HORAI: An integrated management model for historical information

Del Fresno-Bernal Pablo[1], Medina-Gordo Sonia*[2], Travé-Allepuz Esther[3]

[1] Sistemes de Gestió del Patrimoni SCCL – Barcelona, Spain – ORCID ID: 0000-0001-8775-9113
[2] Universitat de Barcelona – Barcelona, Spain – ORCID ID: 0000-0002-3921-3025
[3] Universitat de Barcelona – Barcelona, Spain – ORCID ID: 0000-0002-6769-4487

*Corresponding author
Correspondence: sonia.medina@ub.edu

ABSTRACT

The archiving process goes beyond mere data storage, requiring a theoretical, methodological, and conceptual commitment to the sources of information. We present HORAI as a semantic-based integration model designed to facilitate the development of information systems that promote seamless communication across diverse disciplines within the field of Historical Sciences. This model adopts a data-centric approach, and uses principles and methods derived from Archival Science and Information Studies. Along with the model, we present a few experiences currently developed in different parts of Spain. Within this framework, this research has produced useful insight pertaining to the dynamics of data production and collaborative information management.

Keywords: management system, past construction, slow science, historical knowledge, CRM

Introduction

The archiving process becomes an essential activity in a significant portion of historical studies. Firstly, as part of a comprehensive data management plan, it involves determining what is deemed to be archived; that is, what information is considered relevant and useful for conducting a study on a specific element of the past (Opgenhaffen, 2022: 1). Secondly, it is during this stage that the researcher or research group bears the responsibility of selecting the objects from which knowledge will be generated. Consequently, careful consideration is given to the modes and formats of information recording, employing the necessary standards, tools and techniques accordingly. Hence, we deduce that this practice of information management extends beyond mere data storage, requiring a theoretical, methodological and conceptual commitment to these sources, as they are the foundation of our narratives. This inference is particularly noticeable in contemporary digital contexts, where certain *data imaginaries* shape our approach to historical sources to a greater or lesser extent (Huggett, 2022a: 270-278).

In data-driven approaches, potentially problematic dynamics seem to be emerging; for example, the development of large-scale analysis strategies without considering whether the extracted and cleaned data align to comparable conceptions. C. Chippindale (2000) already addressed this aspect by distinguishing between *data* (raw data) and *capta* (theory-laden data), suggesting the need for a prior theoretical approach to the latter in order to (re)utilize them (Wylie, 2017). In these scenarios, one of the main challenges lies in striking a balance between the interpreted value and the epistemic trust elements (Mickel & Byrd, 2021; Sandoval, 2021) when constructing our narrative about the past, given the substantially artefactual and multifaceted nature of historical data (Owens, 2011; Schöch, 2013). While not exclusive, these observations are indeed noteworthy within disciplines where methodological tradition has been subject to reflection, such as Archaeology, since its parallel development with the use of new technologies has not always been accompanied by theoretical reflection on the implications that these digital means have on the articulation of knowledge (Huggett, 2015; 2021).

Furthermore, when the study of a historical element requires the convergence of multiple disciplines, the aforementioned commitment becomes even more challenging to systematize, especially when the goal is to offer a comprehensive interpretation. In such situations, the knowledge alignment can be hindered by the difficulty to coordinate the approaches and methods used by each specialist for data gathering. This may explain why some research teams manage to aggregate (*complement*) data obtained from diverse sources of information (for example, Costa & Sancho, 2022), but struggle to effectively match (*integrate*) them, despite it being the intended objective. While such results are valuable, they fail to address the necessary understanding among scholars from different fields from a wider perspective. Dealing with this aspect, J. Moreland (2006) provides a historiographic overview of its repercussions, focusing on the relationship between documentalists and archaeologists in recent decades. From this perspective, A. Woolf (2009: 6-7) argues that the challenge lies in the interpretative frameworks employed by each group, and calls for a shift away from perceiving disciplines as autonomous and insurmountable domains of knowledge.

Given the outlined panorama, we believe that the problem does not reside in the data recording process itself, but rather in how we reason about the foundations of its production, particularly to avoid what some scholars have begun to refer to as “data deluge” (Bevan, 2015). As an alternative starting point, F. Niccolucci (2020) has recently suggested a shift in perspective. The author proposes a data-centric viewpoint aimed at enhancing our engagement with different information archiving proposals, whether they come from one or multiple disciplines. Echoing these insights, our research team contends that one possible way for overcoming this issue could be through data modelling. We assert that without a shared conceptual model and a coherent management structure enabling interoperable information recording (including written, archaeological, or heritage-related data, among others), the accessibility and transferability of the generated knowledge could remain challenging for both specialists and other stakeholders. A constructed knowledge, therefore, should transcend the confines of archaeological studies and encourage collaborative synergies with other domains of expertise.

In view of this context, we propose HORAI[1] as a semantic-based integration model aimed at facilitating the creation of information systems that foster smooth communication among the various disciplines

[1] https://horai.es/web/.

encompassed within the so-called Historical Sciences. To achieve this, HORAI adopts a data-centric approach, prioritizing the very data, and at the same time it addresses the processes inherent to a project by drawing upon principles and methods derived from Archival Science and Information Studies. This methodological proposal emerges from previous experiences gained in diverse research contexts, each characterized by its own premises, objectives, and informational foundations. Over time, the refinement of the model has allowed us to set HORAI as a digital platform that underpins the development of other applications, management processes, or technologies within the realms of Historical or Heritage Research. In this regard, two crucial aspects needed to be addressed: (1) the need to mitigate the risks of insular information management and (2) the need to facilitate a robust records management structure that ensures the consistent traceability of the research process along with the preservation of the generated information.

Delving into a key concept: *integration*

The aforementioned problem of insular management is related to the lack of interoperability that turns our knowledge bases into isolated systems. Technically, this management framework corresponds to the challenge of sharing data repositories with third parties. Conceptually, problems arise when a data model cannot employ concepts nor extend its applicability to other contexts. By referring to the concept of *integration*, we intend to highlight the importance of providing an operational inter-system reciprocity both for our repositories and for our models of abstract data representation. In our case, the minimum information units of the HORAI model, namely Unit of Topography (UT), Unit of Stratigraphy (US), and Actor (Ac), play a key role (Mauri, 2006; Travé et al., 2020). In previous publications, we delved into these concepts through various case studies (Del Fresno et al., 2020; Travé et al., 2021a; Travé & Medina, 2021; Medina & Travé, 2021), and now we seek to underscore their significance in the archival processes we are concerned by.

Information exchange between heterogeneous research contexts

In the field of Humanities and Social Sciences, there is a growing number of projects committed to enhancing data exchange. This is evidenced by the proliferation of initiatives that facilitate the coordination of different systems, such as projects like Arachne[2], Arches[3], and PARTHENOS[4], to name just a few. This proliferation is particularly prominent in these contexts due to the complexity and heterogeneity of the information handled, especially in cases that involve multidisciplinary frameworks. However, little attention has been paid to the fine line between compatibility and interoperability. In the former, systems can coexist and share data but have limitations in direct (inter)communication. In the latter, interoperability enables smooth data exchange and ensures data integrity, as long as common conceptual standards are adopted. It is important to consider that each scholar or research group usually adheres to their own management schemes, and data recording often aligns to the specific focus of each project. Matching this diversity of situations and preventing informational isolation is possible through both compatible and interoperable schemes, but it is only through the latter that effective data integration can be achieved.

Considering these premises, designing a digital environment that coordinates the storage and mutual understanding of systems is not misguided. Nonetheless, it is not the ideal option for the integrated management processes we pursue. Data integration also requires a conscious, explicit, and unambiguous association between the value of data and the meaning we attribute to it (Uschold & Gruninger, 1996; Gruber, 1995). This is why works based on ontologies or conceptual models seem to be better received in the development of systems capable of synchronizing knowledge, as the conceptual definition of system elements forms the basis for associating the semantics of our data with those of other environments (Guarino, 1997). This makes them highly competent approaches for information management, with models such as CIDOC CRM[5] serving as widely recognized and adaptable references among heritage and historical

[2] https://arachne.dainst.org/.
[3] https://www.archesproject.org/.
[4] http://www.parthenos-project.eu/.
[5] https://www.cidoc-crm.org/.

documentation managers. However, adopting this research line is not without its challenges, as the tension between the generic and the specific is sometimes not fully taken into account.

In our commitment to data, we must be able to establish a scale of knowledge representation and determine to what extent it satisfies the defined requirements of another system. Indeed, this idea has to do with the abstraction problems that occur when adapting data models, which can blur the integration process. A few years ago, C. González-Pérez and P. Martín-Rodilla reflected on these issues when they presented their Conceptual Reference Model CHARM[6], along with a mechanism they called *gradual refinement of models*, which minimizes the problems we have discussed (González-Pérez & Martín-Rodilla, 2014). If we revisit the question of epistemological equivalence across disciplines, the potential of this methodology allows for versatility in working with data that have rarely been explored in depth. It enables us to access information gathered in different systems, yet aligned to the same conceptual model, without the need to reconstruct each one's organizational or recording architecture. This coincides with the objectives of HORAI, providing a platform that bridges our databases, rather than imposing or adopting specific norms and standards for records management.

A proposal for an integrated Historical Science

We still have to develop how the concepts we presented at the beginning assist us in achieving this objective. Taking a closer look at archaeological (field)work, G. Lucas (2001) reflected on the hermeneutics that enable the construction of archives in this discipline. Acknowledging the argumentative limitations that justify the recording of a portion of this work, the author proposed reconsidering the archaeological archiving process not as a copy of the intervened heritage entity, but as a substitution for it. What is interesting about this idea is that, in this process of conceptualization and information management, the outcome constitutes not just a representation but a *displacement* of the object itself to the archive, ultimately allowing archaeological data to be reinterpreted by others (Lucas, 2001: 44). As mentioned above, although dynamics vary among teams, this archiving procedure typically revolves around a widely recognized minimum unit of information in archaeology: the US. However, the challenge lies in conceptually representing the material data separately from the interpreted one (Martín-Rodilla et al., 2016), and this is where the use of the UT paves the way for data integration.

As suggested by A. Mauri in one of his earlier works on this concept, the management of the past turns around two key notions: time and space (Mauri, 1995). From this perspective, historical logic involves, among other things, recognizing certain actions within space that can be chronologically related (Mathieu, 2021). In our model, a US is an instance of a UT insofar as it informs us about an action that occurred at a specific time and space, leaving a material trace on the territory (Harris, 1989; Travé et al., 2020: 14). This physical evidence, as we mentioned, is characteristic of archaeological work, but it is not the only clue that the discipline searches for to reconstruct the historical narrative. There are other types of vestiges that may lack materiality and can be detected through the combined interpretation of different USs (for example, Carandini, 1997: 139-142), and that we could capture in the form of UTs. Under these terms, the concept of UT expands upon that of US because the identification of the action is possible regardless of its materiality, making it versatile enough for its application in any Historical Science.

As we will see later on, each project can adjust (*refine*) their respective data models using these units according to their aims, and incorporate them into management processes regardless of the nature of the information sources we employ. The interoperability of systems is possible through the use of common concepts, and in domains like History, where epistemic value has traditionally been given greater importance (Topolski, 1992: 36-47), the use of UT represents a significant methodological step forward because it is possible to reproduce its work with usual sources in the discipline, such as written sources. On the other hand, the UT/US/Ac model makes it possible to work with notions such as *spatiality* (Shoorcheh, 2019), and thus to record the interaction between space and society at a given moment. For archives constructed as a result of archaeological work routines, we cannot identify the agency in the analysed past; we could recognise who documented an action, but then we would dig into aspects more related to metainformation (Martín-Rodilla & González-Pérez, 2019; Huvila, 2022). However, identifying

[6] http://www.charminfo.org/.

the social dimension, the Ac, is operable in sources such as texts or photographs, and its conceptual representation through HORAI enables effective integration of the recorded data.

Archival Science applied to heritage-related records management

Along with data integration, HORAI addresses other processes related to records management, thus building on principles and methods of Archival Science. These premises should be connected to the advantages of maintaining traceability throughout the different stages of project progress, allowing for easy access and reproduction of the information generated and stored at any given time. In collaborative contexts, we believe it is important to have control over the sequence of processes carried out before, during, and after data management to trace their development from start to finish. We advocate for its significance because, on the one hand, it facilitates the exchange of files from our archives with different project participants and, on the other hand, it allows us to identify potential inconsistencies in the information and easily address them in all the elements that constitute an archival bond to that information (Stančić & Bralić, 2021: 2-3). In order to define the context and content of what is archived, we propose a structured framework inspired on certain descriptive units (**Fig. 1**), which are defined according to the General International Standard Archival Description.

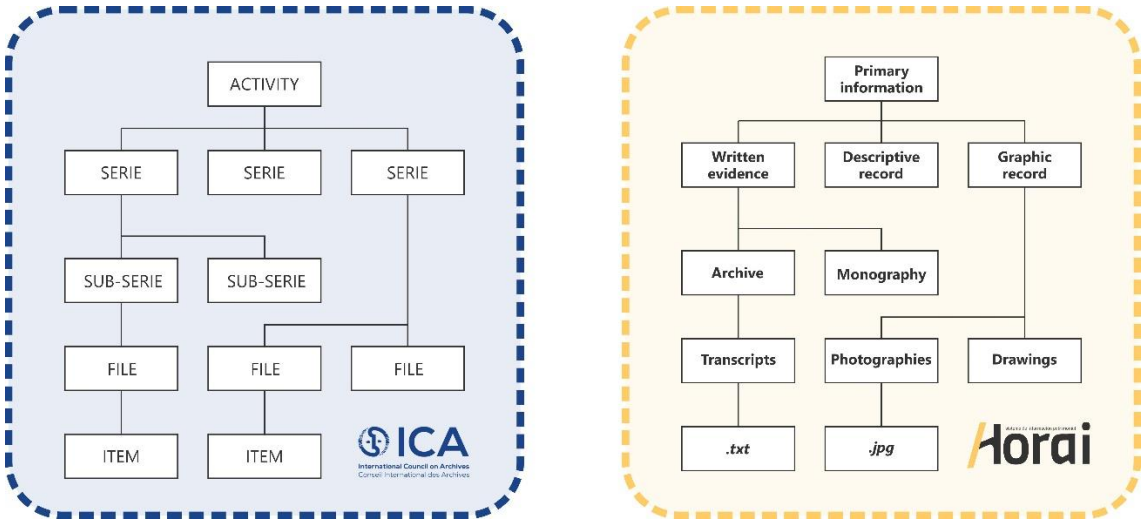

Figure 1 - Comparison chart between the standards of archival description proposed by the International Council of Archives and an example of a classification chart used in HORAI for the *primary information* activity and sub-processes during the datification context.

Within this framework, we can further divide the process into four management phases that roughly correspond to the following:

1. The context of project management. This refers to all the aspects that need to be considered in terms of time, resources, and the number of people involved in the project's execution, among others. As these are the initial steps, this phase also coincides with the beginning of the data capture process, as it involves the design of the study, the obtention of funding and excavation licenses, the access to relevant documents, the establishment of a bibliographic foundation… All of these activities contribute to the datification of our archival collection (**Fig. 1**).
2. Data gathering and management, as a primary form of information management. In this phase, activities such as (systematic) data extraction from sources, development of storage protocols, and insertion of metadata are addressed. It is important to note that we are still in a stage of information extraction, but now overlapped with the initial phase of digitalization. During this process, we observe the *de-contextualization* of data (Leonelli, 2014: 4), their displacement from their original context, while simultaneously reflecting on how we document them (in a digital format).

3. Data processing, as a form of secondary information management. For illustrative purposes, we could mention the processes of exploitation and interpretation of primary data, or those of derived documents, such as statistical exploration, laboratory analyses, computational simulations, etc. Moreover, it is in this phase that the identification of HORAI's units of information converges, and the focus shifts from extracting information to constructing knowledge. Therefore, this stage is distinct from both of the above-mentioned as it involves the *re-contextualization* of data (Leonelli, 2014: 4-5), considering the context in which they were originated and offering (our) new perspectives on them.

4. Data participation, or the final sharing stage where both the data and the information produced are disseminated. Given the target we have been defining within HORAI, later on we will have the opportunity to discuss and assess how, in fact, this sharing can go beyond mere information reuse (Leonelli, 2014: 5-6) and lead to its own evolution (Huggett, 2022b: 104). The involvement with the construction of knowledge is possible through various means (academic, educational, administrative…), formats (open access, under license…), and forms (computer files, interactive computational platforms, web applications…).

As can be seen, the lifecycle of data in the primary management of information extends beyond the project because this work of reflection and commitment allows us to lay the foundations for its potential use in the form of new questions or theoretical contexts. The ability to interoperate this information with other systems ensures such situations, since the direct interaction between them creates scenarios for the evolution of data that, process after process, accumulate experiences and keep them as dynamic elements of management. Furthermore, as we mentioned before, these general guidelines proposed by HORAI are not incompatible with the needs of each research context, which is why the use of specific vocabularies and varied tools is expected (for instance, Del Fresno & Mauri, 2020; Travé et al., 2021b: 7-9). In any case, we are interested in highlighting two concluding ideas from the procedure: on the one hand, this approach to records management ensures organised collaboration among the participants by aligning them all to the proposed description scheme for a project. On the other hand, it promotes information preservation by stimulating the review of data obsolescence, which is characteristic of digital media.

Some current experiences

The issues raised so far are the cornerstone of the systems we introduce below. These three tools have been refined over the past years within a research team characterized by close collaboration between companies, universities, museums, and local cultural entities, creating an environment for sharing diverse experiences and knowledge, while fostering constructive dialogue. Each of these systems was created in response to specific challenges, and entails different levels of complexity and objectives. Nevertheless, the information management processes and archival dynamics are shared and aligned to the previously outlined framework, and provide more detailed and practical insight into our reflections on the subject. For their presentation, we have focused on the elements that help us understand them within the context in which they are developed, as it is not the purpose of this study to delve into the technical aspects.

The archaeological information management system ANATOLE

ANATOLE emerges from a previous experience, SigArq, whose specificities (Del Fresno, 2016), proposed improvements (Fructuoso, 2018), and application to various case studies have suggested avenues for enhancing the tool. This web application was part of an Archaeological Information System aimed at organizing, processing, and standardizing the results obtained from fieldwork and archaeological research (Del Fresno et al., 2021). Among other features, SigArq offered users a protocol for entering archaeological information, allowing them to store and access primary data. However, it was somewhat restrictive in that, in order to incorporate it into other projects, they had to adhere to pre-established standardization guidelines, which ultimately could require reorganizing data recording protocols in cases where studies did not start from scratch. This characteristic hindered its applicability in archaeological sites studied by groups not linked with the SigArq development team. Therefore, being aware of this limitation, we sought to overcome any technological dependency through HORAI.

ANATOLE focuses on the management of heritage-related records. The process of uploading, classifying, and retrieving records involves the development of a project, and the management system is built upon the concept of US in its descriptive, graphic, cartographic, and temporal dimensions, whereas the UT

corresponds in this case to the stratigraphic synthesis. Moreover, ANATOLE enables the spatial exploration of gathered information either through raw or synthetised data. So far, ANATOLE is used by various heritage scholars and curators, thus managing archaeological sites of different chronology, typology, and geographical location, currently in Spain.

The GREYWARE Information System

From a diachronic perspective, the research project GREYWARE [PID2019-103896RJ-I00] explores the processes of change and continuity related to a specific type of material: the reduced-fired pottery or *greyware*. This study proposal builds upon previous experiences we have acquired in periods such as the Medieval and Post-Medieval ages (Padilla, 1984; Travé, 2009; Travé & Vicens, 2018). On these bases, our interest lies in delving deeper into how social changes influence the production and consumption patterns of these materials, particularly during transitional periods (Travé, 2021). This approach has led to the development of a homonymous information system focused on the study of this specific object (Travé et al., 2021b), requiring robust management mechanisms due to the volume and heterogeneity of the information available. This empirical reality encompasses data about the ceramic object itself, as well as information about the agents involved in its production (Vicens & Travé, 2018: 120-124) and other related secondary data (Travé et al., 2014; Travé, 2022). While we have proposals that allow us to systematize some of this data (for instance, Adroher et al., 2016), the challenge lies in processing them within the framework of the conceptual adequacy mentioned in previous sections.

Through HORAI, it has been possible to manage data that aligns to the required abstraction model for each defined entity and establish direct communication with other digital repositories whose information had already been generated prior to the project and needed to be integrated. These entities reveal typologies, ceramic fabrics or productions that, based on the UT concept, can be compared according to the level of precision in our analysis, whereas the Ac concept incorporates the social component.

A historical information management system

The third system is an outcome of an ongoing PhD thesis that pursues a dual objective. On the one hand, we aim at achieving reconciliation and traceability of interpretative frameworks in disciplines such as History, Archaeology, and Remote Sensing (Medina & Travé, 2021). On the other hand, we seek to reflect on the marginalized nature of mountain landscapes during the Early Middle Ages, whose history often seems to start *in media res*, following a sudden, passive integration of the people inhabiting them into the structures of external societies characterized by a certain dominance. The study area considered is the southern region of the Sierra de la Demanda, located between the current provinces of Burgos and Soria. For this region, we have studies that have generated complementary interpretations of written and archaeological evidence (Pastor Díaz, 1996; Escalona, 1996), but they have hardly modified the explanatory foundations of previous works exclusively based on documents (Álvarez Borge, 1991). Besides, the marginality of this area has been considered only in relation to secular forms of domination (Escalona, 1996) and the distribution of known archaeological sites in relation to a few environmental features (Pastor Díaz, 1996: 62-66; Álvaro, 2012: 51-98).

Therefore, we are not starting from scratch, but the work done so far prompts a reflection on the mechanisms used to apprehend this past, both from the point of view of interpretative and information management. To explore the articulation of these landscapes, we identify the notion of power, not as an element of dominance, but as a manifestation of human agency. Precisely, the conceptualization of data by means of HORAI allows us to model the dialogue between the traces of this agency —either material or symbolic— in the form of UT, and the agents who participate in making those traces, in the form of Ac (Medina et al., 2022).

Discussion: data sharing or engaging data?

Each of the aforementioned tools are based on the principles of conceptual reconciliation and efficiency in historical information management that we advocate for. On the one hand, regardless of the technological support employed, all of them allow for a direct exchange of the information and generated knowledge while maintaining their own objectives and study frameworks. These procedures enable an environment of data interoperability as the systems are built upon a conceptual model of data integration,

shifting the combinatorial work of these data towards knowledge matching practices based on semantics. On the other hand, they also enable us to define traceability in our work with information, at least from the perspective of Records Management and administrative organization of archives. This not only ensures transparency in workflows but also their monitoring. At this point, we question to what extent this investment of time in understanding data is allowing us to *decelerate* methodological processes characterized by streamlining and simplification of information (Marila, 2019) and gain awareness of the past from the very foundations upon which we build it (Thibodeau, 2019; 2021).

To address this issue, M. Marila (2019) examines the works that have been thriving within what is known as *fast science*. The author expresses concern regarding the direction this approach is taking, as the necessity to establish a valid way of conducting science, based on an empirical logic, and the systematization of these principles, are leading to an increasing separation between scientific practice and the effective generation of knowledge. Similarly, L. V. Orman (2015) shares related concerns when exploring what he terms the 'information paradox'. Orman delves into issues such as production costs, the sense of obsolescence, and competition across various levels, resulting in a high volume of data with a limited quality of information. Likewise, data modeling can yield a reductionist representation of historical complexity if we fail to consider the narrative within data. The exacerbation of this simplification risks erasing nuances, contexts, and fundamental connections present in the data. Therefore, a more reflective approach to information management should integrate strategies that preserve contextual complexity, thereby enabling a more precise and profound understanding of the past.

All these insights are not isolated concerns, but have also been assessed from different perspectives by other authors (for example, Gero, 2007; Rączkowski, 2020; Lucas & Witmore, 2022; Marila, 2022). From the perspective of working with information, we believe that the recent reflections by J. Huggett (2022b: 103-106) provide a complementary theoretical dimension to the proposal we advocate for. As we have mentioned before, our way of making information available is not solely aimed at its publication, but we hope that our work provides users with the necessary tools to critically evaluate the knowledge generated once it is going to be reused. In these routines, we acknowledge that the usual knowledge management process of paradigms like Big Data, when properly implemented (Liu et al., 2021), facilitates such reuse. However, we must also note that in this increased reusability we may end up neglecting aspects such as how these results were initially signified. Therefore, we believe that the research agenda for those exploring traces of the past in the coming years should promote alternative approaches to mere data sharing, fostering greater reflection and engagement with the data.

Taking as a reference point the juxtaposition of positions regarding data lifecycle outlined by J. Huggett (2022b: 104), and considering the working perspectives we propose, we would like to conclude by assessing three key pillars of the HORAI model, which we have somewhat delineated throughout the previous sections (**Fig. 2**). Firstly, data modelling ensures that we go beyond tasks such as data identification and capture, as the creation of a model brings forth (*reveals*) and adds dimension (*expands*) to the data. In this line, the works that promote the definition of ontologies (Van Helden et al., 2018) or Conceptual Reference Models (González-Pérez & Parcero-Oubiña, 2012) are significant. Secondly, Records Management offers alternative tools for data validation and preservation, as we can discern how they have been constructed (*reflect*) and enhance our involvement (*engage*) with them. Although not exclusive, the tasks of creating taxonomies or file classification systems are two examples within this field, as exemplified in **Fig. 1**. Lastly, interoperability extends beyond data sharing environments, as it allows us to be part of their (re)creation (*participate*) and to dynamize (*evolve*) its use, for which a closer examination of information (Tobalina-Pulido & González-Pérez, 2020) or its semantics (Vlachidis et al., 2013) is relevant, along with the promotion of open standards for data management (Richards, 2009; Huggett, 2012; Kansa, 2012).

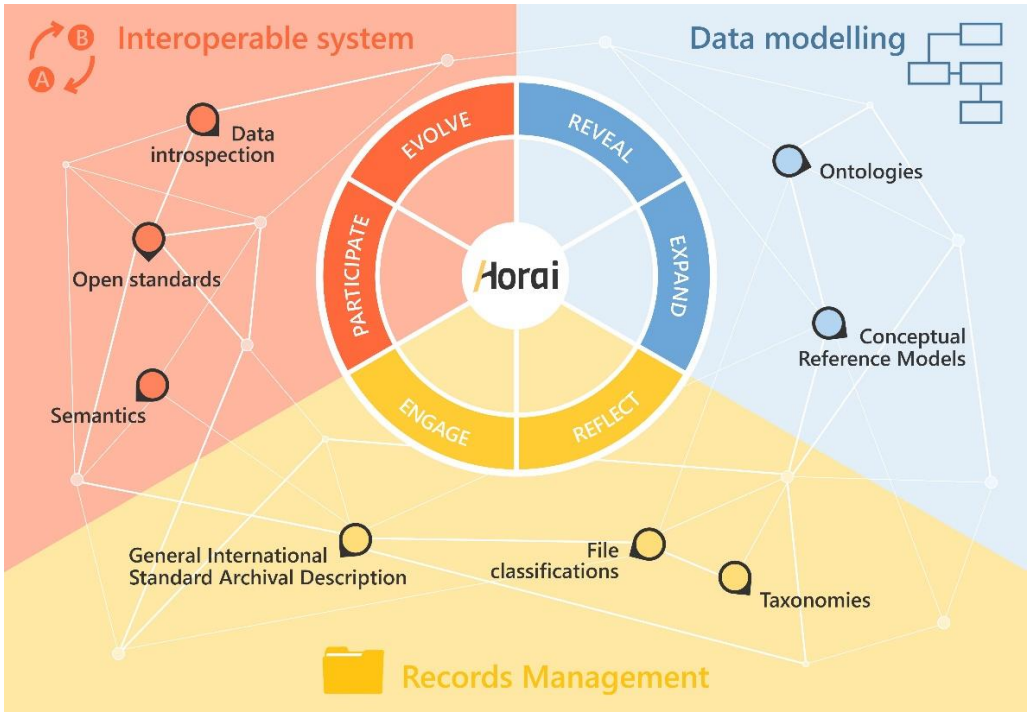

Figure 2 - Pie chart of HORAI's chances and principles in research, inspired in the proposal of J. Hugget (2022b).

Concluding remarks

HORAI allows us to reason, document, and communicate complex phenomena such as historical research and heritage management. In this endeavour involving the construction of the past, we have presented a proposal that is based on conceptual modelling as the foundation to address such complexity, while also providing mechanisms for Records Management of the elements involved in a project's development. As we have observed, both characteristics facilitate the construction of interoperable ecosystems as long as the management systems align with the model's minimum units of information. Within this framework, the research plan has brought to light issues that are related to the dynamics of data production and usage. Based on the results obtained, we conclude that the challenges that lie ahead in this context of information management are diverse, technological constraints being the most prominent. As HORAI does not depend on any specific technology, the way we interconnect each of the digital tools planned for a study development —its virtual environment— is an area that is still being refined, and we expect to obtain results in the upcoming years.

Acknowledgements

The authors wish to acknowledge Ms Noemí Travé for language edition and review. Alfred Mauri provided useful insights and inspiration during the development of this contribution. All collegues involved in the past and still today in the development of the HORAI system deserve our most sincere gratitude. Naming everyone of them singularly will add a never-ending list. We express our deep acknowledgement to them all.

Conflict of interest disclosure

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this paper.

Funding

This piece of research is an outcome of the HORAI Research Project coordinated by the company Sistemes de Gestió del Patrimoni SCCL. One of the information systems has been matured as part of the GREYWARE research project (PID2019-103896RJ-I00), funded by the Spanish Ministerio de Ciencia e Innovación (MCIN/AEI/10.13039/ 501100011033). Another has been developed thanks to a PhD grant supported by the Secretariat for Universities and Research of the Ministry of Business and Knowledge of the Government of Catalonia and the European Social Fund (2022 FI_B1 00021). The whole study is part of the current research tasks carried out by the Medieval and Postmedieval Archaeology Research Group GRAMP.UB (2021SGR-00236-GRC), at the University of Barcelona, and supported by an Archaeological Research Project (CLT/2022/ARQ001SOLC/110) developed in Martorell and co-funded by the local Town Council.

References

ADROHER, A. M. *et al.* (2016): "Registro para la cuantificación de cerámica arqueológica: estado de la cuestión y una nueva propuesta. Protocolo de Sevilla (PRCS/14)". *Zephyrvs*, 78, 87-110; doi: 10.14201/zephyrus20167887110.

ÁLVAREZ BORGE, I. (1991): *Estructura social y organización territorial en Castilla La Vieja. Los territorios entre el Arlanzón y el Duero en el siglo X*. PhD Thesis. León: Universidad de León.

ÁLVARO, K. (2012): *El poblamiento altomedieval y sus manifestaciones funerarias en la cuenca del Alto Arlanza (s. IX y XI)*. PhD Thesis. Barcelona: Universitat de Barcelona.

BEVAN, A. (2015): "The data deluge". *Antiquity*, 89(348), 1473-1484; doi: 10.15184/aqy.2015.102.

CARANDINI, A. (1997): *Historias en la tierra*. Barcelona: Editorial Crítica.

CHIPPINDALE, C. (2000): "Capta and Data: On the True Nature of Archaeological Information". *American Antiquity*, 65(4), 605-612.

COSTA, X; SANCHO, M. (2022): "La cartografía digital como herramienta dinámica e integrativa para el estudio del poblamiento medieval. La propuesta metodológica del proyecto «Muntanya Viva»". *Cuadernos de Arqueología de la Universidad de Navarra*, 30(2), 185-210; doi: 10.15581/012.30.2.009.

DE FRESNO, P. (2016): *Sistema de Información Arqueológica: propuesta de normalización, desarrollo conceptual e informático*. PhD Thesis. Vitoria: Universidad del País Vasco – Euskal Herriko Unibertsitatea.

DEL FRESNO, P.; MAURI, A.; (2020): "Una propuesta de sistematización para la gestión de la documentación e información arqueológica (SIA)", in J. L. Lerma, A. Maldonado, V. M. López-Menchero (coord.). *I Simposio Anual de Patrimonio Natural y Cultural: ICOMOS España*. Valencia: Editorial Universitat Politècnica de València, 585-593.

DEL FRESNO, P.; MAURI, A.; TRAVÉ, E. (2020): "Documentació d'intervencions de restauració en jaciments arqueològics i patrimoni edificat: cap a la construcció d'un sistema d'informació". *Documentar en la conservació-restauració. Mètodes i noves tecnologies: XVI Reunió Tècnica de conservació-Restauració. 12-13 novembre 2020*. Barcelona: CRAC, 75-90.

DEL FRESNO, P.; TRAVÉ, E.; MAURI, A. (2021): "Sistemas de Información Arqueológica: procesos de gestión integrada del patrimonio arqueológico para la investigación interdisciplinar", in M. Á. Celís (coord.). *Las Humanidades Digitales como expresión y estudio del patrimonio digital*. Cuenca: Ediciones de la Universidad de Castilla-La Mancha, 227-235.

ESCALONA, J. (1996): *Transformaciones sociales y organización del espacio en el Alfoz de Lara en la Alta Edad Media*. PhD Thesis. Madrid: Universidad Complutense de Madrid.

FRUCTUOSO, X. (2018). *Proposta de classificació terminològica per a l'elaboració d'un vocabulari arqueològic normalitzat*. Bachelor dissertation. Bellaterra: Universitat Autònoma de Barcelona.

GERO, J. M. (2007): "Honoring Ambiguity/Problematizing Certitude". *Journal of Archaeological Method and Theory*, 14, 311-327; doi: 10.1007/s10816-007-9037-1.

GONZÁLEZ-PÉREZ, C.; MARTÍN-RODILLA, P. (2014): "Integration of Archaeological Datasets through the Gradual Refinement of Models", in F. Gilingy, F. Djindjian, L. Costa, P. Moscati, S. Robert (eds.). *CAA2014: 21st*

Century Archaeology. Concepts, methods and tools. 42nd Annual Conference on Computer Applications and Quantitative Methods in Archaeology. Oxford: Archeopress, 193-204.

GONZÁLEZ-PÉREZ, C.; PARCERO-OUBIÑA, C. (2012): "A Conceptual Model for Cultural Heritage Definition and Motivation", in M. Zhou, I. Romanowska, Z. Wu, P. Xu, P. Verhagen (eds.). *CAA2011: Revive the Past. 39th Annual Conference on Computer Applications and Quantitative Methods in Archaeology.* Amsterdam: Pallas Publications, 234-244.

GRUBER, T. R. (1995): "Toward principles for the design of ontologies used for knowledge sharing". *International Journal of Human-Computer Studies*, 43(5-6), 907–928; doi: 10.1006/ijhc.1995.1081.

GUARINO, N. (1997): "Semantic matching: Formal ontological distinctions for information organization, extraction, and integration", in: M. T. Pazienza (ed.). *Information Extraction. A Multidisciplinary Approach to an Emerging Information Technology*. Berlin: Springer, 139-170; doi: 10.1007/3-540-63438-X_8.

HARRIS, E. C. (1989): *Principles of archaeological stratigraphy*. Academic Press Limited.

HUGGETT, J. (2012): "Lost in information? Ways of knowing and modes of representation in e-archaeology". *World Archaeology*, 44(4), 538-552; doi: 10.1080/00438243.2012.736274.

HUGGETT, J. (2015): "A Manifesto for an Introspective Digital Archaeology". *Open Archaeology*, 1(1); doi: 10.1515/opar-2015-0002.

HUGGETT, J. (2021): "Algorithmic Agency and Autonomy in Archaeological Practice". *Open Archaeology*, 7(1); doi: 10.1515/opar-2020-0136.

HUGGETT, J. (2022a): "Data Legacies, Epistemic Anxieties, and Digital Imaginaries in Archaeology". *Digital*, 2, 267-295; doi: 10.3390/digital2020016.

HUGGETT, J. (2022b): "Is less more? Slow Data and Datafication in Archaeology", in K. Garstki (ed.). *Critical Archaeology in the Digital Age*. California: The Cotsen Institute of Archaeology Press, 97-110.

HUVILA, I. (2022): "Improving the usefulness of research data with better paradata". *Open Information Science*, 6(1), 28-48; doi: 10.1515/opis-2022-0129.

KANSA, E. (2012): "Openness and archaeology's information ecosystem". *World Archaeology*, 44(4), 498-520; doi: 10.1080/00438243.2012.737575.

LEONELLI, S. (2014): "What difference does quantity make? On the epistemology of Big Data in biology". *Big Data & Society*, 1(1); doi: 10.1177/2053951714534395.

LIU, P.; LOUDCHER, S.; DARMONT, J.; NOÛS, C. (2021): "ArchaeoDAL: A Data Lake for Archaeological Data Management and Analytics", in B. C. Desai, J. Ullman, R. McClatchey, M. Toyoma (eds.). *25th International Database Engineering & Applications Symposium (IDEAS 2021)*. New York: Association for Computing Machinery, 252-262; doi: 10.1145/3472163.3472266.

LUCAS, G. (2001): "Destruction and the Rhetoric of Excavation". *Norwegian Archaeological Review*, 34(1), 35-46; doi: 10.1080/00293650119347.

LUCAS, G.; WITMORE, C. (2022): "Paradigm Lost: What Is a Commitment to Theory in Contemporary Archaeology?". *Norwegian Archaeological Review*, 55(1), 64-77; doi: 10.1080/00293652.2021.1986127.

MARILA, M. M. (2019): "Slow science for fast archaeology". *Current Swedish Archaeology*, 27(1), 93-114; doi: 10.37718/CSA.2019.05.

MARILA, M. M. (2022): "A Theoretically Committed Archaeology is a Civilised Archaeology". *Norwegian Archaeological Review*, 55(1), 78-80; doi: 10.1080/00293652.2021.2010123.

MARTÍN-RODILLA, P; GONZÁLEZ-PÉREZ, C. (2019): "Metainformation scenarios in Digital Humanities: Characterization and conceptual modelling strategies". *Information Systems*, 84(C), 29-48; doi: 10.1016/j.is.2019.04.009.

MARTÍN-RODILLA, P.; GONZÁLEZ-PÉREZ, C.; MAÑANA-BORRAZAS, P. (2016): "A Conceptual and Visual Proposal to Decouple Material and Interpretive Information About Stratigraphic Data", in S. Campana, R. Scopigno, G. Carpentiero, M. Cirillo (eds.). *CAA2015: Keep the revolution going. Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods In Archaeology*. Oxford: Archeopress, 201-211.

MATHIEU, J. (2021): "Is Historical Temporality "Heterogeneous" and "Contingent"? William H. Sewell's Cultural Turn". *Histories*, 1, 12-21; doi: 10.3390/histories1010005.

MAURI, A. (1995): "La aplicació del Mètode Harris a l'estudi del territori". *Actes del 3r curs d'arqueologia d'Andorra, del 30 de setembre al 4 d'octubre de 1991. La vida medieval als dos vessants del Pirineu.* Andorra: Patrimoni Cultural d'Andorra – Servei de Recerca Històrica, 8-24.

MAURI, A. (2006): *La configuració del paisatge medieval: el comtat de Barcelona fins al segle XI.* Tesis doctoral. Barcelona: Universitat de Barcelona.

MEDINA, S.; ÁLVARO, K.; TRAVÉ, E. (2022): "Integración de datos y procesos de registro de la información: un estudio de caso en el Alto Arlanza". *Cuadernos de Arqueología de la Universidad de Navarra*, 30(2), 161-184; doi: 10.15581/012.30.2.008.

MEDINA, S.; TRAVÉ, E. (2021): "Developing integrated research in Historical Science: Transdisciplinary strategies for information management and exploitation". *16th Iberian Conference on Information Systems and Technologies (CISTI)*. Portugal: 1-6; doi: 10.23919/CISTI52073.2021.9476336.

MICKEL, A; BYRD, N. (2021): "Cultivating trust, producing knowledge: The management of archaeological labour and the making of a discipline". *History of the Human Sciences*, 35(2), 1-26; doi: 10.1177/09526951211015855.

MORELAND, J. (2006): "Archaeology and Texts: Subservience or Enlightenment". *Annual Review of Anthropology*, 35, 135-151; doi: 10.1146/annurev.anthro.35.081705.123132.

NICCOLUCCI, F. (2020): "From Digital Archaeology to Data-Centric Archaeological Research". *magazén*, 1(1), 35-54; doi: 10.30687/mag//2020/01/002.

OPGENHAFFEN, L. (2022): "Archives in action. The impact of digital technology on archaeological recording strategies and ensuing open research archives". *Digital Applications in Archaeology and Cultural Heritage*, 27, e00231; doi: 10.1016/j.daach.2022.e00231.

ORMAN, L. V. (2015): "Information Paradox: Drowning in Information, Starving for Knowledge". *IEEE Technology and Society Magazine*, 4, 63-73; doi: 10.1109/MTS.2015.2494359.OWENS, T. (2011). "Defining Data for Humanists: Text, Artifact, Information or Evidence?". *Journal of Digital Humanities*, 1(1).

PADILLA J. I. (1984): "Contribución al estudio de las cerámicas grises catalanas de época medieval: El taller, los hornos y la producción de Casampons". *Acta historica et archaeologica mediaevalia*, 2, 99-143.

PASTOR DÍAZ, E. (1996): *Castilla en el tránsito de la Antigüedad al Feudalismo. Poblamiento, poder político y estructura social: del Arlanza al Duero (siglos VII-XI)*. Valladolid: Junta de Castilla y León.

RĄCZKOWSKI W. (2020): "Power and/or Penury of Visualizations: Some Thoughts on Remote Sensing Data and Products in Archaeology". *Remote Sensing*, 12(18), 2996; doi: 10.3390/rs12182996.

RICHARDS, J. D. (2009): "From anarchy to good practice: the evolution of standards in archaeological computing". *Archeologia e Calcolatori*, 20, 27-35.

SANDOVAL, G. (2021): "Single-Context Recording, Field Interpretation and Reflexivity: An Analysis of Primary Data In Context Sheets". *Journal of Field Archaeology*, 46(7), 496-512; doi: 10.1080/00934690.2021.1926700.

SCHÖCH, C. (2013): "Big? Smart? Clean? Messy? Data in the Humanities". *Journal of Digital Humanities*, 2(3).

SHOORCHEH, M. (2019): "On the spatiality of geographic knowledge". *Asian Geographer*, 36(1), 63-80; doi: 10.1080/10225706.2018.1463854.

STANČIĆ, H.; BRALIĆ, V. (2021): "Digital Archives Relying on Blockchain: Overcoming the Limitations of Data Immutability". *Computers*, 10(91); doi: 10.3390/computers10080091.

THIBODEAU, K. (2019): "The Construction of the Past: Towards a Theory for Knowing the Past". *Information* 10(11), 332; doi: 10.3390/info10110332.

THIBODEAU, K. (2021): "Discerning Meaning and Producing Information: Semiosis in Knowing the Past". *Information* 12(9), 363; doi: 10.3390/info12090363.

TOBALINA-PULIDO, L.; GONZÁLEZ-PÉREZ, C. (2020): "Valoración de la calidad de los datos arqueológicos a través de la gestión de su vaguedad. Aplicación al estudio del poblamiento tardorromano". *Complutum*, 31(2), 343-360; doi:10.5209/cmpl.72488.

TOPOLSKI, J. (1992): *Metodología de la Historia*. Madrid: Cátedra.

TRAVÉ, E. (2009): *Producció i distribució d'una terrisseria medieval: Cabrera d'Anoia*. PhD Thesis. Barcelona: Universitat de Barcelona.

TRAVÉ, E. (2021): "Building traditional craftsmanship. Some thoughts about endurance and change". *Academia Letters*, 4309; doi: 10.20935/AL4309.

TRAVÉ, E. (2022): "Statistical Analysis of Morphometric Data for Pottery Formal Classification: Variables, Procedures, and Digital Experiences of Medieval and Postmedieval Greyware Clustering in Catalonia (Twelfth-Nineteenth Centuries AD)". *Open Archaeology*, 8(1), 1269-1285; doi: 10.1515/opar-2022-0269.

TRAVÉ, E.; DEL FRESNO, P.; MAURI, A. (2020): “Ontology-Mediated Historical Data Modeling: Theoretical and Practical Tools for an Integrated Construction of the Past”. *Information*, 10(4), 182; doi: 10.3390/info11040182.

TRAVÉ; E.; DEL FRESNO, P.; MAURI, A.; MEDINA, S. (2021a): “The Semantics of History. Interdisciplinary Categories and Methods for Digital Historical Research”. *International Journal of Interactive Multimedia and Artificial Intelligence*, 6(5), 47-56; doi: 10.9781/ijimai.2021.02.002.

TRAVÉ, E.; MEDINA, S.; DEL FRESNO, P.; VICENS, J.; MAURI, A. (2021b): "Towards an Ontology-Driven Information System for Archaeological Pottery Studies: The Greyware Experience". *Applied Sciences*,11(17), 7989; doi: 10.3390/app11177989.

TRAVÉ, E.; MEDINA, S. (2021): “Explotación de documentación contable para el estudio de las sociedades mercantiles en la Baja Edad Media: Ontología del sistema FENIX y gestión integrada de Big Data”, in M. Á. Celís (coord.). *Las Humanidades Digitales como expresión y estudio del patrimonio digital*. Cuenca: Ediciones de la Universidad de Castilla-La Mancha, 295-272.

TRAVÉ, E.; QUINN, P. S.; LÓPEZ, M. D.; PADILLA, J. I. (2014): “One hundred sherds of grey: compositional and technological characterization of Medieval greyware pottery production at Cabrera d’Anoia, Catalonia, Spain”. *Archaeological and Anthropological Science*, 6, 397-410; doi: 10.1007/s12520-014-0179-2.

TRAVÉ, E.; VICENS, J. (2018): “Terrissa negra i canvi social: pervivències i transformacions en l’ús de ceràmica de cocció reductora a Osona i les comarques gironines a partir de la recerca arqueològica i documental”. *Ausa*, 28, 829-850.

USCHOLD, M.; GRUNINGER, M. (1996): “Ontologies: Principles, methods and applications”. *The Knowledge Engineering Review*, 11(2), 93-136; doi: 10.1017/S0269888900007797.

VAN HELDEN, D.; HONG, Y.; ALLISON, P. (2018): “Building an Ontology of Tablewares using 'Legacy Data'”. *Internet Archaeology*, 50; doi: 10.11141/ia.50.13.

VICENS, J.; TRAVÉ, E. (2018): “La terrissa popular de Josep Escortell i Cerqueda: la tipologia de Quart”. *Estudis del Baix Empordà*, 37, 97-129.

VLACHIDIS, A.; BINDING, C.; MAY, K.; TUDHOPE, D. (2013): “Automatic Metadata Generation in an Archaeological Digital Library: Semantic Annotation of Grey Literature”, in: A. Przepiórkowski, M. Piasecki, K. Jassem, P. Fuglewicz (eds.). *Computational Linguistics*. *Applications.* Berlin: Springer, 187-202; doi: 10.1007/978-3-642-34399-5_10.

WYLIE, A. (2017): “How Archaeological Evidence Bites Back: Strategies for Putting Old Data to Work in New Ways”. *Science, Technology, & Human Values*, 42(2), 203–225; doi: 10.1177/0162243916671200.

WOOLF, A. (2009): “A dialogue of the deaf and the dumb: Archaeology, history and philology”, in Z. Delvin, C. Holas-Clark (eds.). *Approaching Interdisciplinarity: Archaeology, History and the Study of Early Medieval Britain, c. 400-1100*. Archaeolopress, 3-9.