From dBase III+ to the semantic web: twenty-five years of the Coin Hoards of the Roman Republic database.

Ethan Gruber* and Kris Lockyear**

* American Numismatic Society. gruber@numismatics.org
** Institute of Archaeology, University College London. noviodunum@hotmail.com
[corresponding author]

Abstract

The CHRR database began life in 1988 and has grown from 24 to over 700 hoards. It has been successfully moved between software packages and has been used as a teaching resource as well as for research. Over the last year the database has been incorporated into the online numismatic databases hosted by the American Numismatic Society. This paper addresses the issues raised in turning a long-term personal research database into a publicly available resource database built on the principles of linked open data. We will address both the technical challenges and the problem of the expectations of the target audience.

Key words

numismatics, hoards, databases, linked data, Roman

1. Background to the project

Coin hoards are an important source of information for numismatists. As sealed assemblages of coins they can be used in the relative dating of coin issues by seriation, a technique pioneered by numismatists prior to Petrie’s famous study (Crawford 1990). The contents of hoards can be examined to show patterns of production and supply and the distribution of hoards indicate the regions and periods in which coins circulated. As such, numismatists regularly publish both detailed reports on individual hoards, as well as corpora of hoards of a particular period and issuer, for example Crawford’s Roman Republican Coin Hoards (1969).

The Coin Hoards of the Roman Republic (CHRR) database was started as part of an MSc dissertation by the second author (Lockyear 1989). At that time, the data it contained was solely that published in Roman Republican Coinage (RRC) by Crawford (1974). The principal, and at the time unrealized, aim was to test Crawford’s estimates of the size of issues of coinage during the later Republic. The data, however, were also used to examine the factors that influence the structure of coin hoards via a computer simulation, a study that was published in the 1990 CAA proceedings (Lockyear 1991). This initial research led to the second author to embark upon a doctorate which analysed a much larger set of coin hoards using a variety of statistical techniques including Dmax based cluster analysis (Lockyear 1995, 1996a ) and Correspondence Analysis (Lockyear 1993, 2000, 2007a). Further simulation work also undertaken to examine the original question regarding the veracity of Crawford’s die estimates (Lockyear 1999). The analytical part of the resultant thesis (Lockyear 1996b) was substantially revised ten years later (Lockyear 2007a).

The CHRR database was redesigned as part of the doctoral project, using a simplified version of Ryan’s (1988) database schema (Fig. 1) and implemented in dBASE III+. This was necessary as data collection in the British Museum was undertaken using a double 5 ¼ inch floppy “laptop”. A suite of programs was written to ease data input, conversion, checking and manipulation (Lockyear 1996b: 120–7).

Following completion of the thesis, the database was used in teaching at the Institute of Archaeology, UCL, and for this it was imported into Microsoft Access. It was quickly found that the relation and attribute names were confusing to students and these were renamed (Lockyear 2007a: 12–9). No new
data were added to the database until 2006 when a second phase of data collection was undertaken prior to a complete reworking of the analyses (Lockyear 2007a).

Since 2007 new data have been added to the database in a piecemeal fashion, largely to enable specific pieces of research to be undertaken.

Following the publication of Patterns and Process the second author was approach by Rick Witschonke of the American Numismatic Society (ANS) with a view to creating an online version of the database which is the subject of this paper. CHRR Online is based on the principles of open linked data.

2. Research v. resource?

Databases can be divided into three broad but overlapping types (Fig. 2). The first type consists of management databases which encompasses library catalogues or the ubiquitous sales databases beloved by writers of manuals. The second category consists of personal research databases. These are created by scholars for a particular project or interest with no intention, at least initially, of making the information publicly available. The database application, if any, is basic and task orientated, the data included are restricted to that immediately relevant to the project at hand and it is unlikely that any formal data definition document or manual exists. The creator and user of the database are the same person, and that person understands the limitations of the database, and how to use it. The last type, resource databases, are those created with the express intention of being a publically available resource for scholars, such as the Celtic Inscribed Stones Project database.¹ These databases suffer from having to try and predict what sort of data scholars will want, and in having to provide those data in a manner that academics not conversant with SQL and entity relationship diagrams can use.

The CHRR database was conceived and created solely for the purposes of the second author. As such there are gaps in the coverage. For example, few bronze hoards are included and early hoards are largely omitted. On the other hand, some areas, such as hoards from Romania, are very well represented. One major hurdle, therefore, was to create an online resource database from a personal research database in a way that the potential audience will appreciate its limitations and view the product as a glass half full, rather than a glass half empty. These challenges have been met by a process of informing, enhancing, and enabling.

The users of the database have a variety of sources to consult to inform them about the database. CHRR Online includes a FAQ section and help, a paper written for the American Journal of Numismatics (Lockyear forthcoming) is also available from the site as well as links to the original PhD thesis and the relevant section of Patterns and Process (Lockyear 1996b, 2007a). The creator of the database is available to answer questions either via email or via various relevant online Bulletin boards.

The process of creating the online version of the database has already enhanced the content by the addition of more detailed information regarding coin types and by the provision of geographical data. In addition, new hoard data added to the CHRR database will be uploaded to the online version on a regular basis. It is planned to extend the coverage of the database as well as to add new hoards as the information becomes available.

Lastly, the web interface enables users to access the data via a variety of searches, to map the results, to plot graphs of the contents of the hoards and so on. Most importantly, it allows users to download the results of their searches to allow them to conduct more detailed analyses if they so wish.

The remainder of this paper examines the technical issues involved in converting a personal research database structured along relational lines into a open access resource database using linked data.

¹ http://www.ucl.ac.uk/archaeology/cisp/
3. NUDS for Hoards

With no standard approach to encoding coin hoards electronically, the American Numismatic Society endeavored to develop a draft XML schema in order to address inadequacies in the printed format of hoard catalogues and to serve as an interchange format between existing relational database systems designed for the recording of hoards and finds. To inform the data modeling process, a study of printed hoard publications was undertaken to create a flexible schema capable of representing relatively simple catalogues as well as those hoards published following modern archaeological excavations. To fulfill the former example, i.e., a simple hoard catalogue, An Inventory of Greek Coin Hoards (IGCH) (Thompson et al. 1973) was consulted. Listing more than 2,000 Greek coin hoards, a reference in IGCH rarely exceeds a few dozen lines, with the quantity of coins identified by merely mint or ruler. On the opposite end of the spectrum are the many hoards published in the last three decades, such as Bland and Burnett’s The Normanby Hoard (1988b). In this monograph, each coin is listed with its associated coin type in addition to physical attributes, such as weight and diameter.

CHRR lies somewhere in the middle of this spectrum, with the majority of coins defined by the type corpora, Roman Republican Coinage (RRC; Crawford 1974) and Roman Imperial Coinage (RIC), but the records contain no specific physical characteristics. The heavy use of RRC and RIC to identify coins in CHRR makes it possible to extract data in the form of RDF directly from Nomisma.org, defined as a “collaborative effort to provide stable digital representations of numismatic concepts and entities” (Gruber et al. forthcoming b).2 The use of data from Nomisma is especially important in this case because the Access database contained a table for coin types, which on one hand was partially duplicated Nomisma’s data, but was also incomplete for it lacked the mints, issuers, legends, and type descriptions that Nomisma’s RDF contains. Other coins listed in Lockyear’s database which were not identified by existing catalogue numbers are described by a general typology (e.g., “miscellaneous Republican denarius”).

The schema for hoards is based largely on the Numismatic Description Standard (NUDS) XML schema, discussed in greater detail in “Linking Roman Coins: Current Work at the American Numismatic Society,” presented at CAA 2012 (Gruber et al. forthcoming a). The basic structure of this NUDS-Hoard document is as follows:

- nudsHoard (root element)
  - nudsHeader (header)
  - descMeta (descriptive metadata)
  - title (titles in multiple languages accepted)
  - hoardDesc (hoard description)
    - findspot
    - discovery
    - deposit
    - disposition
  - contentsDesc (contents description)
    - contents (coins)
    - containers
    - otherObjects (objects other than coins—ingots, jewelry, etc.)

NUDS-Hoard includes a <nudsHeader> which contains metadata about the electronic record itself. The content and structure of this element are identical to that of the <nudsHeader> in NUDS documents. Below the header is Descriptive Metadata which contains two subsections, in contrast to the same section in NUDS, which may contain six descriptive subsections in addition to several sets of descriptors.

The Hoard Description contains information about the find spot: the name and/or coordinates. The find spot can link to URIs defined by Nomisma or the Pleiades Gazetteer (http://pleiades.stoa.org) for

---

2See nomisma.org for more information.
ancient places, or Geonames (http://www.geonames.org) for modern ones. CHRR denotes the find spot by modern place name, and therefore Geonames URIs were adopted. The Geonames API allows programmatic extraction of geographic coordinates, facilitating the generation of KML for the mapping of hoards.

Linking to URIs will likely be the most common method of defining the find spot for a hoard, but geographic coordinates can also be explicitly input. These coordinates may be public or private (withheld from the public display of the record on the web). This is particularly advantageous when used in conjunction with URI-defined find spots. The public may see that a hoard was discovered near Lincoln, but specific and highly accurate GPS coordinates of the find spot may be recorded for the benefit of archaeologists and administrators, but blocked from appearance on the web to prevent looting of the site.

The Hoard Description contains information about the discovery of the hoard—the date, type of discovery (excavation, agricultural activity, metal detecting, etc.), the finder(s), or other notes about the circumstances of the discovery. After the discovery section is the deposit date. If this date is not explicitly expressed in the NUDS-Hoard record, the date can be derived programatically by analyzing the contents of the hoard. The date should be explicit when archaeology provides more accurate context than can be derived from the dates of the coinage, although this is unlikely for Republican hoards. Finally, the Hoard Description contains descriptors for the disposition of the hoard described generally as a whole. The disposition can also be defined on the coin level.

Below the Hoard Description section is the Contents Description, which comprises the bulk of the hoard record. The <contentsDesc> contains three subsections: contents (for listing coinage), containers (for defining the materials, dates, and typologies of the containers in which the hoards were placed), and a listing of other objects found in the hoard. It is possible to internally link containers defined in the hoard record to coins and other objects listed in the Contents Description in order to maintain the relationship between the container and its contents, which may be useful in large hoards which consist of more than one container.

With the structure of the NUDS-Hoard document model generally described, attention must be focused more directly on the structure of the numismatic contents section of the record. Two XML elements may appear in the <contents>: coin and coin group (<coinGrp>). The difference between these two elements is that the latter accommodates a “count” element to denote the quantity of coins which adhere to the descriptors listed within the group. Both elements allow any Descriptive Metadata elements found in NUDS—Physical Description, Typological Description, etc., defined by the NUDS XML namespace. A hoard very generally described by denomination may include one denarius and two quinarii in the structure outlined below:

```xml
<coin>
    <nuds:typeDesc>
    </nuds:typeDesc>
</coin>
<coinGrp count="2">
    <nuds:typeDesc>
    </nuds:typeDesc>
</coinGrp>
```

Although this structure is simple, statistical analyses can be supported because the data is machine-readable by processors capable of interpreting XML. Denominations are explicitly captured by XML tags which also link Nomisma URIs. The quantity of quinarii is captured in an XML “count” attribute.
One may expand the above XML fragment further by differentiating the two quinarii geographically by separating the <coinGrp> into two coins, each with a Typological Description including the denomination of quinarius, but one minted in Rome and the other Emerita, for example. Legends and type descriptions may also be inserted into the <typeDesc>. A <physDesc> element may be added for capturing the weight, diameter, and die axis of each coin. Thus, this approach to listing coins and coin groups within a hoard can accommodate the granular descriptive practices applied by Bland and Burnett in The Normanby Hoard (1988a) and other similar publications.

The <typeDesc> within a coin or coin group can also utilize the xlink:href attribute to link to Nomisma URIs:

```xml
<coinGrp count="5">
  <nuds:typeDesc xlink:href="http://nomisma.org/id/rrc-44.5"/>
</coinGrp>
<coin>
  <nuds:typeDesc xlink:href="http://nomisma.org/id/rrc-88.2a"/>
</coin>
```

The fragment above indicates that a hoard comprises of five instances of RRC 44/5 and one of RRC 88/2a. The majority of entries from the Access database were migrated into NUDS-Hoard XML records which contain little more than a reference to a Geonames URI for the find spot and a list of Nomisma URIs for Roman Republican Coinage numbers and their associated quantities. The data model is relatively simple, especially compared to the organization of Lockyear's original database schema, but the analyses which may be conducted by users of the public interface are quite sophisticated. CHRR is built upon Numishare, an open-source application for managing and publishing numismatic data, which was heavily modified to support the publication and dissemination of hoard data.3

4. Numishare for Hoards

From a presentational standpoint, CHRR does not differ significantly from Online Coins of the Roman Empire (OCRE, http://numismatics.org/ocre/) or other Numishare-based digital numismatic collections. There is an HTML interface for searching and browsing the collection of hoards with keyword search and filtering by facets derived from typological attributes. Search results can be delivered in the form of Atom, KML, and CSV. It also contains a Solr facet-based mapping interface. CHRR’s most significant departure from other Numishare collections for coins or coin types lies in the HTML representation of the NUDS-Hoard record. While the HTML resource for a hoard is not wholly dissimilar to a coin record in that it contains a textual description of the record (including a list of hoard contents), a mapping component, and a quantitative analysis component, the processes by which these aspects of the HTML view are generated differ substantially.

Before the transformation from XML into HTML can begin, the subject referenced by each unique xlink:href attribute associated with Typological Descriptions are imported into an XSLT variable. By default, the script will look for a NUDS datastream by appending '.xml' to the URI. If the <typeDesc> links to http://nomisma.org, '.nuds' is appended to extract the NUDS/XML model provided by Nomisma. A table row for each coin or coin group in the hoard contents is created. The denomination and date are displayed by default. For those coins or groups which link to coin types on Nomisma or OCRE, the <typeDesc> stored in the XSLT variable associated with that particular URI is passed through an XSLT template to generate a block of HTML which is hidden by default, but may be shown by Javascript with a mouse click. The XSLT template for rendering the Typological Description in hoard records is shared with XSLT for generating HTML views for coins and coin types, and therefore the HTML view of typologies in CHRR is identical to OCRE, complete with links to Nomisma or predefined search results (Fig. 3).

3See CAA 2012 paper on Numishare (Gruber et al. forthcoming b).
5. Visualization and Analysis

CHRR’s fundamental mode of visualization is geographic: the record page for a hoard displays a map combined with SIMILE Timeline through the intermediary Javascript library, Timemap. The map for a hoard in CHRR not only shows a point for the findspot and all of the mints associated with it, but the timeline shows every datable coin, which makes it easier for users to get a sense of the temporal, in addition to geographic, distribution of coins found within the hoard (Fig. 4).

One of the most important features of this project is its analytical functionality. Quantitative analysis of hoards provides insight to scholars on the social and economic conditions in the time and place where the hoard was buried. While IGCH (Thompson et al. 1973) is merely a catalogue of hoards containing very little consideration of their contexts, modern hoard statistics are quantified into charts. Further historical context can be ascertained about a particular hoard by comparing its contents to hoards in neighboring geographic areas. A table in The Normanby Hoard shows a listing of the percentage of coins per mid-late third century emperor across fourteen different hoards (Bland and Burnett 1988a: 116). The format of the data in The Normanby Hoard is purely tabular, but more recent publications, such as Coin Hoards X (Hoover et al. 2010) include computer-generated graphs. Wright (2010) includes a column chart showing the percentages of coins in two hoards originating from various mints. The following page in this article includes a representation of these data in a table. In an article about the Gaziantep Hoard in the same volume, Meadows and Houghton (2010: 147) include pie charts showing the geographic origins and general categories (Seleucid, Alexanders, etc.) of its contents. According to Hoover, many charts published in this volume were generated in Microsoft Excel and submitted as images exported directly from that program or post-processed in Adobe Illustrator to improve the layout for publication (Fig. 5, which is based on the data reproduced in Fig. 6). Thus, each chart required manual entry into a spreadsheet the categories and values desired for visualization. The scholar is limited by the time available to create new spreadsheets in the categories desired for statistical analysis. A comprehensive database solution like CHRR dramatically improves upon the spreadsheet-based method of chart generation used even in current hoard publications.

With these charts and tables as a model for the arrangement and display of data to which scholars are accustomed, it was immediately apparent that Numishare should be adapted to replicate these statistical analyses and visualizations. With a broader set of typological fields available through Nomisma-defined coin types, it is possible to quantify hoards not only by ruler or denomination, but by mint, region, material, dynasty, date, or even deity. The final category, deity, is rarely, if ever, analyzed in printed hoard catalogues, which is striking considering the distribution of deities may lend insight into a hoard's possible religious context (e.g., votive deposits of coins with in a sanctuary).

Hoard analyses are available through two avenues in CHRR: on the record page for a particular hoard, under the “Quantitative Analysis” tab, and through the “Analyze Hoards” page, which is accessible from the navigation menu. The user may select from a variety of options to generate results which may be visualized in the form of charts using the Javascript library, Highcharts (Fig. 7), or downloaded in a CSV file. For chart visualizations, the user may select the numeric response type between the options of count and percentage (of total coins in a hoard), as well as cumulative percentage when analyzing dates. Next, the user selects the chart format: bar and column for most typologies, but a variety of linear graph options are available for date visualizations. Next, the user may select one or more typological categories listed above, such as denomination or mint. Finally, the user may select hoards for comparison. When visualizing data in chart form, the user may select up to 8 hoards for comparison or up to 30 for generating a CSV file. It may take up to 10 or 15 seconds to process 30 hoards for comparison, which is a significant improvement over manual methods.

---

4SIMILE Timeline is an open source chronological Javascript library developed by MIT [http://www.simile-widgets.org/timeline/]. Timemap is developed by Nick Rabinowitz, which is a library that interlinks the timeline function with Google or OpenLayers maps. It is open source and available through Google Code at [http://code.google.com/p/timemap/].

5See figures on page 258 and 259 of Hoover et al. (2010). Fig. 6 shows the tabular format of the table represented graphically by Fig. 5.
employed before the availability of computers for calculating the data. The scholar is no longer restricted to the presentation of data provided by Bland and Burnett (1988b), for example, but now has greater control over setting the parameters of his or her own research questions.

With respect to CHRR, Numishare introduces an option to generate linear graphs based on the cumulative percentage of issue dates within a hoard. This type of graph shows the temporal extent of the hoard, with particular focus on the most vigorous periods of hoarding activity. Perhaps no Republican hoard illustrates this more than the one found at Actium, where nearly 80% of its contents are denarii issued in 31 B.C. by Mark Antony (all from uncertain mints) and buried shortly thereafter (Fig. 8).

6. Integrating CHRR Data into OCRE

Of the 694 hoards currently cataloged in CHRR, 72 contain references to Augustan types defined in RIC Volume I. These are 72 findspots which can be incorporated into the maps of OCRE for perhaps one-quarter of all Augustan coin types. As part of its architecture, OCRE employs an Apache Fuseki-based RDF database and SPARQL endpoint for associating objects (whether physical coins or coin hoard records) with imperial coin types.6 The RDF describing a hoard associates its own URI to URIs of coin types by means of the Nomisma-defined nm:type_series_item. There may be many nm:type_series_item references in the RDF for each hoard. In the case of CHRR, most types will be attributed to RRC numbers in Nomisma, but a significant minority link to types in OCRE. See below, an abbreviated RDF fragment (numerous RRC type references have been omitted):

```xml
<rdf:RDF>
  <rdf:Description rdf:about="http://numismatics.org/chrr/id/CAP">
    <dcterms:title xml:lang="en">Capilna (Romania; CAP)</dcterms:title>
    <dcterms:publisher>American Numismatic Society</dcterms:publisher>
    <nm:findspot>
      <rdf:Description rdf:about="http://www.geonames.org/682812/">
        <geo:lat>46.7333</geo:lat>
        <geo:long>22.1</geo:long>
      </rdf:Description>
    </nm:findspot>
    <nm:closing_date rdf:datatype="xs:Year">0004</nm:closing_date>
    <nm:type_series_item rdf:resource="http://nomisma.org/id/rrc-44.5"/>
  </rdf:Description>
</rdf:RDF>
```

This RDF associates the Căpîlna (CAP)7 hoard in CHRR with two coin types listed amongst its contents: RRC 44/5 and RIC Augustus 287. Additionally, the find spot is defined by the Geonames place, http://www.geonames.org/682812/, which defines Căpîlna, Romania.8 Once this RDF has been inserted into the triplestore, the OCRE record for Augustus 287 will immediately show a point on its map for this small Romanian town. The XSLT stylesheet which generates the KML file for this coin type (http://numismatics.org/ocre/id/ric.1(2).aug.287.kml) uses SPARQL to query its findspots from the triplestore and process them into KML Placemarks. The closing date, defined by Nomisma as http://nomisma.org/id/closing_date and shown in the fragment above as nm:closing_date, can be used to insert dates into the KML, making it possible to render the temporal extent of a coin type’s circulation in OCRE using the TimeMap library (Fig. 9).

---

6See http://jena.apache.org/documentation/serving_data/ for more information on Apache Fuseki.

7 Each hoard is given a three-letter primary key which is hoped will become the standard identifier for Roman Republican coin hoards.

8 CHRR uses place names as given in the original publications. Romanian orthography was changed in the early 1990s leading to most instances of î being replaced by à.
7. Conclusions

In general, numismatists have quickly embraced the benefits of the world wide web, and of online databases. The majority of these databases, however, are of collections, either personal or museum-based such as those of the ANS or the British Museum. Surprisingly, data about coin assemblages, be they hoards or site finds, are harder to find online despite the long history of the use of computer software to manipulate those data (e.g., Reece 1991, see also Lockyear 2007b). Exceptions include the Iron Age and Roman Coins from Wales data which can be downloaded from the ADS website (Guest and Wells 2007), and metal detected finds from the UK available from the Portable Antiquities Scheme website.\(^9\) Two new coin hoard projects, one looking at Romano-British finds and one looking at other Roman hoards, are due to begin in 2013. It is hoped that the work of the European Coin Find Network and the Nomisma project in defining numismatic concepts and creating stable digital representations of them, will enable the various projects to employ compatible terminology and allow the development of truly useful cross-database search and analysis tools.

Acknowledgments

The authors would like to thank Rick Witschonke for providing the initial impetus for the project and for coordinating its development, and to Arianna Traviglia for organizing the CHRR Online launch at CAA2013. The second author would also like to thank Michael Crawford for allowing access to his records now housed in the Department of Coins and Medals, and to the many members of the Department past and present for facilitating that access.

Bibliography


\(^9\) http://finds.org.uk/


**List of Figures**
Fig 1. The entity relationship diagram for the CHRR database. The dashed line represents a conceptual link between the dBase/Access database and the bibliographic information stored in the second author’s BibTeX files. Both sets of data have been included in CHRR Online.

Fig 2. The relationship between database types.

Fig 3. Screenshot of Hoard Record Page.

Fig 4. Temporal and geographic distribution of the Pachino hoard.

Fig 5. Comparison of mints between two hoards. From Wright 2010, fig. 2.

Fig 6. Tabular results for mints in Fig. 5. From Wright 2010, Table 1.

Fig 7. Visualization of deities in a selection of Romanian hoards.

Fig 8. Cumulative percentage by date comparison.

Fig 9. Distribution of RIC 1(2) Augustus 2b.