**Identification of studies via databases and registers**

**Manual Screening**

**Automatic Screening**

Records identified from:

*Web of Sciences*

*PubMed*

*Tübingen University Library*

*German Archaeological Institute*

*German National Library*

Databases (n = **1460**)

Records removed *before screening*:

Duplicate records removed (n = **902**)

Total unique records (n = **558**)

Records excluded:

Reason 1 (n = **4**)

Reason 2 (n = **3**)

Reason 3 (n = **6**)

Records screened first time (full text) with the word list

(n = **545**)

Records excluded:

Reason 4 (n = **119**)

Reason 5 (n = **4**)

Reason 6 (n = **102**)

Records screened second time (abstract or tittle) with word list

(n = **320**)

Studies included in review (n = **147**)

Convert into .txt; files and removed the reference part

Records excluded:

Reason 2 (n = **2**)

Reason 7 (n = **171**)

Records identified from:

*GoogleScholar*

Database (n = **300**)

**Identification**

Duplicate records removed (n = **18**)

Total unique records (n = **282**)

**Screening**

Records excluded:

Reason X (n = **91**)

Records screened first time (n = **282**)

Records excluded:

Reason Y (n = **62**)

Reason 6 (n = **36**)

Records screened second time (n = **191**)

Studies included in review (n = **93**)

**Included**

Total articles treating machine learning approach in archaeology (n = **140**)

Total cases with machine learning approach in archaelogy (n = 157)\*

Records excluded:

Reason 8 (n = **4**)

Reason 9 (n = **52**)

Reason 10 (n = **2**)

Total articles included in the review (n = **198**)

**Analysed**

Reason 1 = Ineligible with automatisation tool; Reason 2 = Non English papers; Reason 3 = Pdfs not accessible; Reason 4 = archaeological & ML words from the list are not present in the text ; Reason 5 = Absence of abstract; Reason 6 = Non journal articles records; Reason 7 = archaeological & ML words from the list are not present in the abstract or in the title; Reason 8 = Not dealing with archaeological problematics/data set; Reason = 9 Not using machine learning methods as defined in ours article; Reason 10 = Conflicts of interests.

\*Consider, if feasible to do so, reporting the number of records identified from each database or register searched (rather than the total number across all databases/registers).

\*\*If automation tools were used, indicate how many records were excluded by a human and how many were excluded by automation tools.

Query Jordy with doubles (300); All Jordy papers = 282; to filter Jordy = 191; included Jordy = 129; Paper included Jordy = 93

198

*From:*  Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71

For more information, visit: <http://www.prisma-statement.org/>