

## DATA PAPER

# A Relational Database of Roman Law Based on Justinian's *Digest*

Marton Ribary

University of Surrey School of Law, GB  
m.ribary@surrey.ac.uk

The *Digest* is the definitive historical sourcebook of Roman law compiled under the Byzantine emperor Justinian I (533 CE). This relational database includes the text of Theodor Mommsen's authoritative edition of the *Digest* with accompanying information about its compositional structure and its featured jurists. The data was collected from raw text files and transformed to structured machine-readable form in a Python coding environment. Preprocessed data stored in flat files were loaded to a single super lightweight SQLite database (<7Mb) which chains six tables to each other in many-to-one relationships. The database can be easily browsed, queried and expanded in a variety of free-to-use interfaces. Apart from reliable, efficient and structured information retrieval, the database assists large scale quantitative analyses and it can be used as the starting point for further computer-assisted Roman law projects.

**Keywords:** legal history; databases; Roman law; Justinian; Latin; SQLite

**Funding statement:** The research was carried out as part of an Early Career Research Fellowship funded by the Leverhulme Trust (ECF-2019-418).

## (1) Overview

### Repository location

DOI: 10.6084/m9.figshare.12333290

URL: <https://doi.org/10.6084/m9.figshare.12333290.v1>

### Context

The data was produced as part of a three-year research project on the topic of "Computational modelling of law - Sustainable legal AI from Roman legal sources" conducted at the University of Surrey School of Law since November 2019. The project aims to create a computational model in the laboratory conditions of a historical legal system based on Justinian's corpus of Roman law (533–535 CE) and focuses on three cascading layers: (1) the compositional and conceptual structure of legal texts; (2) the network of legal concepts and axioms; and (3) the logic of legal rules. Developed functionalities will be adapted to address challenges of modern law and legal technology. The relational database of the *Digest* is the result of investigations at the first layer.

## (2) Methods

Information was pulled from three types of sources: (1) the text of the *Digest*, (2) research papers reconstructing the *Digest*'s compositional structure, and (3) encyclopaedia and dictionary articles, and other sources about the jurists quoted in the *Digest*.

### Steps

The text of Justinian's *Digest* in the database is based on the authoritative edition by Theodor Mommsen and Paul Krüger [1] which largely reproduces the text of the *Littera Florentina*, an extraordinarily early and full manuscript created right after the official publication of the *Digest* in 533 CE [2, 3]. In the 1970s, the ROMTEXT project by the University of Linz transformed the Mommsen text into digital form, which was later migrated to DOS format with extensive search functions in a command line interface (CLI) [4]. The Amanuensis software provided ROMTEXT with a graphical user interface (GUI) to assist in the browsing of the text and in conducting simple search queries [5]. For the purpose of the current database, the raw text of the *Digest* was pulled from Amanuensis with titles of sections, bibliographical inscriptions of text units, and text units all separated on individual lines. The raw text is transformed into flat files (.csv) in a processing pipeline including Python scripts and manual steps, as outlined in a flowchart in the documentation of the project's GitLab page [6].

Friedrich Bluhme's compelling theory about the *Digest*'s compositional history [7] and its revision by Tony Honoré [8] incorporating insights from Dario Mantovani [9] is presented in a structured format in the Bluhme-Krüger Ordo (bko) table. Linkage with the text of the *Digest* was created by aligning inscriptions of the *Digest*'s

text units with the entries in the tabular presentation of the Bluhme-Krüger Ordo pulled from Honoré [8]. Full correspondence between the two datasets was achieved in a series of Python-assisted and manual steps, described in the project's GitLab page [6].

The date information about the jurists quoted in the *Digest* was taken from Adolf Berger's Roman law dictionary [10] and Paulys encyclopaedia of the classical world [11] with an eye on Jop Spruit's *Enchiridium* [12] which is incorporated in the online appendix of Borkowski's Roman law textbook revised by Paul du Plessis [13]. Full correspondence between jurists, dates and other datasets was achieved by aligning values (i.e. names of jurists) with the bibliographic inscriptions in the *Digest* text and entries in the bko table.

The SQLite relational database was built in Python with the imported sqlite3 package. An empty ("skeleton") database was initialized with primary and foreign keys as well as data type and value restrictions before loading the data into predefined tables row by row. While typographical errors in particular cells of particular tables may be present, the solid structure is ensured by enforcing the restrictions right at the point of creating the database.

The database is published on Figshare with documentation, a SQL schema graph (see **Figure 1**), and a set of sample SQL queries which are based on consultation with colleagues carrying out legal and historical research on Roman law as presented in the *Digest*.

### Quality Control

Typographical errors and data inconsistencies in text units and bibliographic inscriptions of ROMTEXT were corrected by hand and by Python scripts according to the published text of the *Digest* in Mommsen [1]. Further errors and inconsistencies were corrected when aligning datasets along bibliographic inscriptions and names of jurists. The exercise was carried out in multiple rounds until the designated Python script captured no errors, indicating that the 37 quoted jurists, the 300 bibliographic headings, the 432 thematic sections and the 21,055 text units are all perfectly aligned in the flat files. The project's GitLab page includes detailed documentation of any and all corrections performed on raw data [6]. Typographical errors in the text of the *Digest* inherited from ROMTEXT may remain, but this does not affect the quality of structured (SQL) queries carried out on the database. Typographical errors will be continuously corrected in future releases which will also include additional sample queries in response to user feedback.

### (3) Dataset description

#### Object name

digest.db

#### Format names and versions

Version 1, SQLite database format (.db)

#### Creation dates

2019-11-01–2020-05-20

### Dataset Creators

Marton Ribary (data curation, investigation, formal analysis, software, conceptualisation, methodology).

### Language

Latin, Greek, English.

The core "text" table of the SQLite database includes the 21,055 text units in Justinian's *Digest*. The text is primarily in Latin with occasional text units and embedded quotations in Greek. The "section" table includes the titles of the *Digest*'s 432 thematic sections, all in Latin. Additional tables with "note" or "reference" columns include information about the manual editing process, all in English. Column labels and documentation are in English.

### License

CC BY 4.0

### Repository name

Figshare

### Publication date

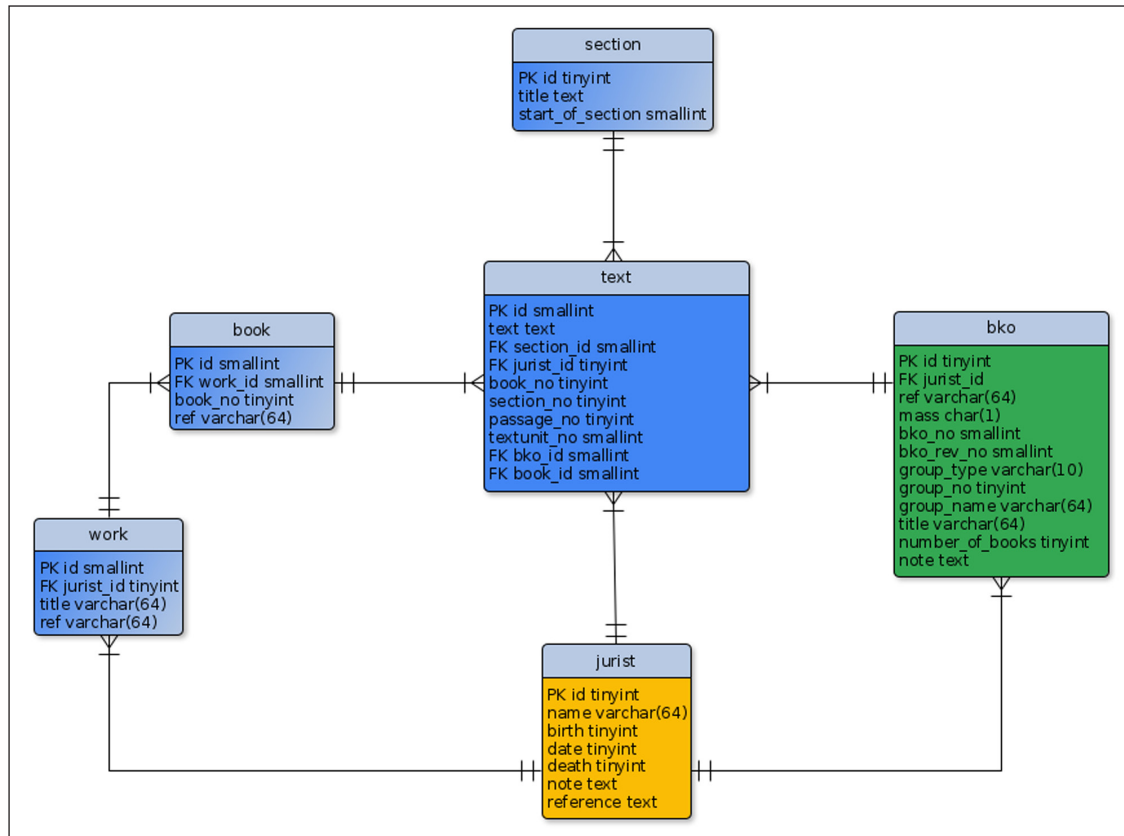
2020-05-20

### (4) Reuse potential

The relational database based on Justinian's *Digest* provides a tool for consulting Roman law sources at scale which goes beyond browsing and keyword searches. Texts are interlinked with information about jurists, thematic sections and compositional structure which allows to filter and layer results as well as identify hidden connections. The database opens up the *Digest* for structured quantitative analyses adding a new perspective to Roman legal scholarship which is primarily based on the intimate knowledge of legal issues and key sources. The database will also benefit historians, linguists and literary scholars working with textual data from the Roman world.

Presenting the text of the *Digest* only would have no added value. Theodor Mommsen's tested and respected text [1] is already available in many forms. It can be found in William L. Carey's online Latin Library [14], on the Perseus website [15], or in the Amanuensis software in its ROMTEXT version [5]. The text is also part of the Packard Humanities Institute's Latin digital text archive [16] which can be navigated, for example, with the Diogenes application developed by Peter Heslin at the University of Durham with abundant help from dictionaries and lexicons [17].

While the text is easily accessible, a structured presentation has been hitherto missing. The compilers of Justinian's *Digest* meticulously recorded bibliographic information in the inscriptions of excerpted text units which they carefully arranged in 432 thematic sections. Speaking in modern terms, this valuable metadata is not fully exploited in the raw text repositories mentioned above. The relational database approach presented here will not only provide a new angle for researching the *Digest*, but it will also hopefully inspire others to invest in normalizing and structuring the ancient historical data they work with. A linked data universe of the ancient



**Figure 1:** SQL schema of *digest.db*.

world [18] including structured data repositories of documents [19], papyrological [20] and numismatic evidence [21] as well as prosopographical [22] and geographical [23] data, among many others, will ultimately break down the walls between disciplinary silos and steer scholarship towards a systemic understanding of the ancient world. It will never replace close reading as the cornerstone of historical (legal) research, but it will help to bring together remote and seemingly unrelated pieces of information for more nuanced and deeper insights.

By including Honoré's revision of the Bluhme-Krüger Ordo in the *bko* table, the database allows to assess the plausibility of a much-debated theory about the *Digest's* compositional structure and compositional history. Its inclusion is not meant to promote the theory over that of David Pugsley [24] or other more agnostic scholars. Honoré's quantitative research naturally lends itself for database presentation. It was largely based on manual aggregation of ROMTEXT queries which, among others, led to the creation of a *Digest* concordance [25] and the statistical tables presented in numerous books. Honoré wanted to identify objective markers of style of prominent jurists who, according to Honoré, shaped and reworked the texts presented in the *Digest* [26]. He presented a compositional theory for the *Digest* based on statistical analysis in so-called biographies for the editor-in-chief Tribonian [27] and for two main jurists quoted in the *Digest* from the early and late classical period of Roman law, Gaius [28] and Ulpian [29]. Honoré's theory sparked a heated debate with Alan Watson being one of the main contesters [30]. The "battle of the Atlantic" [31]

was staged on the pages of the *Rechtshistorisches Journal* which published Watson's [32] and Honoré's [33] views side by side. The "battle" was eloquently summarised by Peter Birks [34] who pointed to the crucial contributions of Honoré's quantitative approach while acknowledging that his grand theory of textual interventions by a handful of jurists is probably unfounded. An alternative and similarly controversial theory about the *Digest's* compositional history was developed by David Pugsley who argued that Tribonian had discovered a historical sourcebook of Roman law and recycled its material according to the existing law school practice which was largely following Ulpian's commentaries [24]. Pugsley's theory is partly based on the arrangement of 432 thematic sections in the 50 books of the *Digest* which could be translated to a database presentation in a future release. The database allows to recreate and expand quantitative analyses by automating significant aspects of what Honoré, Pugsley and others achieved by labour-intensive and largely manual data aggregation.

Apart from compositional structure, one may also conduct quantitative analyses on the text of the *Digest* according to custom-defined time slices. The "date" column in the "jurist" table includes the year in which a particular jurist was the most active based on assumptions derived from demographic studies of the Roman world [35, 36]. Courtesy of this date, jurists and the text units preserved from them can be grouped in custom-defined periods to generate aggregate statistics. The corresponding SQL query reveals that there are 751 text units in the early and republican period of Roman law (until 27 BCE), 4,169 in

the early classical (until ca. 190 CE), 15,904 in the late classical (until ca. 240 CE) and 231 in the post-classical period. Another SQL query tells us that the size of the late classical group is due to the number of text units excerpted from the works of Papinian (1,156), Ulpian (8,979) and Paul (3,954) who collectively take up 67% of the entire *Digest* corpus. This suggests that one will need to downsample these authors for a representative corpus-based analysis.

Filtered term search is another example demonstrating the benefits of the database approach. Let's say one is interested in the term "proprietas" and runs a search with the "%" wildcard to locate the term with potential morphological variations. By linking the "text" and the "jurist" table, the initial 255 hits can be narrowed down to 18 for someone who is only interested in how the jurist Papinian uses the term. The 255 hits are distributed to the four custom periods of Roman law with 7 text units in the early and republican period (0.93% of all text units in the period), 30 in the early classical (0.72%), 217 in the late classical (1.36%) and 1 in the post-classical period (0.40%). Even though the numbers are too small to draw conclusions from them, they encourage to examine the hypothesis that the concept of "ownership" received abstract formulation in the late classical period of Roman law.

Combining the search functions in Amanuensis with the *Digest* concordance [25] and Otto Lenel's *Palingenesia* [37] could achieve similar aggregate statistics, but the database approach has at least two major advantages. First, it requires a structured search query which saves the effort of aggregating data manually and promotes the open science virtues of transparency and reproducibility. Second, the database approach allows to plug in advanced quantitative methods such as distributional semantics. For example, word embeddings models [38] trained on general and genre-specific (legal) corpora allow to extract words with vector representations most similar to that of "proprietas". The qualitative, that is, "close reading" inspection of the appropriate text units would then reveal whether the concept of "ownership" is indeed encoded in many terms and phrases. The "development" of the idea could be also mapped to time slices, if we have an appropriate amount of relevant textual data in the corpus. Such hybrid investigations would provide quantifiable support for the argument that even though Roman law does not define "ownership" as such [39, 40] the idea is encapsulated in ancient formulas such as "the thing is mine" (*res meum esse*) [41, 42]. The idea may be undefined and dispersed, but it is very much present. The database approach combined with advanced quantitative text analytical methods such as distributional semantics would support semantic mapping and point to additional passages for "close reading" which are sometimes missed when research is principally term-driven.

This first release of the *Digest's* SQLite relational database includes six tables chained together by keys in one-to-many relationships (see **Figure 1**). This core release will be continuously supplemented with additional information either by adding columns to current tables or by adding and chaining new tables to the database. One planned addition is to add keyword tags to thematic sections based

on computer-assisted text analyses such as hierarchical clustering. Keywords will assist topic search and the navigation across legal disciplines. These keywords will eventually be transformed into a semantic web ontology in, for example, Resource Description Framework (RDF) which will provide a systematic map of (Roman) legal themes. Another planned addition is to include the dictionary form (lemma) and part of speech of word tokens which constitute the text units of the *Digest*. This will enable improved information retrieval by returning morphological and semantic variations of the search term. It will also provide a starting point for linguistic and stylistic analyses of texts grouped by user-defined values such as jurists, themes or periods.

The database does not currently have a custom interface. It can be viewed in a command line, desktop or online application environment which are all free of charge. The release comes with a set of sample SQL queries to give an idea about what kind of questions a relational database of the *Digest* can answer. Users are encouraged to get in touch for assistance with translating their research questions to SQL queries. These queries will be added to further minor releases. A future major release will include a custom interface sitting on top of the database and the SQL queries. User feedback will play a key role in correcting typographical errors inherited from raw text, adding functionalities to the database, expanding it with linked information, and designing the custom interface.

#### Additional File

The additional file for this article can be found as follows:

- **Supplement File.** A Corpus Approach to Roman Law Based on Justinian's Digest. DOI: <https://doi.org/10.5334/johd.17.s1>

#### Acknowledgements

The raw text data was pulled from the graphical interface of the Amanuensis V5.0 software developed by Peter Riedlberger and Günther Rosenbaum. Amanuensis incorporates the ROMTEXT database created by the University of Linz under the supervision of Josef Menner. Barbara McGillivray offered advice during the database creation process.

#### Competing Interests

The author has no competing interests to declare.

#### References

1. **Mommsen T, Krüger P.** (Eds.) *Corpus Iuris Civilis. Vol 1: Institutiones. Digesta.* 5<sup>th</sup> edition. Berlin: Weidmann; 1889.
2. **Radding C, Ciaralli A.** *The Corpus Iuris Civilis in the Middle Ages: Manuscripts and transmission from the sixth century to the juristic revival.* Leiden: Brill; 2007. DOI: <https://doi.org/10.1163/ej.9789004154995.i-277>
3. **Baldi D.** Il Codex Florentinus del Digesto e il Fondo Pandette della Biblioteca Laurenziana (con un'appendice dei documenti inediti). *Segno e Testo.* 2010; 8: 99–186.
4. **Klingenberg G.** Die ROMTEXT-Datenbank. *Informativa e diritto.* 1995; 4: 223–232.



5. **Riedlberger P, Rosenbaum G.** (Eds.) *Amanuensis V5.0*. München. 2020 [Last accessed: 20 May 2020]. URL: <http://www.riedlberger.de/08amanuensis.html>
6. **Ribary M.** Documentation files of the pyDigest project on GitLab. 2020. URL: <https://gitlab.eps.surrey.ac.uk/mr0048/pydigest>.
7. **Bluhme F.** Die Ordnung der Fragmente in den Pandectentiteln. Ein Beitrag der Entstehungsgeschichte der Pandecten. *Zeitschrift der Savigny-Stiftung für Rechtsgeschichte*. 1820; 4: 257–472.
8. **Honoré T.** Justinian's Digest. The distribution of authors and works to the three committees. *Roman Legal Tradition*. 2006; 3: 1–47. DOI: <https://doi.org/10.2139/ssrn.1270181>
9. **Mantovani D.** *Digesto e masse Bluhmiane*. Milan: Giuffrè; 1987.
10. **Berger A.** Encyclopedic dictionary of Roman law. *Transactions of the American Philosophical Society*. 1953; 43: 333–809. DOI: <https://doi.org/10.2307/1005773>
11. **Wissowa G, Kroll W, Mittelhaus K, Ziegler K, Gärtner H.** (Eds.) *Paulys Realencyclopädie der classischen Altertumswissenschaft: Neue Bearbeitung*. Stuttgart: Metzler; 1893–1980.
12. **Spruit JE.** *Enchiridium. Overzicht van de geschiedenis van het Romeinse privaatrecht*. 3<sup>rd</sup> edition. Deventer: Wolters Kluwer; 1992.
13. **du Plessis P.** *Borkowski's Textbook of Roman Law*. 5<sup>th</sup> edition. Oxford: Oxford University Press; 2015. Online supplements: <https://global.oup.com/uk/orc/law/roman/borkowski5e/>
14. **Carey WL.** (Ed.) *The Latin Library*. Fairfax, VA: George Mason University; 2020. URL: <https://www.thelatinlibrary.com/>.
15. **Crane GR.** (Ed.) *Perseus Digital Library v.4.0*. Medford, MA: Tufts University; 2019. URL: <http://www.perseus.tufts.edu/hopper/>.
16. **The Packard Humanities Institute.** *Classical Latin Texts*. 2020. URL: <https://latin.packhum.org/browse>.
17. **Heslin P.** (Ed.) *Diogenes v.4.5*. Durham: Department of Classics and Ancient History; 2020. URL: <http://d.iogen.es/>.
18. **Clayless HA.** Sustaining Linked Ancient World Data. In Berti M (Ed.), *Digital Classical Philology: Ancient Greek and Latin in the Digital Revolution*. Berlin/Boston: De Gruyter Saur; 2019. pp. 35–50. DOI: <https://doi.org/10.1515/9783110599572-004>
19. **Depauw M, Gheldof T.** Trismegistos. An interdisciplinary Platform for Ancient World Texts and Related Information. In Bolikowski Ł, Casarosa V, Goodale P, Houssos N, Manghi P, Schirrwagen J (Eds.), *Theory and Practice of Digital Libraries – TPDL 2013 Selected Workshops*. Cham: Springer; 2014. pp. 40–52. URL: [https://link.springer.com/chapter/10.1007/978-3-319-08425-1\\_5](https://link.springer.com/chapter/10.1007/978-3-319-08425-1_5).
20. **The Duke Collaboratory for Classics Computing & The Institute for the Study of the Ancient World.** *Papyri.info*. 2020. URL: <https://papyri.info/>.
21. **Ackermann R,** et al. (Eds.) *Nomisma*. 2020. URL: <http://nomisma.org/>.
22. **King's College London Classics and Digital Humanities.** *Digital Prosopography of the Roman Republic (DPRR)*. 2020. URL: <http://romanrepublic.ac.uk/>.
23. **Ancient World Mapping Center and Institute for the Study of the Ancient World.** *Pleiades*. 2020. URL: <https://pleiades.stoa.org/home>.
24. **Pugsley D.** *Justinian's digest and the compilers*. 2 vols. Exeter: University of Exeter; 1995–2000.
25. **Honoré T, Menner J.** (Eds.) *Concordance to the Digest Jurists*. Oxford: Clarendon Press & Oxford Microform Publications; 1980.
26. **Honoré T.** *Emperors and lawyers: With a Palingenesia of Third-Century Imperial Rescripts 193–305 AD*. 2<sup>nd</sup> edition. Oxford: Clarendon Press; 1994.
27. **Honoré T.** *Tribonian*. London: Duckworth; 1978.
28. **Honoré T.** *Gaius: A Biography*. Oxford: Clarendon Press; 1962.
29. **Honoré T.** *Ulpian: Pioneer of Human Rights*. 2<sup>nd</sup> edition. Oxford: Oxford University Press; 2002. DOI: <https://doi.org/10.1093/acprof:oso/9780199244249.001.0001>
30. **Watson A.** Review of Honoré's Ulpian. *Times Literary Supplement*. 18 February 1983. p. 104.
31. **Cohen D.** The Battle of the Atlantic. *Rechtshistorisches Journal*. 1983; 2: 33–36.
32. **Watson A.** An Open Letter to D.C. *Rechtshistorisches Journal*. 1984; 3: 286–290.
33. **Honoré T.** New Methods in Roman Law. *Rechtshistorisches Journal*. 1984; 3: 290–305.
34. **Birks P.** Honoré's Ulpian. *Irish Jurist NS*. 1983; 18: 151–181. DOI: <https://www.jstor.com/stable/44027631>
35. **Scheidel W.** Roman age structure: Evidence and models. *The Journal of Roman Studies*. 2001; 91: 1–26. DOI: <https://doi.org/10.2307/3184767>
36. **Bruce F.** Roman life expectancy: Ulpian's evidence. *Harvard Studies in Classical Philology*. 1982; 86: 213–251. DOI: <https://doi.org/10.2307/311195>
37. **Lenel O.** (Ed.) *Palingenesia iuris civilis*. Leipzig: Teubner; 1889.
38. **Sprugnoli R, Passarotti M, Moretti G.** Vir is to Moderatus as Mulier is to Intemperans: Lemma Embeddings for Latin. In *Proceedings of the Sixth Italian Conference on Computational Linguistics, CEUR Workshop Proceedings, Bari, Italy, 13–15 November 2019*. 2019. DOI: <https://doi.org/10.5281/zenodo.3565572>
39. **Birks P.** The Roman Law Concept of Dominion and the Idea of Absolute Ownership. *Acta Juridica*. 1985; 1–38.
40. **Schermaier M.** Dominus actuum suorum. *Zeitschrift der Savigny-Stiftung für Rechtsgeschichte: Romanistische Abteilung*. 2017; 134: 49–105. DOI: <https://doi.org/10.26498/zrgra-2017-0104>
41. **Ankum H, Pool E.** Traces of the Development of Roman Double Ownership. In Birks P (Ed.) *New Perspectives in the Roman Law of Ownership. Essays for Barry Nicholas*. Oxford: Clarendon Press; 1989. pp. 5–41.
42. **Giglio F.** The concept of ownership in Roman law. *Zeitschrift der Savigny-Stiftung für Rechtsgeschichte: Romanistische Abteilung*. 2018; 135: 76–107. DOI: <https://doi.org/10.26498/zrgra-2018-1350106>

**How to cite this article:** Ribary M 2020 A Relational Database of Roman Law Based on Justinian's *Digest*. *Journal of Open Humanities Data*, 6: 5. DOI: <https://doi.org/10.5334/johd.17>

**Published:** 05 August 2020

**Copyright:** © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 Unported License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.