

# AI-based approaches to automatic and computer supported resource description in archaeology

Isto Huvila and Ying-Hsang Liu

Department of ALM | Uppsala University

@CAPTURE\_ERC | [www.uu.se/en/research/capture](http://www.uu.se/en/research/capture)

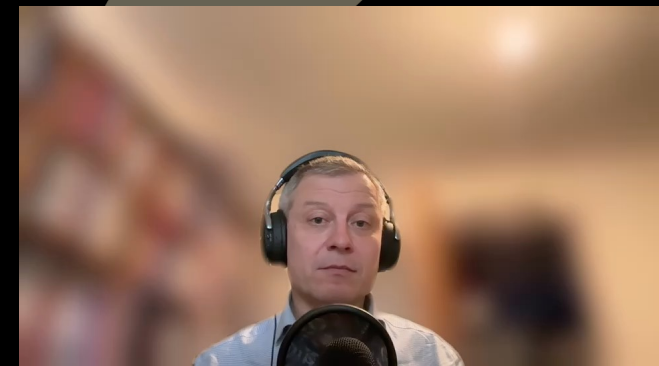


UPPSALA  
UNIVERSITET



European Research Council  
Established by the European Commission

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme grant agreement No 818210



# Long history of AI in archaeology

CAPTURE



# 22

## *Automatic Archaeology: Bridging the Gap between Virtual Reality, Artificial Intelligence, and Archaeology*

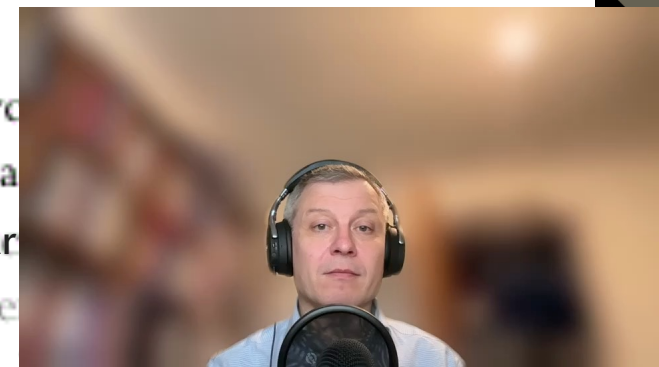
Juan Antonio Barceló

---

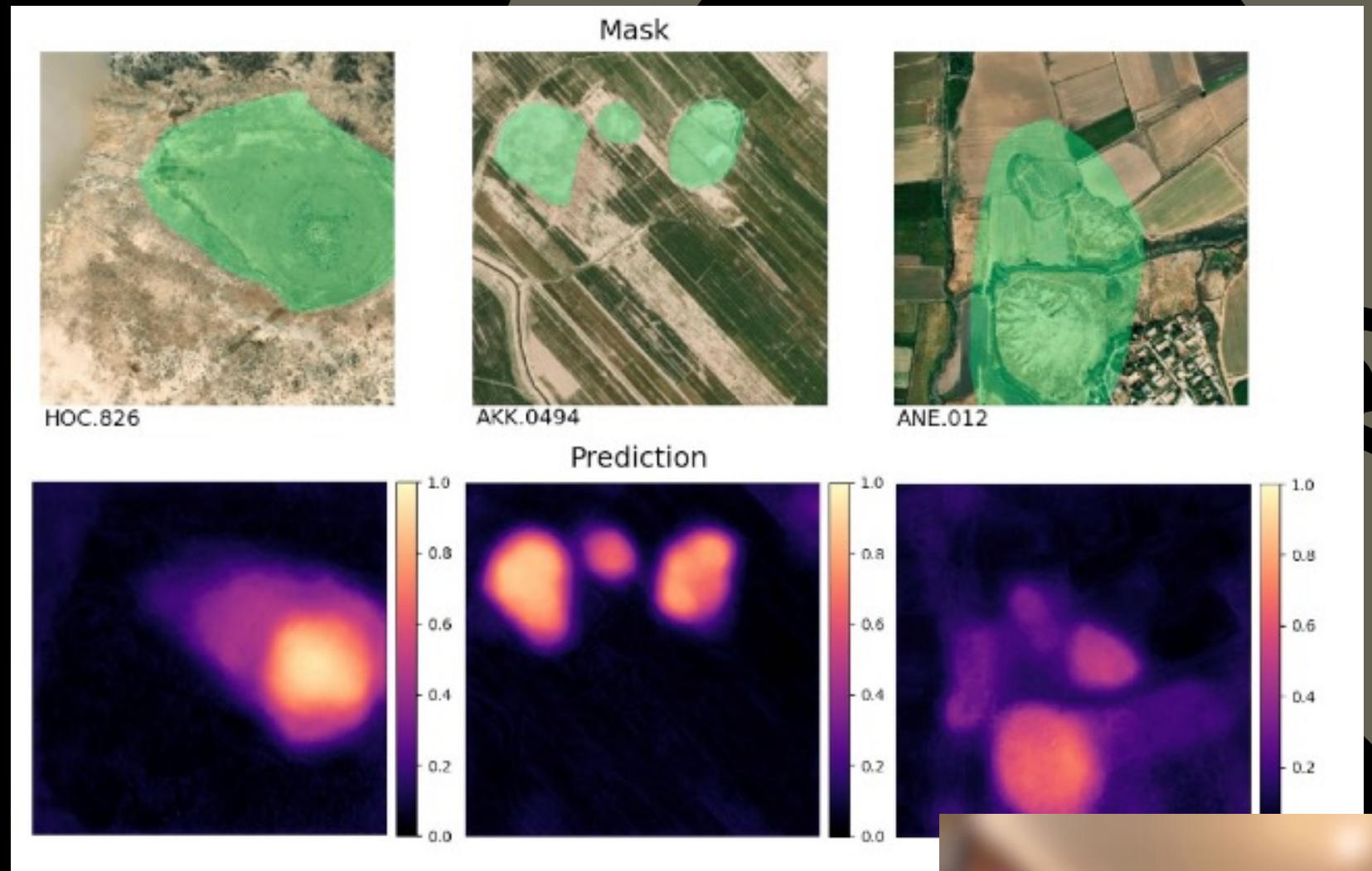
### *The Nature of Archaeological Problems*

What is archaeology? In contrast to most usual descriptions of the field, archaeologists do not study artifacts as mute witnesses of the past, but analyze social and cultural processes that have shaped the dynamic nature of present society. We need ancient history to understand the present, and the present to understand the past.

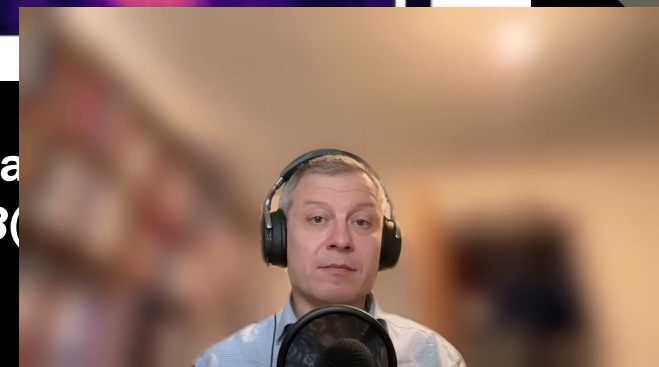
Barceló, J. A. (2007). *Automatic Archaeology: Bridging the Gap between Virtual Reality, Artificial Intelligence, and Archaeology*. <https://doi.org/10.7551/mitpress/6751.003.0028>



# Remote sensing



Casini, L., Marchetti, N., Montanucci, A., Orrù, V., & Roccetti, M. (2023). A human-in-the-loop workflow for archaeological sites detection. *Scientific Reports*, 13(1), 1-12. <https://doi.org/10.1038/s41598-023-36015-5>





SEARCHING IN ARCHAEOLOGICAL TEXTS  
PROBLEMS AND SOLUTIONS USING AN ARTIFICIAL INTELLIGENCE  
APPROACH

*Hans Pajmans\* & Alex Brandsen#*

\*Tilburg University (Warandelaan 2, 5000 LE Tilburg, The Netherlands)

#Corresponding author: Leiden University, Faculty of Archaeology (Reuven  
2311 BE Leiden, The Netherlands) - Current address: University of York, Faculty  
of Archaeology (The King's Manor, York, YO1 7EP, UK) alex.brandsen@gmail.c

**Pajmans, J. J., & Brandsen, A. (2010). Searching in Archaeological Texts: Problems and Solutions Using an Artificial Intelligence Approach. *PalArch's Journal of Archaeology of Egypt/Egyptology*, 7(2), 1-10.**



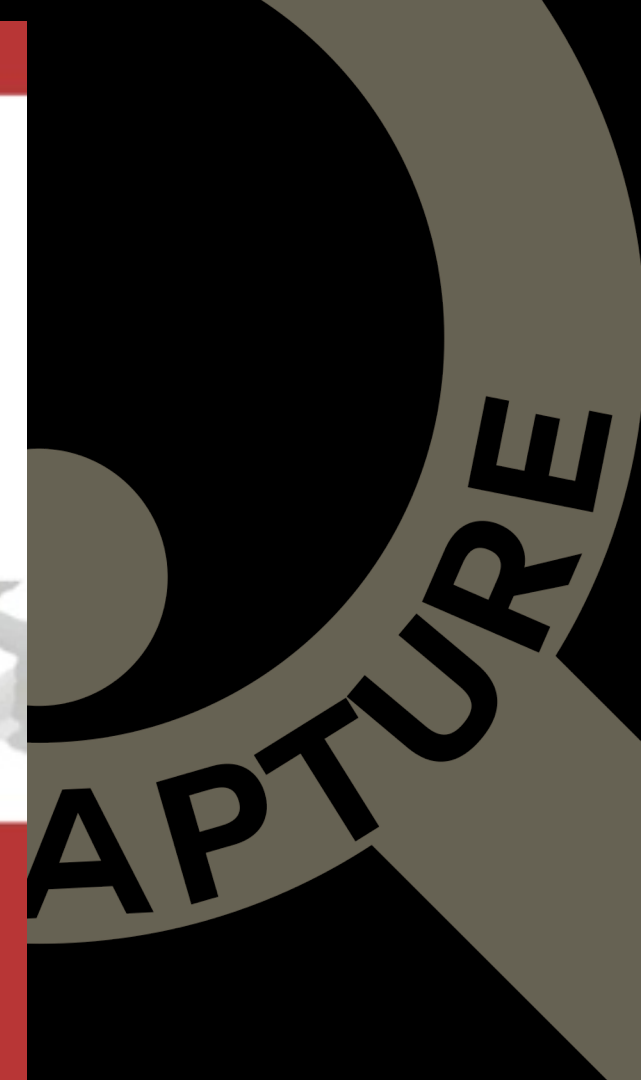




## D16.4: Final Report on Natural Language Processing

### Authors:

Andreas Vlachidis, USW  
Douglas Tudhope, USW  
Milco Wansleben, LU  
Jeremy Azzopardi, SND  
Katie Green, ADS  
Lei Xia, ADS  
Holly Wright, ADS



Annotation Sets Annotations List Annotations Stack Co-reference Editor OAT Text

Index ----- NORTH OF WESTENHANGER CASTLE (ARC WGC 97) EVALUATION REPORT

APPENDIX 1 POTTERY By Roy Stephenson . 1 Introduction 1.1 The evaluation produced a total of 45 sherds (383g) of medieval pottery and one sherd (7g) of residual Roman pottery. Sherds were generally abraded with shell and chalk inclusions leached out, and glazed sherds have largely partially lost their glaze through abrasion. The pottery was examined using a x 20 binocular microscope and recorded using standard MoLAS codes on pro-forma sheets. Quantification of the assemblage was by sherd count and weight . Pottery was recorded from unstratified and seven contexts. 2 Fabrics 2.1 The fabrics identified have been identified in broad fabric type. The fabrics have been identified on the basis of their main inclusions, coupled to a partial overview of the literature on pottery from this region of this date, although some can be divided into defined sources. It should be stressed that some of these identifications are tentative. The fabric codes used where possible those used at Canterbury Archaeology Trust (CAT ) see (Blackmore 1988), and where not possible MoLAS fabric codes are used. The CAT codes are marked below + thus. 2.2 The Fabric Group consists of: Roman Grog tempered ware ( GROG ) 1 sherd 7g Medieval Rough Medieval Red-ware (A)+ 24 sherds 126g Rough Medieval Red-ware ( B ) + 4 sherds 20g Early medieval flinty ware (EMFL) 4 sherds 20g Early Surrey ware ( ESUR ) 1 sherd 1 0g Local greyware (LOGR) 4 sherds 20g Ryde ware 1 sherd 105g Tyler Hill ware 1 sherd 1 0g Late Saxon/Medieval Sand-and-shell-tempered Ware (SHS) + 6 sherds 57g 3 Forms 3.1 The medieval pottery sherds present are mostly unassignable, due to small size and abrasion. Despite this there are a number forms represented among the assemblage, these are an unstratified stabbed strap handle from a Ryde ware jug , a sand and shell tempered ware (SHS) cooking pot fragment from [17], and a possible early Surrey (c) UNION RAILWAYS LIMITED 1997 19

----- NORTH OF WESTENHANGER CASTLE (ARC WGC 97) EVALUATION REPORT ware (ESUR) bowl from [1]. 4 Chronology 4.1 Most of the sherds are non-diagnostic small abraded sherds , that return only very broad 11th - 13th century date ranges. They appear to conform to existing fabric types in the area, although there appears to be to a lack of wares manufactured in nearby Sussex, apart from an unstratified, abraded, stabbed strap handle from a Ryde ware jug. 5 General Comments 5.1 The range of vessels are indicative of domestic activity on or near the site. 6 Assessment of potential and further work 6.1 The fabric identifications should be confirmed by comparison with an existing type-series, allowing the use of existing fabric codes. 7 Table 3: Bulk dataset, pottery

TRENCH	CONTEXT	MATERIAL	COUNT	WEIGHT	COMMENTS
0	POTTERY	1	105	1275-1375	
1823TT	1	POTTERY	2	20	1150-1250
1821TT	5	POTTERY	5	15	1000-1250
1821TT	7	POTTERY	10	40	1050-1150
1822TT	9	POTTERY	15	120	1000-1250

E19  
E19Negation  
E49  
E53  
E57  
E57Negation  
EHE0007  
EHE0009  
EHE0030  
EHE0039  
EHE1002  
EHE1004  
Lookup  
NP  
Negation  
P45  
Section  
StopNeg  
Token  
VG  
Original markups

<https://github.com/avlachid/Multilingual-NLP-for-Archaeological-Reports-Aria>



# Automatic Metadata Generation in an Archaeological Digital Library: Semantic Annotation of Grey Literature

Andreas Vlachidis, Ceri Binding, Keith May, and Douglas Tudhope

**Abstract.** This paper discusses the automatic generation of rich metadata from excavation reports from the Archaeological Data Service library of grey literature (OASIS). The work is part of the STAR project, in collaboration with English Heritage. An extension of the CIDOC CRM ontology for the archaeological domain acts as a core ontology. Rich metadata is automatically extracted from grey literature, directed by the CRM, via a three phase process of semantic enrichment employing the GATE toolkit augmented with bespoke rules and knowledge resources. The paper demonstrates the potential of combining knowledge based resources (ontologies and thesauri) in information extraction, and techniques for delivering the automatically extracted metadata as XML

CAPTURE

Binding, C., Tudhope, D., & Vlachidis, A. (2018). A study of semantic integration and data and reports in different languages. *Journal of Information Science*, 42(1), 1-15. <https://doi.org/10.1177/0165551518789874>





# 31

## **Beyond classification: the use of artificial intelligence techniques for the interpretation of archaeological data**

Jitu Patel\*

Arthur Stutt†

### 31.1 Introduction

Archaeology, despite its relatively short history, is already faced with the problem of a large accumulation of data. This will inevitably become more pressing in the future. As a means of dealing with the problem, archaeologists have naturally turned to the

**Patel, J., & Stutt, A. (1989). Beyond Classification: The Use of Artificial Intelligence in the Interpretation of Archaeological Data. In S. Rahtz (Ed.), *Computer Applications in Archaeology: Methods in Archaeology 1989*. CAA89 (Vol. 548, pp. 338–347). [https://proceedings.caaconference.org/paper/30\\_patel\\_stutt\\_caa89](https://proceedings.caaconference.org/paper/30_patel_stutt_caa89)**



**(Ethical) concerns?**

**Bias, inaccuracies.**



# Generative AI? Synthetic KO?

CAPTURE



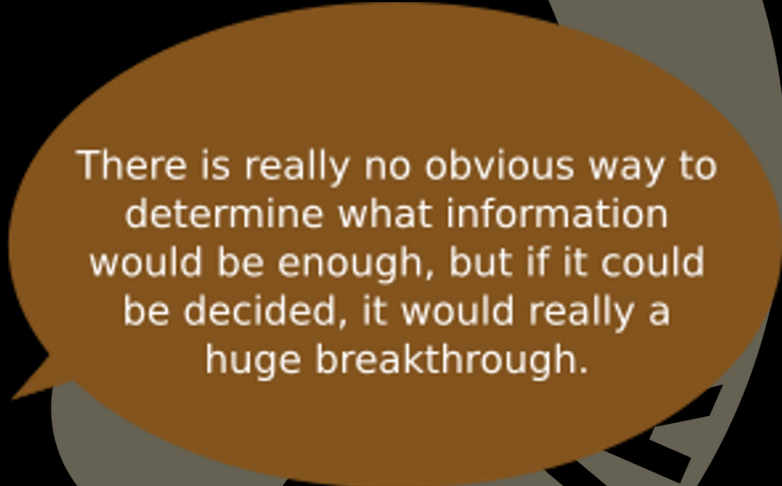
# Example

CAPTURE





There is really no obvious way to determine what information would be enough, but if it could be decided, it would really a huge breakthrough.



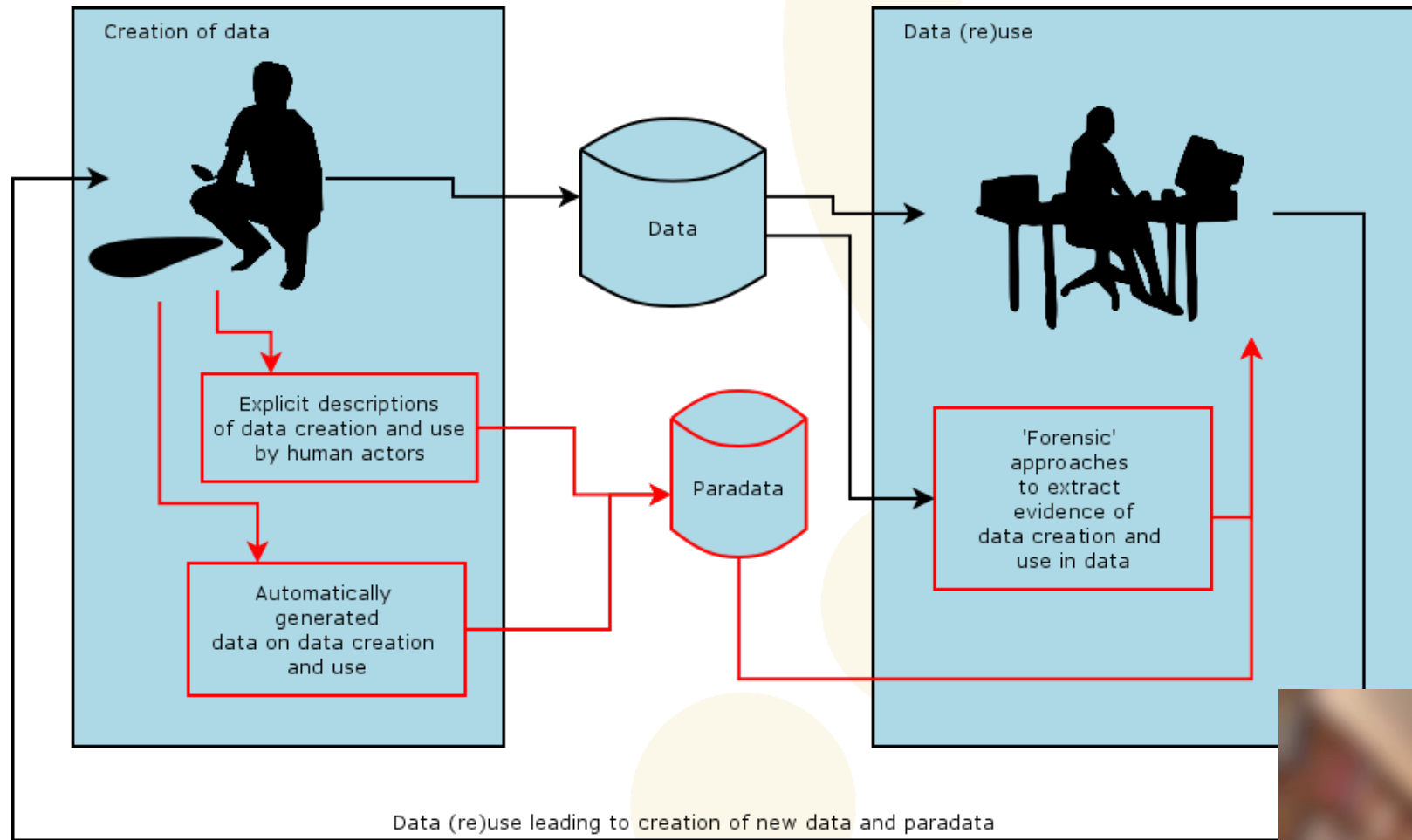
CAPTURE



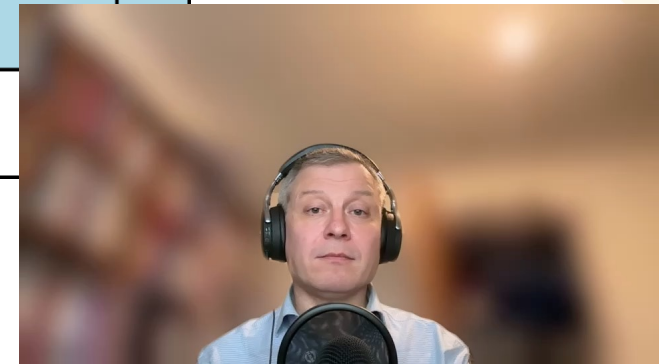
This data is pretty unusable because I don't know how it was created. But I kind of understand that it is difficult to know at the time of creation what information will be needed in the future. And documenting all possible things done when creating it is also pretty much impossible.



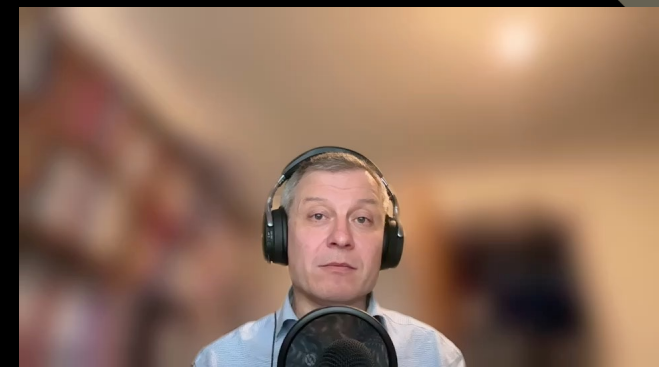
# Paradata describes processes



PTURE



**Extracting paradata (or information for paradata) using named entity recognition and object identification.**



**Trials on archaeological  
investigation reports.**





# Automatic identification of archival paradata using artificial intelligence techniques

Isto Huvila<sup>a</sup>, Ekta Vats<sup>a,b</sup>, Zanna Friberg<sup>a</sup>, Lisa Börjesson<sup>a</sup>, Jessica Kaiser<sup>a</sup>, Olle Sköld<sup>a</sup>

(a) Department of ALM | Uppsala University

(b) Centre for Digital Humanities Uppsala | Uppsala University

@CAPTURE\_ERC | [www.uu.se/en/research/capture](http://www.uu.se/en/research/capture)



UPPSALA  
UNIVERSITET



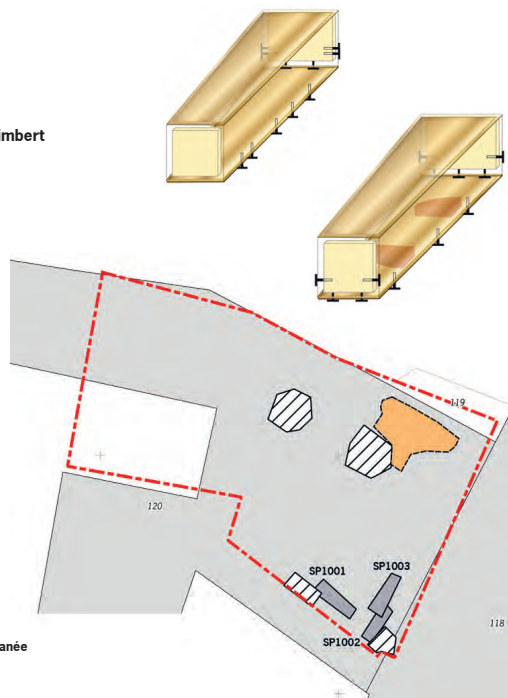
European Research Council  
Established by the European Commission

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme grant agreement No 818210



Occitanie, Tarn-et-Garonne, Montauban

## Rue Michelet

sous la direction de  
Laurent GrimbertInrap  
Midi-Méditerranée  
février 2020

des dépôts sableux, puis limoneux, d'épaisseur très variable. L'épaisseur totale des alluvions, des limons aux cailloux, varie avec la rivière qui les a déposées; elle est en moyenne de 6,5 mètres pour le Tarn, de 5 mètres pour l'Aveyron.

Les vestiges de la haute terrasse (supérieur à 100-110 mètres d'altitude au-dessus de l'étiage) correspondent aux dépôts alluviaux les plus anciens ; La moyenne terrasse se situe entre 65-80 mètres au-dessus de l'étiage de l'Aveyron, et se présente sous forme de replats fragmentés par l'actuel réseau hydrographique qui descend des coteaux molassiques ; la basse terrasse, sur laquelle est localisée la zone d'emprise de l'intervention (fig. 6), est large de 4 à 5 km. Elle présente un dénivelé global d'environ 20 mètres (entre 18 et 35 m au-dessus de l'étiage), de son bord externe jusqu'au talus qui la sépare de la basse plaine. Située en zone de confluence, elle a certainement été construite par les deux rivières. En effet, on détecte sur photo aérienne des paléoméandres attribuables à l'Aveyron mais aussi au Tarn, de par leur axe de courbure ;

La basse plaine du Tarn, elle, se développe entre 80 et 83 m NGF. Le degré d'altération du sédiment graveleux est différent selon l'ancienneté de la terrasse : le matériel de la basse terrasse est assez frais ; les granites, grès, schistes y sont décomposés jusqu'au centre du caillou roulé, mais l'ensemble demeure de couleur grise. Sur la moyenne terrasse, les quartzites, les gneiss, les schistes silicifiés se décomposent aussi ; les alluvions prennent une teinte ocre plus ou moins foncée et présentent des traînées bleutées dans chaque lentille argileuse. Sur la haute terrasse, seuls les cailloux de quartz demeurent intacts; encore sont-ils très rubéfiés sur leur surface et aussi le long de fentes qui les affectent ».

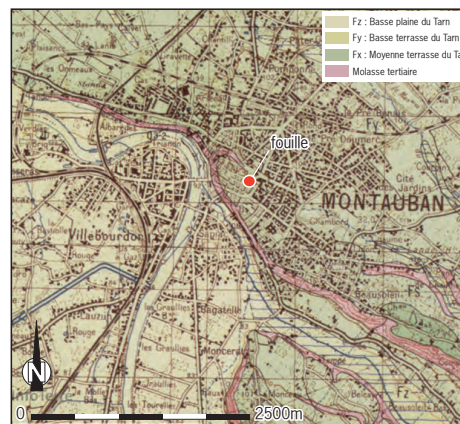


Fig. 6 Contexte géologique  
© BRGM Info Terre

## 1.3. Contexte archéologique

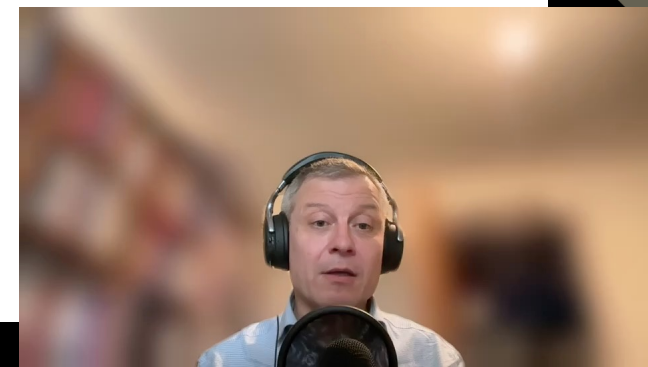
La zone de l'intervention se situe dans le secteur défensif extra-muros, au-devant de l'enceinte défendant le flanc oriental de la ville (fig. 7). L'emplacement de cette enceinte, *a priori* médiévale, correspond théoriquement aux façades ouest des allées de l'Empereur, situées à 70 m à l'ouest de la fouille.

Plusieurs opérations archéologiques ont eu lieu dans ce secteur depuis le début des années 2000 (fig. 8). Ces travaux ont principalement mis en évidence des éléments défensifs modernes (fossés, bastions) mais des occupations plus anciennes, liées ou pas aux problématiques des fortifications, ont également été observées, notamment sur la fouille du 8 Allées de l'Empereur (Grimbert 2018b) située exactement en vis-à-vis de l'opération de l'autre côté de la rue Michelet. Cette intervention a notamment mis en évidence différents dispositifs défensifs échelonnés entre le XIII<sup>e</sup> et la fin du XVII<sup>e</sup> siècle.

Concernant la problématique funéraire, trois interventions de diagnostic ont révélé la présence d'inhumations dans le secteur du front oriental de la ville (fig. 9). Durant l'aménagement de la Préfecture en 2001 (Martineau 2001) une dizaine de tombes avait été observée dans le comblement supérieur d'un fossé qui pourrait constituer le prolongement méridional d'un fossé observé durant le chantier du 8 allées de l'Empereur et dont le comblement semble se situer au XIV<sup>e</sup> siècle. Ces tombes, non datées, ont été attribuées à la période moderne. Une autre intervention avait également mis au jour une sépulture à l'angle des rue Michelet et Gambetta (Nibodeau 2001). Cette sépulture se situe à moins d'une quinzaine de mètres de la fouille et il est plus que probable qu'elle appartienne au même ensemble funéraire (cf. analyse).

Une autre inhumation, plus anecdotique a également été observé en 2017 durant le diagnostic des allées Mortarieu (Grimbert 2018a). Cette sépulture se trouvait au-devant de l'Institut familial, ancien couvent des Cordeliers et sa présence est très certainement liée, d'une manière ou d'une autre, à cet ensemble religieux.

Depuis une vingtaine d'années les interventions dans ce secteur des allées de l'Empereur ont donc démontré la présence de plusieurs dispositifs défensifs au sein desquels se mêlent des zones funéraires dont l'ampleur et la ou les chronologies restent cependant assez floues. La présente opération de la rue Michelet était donc l'occasion de préciser cet état des lieux.



## Bilaga 1. Fyndlista

# Fyndlista

**Kv Sunaman 2 2018**  
**RAÄ 170 Växjö sn, Småland**  
Lst dnr: 431-130-2018  
KLM dnr: 33-67-2018

Fynd-nr	Kontext	Föremålstyp	Vikt (g)	Antal fragm.	Kommentar
F1	A14	Kärl, fat/skål	606	6	Keramiskärvo mindre fat, elle gods. Dekor m grönt på botte med hemring l
F2	A14	Kärl, fat	189	5	Keramiskärvo med kritpipsde
F3	A14	Kärl, flaska	53	1	Botten till rund rakt, räfflat mö genom formbl botten.
F4	A14	Kritpipa	10	3	Fragment av sk kasserade.

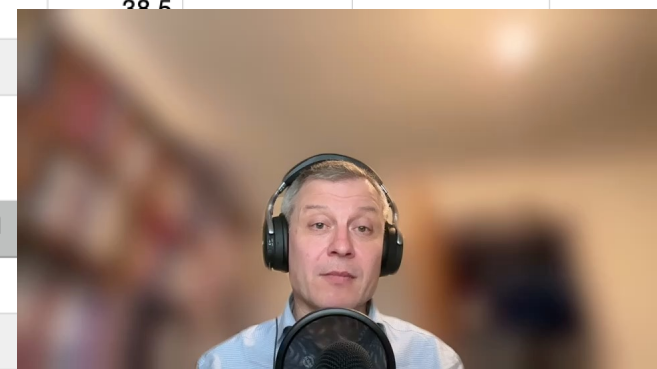
## Bilaga 1. Fyndlista

### A. Kastalruinen

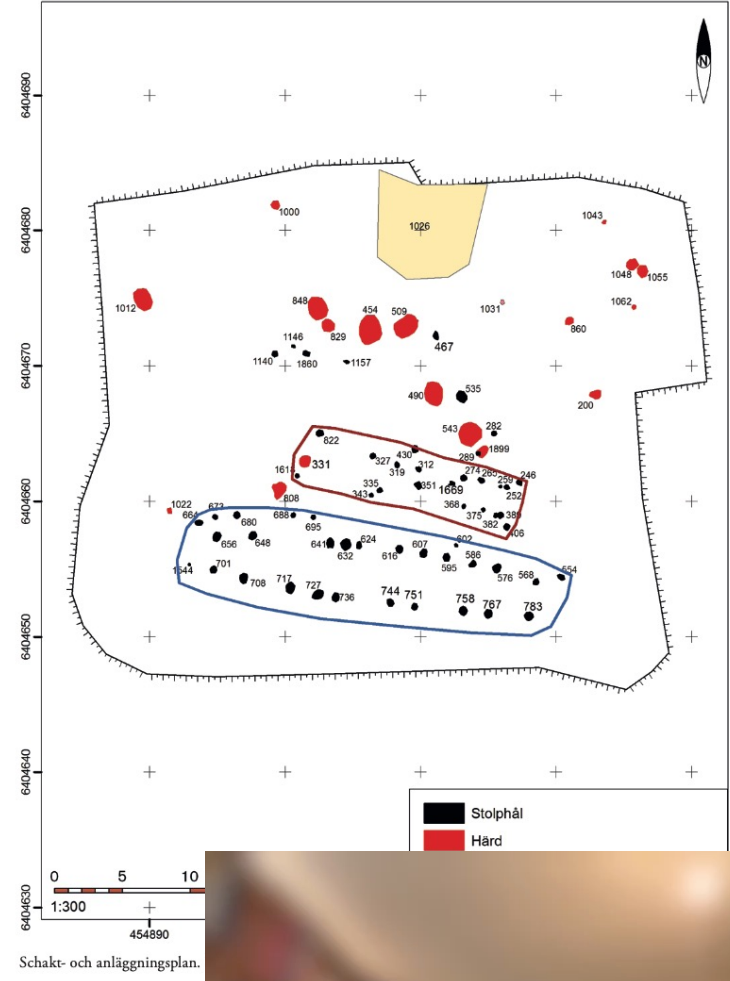
Fyndnr.	Typ	Material	Antal	Vikt (g)	Längd (mm)	Övrigt
1	Gjutform?		1	9,5	40	
2	Sporre med skänklar och stjärntrissa	Järn	1	145	86	
3	Pärta	Ben/horn	1	2	0,9	
4	Beslag	Ben/horn	1	6	23	
5	Spik/nit	Järn	7	88	67	
6	Remsölja	Brons	1	3	14	Sållfynd
7	Pärta	Glas	1	4	10	Sållfynd
8	Pärta (halv)	Glas?	1	2	9	
9	Pärta	Ben/horn	1	2	4	Sållfynd
10	Pärta (ornerad)	Glas	1	2	7	
11	Kam	Ben/horn	1	19	75	
12	Keramik	Lera	1	47		
13	Obest	Järn	8	28,5		
14	Armborstpil	Järn	1			

### B. "Klockargården"

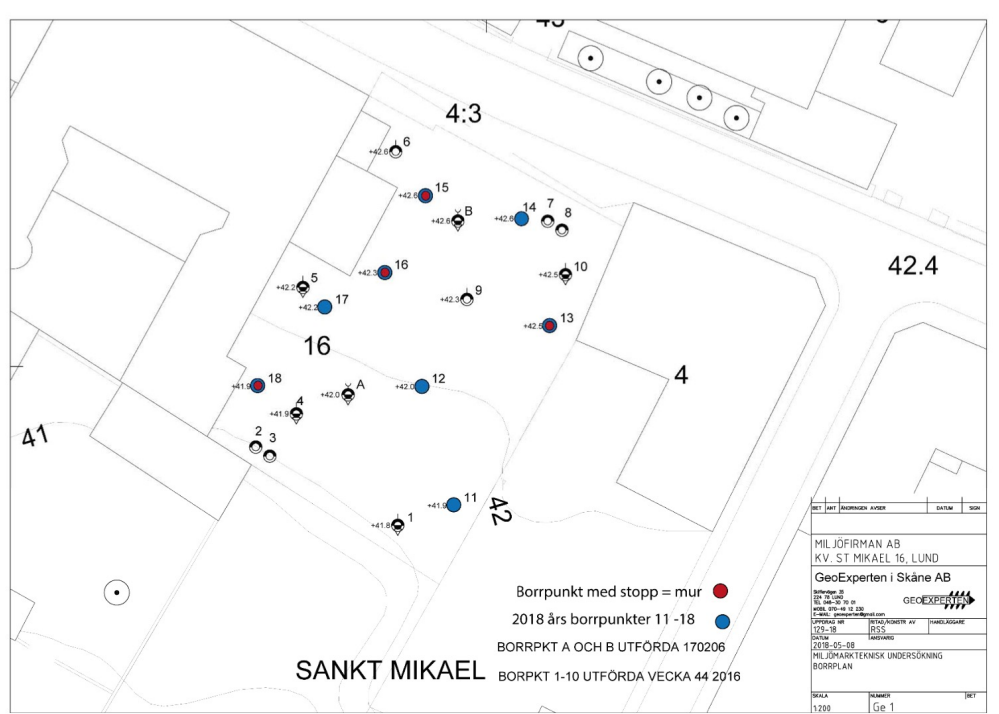
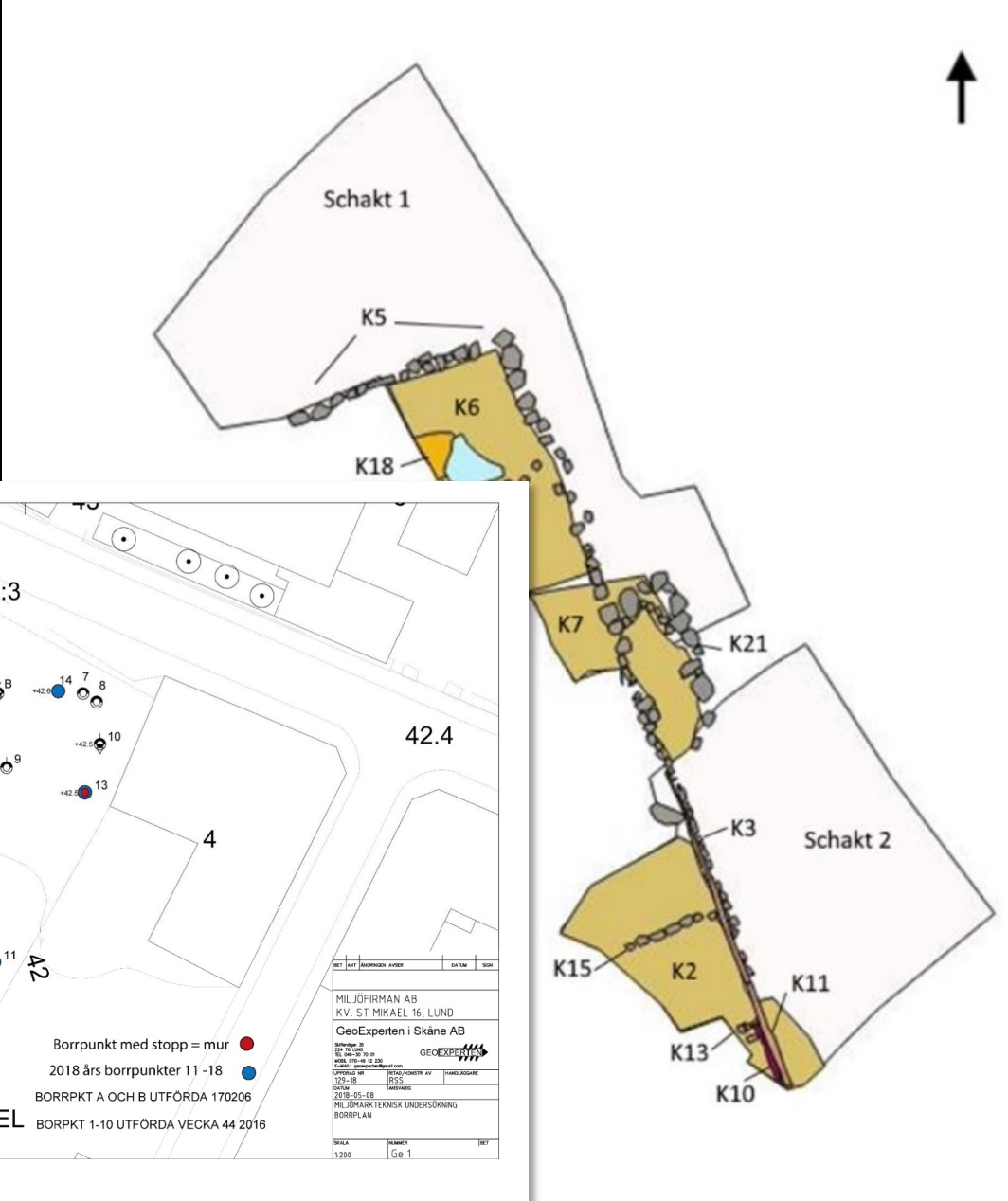
Fyndnr.	Typ	Material	Antal
1	Spikar	Järn	
2	Ben, animalt	Ben	



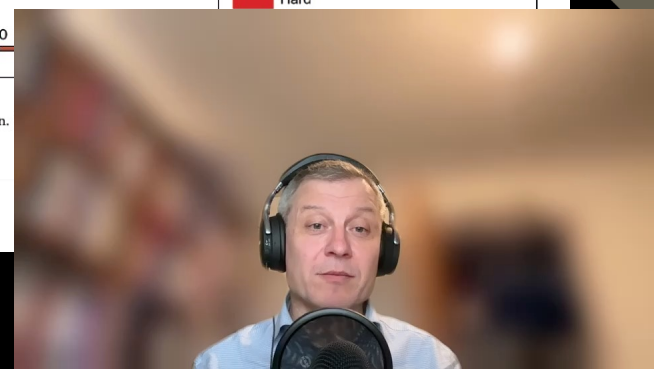
Bilaga 6. Planritningar



Schakt- och anläggningsplan.



1. Planritning borrhöjningar 2018 BP 11-18





**PDF → TXT**

**OCR**

**BERT-based NER**

**CAPTURE**



## ARKEOLOGISK RAPPORT 2018:04 •

### Inledning

Under hösten 2016 genomfördes en arkeologisk undersökning av RAÄ-nr 175 Nykyrka socken i Mullsjö kommun. Fornlämningen är registrerad som boplats. En förundersökning gjordes ett år tidigare, hösten 2015 och då påträffades boplatser av förhistorisk karaktär inom det nu aktuella undersökningsområdet. Området som undersöktes var 1 200 m<sup>2</sup> stort och anledningen till arbetet var Mullsjö kommuns planer på att exploateras för bostadsändamål på vilken den ovan nämnda fornlämningen belägen.

Anvarig för fältarbetet var Ann-Marie Nordman och för laboratoriet Ödeén, båda antikvarier vid Jönköpings läns museum. Vedartsanalysen har genomförts av Erik Danielsson vid Umeå universitet. 14C-analysen var Göran Possnert, Ångströmlaboratoriet vid Uppsala universitet och för makrofossilanalysen ansvarig var Ahlqvist vid Miljöarkeologiska laboratoriet vid Umeå universitet.

### Målsättning och metod

#### Målsättning

Den övergripande målsättningen i den arkeologiska undersökningen

Under is a TME with score 0.9995560050010681  
hösten is a TME with score 0.9997100234031677  
2016 is a TME with score 0.999579131603241  
Nykyrka is a LOC with score 0.9955717325210571  
socken is a LOC with score 0.9788461923599243  
Mullsjö is a LOC with score 0.998978853225708  
kommun is a LOC with score 0.9980702996253967  
ett is a TME with score 0.9988483190536499  
år is a TME with score 0.9995101690292358  
tidigare is a TME with score 0.9849643111228943  
hösten is a TME with score 0.9995123147964478  
2015 is a TME with score 0.9996029734611511  
1 is a MSR with score 0.9976398944854736  
200 is a MSR with score 0.9993332028388977  
m2 is a MSR with score 0.9993332028388977  
Mulleis is a LOC with score 0.9993332028388977



**Image annotation using Labelme**  
**Python/Pytorch based implementation for**  
**object detection**  
**Transfer learning using Imagenet models**





Annotations

Group: objects

- image 3
- people 2
- shovel 1
- tool 1
- unknown 1

UNUSED CLASSES

- excavator
- measuring stick

image



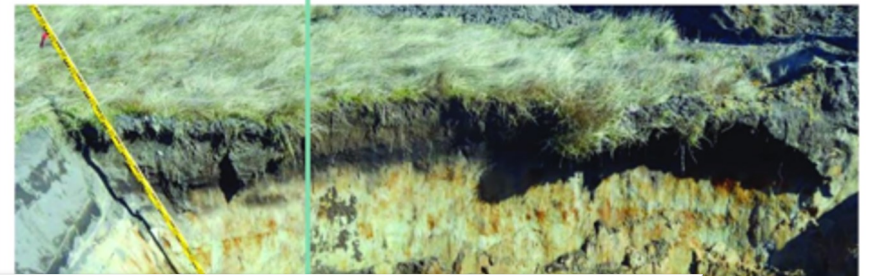
people

excavator



Figur 10. Schakt 3. I bakgrunden isbaktas for ny elkabel. Foto från nordis

tool



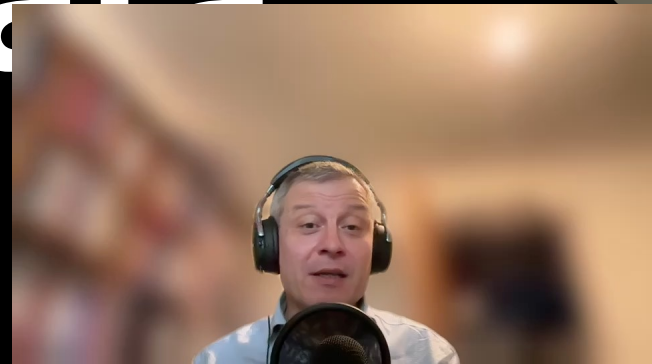
# Conclusions?

CAPTURE





**Possible to identify  
information relevant as  
(a basis for) paradata**



**Much of the relevant  
information is contextual  
and inferential.**

CAPTURE

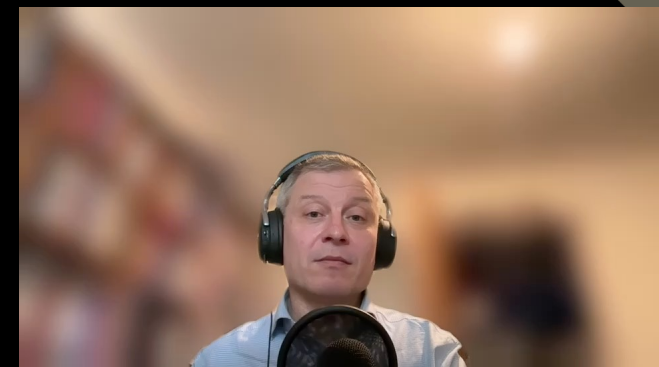


**Better understanding of  
paradata phenomenon and  
its instances needed.**

CAPTURE



**But still promising...**



# Automatic identification of archival paradata using artificial intelligence techniques

Isto Huvila<sup>a</sup>, Ekta Vats<sup>a,b</sup>, Zanna Friberg<sup>a</sup>, Lisa Börjesson<sup>a</sup>, Jessica Kaiser<sup>a</sup>, Olle Sköld<sup>a</sup>

(a) Department of ALM | Uppsala University

(b) Centre for Digital Humanities Uppsala | Uppsala University

@CAPTURE\_ERC | [www.uu.se/en/research/capture](http://www.uu.se/en/research/capture)



UPPSALA  
UNIVERSITET



European Research Council  
Established by the European Commission

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme grant agreement No 818210

