SEVENHOVEN 1681-1705	(384)	

20 notaries, circa 11.000 scans controlled. 12 good specialist models
First test with 40.000 automatically transcribed scans

General 18th century model

15 notaries, 2671 scans training, Character Error Rate (CER) of 2%

Op dinsdag den Seevenden **Septembe** **September**
des Jaars Seeventienhonderd negentig
compareerden voor mij Mr. **Sponard** **Leonard**
Beels, Notaris by den Edelen Hove van
Holland geadmitteerd in Amsterdam
resideerende in praesentie van de nagenoemde
getuygen **getuygen**.

No. 2.
Anno 1790
Dingsdag
7 september

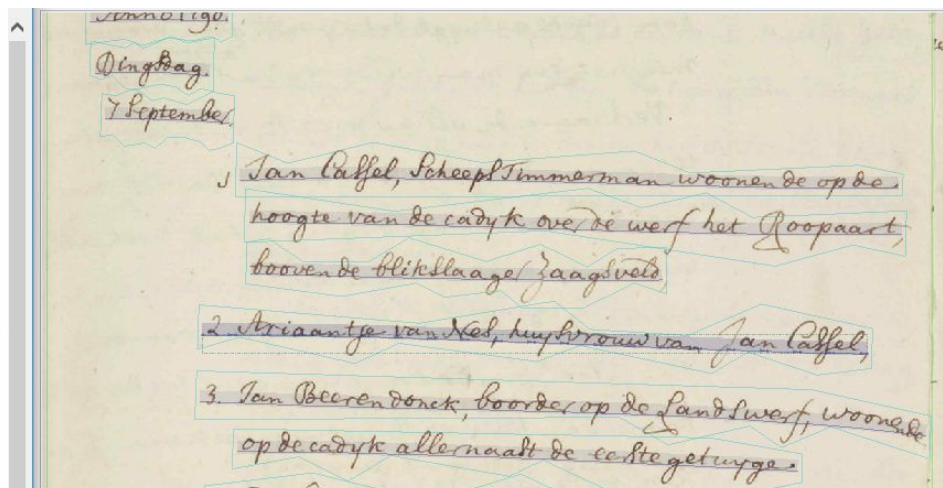
Jan Cassel, ScheepsTimmerman woonende op de
hoogte van de cadijk over de werf het Roopaart,
boovende **blikslaage** **Jaagsveld**, **blikslaager** **Zaagsveld**,

2 **Atriaantje** **Ariaantje** van Nes, huisvrouw van Jan Cassel,
3. Jan **Beerendorck**, **Beerendonck**, boorder op de Landswerf, woonende
op de cadijk allernaast de eerste getuyge

^ En Catharina Boshoff, huisvrouw van Cornelis van
Nes, woonende in de **kelde**, **kelder**, onder het huys van de
derde getuyge.

alle getuygen van competenten ouderdom, dewelke
ter requisitie van den Wel Ed. Gestr. Heer en Mr.
W. E Backer Hoofdofficier dezer Stad, R.O. en ten
behoefte van de geene die zulks verder **soude** **zoude** mogen
aangaan, voor de oprechte waarheid hebben **getuyg** **getuygd**
en verklaard

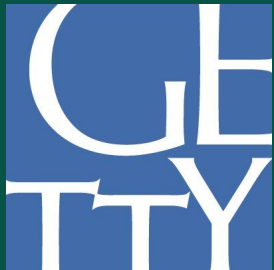
En wel eerstelijk de tweede getuyge, dat op den
dertigsten Augustus dezes jaars 1790, s' **middagh** **middags** omtrent
half een uur by haar op de kaamer gekoomen is zekere
Hannes Davids, woonende op de cadijk boven de **Horolen** **Horolo**
giemaaker Houtman, en dat hij haar getuyge, ten
aanhooren van de vierde getuyge, die als nu meede
verklaart, onder het uijtbraeken van diversche **schald** **scheld**
woorden **uijtdaagde** **uijtdaagde** om de trap afte komen en met



- 5 Jan Cassel, ScheepsTimmerman woonende op de
- 6 hoogte van de cadijk over de werf het Roopaart,
- 7 boovende blikslaage Jaagsveld,
- 8 2 Atriaantje van Nes, huisvrouw van Jan Cassel,
- 9 3. Jan Beerendorck, boorder op de Landswerf, woonende
- ^ op de cadijk allernaast de eerste getuyge

Working with the raw output of the HTR-model

A case study in linking items and inventories from the Getty Provenance Index (GPI) to the archival source



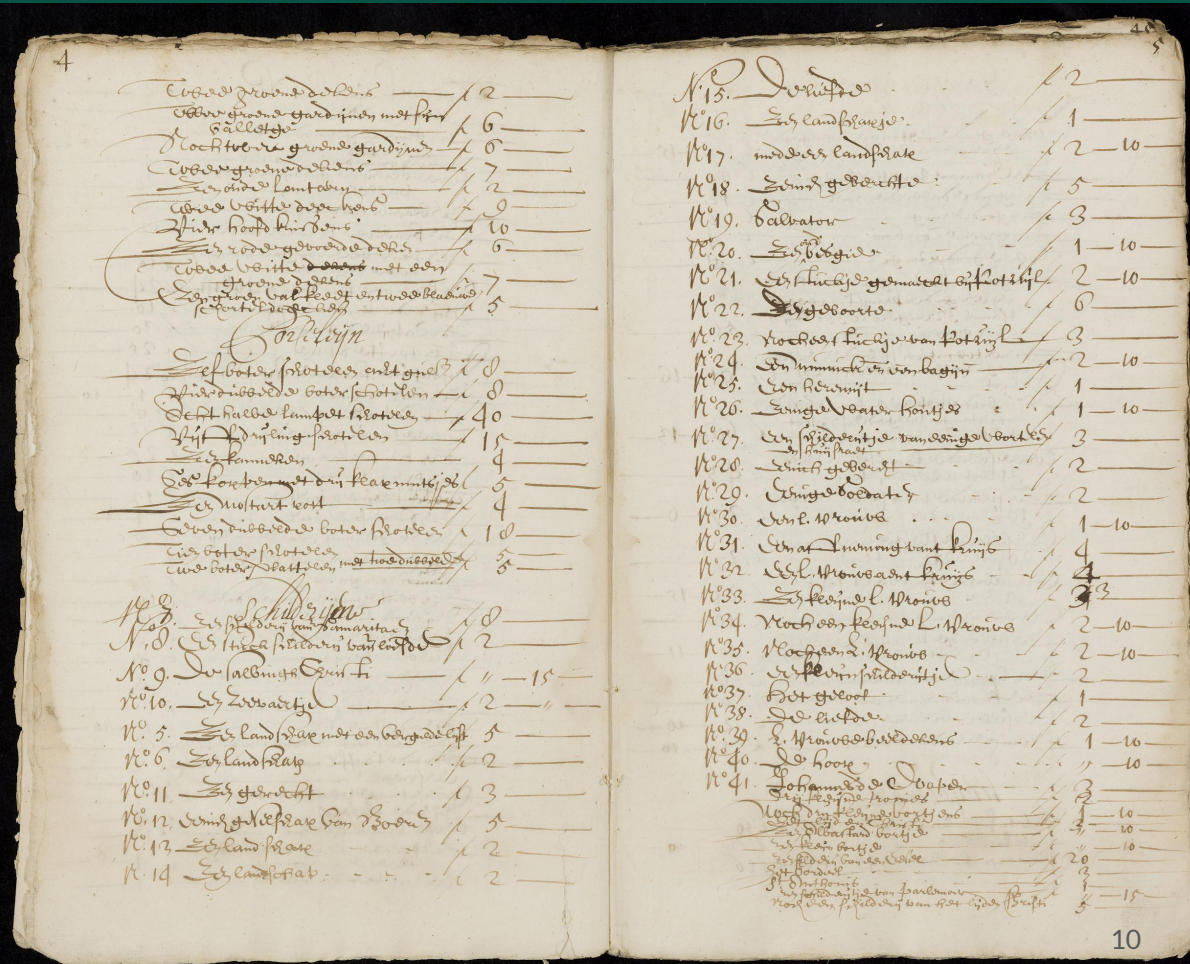
golden
agents

From scan to item

Inventory: 2408

Notary: Jacob de Winter

One scan, two pages





Transkribus XML: regions and lines

- Per scan (A16098000033.XML)
- Per region (recto / verso)
- Per line

Problems

- Record segmentation?
 - Textual or optical
- Line segmentation?
- Parsing (+ interpreting)

*Inventaris ende Specificatie
van allen den goederen nagelaten
bij zal. Maria koerten sulcx
zij die met haer man Niclaes felt
int gemeen beseten ende metter
dood deser werelt ontruijmt
ende nagelaten heeft.
Eerstelijck den huijsraet, Imboel
klederen, kleijnodien ende Juwelen
getaxeert door Annetje hendricx en Susannetje
Anthonis gesworen Taxeersters deser stede
tot alsulcken prijse als volcht
Een wiegh met een baecker mat, ende
een matstoel, een bakermant ende
4
eenige kleijne mantjes
Een scharfbort met een doos ende eenigh
- 10
withoutwerck en rommeling
Een ijsere rooster, brandijser en een
Een kleijne kopere doofpott
Twee kopere wateraeckers f 5*

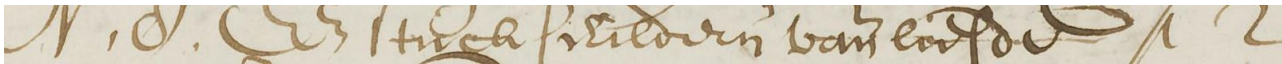
Left page,

<https://archieff.amsterdam/inventarissen/inventaris/5075.nl.html#A16098000033.JPG>

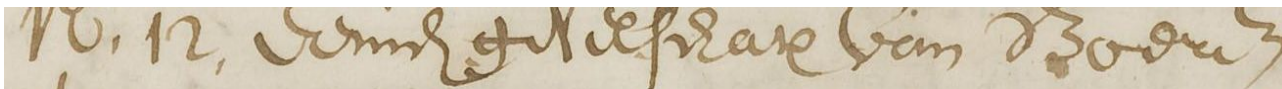
Left page (region + lines)



No. 7. Een schilderij van de Samaritane

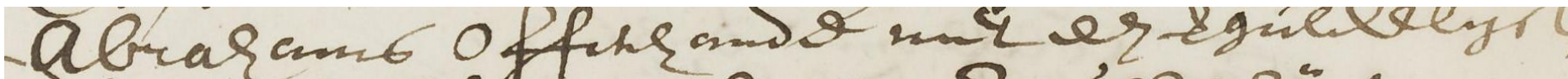


No. 8. Een stuk schilderij vande liefde f 2



No. 12. eenich geselschap van Boeren

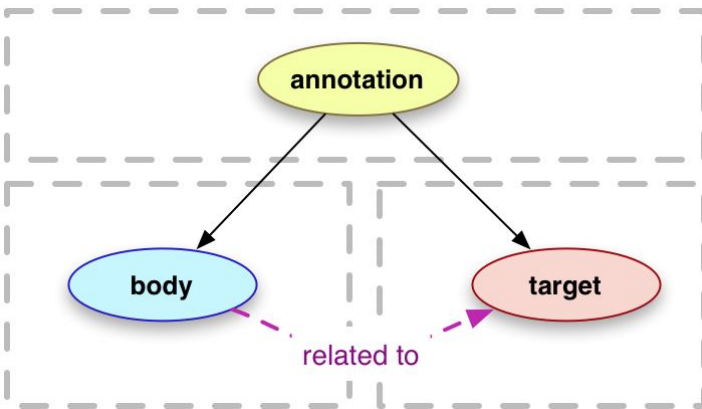
Uwe groene daken	f 12
Uwe groene daken met pin	f 6
Uwe groene daken met pin	f 6
Uwe groene daken met pin	f 7
Uwe groene daken met pin	f 2
Uwe groene daken met pin	f 9
Uwe groene daken met pin	f 10
Uwe groene daken met pin	f 6
Uwe groene daken met pin	f 7
Uwe groene daken met pin	f 5
Uwe groene daken	
Uwe groene daken met pin	f 8
Uwe groene daken met pin	f 8
Uwe groene daken met pin	f 40
Uwe groene daken met pin	f 15
Uwe groene daken met pin	f 4
Uwe groene daken met pin	f 5
Uwe groene daken met pin	f 4
Uwe groene daken met pin	f 10
Uwe groene daken met pin	f 5
Uwe groene daken	
N ^o 7. Een schilderij van de Samaritane	f 8
N ^o 8. Een stuk schilderij vande liefde f 2	f 2
N ^o 9. De jaalings dinst	f 11 - 15
N ^o 10. De zwaarte	f 12 - 11
N ^o 5. De landpreze met een berg	f 5
N ^o 6. De landpreze	f 12
N ^o 11. De landpreze	f 3
N ^o 12. Eenich geselschap van Boeren	f 5
N ^o 13. De landpreze	f 12
N ^o 14. De landpreze	f 12



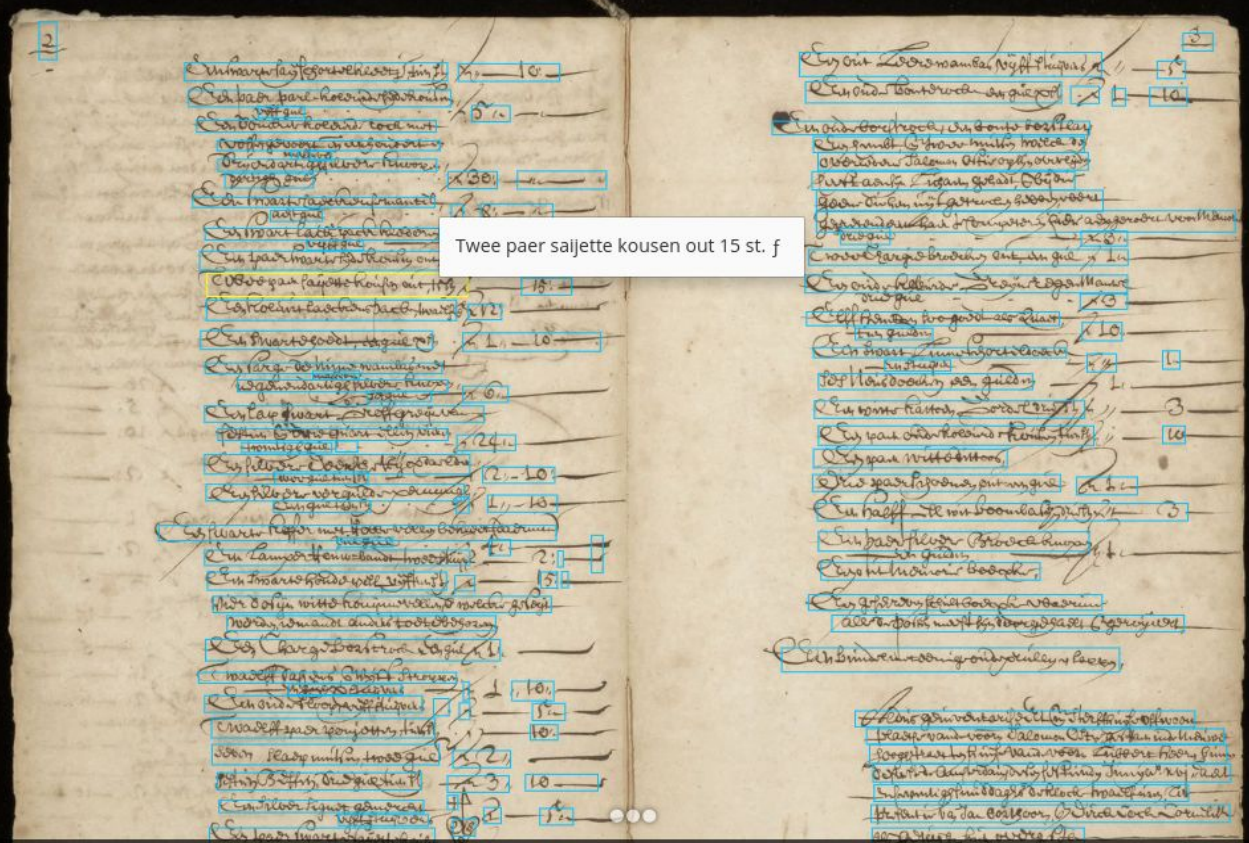
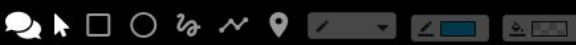
Conversion to Annotations

- Storing the text as RDF / Linked Data
- Web Annotation Model
- IIIF Presentation API

```
saa:A16098000075r2l43 a oa:Annotation ;  
  oa:bodyValue "Abrahams Offerhande met een vergulde lijst," ;  
  oa:hasTarget [ oa:hasSelector [ a oa:FragmentSelector ;  
    dcterms:conformsTo <http://www.w3.org/TR/media-frags/> ;  
    rdf:value "700,2975,1846,86" ] ;  
    oa:hasSource <https://archieff.amsterdam/inventarissen/inventaris/5075.n1.html#A16098000075.JPG> ] ;  
  prov:wasDerivedFrom "A16098000075.xml" .
```



Example iiif: <https://iiif.leonvanwissen.nl/iiif/2/a16098000075.jpg/700,2975,1846,86/full/0/default.jpg>



Twee paar saljette kousen out 15 st. f



Page 1

Page 2

Page 3

Page 4

Page 5

Page 6

Page 7

Page 8

Page 9

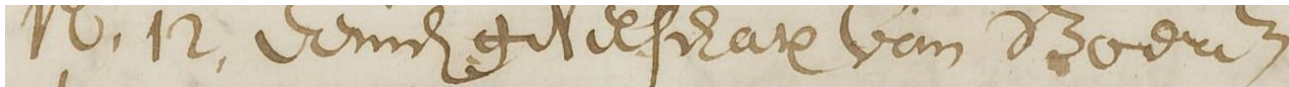
Page 10

Page 11

Page 12

The annotation

```
<https://archieff.amsterdam/inventarissen/inventaris/5075.nl.html#A16098000033r2i56> a oa:Annotation ;
  saa:partOf <https://archieff.amsterdam/inventarissen/inventaris/5075.nl.html#A16098000033r2> ;
  oa:bodyValue "No. 12. eenich geselschap van Boeren" ;
  oa:hasTarget [ oa:hasSelector [ a oa:FragmentSelector ;
    dcterms:conformsTo <http://www.w3.org/TR/media-frags/> ;
    rdf:value "504,3403,1316,85" ] ;
    oa:hasSource <https://archieff.amsterdam/inventarissen/inventaris/5075.nl.html#A16098000033.JPG> ] ;
  prov:qualifiedDerivation _:N07d8191c46484ff28d5696c2285072ed ;
  prov:wasDerivedFrom "A16098000033.xml" ;
  foaf:depiction <https://iiif.leonvanwissen.nl/iiif/2/a16098000033.jpg/504,3403,1316,85/full/0/default.jpg> .
```



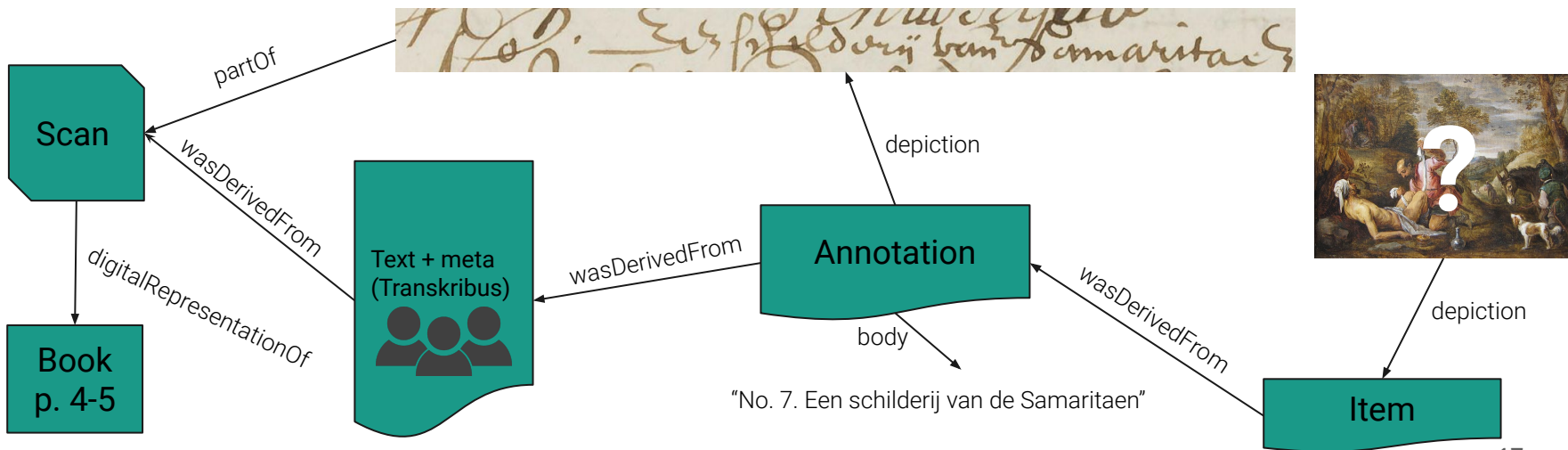


The item

```
<http://goldenagents.org/uva/SAA/Inventory/Item/N-2097_0008> a saa:Item ;  
  rdfs:label "Eenich geselschap van boeren"@nl ;  
  prov:wasDerivedFrom <https://archieff.amsterdam/inventarissen/inventaris/5075.nl.html#A16098000033r2I56> ;  
  saa:artist "[Anonymous]" ;  
  saa:identifier "DUTCHINV-17280" ;  
  saa:index "8" ;  
  saa:isInRecord <http://goldenagents.org/uva/SAA/Inventory/N-2097> ;  
  saa:transcription "Eenich geselschap van boeren 5:--:--"@nl ;  
  saa:valuation "5:--:--" ;  
  saa:iconclass <http://iconclass.org/46A14> ;  
  saa:workType "Schilderij"@nl .
```


Provenance

An interpretation [=item] is derived from a particular piece of text in a book



Contact

Jirsi Reinders

Stadsarchief Amsterdam

jirsi.reinders@amsterdam.nl

Leon van Wissen

Universiteit van Amsterdam

l.vanwissen@uva.nl

www.alleamsterdamseakten.nl

www.goldenagents.org