



HAL
open science

From Calculations to Reasoning: History, Trends and the Potential of Computational Ethnography and Computational Social Anthropology

Manolis Peponakis, Sarantos Kapidakis, Martin Doerr, Eirini Tountasaki

► **To cite this version:**

Manolis Peponakis, Sarantos Kapidakis, Martin Doerr, Eirini Tountasaki. From Calculations to Reasoning: History, Trends and the Potential of Computational Ethnography and Computational Social Anthropology. *Social Science Computer Review*, 2023, pp.1-20. 10.1177/08944393231167692 . hal-04072818

HAL Id: hal-04072818

<https://hal.science/hal-04072818v1>

Submitted on 18 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

From calculations to reasoning: history, trends, and the potential of Computational Ethnography and Computational Social Anthropology

Manolis Peponakis ; Sarantos Kapidakis ; Martin Doerr ; Eirini Tountasaki

Abstract: The domains of *computational social anthropology* and *computational ethnography* refer to the computational processing or computational modelling of data for anthropological or ethnographic research. In this context, the article surveys the use of computational methods regarding the production and the representation of knowledge. The ultimate goal of the study is to highlight the significance of modelling ethnographic data and anthropological knowledge by harnessing the potential of the semantic web. The first objective was to review the use of computational methods in anthropological research focusing on the last 25 years, while the second objective was to explore the potential of the semantic web focusing on existing technologies for ontological representation. For these purposes, the study explores the use of computers in anthropology regarding data processing and data modelling for more effective data processing. The survey reveals that there is an ongoing transition from the instrumentalisation of computers as tools for calculations, to the implementation of information science methodologies for analysis, deduction, knowledge representation, and reasoning, as part of the research process in social anthropology. Finally, it is highlighted that the ecosystem of the semantic web does not subserve quantification and metrics but introduces a new conceptualisation for addressing and meeting research questions in anthropology.

Keywords: computational anthropology, computational ethnography, ethnographic data, semantic web, knowledge representation, data modelling, ontology

This is the accepted for publication version of the article:

Peponakis, M., Kapidakis, S., Doerr, M., & Tountasaki, E. (2023). From calculations to reasoning: History, trends and the potential of Computational Ethnography and Computational Social Anthropology. *Social Science Computer Review*.

<https://doi.org/10.1177/08944393231167692>



This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

1 Introduction

At the beginning of the 21st century, Information and Communication Technologies (ICT) have been integrated into everyday life, especially in the Western world, and they have significantly affected human activities. The plethora of available devices, in conjunction with broadband technology and cloud computing, has been ongoingly transforming both the means through which reality is perceived and the way people interact with it.

More specifically, in the domain of social anthropology, the revolutions brought by computer science had a substantial, twofold impact. On one side, they have affected the way anthropologists study technology as a cultural process. On the other side, they have affected the way technology is used as a tool for anthropological studies. Hakken expresses the situation eloquently: *One must [...] distinguish anthropologists who study computing as a cultural process [...] from those who are mostly interested in computers as a tool or computing as a methodology in anthropology* (Hakken, 1993, p. 112). The second perspective is of interest to this study.

Previous studies reveal that the dominant medium for producing and representing ethnography and anthropology has been natural language. Mueller notices that ethnography, both as a process and as a product, is firmly verbal and scriptocentric (Mueller, 2016, p. 99), while Fischer adds that script is still the preferred carrier for representing ethnographic research, mostly in the form of a narrative (M. D. Fischer, 2006, p. 4). This study discusses alternatives based on current ICT infrastructures, regarding both the processes of anthropological studies as well as their outcome.

To further elaborate, the domains of interest here are *computational social anthropology* and *computational ethnography* which refer to the computational processing or computational modelling of data for anthropological or ethnographic research. In this context, the authors survey the use of computational methods regarding the production and the representation of knowledge. During literature review, the quest for a precise definition of computational anthropology and computational ethnography has put in the spotlight varying perspectives which are presented in detail in section 4.1. The survey reveals an ongoing transition from the instrumentalisation of computers as tools for calculations to the implementation of information science methodologies for analysis, deduction, knowledge representation, and reasoning, as part of the research process in social anthropology.

In the following sections, the history of using computers in the field of anthropology is deployed, and the potential of using computational processing is highlighted, viewed as the result of “computational thinking” according to Wing (Wing, 2006). This means that the computational processing discussed here is not perceived just as a programming process but as a way of conceptualising; it is not about introducing thinking machines, but about using an alternative way to represent how people think while addressing and solving problems. Seaver, an anthropologist who claims that algorithms are inherently cultural, states that social scientists are not distant observers but active participants who produce algorithms through their research (Seaver, 2017, p. 5).

2 Aims and objectives

After reviewing existing definitions regarding what *computational* stands for in *computational social anthropology* and *computational ethnography*, the main aim of this study is to explore the intersection of computer science with these two domains. For this purpose, we demonstrate the transition from computers as tools for calculations to information science methodologies

for knowledge representation and reasoning. For fulfilling this aim, two objectives are identified. First, the study explores how anthropologists have been using computers to conduct their research over the last 25 years. Second, the study underlines the potential arising from the dynamics of the contemporary technological context, i.e. the semantic web, to model data toward data re-use and knowledge representation.

For meeting the first goal, we have overviewed and documented the related bibliography, and the types of software used by ethnographers and social anthropologists for conducting research. The effort was not towards creating an inventory of software solutions or compiling an exhaustive bibliography on the subject. Instead, this study aimed to outlining the wider context of computer science and the applied methodologies in a way that could support the implementation of these methodologies for processing and analysing anthropological data, i.e. report on generic functionalities of specific types of software and not on how specific software processes data.

For meeting the second goal, and after having reviewed related literature, the article presents data structures supported by certain software types, focusing on the expressive power of these structures. In this context, the structures mentioned are suitable for knowledge representation, since they are more appropriate both for qualitative analysis and data reuse. Regarding knowledge representation, this study presents efforts that intend to the codification and representation of sociocultural data, even if they are not applied by anthropologists. Besides, as section 6.2.3 reveals, few efforts can be attributed to them.

3 Methodology

As already mentioned, this study comprises an extensive literature review as well as a review of certain software solutions, ontologies and web services. For the analysis of the collected resources (i.e. studies, software and relevant data structures, web services, ontologies and datasets)¹ we have built on the fundamentals of grounded theory, which is based on inductive reasoning where any research question(s) occur during analysing available data. Eventually the analysis leads to conclusions. In our study, the collected resources were analysed using the method of meta-synthesis (Walsh & Downe, 2005). Abiding by the meta-synthesis methodology and having an interpretive rather than aggregating approach, we have integrated results from different but interrelated resources. The method of meta-synthesis deploys in six steps: definition of the research question, selection of relevant studies, assessment of their quality, extraction of relevant information, information analysis, and production of the critical interpretive synthesis.

To perform the second step (and ancillary to step three), i.e. to locate the relevant resources and assess their quality, the search techniques aimed on retrieving the maximum number of relevant records, i.e. keep a high recall rate while precision was considered of minor importance. During the initial phase of bibliographic research, variant searches were performed in certain databases like the Web of Science, Scopus, and Google Scholar. Additionally, alert services were activated for continuous and systematic monitoring of new publications. The software and web services were detected either by following references mentioned in the bibliography or by searching in specialized software repositories like Git-Hub.

¹ We must note that 'ready to use' databases and datasets were out of the scope of this study. Therefore, services like Human Relations Area Files (HRAF) <https://hraf.yale.edu/> or D-place <https://d-place.org/> (despite their crucial role in comparative anthropology) are not discussed here.

4 Background analysis

While introducing the readers to the overall context of what this study discusses, in this section we address two issues in the following subsections. The first subsection presents definitions for the domains of computational anthropology and ethnography, before concluding to the definition finally adopted for the purposes of this study. As it becomes evident, the crucial factor in defining what exactly *computational* refers to when associated with anthropology and ethnography, is the concept of data. Therefore, the second subsection discusses about what is considered raw and what is considered processed data.

4.1 Contradicting definitions for computational anthropology and computational ethnography

In this subsection, indicative definitions are deployed regarding the domains of computational anthropology and computational ethnography. The definitions are derived from varied viewpoints and serve as the basis for eventually adopting the definition which best serves our perspective and purposes.

Bharwani underlines that anthropologists are, in general, hesitant in engaging themselves with multidisciplinary methods; and they are less willing to combine these methods with quantitative and formal ethnographic approaches (Bharwani, 2006, p. 79). Nevertheless, some anthropologists deal with formalisms and are processing them using computers. In addition, there are many researchers from other domains, like the domain of computer science, who are involved with anthropological data and their computational analysis. This scientific pluralism raises more barriers in terms of defining these emerging fields, making it especially difficult to agree on the semantics of the adjective “computational” as part of the terms “computational anthropology” and “computational ethnography”.

In the collective work “21st Century Anthropology: A Reference Handbook”, Artmann brings together artificial intelligence and computational anthropology. He advocates that when artificial intelligence manages to describe human cognition in the form of computationally processable digital information, then we can talk about computational anthropology (Artmann, 2010). In another collective work, “The Routledge companion to digital ethnography” (Hjorth et al., 2017), Beaulieu sees computational ethnography as Hine sees virtual ethnography, which is still considered ethnography but performed in the virtual digital world. Beaulieu considers that, in this case, the status of ethnographers becomes problematic regarding their role as a tool in the ethnographic research, because two essential factors of implementing traditional ethnography are missing: presence and engagement (Beaulieu, 2017). Recently, Brooker, who uses interchangeably the terms *digital* and *computational*, saw computational ethnography as the meeting point of anthropology and sociology (Brooker, 2022). Munk & Winthereik, also use interchangeably the terms digital and computational but distinguish *digital* as a field from *digital* as a method (Munk & Winthereik, 2022). While Seo et al define a field by the name “computational ethnographic research”, which is comprised of virtual ethnography, quantitative ethnography, and computational ethnography. They do not define, though, in detail what computational ethnography is, but they consider it as a method that encompasses -among others- social network analysis, natural language processing, and data visualisation (Seo et al., 2022).

Different perspectives, regarding the definition of computational ethnography, are expressed by researchers in the computer science domain. For example, Zheng et al relate the definition of computational ethnography to the collection of data with automated and electronic means (Zheng et al., 2015, p. 114). The same approach is followed by Tallyn et al who examine

the potential and the challenges of collecting data using a chatbot without the involvement of an ethnographer (Tallyn et al., 2018). While Arnold and Fuller believe that computational ethnography is about using big datasets to provide insight into users' habits for redesigning (Arnold & Fuller, 2018).

After examining varying perspectives and proposed definitions of computational anthropology and ethnography, it is obvious that there is a different starting point which mandates the course of forming a definition. In some cases, the starting point is the data acquisition method, while in other cases it is about the processing of the acquired data. For our analysis, we define that the domains of *computational social anthropology* and *computational ethnography* are about the computational processing of data or the computational modeling of data in a way suitable for conducting anthropological or ethnographic research. Consequently, it is not about doing ethnography in the digital environment, which is often called *digital ethnography*; nor about disseminating a questionnaire digitally or studying the digital trail of a person whose location has been tracked using a mobile device. It is primarily about whether the data are going to be computationally processed or computationally modelled for anthropological analysis. Respectively, modelling the data acquisition method using computational principles is also considered as computational ethnography. Following this approach, the object of these emerging fields is placed under the broader class of computational social science, where computers are used to model, simulate, analyse, and represent social phenomena. At this point, it is worth mentioning that Fischer, despite his long-term and ongoing research in the field of computational anthropology, refrains from explicitly using this term. Instead, he refers to computational anthropologists who do not just create tools for facilitating the research process, but who indulge in computer-based methods as part of the research process (M. D. Fischer, 2006, p. 12).

Finally, the wider scientific field, within which computational anthropology and ethnography are signified, needs to be framed. Although it is evident that this action calls for an interdisciplinary approach not restricted by the boundaries of disciplinary jurisdiction, we agree with Cioffi who claims that computational social science is a social science, in a way that any computational X-science is part of X-science, i.e. computational astronomy, computational biology, and computational linguistics are part of astronomy, biology, and linguistics, respectively (Cioffi-Revilla, 2016, p. 3).

4.2 Types of data in anthropological analysis: raw and processed

It has already been discussed that the adjective *computational* -when associated with the field of ethnography and anthropology- does not refer to the way data are acquired but to the way data are processed or modelled. In anthropological research, the question arising is what type of data can be computationally processed. Bartlett et al distinguish raw data and processed data on a sophisticated basis, classifying data into two basic categories, namely “crafted” and “found”. “Crafted” data are produced by the scientific community that uses them to document the produced knowledge. In other words, they are data produced within a framework defined by the scholars of certain scientific domain and are signified within this framework. On the other hand, the existence of “found” data is irrespective of goals or any prior plan by the community that analyses them, like in the case of studying transaction log files (Bartlett et al., 2018, p. 4). The previous distinction clearly suggests that “crafted” data produced within a certain scientific field may be considered “found” data for another.

According to a different viewpoint, Pool highlights the hybrid nature of ethnographic data and classifies them as “soft” and “hard”, so that verbatim transcripts are considered “hard data” that can be archived and analysed anew, while memories and impressions are considered

“soft data” that can hardly be the subject of any further analysis. And he makes an interesting remark: hard ethnographic data are partial and personal, making simple checking of the basic ‘facts’ feasible, but any re-use or re-interpretation of findings difficult in the absence of the soft component (Pool, 2017).

Although each data category may be prone to different ways of processing, yet, the category does not mandate the final choice of their processing method. Like in the case of myths, when considered as data and become the subject of research; Levi-Strauss consulted them to draw conclusions, while Weingart used software for counting word occurrences to empirically confirm research hypotheses proposed earlier by feminist scholars (Weingart & Jorgensen, 2013). Stubbersfield and Tehrani followed a similar approach to Weingart’s for studying the myth of “Bloody Mary” (Stubbersfield & Tehrani, 2013).

5 Data processing and deduction

In the previous section, we clarified that processable data may derive either from raw data or from a wide range of “crafted” data, the latter of which may derive from “mainstream” anthropological resources. In this section we explore the ways of processing these data, taking into consideration that, in the digital era, data are everywhere. What is of interest to us is Anthropology by Data Science as defined by Paff, meaning using data science tools to conduct anthropological work (Paff, 2022). In this context, we distinguish two basic approaches. One of them is about the use of computational methods for assisting the task of data analysis, often by using a solution from the computer assisted qualitative data analysis software (CAQDAS) category. The other approach is about direct deduction from data (reasoning), which is the first substantial step toward data science and artificial intelligence. Both approaches are open to different types of data visualisation, which may unlock new perspectives of data interpretation as discussed later in this article.

5.1 Computer-assisted analysis of data: statistics, patterns, and tagging

As introduced right above, this subsection covers data processing concerning data analysis for producing or documenting anthropological knowledge. The approach deployed here may, initially, lead the reader to assume that the study refers to methods of quantitative analysis. The attributes “quantitative” and “qualitative” are, indeed, found in this text, assigned to respective methods of research analyses, since they comprise widely accepted research methods. But we believe that such a distinction should be approached with scepticism, corroborated by the position of Behrens, who declares that such a distinction is not inherent in anthropological data; instead, he finds that the categorization was articulated based on former methodologies that required classification of anthropological data into convenient conceptual categories (Behrens, 1990, p. 305). Similarly, in the introduction of the collective work “Ethnography for a data-saturated world”, the authors mention that quantitative and qualitative knowledge is not inherently separate, but the distinction between the two has been a longstanding Western cultural cleavage that has had the effect of separating them; this approach is addressed and documented in the individual studies of this collective work (Knox & Nafus, 2018).

A typical example of anthropologists using metrics can be found in the classical work “Culture as Consensus: A Theory of Culture and Informant Accuracy” (Romney et al., 1986) where a formalistic mathematical model is introduced: culture is defined as the consensus and can, therefore, be quantified. The creators of the model encourage researchers to opt for the probabilistic character of the informants’ statements based on the understanding that some

informants could not be considered as sources of valid and reliable information. Because of this and to ensure a more reliable outcome, they define reliability weights of each informant. Such factors can efficiently be handled by current software solutions (Purzycki & Jamieson-Lane, 2017).

An important aspect of the ethnographic fieldwork is the recording and management, by computational means, of the fieldnotes. Various software tools have been developed to assist recording and handling this material. Properly coding and enriching of the initial data with suitable tags, attributes etc could create what Albris et al call ‘computational fieldnotes’ (Albris et al., 2021).

One of the most basic uses of computer science in the context of social sciences and the humanities, is in pattern recognition. This process is about locating regularities in data by implementing statistical analysis. The analysis of identified patterns and their metrics are at the centre of what is called “Computational Grounded Theory” (Nelson, 2017), (Carlsen & Ralund, 2022). Seedbed for such implementations is the arising domain of big data, but anthropologists seem quite hesitant about adopting data science methods. Nevertheless, there are few voices suggesting that anthropology should come closer to big data (Beuving, 2020).

It is widely accepted that managing large volumes of data is a fruitful ground for computational processing. The case of the Google Books project proves the statement (Michel et al., 2011). Involved stakeholders explained the project’s rationale during a TEDx Talk: millions of books store the world’s knowledge and it would be an extraordinary experience to be able to read it all, yet rather impractical for a human. On the other hand, “reading” these books through computational methods would be a very realistic scenario, and equally extraordinary (TEDx Talks, 2011). In the context of digital humanities, Moretti calls the process of non-human reading “distant reading” (Moretti, 2013).

Distant reading was also used for ethnographic data. As already mentioned in section 4.2, Weingart counted certain word occurrences to empirically confirm research hypotheses proposed earlier by feminist scholars (Weingart & Jorgensen, 2013) and so did Stubbersfield and Tehrani for studying the myth of “Bloody Mary” (Stubbersfield & Tehrani, 2013). On the other hand, Bakharia and Corrin have recommended more sophisticated methods for natural language processing (NLP), emphasising on the impact it can have for the analysis of a dataset when applying semantic search versus simple string matching (Bakharia & Corrin, 2019).

Beyond analysing texts, pattern recognition applies to other formats as well. Wilf, in her paper “Toward an Anthropology of Computer-Mediated, Algorithmic Forms of Sociality”, reports on how music styles are transformed into statistical patterns (Wilf, 2013). And Leetaru et al analysed 959,000 scholarly papers about Africa to highlight geographical distribution, study the Nuer people, spot influential authors, etc (Leetaru et al., 2014). Interesting is the work of Alvard and Carlson, published in *Current Anthropology*, one of the most respected journals in the field of anthropology, where they analysed geo-location data collected from a Global Positioning System (GPS) of fishermen (Artisanal Fishers) at the Commonwealth of Dominica. They have concluded that computational processing methods for analysing the increasing volume of big datasets produced may reduce the cost of analysis and save time toward finding answers to anthropological questions which would, otherwise, be very difficult to address (Alvard & Carlson, 2020). Under the same spirit, Sosna et al test how the graph theory could unveil burial customs and they conclude that the computational approach is an asset for revealing patterns that would be challenging for human reasoning to reveal (Sosna et al., 2013, p. 57).

5.2 Direct deduction from data

The efforts presented in the previous subsection deal with the semi-automated processes, i.e. human reasoning assisted by computational data processing. But computers can perform at a more abstract level than that. The most indicative example of this type of reasoning is the use of a logic programming language named Prolog, which is a declarative programming language based on first-order predicate logic. The language logic is expressed in terms of relations, represented as facts and rules.

Prolog knowledge bases comprise facts and rules for reasoning about these facts, so they can be queried to deduce new facts or to check the consistency of the existing ones. In Prolog's terminology, the facts represent statements about reality while the rules may be used as the 'inference mechanism' that allows the inference engine to answer the queries. This makes Prolog fit for the representation and manipulation of relationships and structures (M. D. Fischer, 1994, p. 9). A simple but indicative example is the following: if X is the father of Y (fact), and fathers are males (rule), then the system could deduce that X is male, without any explicit declaration of the gender of X. The attribution of X's gender is the result of a reasoning process.

Prolog has been applied in many cases for analysing ethnographic data, especially kinship and genealogical data, but, in any case, research questions in anthropology do not revolve around the identification of a father's gender nor the location of the lost cousin! Anthropologists working on kinship had to develop ad hoc software since commercial solutions were insufficient, in a sense that existing kinship models only considered a narrowly defined nuclear family model as the reference entity (Lyon & Magliveras, 2006, p. 31). One of the earliest and most profound tools for algorithmic analysis of kinship, namely the Kinship Algebra Expert System (KAES), was built by Read using Prolog (D. W. Read & Behrens, 1990). The program was rewritten by Read and Fischer, ending up with the Kinship Algebra Modeler (KAM) (D. Read et al., 2013). Read mentions that the use of this software provided a framework for potentially constructing a generative algebraic model that relates properties and structure of kinship terminologies to an underlying logic which the KAES program helps uncover and model as a generative structure (D. W. Read, 2006, p. 43). The underlying logic of KAES could also be used for the construction of ontologies, in a way that human judgements may be used to construct a logic that predicts the remaining relationships and interactions within a wholly symbolic system. As we point in the section 6.2.3, KAES's algebraic structures were actually used to formalize kinship relationships as definable relations, for constructing a context specific ontology.

5.3 Data visualisation

Whether using computer-assisted methods for analysis or methods for direct deduction from data, there is no escape from the hermeneutics of interpretation, although these techniques may minimize the existence -and sometimes practice- of arbitrariness (Abramson et al., 2018, p. 259). But these methods provide an additional advantage: various types of visualisation are possible, both for data and for their analysis.

Visualisation types offer both a friendlier way of presenting data and sometimes a more neutral -less culturally biased- way for the transmission of meaning, like in the kinship domain. Read et al note that the best way to capture the conceptual structure of systems of kinship definitions, without using one's cultural conceptions as an obscuring filter, is usually to ask for diagrams, not lists. Comparing the English kinship system with the one of Punjabi, Read et al. state that the most important is the point of most apparent contrast. The shapes are not the same. English looks like a Christmas tree and Punjabi looks like a butterfly. Therefore, they consider

that, in essence, kinship terminologies are not like sets of labels but like geometries. They are not collections of lexemes but systems of ideas (D. Read et al., 2013).

In the context of anthropological research, another example of benefits from data visualisation is given by Abramson et al by using patterns for what they call “ethnographic heatmaps (ethnoarrays)” (Abramson et al., 2018). The emphasis given on the role of patterns for visualising ethnographic data, for example through charts, is deployed in their work *Beyond Text: Using Arrays to Represent and Analyze Ethnographic Data* (Abramson & Dohan, 2015). Yet, despite the recorded advantages of data visualisation, there are only few, segmental, efforts. Unfortunately, anthropologists make very little use of diagrams for modelling the social processes they study (Carlson & Anderson, 2007, p. 642).

6 Modelling the data and the knowledge

Geertz, in his classical work “The Interpretation of Cultures”, states that the ultimate goal of anthropology is the construction of a consultable record of what has ever been said in the human history (Geertz, 1973, p. 30), while Marcus wonders whether anthropological knowledge could be accumulated in the form of an archive (Marcus, 1998). To sketch a potential answer, this section explores ways to model Geertz’s “consultable record” or Marcus’s “archive”, to make possible the processing and re-use of this record in the digital environment beyond its textual form.

Based on Sahlins’s statement that “ethnography is anthropology, or it is nothing” (Sahlins, 2002, p. 12), we suggest that the ethnographer’s knowledge has to exist beyond the ethnographer and should be interoperable with knowledge created by others; unless it does so, it is nothing. If researchers are not able to share what they observed, then, in the context of science, this observation never existed. If data and knowledge are modelled in a formalistic and precise description, they can be disseminated and can interoperate efficiently with other data. Even more, if these models are compatible with current digital infrastructures, it is highly likely that this will escalate the impact of anthropological knowledge both toward inter-disciplinary scientific endeavours and to a wider audience.

6.1 Software-agnostic data

Modelling data, which is extremely important for anthropological research, is a prerequisite for effectively disseminating information. As Van Der Leeuw notes, data models allow researchers to describe a wide spectrum of relations with such a precision that cannot be achieved by using natural language, which is the usual tool for describing them (Van Der Leeuw, 2004, p. 122). Before proceeding to the deployment of certain issues regarding data modelling, it is essential to present another significant role that information modelling holds: it contributes to distinguishing and detaching data from the software that manages them. Due to the rapidly changing ICT infrastructures for disseminating information, any software’s life-cycle is short-termed. Therefore, for the long-term preservation and sustainability of data, it is necessary to detach it from specific software solutions, meaning that software-agnostic data must be created. Focusing on data modelling can lead the developments of the domain.

In this process, interoperability is a key issue. Interoperability is defined as the ability of a system to effectively communicate with other systems. Usually, it is implemented in two basic levels: syntactic and semantic². Syntactic interoperability regards data modelling in terms of

² Several researchers suggest another, more abstract, level of interoperability, i.e. “pragmatic interoperability”. This level requires the other two levels, and it is mainly about the intended effect

the transmission of information and it is achieved if two systems use common data formats and communication protocols. The other level, semantic interoperability, is about information systems exchanging data without ambiguousness, which constitutes one of the greater challenges in the field of data integration.

Many anthropologists keep their distance from these developments (Pels et al., 2018), yet, transparency, data sharing, and data re-use are vital issues for producing reliable results; and interoperability is a significant condition in this context. Data re-use signifies that available data are used by another researcher who may have different research questions to answer; re-analysing and re-interpreting data are known as integral parts of the research process. In these terms, the value of re-using ethnographic data is high, even though it is challenging to achieve it (M. D. Fischer & Ember, 2018, p. 329). At this point, we must clarify that modelling data for allowing their re-use does not mandate data sharing.

Currently, the modelling framework for encoding interoperable data is the ecosystem of the semantic web as it is implemented through linked data. In this context, we explore the role of ontologies as a fundamental interoperable tool of the semantic web for modelling anthropological knowledge.

6.2 Knowledge Representation and ontologies

The following text deals with the notion of knowledge representation, ontologies, knowledge graphs, and what they have to offer within the aforementioned framework. There is also an overview of the process for developing an ontology and an illustration of the aspect of ontologies as cultural constructions. Finally, the study considers several projects which build on the ontological representation of ethnographic data.

Knowledge representation is a field of artificial intelligence focusing on the encoding of information for computer use so that information can be processed for performing complicated functions and problem-solving. For this purpose, three components are necessary: syntax, semantics, and reasoning. The syntax allows for well-formed statements in a way similar to the syntax of natural languages; semantics control the transmission of unambiguous meaning, and reasoning is the data analysis process for inference.

Knowledge representation is central to the core interests of social anthropology (M. D. Fischer, 1994, p. 3). Acknowledging that the subject matter of anthropology is intrinsically messy (Fuentes & Wiessner, 2016, p. S3), formal representations for constructing descriptions, allow us to overcome many barriers toward clearly representing social knowledge. Such barriers bring to the surface -if not resolving them- ambiguous statements and validate a description in terms of its completeness or consistency (M. D. Fischer & Finkelstein, 1991, p. 120).

The basic mechanism for knowledge representation in the semantic web is ontologies. Gruber, the pioneer in ontological representation, defines an ontology as “an explicit specification of a conceptualization. The term borrowed from philosophy, where Ontology is a systematic account of Existence. For AI systems, what ‘exists’ is that can be represented” (Gruber, 1995, p. 908). From this perspective, the ontologist does not question reality but its representation (Kohne, 2014, p. 88).

According to Gruber, ontologies are the required technological means for actually developing the semantic web, since they are suitable for clearly declaring the intended semantics of the terms used in data (Gruber, 2007). Ribes and Bowker highlight another aspect:

during systems’ communication (Asuncion & van Sinderen, 2010; Pokraev et al., 2005); Therefore, this level is quite generic and deals with the interpretation of models and strategic planning.

ontologies can be perceived as a kind of semantic gateway technology that facilitates communication and coordination across scientific domains, languages, classes, or database schemas, which are fundamental in cyber-infrastructure ventures. Otherwise, data and resources could not be moved freely across the institutions of science (Ribes & Bowker, 2009). In essence, ontologies contribute to dealing with the problem Fischer has illustrated more than twenty-five years ago, that computer uses for analysing complex situations are mostly ad-hoc constructions, generally hand-crafted for each application (M. D. Fischer, 1994, p. 10). With the advent of ontologies, and by achieving interoperability, the problem is not about applications anymore; the problem is, now, placed at a more generic level.

6.2.1 Developing an ontology

Representing the knowledge of a researcher as an ontology requires the implementation of a modelling method. Ribes and Bowker used ethnographic methods for keeping up with the development of the GEON ontology, which manages geographic data. They have concluded that researchers who engage with knowledge representation in any field must educate themselves in expressing the acquired knowledge in the coding language, often called *First Order Logic*, which is used for knowledge representation and is represented through the Ontology Web Language (OWL). To use this language, researchers need to be able to express their knowledge using entities, operators, and logical relations (Ribes & Bowker, 2009), which is a sophisticated process, requiring special skills and interdisciplinary knowledge. What needs to be defined and expressed is not always obvious or unambiguous, even for the expert researcher of the domain; so, its development requires the contribution from various experts to consolidate it into a formal logic (Doerr & Iorizzo, 2008, p. 3). The fact that concepts found in ontologies are sometimes not contained in specialized terminological dictionaries advocates for the aforementioned observation, meaning how far from obvious are the required classes for modelling a domain (L'Homme & Bernier-Colborne, 2012).

6.2.2 Ontologies as a cultural perspective

Despite the difficulty to articulate, an ontology reflects the thoughts and perspectives of its creators. Anticoli and Toppano comment that an ontology is both a socio-technical and cultural construction (Anticoli & Toppano, 2011b, p. 1) and compare it to a pair of glasses. The ontology, like the pair of glasses, intervenes to the process of apprehending the conceivable and defines types of concepts and relations used for interpreting reality. Additionally, building an ontology, calls for answers to certain questions about the represented conceptualisation, which makes us realize that ontologies affect both interpretation and implementation (Anticoli & Toppano, 2011a, p. 10). Seaver defines algorithms generally as any type of rules required for computing, and places them in more abstract level of thinking, recommending that we should think of algorithms not as *in* culture but *as* culture (Seaver, 2017, p. 1), while Papakyriakopoulos and Mboya present an example from the Google image search algorithm. In their study, they discovered that the algorithm reproduces patriarchal stereotypes, while also performing racial discrimination (Papakyriakopoulos & Mboya, 2022).

Ontologies explicitly define categories, relations, individuals, etc aiming to unambiguously represent the intent of their creator, allowing to depict personal views through the analysed data since the provided definitions of categories, relations, individuals, etc are selected based on the beliefs and viewpoints of their creator. To elaborate on this, ontologies enable the disambiguation of the creator's intent, and not, necessarily, the representation of "objective statements" about reality. This feature allows us to consider an ontology as a cultural

construction reflecting a specific epistemological approach; as such, it is far from being itself an objective reality or even an objective representation of reality.

6.2.3 Projects on ontological representation

Some of the attempted efforts for knowledge representation relate to the domain of anthropology, but the research projects actively engaging anthropologists are significantly limited. In this subsection, where we explore projects on ontological representation, only one of them, the last project presented here, is conducted by anthropologists.

Among these projects, Phefo et al had the ambitious plan to represent all cultures in one ontology, but this effort resulted to oversimplified approaches (Phefo et al., 2015). Another project, focusing on the indigenous knowledge domain, targeted to framing the object of description, although it was not very warmly received neither from the anthropological nor the knowledge representation perspective. Yet, an innovative approach was introduced, which makes it an interesting case. The researchers used OWL (language) and Protégé (software) for developing an ontology, but they did not adequately document their selection of classes or the defined properties (Haron & Hamiz, 2014a) (Haron & Hamiz, 2014b).

Another case aiming at the representation of indigenous knowledge, namely traditional medicine, is the one reported by Ayimdji et al. They deploy the rationale that served as a basis for developing an ontology, and they recommend the use of Description Logic (DL) for an ontology about African traditional medicine (ATM). But they do not recommend specific classes or properties due to their focus on the required formalism for implementing DL. Even so, their conclusion is important and intends to guide other researchers in the field since they suggest that it is necessary to gather a multidisciplinary team to build a complete ontology for ATM (Ayimdji et al., 2011, p. 250).

Chi et al collected numerous open datasets about Taiwanese indigenous populations; the datasets that were not already in RDF were migrated to RDF, and they developed SPARQL endpoints to integrate searching among the heterogeneous resources (Chi et al., 2020). Their result is not a single ontology; still, the technologies implemented provide the infrastructure to eventually create one. During another effort, a kinship ontology was developed by Chui et al (2020) building on the extended work of Read and Fischer on kinship as presented above, in section 5.2. Other attempts have a far more restricted focus, like the one representing dance styles (Chantas et al., 2018), which falls into the category of Intangible Cultural Heritage.

As we introduced earlier, this section concludes with a knowledge representation project involving anthropologists. Although it is a work by Fischer and Finkelstein dating back to the early '90s, it was largely implemented in similar terms to what knowledge representation is currently about and it is described in their work "Social knowledge representation: A case study" (M. D. Fischer & Finkelstein, 1991). This project was based on Modal Action Logic (M[A]L), a formal language based on deontic logic, comprising a set of axioms and declarations used for representing situations related to structural relationships and handling the potential implications these relations might have when triggered by actions taken by involved subjects. The critical thing about deontic logic is that it replaces the imperative 'possibly' modal operator, with 'permitted', and 'necessarily' with obliged. The authors' goal was to define the pieces of information needed to make sense of what people were reporting to researchers by establishing a "logical consistency", using M[A]L to develop and implement a formalistic method for describing social behaviour -and social behaviour theories- while conducting an ethnographic study about marriages in a Punjabi community in Pakistan.

7 Discussion

Many researchers, indicative views of whom are presented in the current study, see computers -if not as rivals- more as autonomous entities which function somewhat differently from the human nature. Significantly opposed to this approach, we perceive computers as human constructions; in that sense, the way computers function reflects human thinking. Integrating computers in so many domains of everyday life is aligned with the evolvement of thinking in western societies over the last few centuries. Fauconnier and Turner introduce us to “The Way We Think”, their highly influencing work in terms of references, by stating that we live in the age of the triumph of form (Fauconnier & Turner, 2003, p. 3). And, according to Totaro and Ninno, this dominance of formalism leads to technology, explaining that the culture of mechanization did not arise as a ‘superstructure’ generated by the existence of machines, but it was the rise of a mentality -oriented towards process formalization- that facilitated the designing of mechanical equipment and their spread (Totaro & Ninno, 2014, p. 33).

A fundamental limitation of our study is that it builds its hypothesis on the position of Fortun et al concerning the necessity to re-consider digital forms so that ethnography is approached under a new perspective and the ethnographic archive is re-written, imposing the redesign of the digital infrastructure on which they will exist (Fortun et al., 2017, p. 13). Additionally, it should be clear that any encoding or representation efforts described in this study, are considered cultural constructions, which reflect a specific epistemological approach; they are not an objective reality, nor do they represent one. Having this clarification in mind, the study is not based on simplistic positivism; instead it nourishes the idea of developing encodings that are comprehensive and internally coherent, so that ethnographic data can be processed and effectively preserved. Moreover, the study supports current efforts which create data to serve as the ground for developing representations that ensure sharing of research results in the most reliable and interoperable way. As already mentioned in this study, if researchers cannot share their observations, then -in the scientific context- it is as if these observations have never existed.

One of the main strengths of this study is the recommendation of introducing formalistic methods for producing and representing anthropology. We have already mentioned that natural language has been established as the only way to disseminate anthropological knowledge so far. But now, there are alternative structures available, too. Lyon comments on these structures that formalism in ethnographic data representation is vital for their sustainability and for anthropology as a social science because they support transparency regarding the researchers’ claims and facilitate verifiability, which leads to documented reliability of the findings (Lyon, 2013, p. 46). We acknowledge the significance of this approach, but as Fischer noticed, such methods are not a panacea for all kinds of problems, even though they are widely implemented across different sub-domains of anthropology ranging from the Humanities perspective to pure Science approaches (M. D. Fischer, 2006, p. 5). In addition to that, by adapting techniques drawn from distributed artificial intelligence, system specification, anthropology, and formal semantics it is possible to avoid the triviality often associated with formal descriptions of social domains on the one hand, and arid abstraction on the other (M. D. Fischer & Finkelstein, 1991, p. 119).

An additional strength of the study is that its extended-in-time literature review reveals low acceptance rates by anthropologists regarding the use of computers as tools in their research. Even though our study refers to contemporary approaches, the idea of using computers in anthropology dates back to the ‘60s and ‘70s (Burton, 1973), but the documented

efforts were scattered and non-systematic, except for a very limited number of them. Wagner described that there was, in the late 80's, substantial use of computing within anthropology, but there was relatively little anthropological computing (Wagner, 1989, p. 419), leading other researchers, like Behrens and Read, to raise their voices inviting anthropologists to use computers not only for automating known processes, but to think about their research in a different way (Behrens & Read, 1993, p. 429). Despite the calls for action, the anthropological community has hardly been involved in such efforts, except in the kinship field where important progress has been demonstrated. Yet, the pioneers of the domain keep highlighting its potential. Fischer denotes that the possibilities of the domain are, in principle, compatible with scientific as well as postmodern approaches (M. Fischer, 2004, p. 154), also arguing that computational methods *are* theory-building tools and not just tools built *for* and *by* theories (M. D. Fischer et al., 2013, p. 12).

Regarding any practical implications and future directions of the study, we must note that, without much doubt, the required shift to computational methods constitutes a difficult process. The introduction of advanced technologies into a cultural context demands a new way of thinking, a new language to express the findings and new processes of interpreting and establishing an order in the world (Combi, 1992, p. 43). Consequently, it is essential to create a shared viewpoint through interdisciplinary approaches, through which the field of computational social science will emerge, along with a new paradigm for training new scholars (Lazer et al., 2009, p. 722). New scholars must develop computational thinking in the sense that Wing defined it, namely as a required skill that is not only a matter for computer scientists; this way of thinking focuses on the conceptualisation and not on programming, therefore, in principle, it is about how people think and about ideas, and not about computers or objects (Wing, 2006).

8 Conclusions

Intersecting anthropology and computer science guides the formation of new conceptualisations, since computational tools unfold new potential for the domain of ethnography on the basis that they set questions about generalisability (Abramson et al., 2018). The current study has outlined the contemporary technological environment as a vehicle for producing and disseminating anthropological knowledge. Two basic dimensions are discussed. One of them deals with the degree to which computer science contributes to the work of anthropologists regarding data processing and reasoning. The other dimension examines ways of representing these conclusions in a machine-processable way. In this framework, the study challenges the use of natural language as the dominant tool for representing results of anthropological research and investigates computational encodings and representations appropriate for ethnographic data and anthropological knowledge.

In compliance with its main objectives, the study identifies three basic stages regarding the use of computers in anthropology. The first stage is defined by using computers as tools for calculations to assist quantification studies. The second stage is defined by using computers to assist qualitative analysis of data. And the third stage is about knowledge representation and reasoning. Although the first stage is about primary uses of computers (past), the second is about the contemporary transitional era (present), and the third about the potential (future), still, there is no strictly linear time sequence regarding their adoption and implementation.

To conclude, this study draws attention to a paradigm shift flagged by the semantic web in terms of the methods used for conducting anthropological research and how research results become publicized. The shift is observed in information science regarding moving away from

the earlier focus, which was the development of algorithms for enabling the processing of unstructured data, toward the current focus, which is placed on data structures. In that sense, the semantic web, which combines all three stages of computers use, is at the top of the pyramid of data modelling, allowing interoperability and effective data re-use, and can be the means through which ethnographic data and anthropological analyses could form a cumulative archive and, as such, be open to new analyses and interpretations.

References

- Abramson, C. M., & Dohan, D. (2015). Beyond Text: Using Arrays to Represent and Analyze Ethnographic Data. *Sociological Methodology*, 45(1), 272–319. <https://doi.org/10.1177/0081175015578740>
- Abramson, C. M., Joslyn, J., Rendle, K. A., Garrett, S. B., & Dohan, D. (2018). The promises of computational ethnography: Improving transparency, replicability, and validity for realist approaches to ethnographic analysis. *Ethnography*, 19(2), 254–284. <https://doi.org/10.1177/1466138117725340>
- Albris, K., Otto, E. I., Astrupgaard, S. L., Gregersen, E. M., Jørgensen, L. S., Jørgensen, O., Sandbye, C. R., & Schønning, S. (2021). A view from anthropology: Should anthropologists fear the data machines? *Big Data & Society*, 8(2), 20539517211043656. <https://doi.org/10.1177/20539517211043655>
- Alvard, M., & Carlson, D. (2020). Identifying Patch Types Using Movement Data from Artisanal Fishers from the Commonwealth of Dominica. *Current Anthropology*, 61(3), 380–387. <https://doi.org/10.1086/708720>
- Anticoli, L., & Toppano, E. (2011a). How culture may influence ontology co-design: A qualitative study. *International Journal of Information Technology and Web Engineering*, 6(2), 1–17. <https://doi.org/10.4018/jitwe.2011040101>
- Anticoli, L., & Toppano, E. (2011b). The role of culture in collaborative ontology design. *Proceedings of the 2011 International Conference on Intelligent Semantic Web-Services and Applications*, 4:1-4:9. <https://doi.org/10.1145/1980822.1980826>
- Arnold, T., & Fuller, H. J. A. (2018). In Search of the User’s Language: Natural Language Processing, Computational Ethnography, and Error-Tolerant Interface Design. *Advances in Usability, User Experience and Assistive Technology*, 36–43. https://doi.org/10.1007/978-3-319-94947-5_4
- Artmann, S. (2010). Computers and Anthropology. In J. H. Birx (Ed.), *21st century anthropology: A reference handbook* (pp. 915–924). Sage.
- Asuncion, C. H., & van Sinderen, M. J. (2010). Pragmatic Interoperability: A Systematic Review of Published Definitions. In P. Bernus, G. Doumeings, & M. Fox (Eds.), *Enterprise Architecture, Integration and Interoperability* (pp. 164–175). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-15509-3_15
- Ayimdji, A., Koussoubé, S., Fotso, L. P., & Konfé, B. O. (2011). Towards a “Deep” Ontology for African Traditional Medicine. *Intelligent Information Management*, 03, 244–251. <https://doi.org/10.4236/iim.2011.36030>
- Bakharia, A., & Corrin, L. (2019). Using Recent Advances in Contextual Word Embeddings to Improve the Quantitative Ethnography Workflow. In B. Eagan, M. Misfeldt, & A. Siebert-Evenstone (Eds.), *Advances in Quantitative Ethnography* (pp. 299–306). Springer International Publishing.
- Bartlett, A., Lewis, J., Reyes-Galindo, L., & Stephens, N. (2018). The locus of legitimate interpretation in Big Data sciences: Lessons for computational social science from -

- omic biology and high-energy physics. *Big Data & Society*, 5(1), 2053951718768831. <https://doi.org/10.1177/2053951718768831>
- Beaulieu, A. (2017). Vectors for Fieldwork: Computational Thinking and New Modes of Ethnography. In L. Hjorth, H. A. Horst, A. Galloway, & G. Bell (Eds.), *The Routledge companion to digital ethnography* (pp. 29–39). Routledge, Taylor & Francis Group. <https://www.routledge.com/The-Routledge-Companion-to-Digital-Ethnography-1st-Edition/Hjorth-Horst-Galloway-Bell/p/book/9781138940918>
- Behrens, C. A. (1990). Qualitative and quantitative approaches to the analysis of anthropological data: A new synthesis. *Journal of Quantitative Anthropology*, 2(1), 305–328.
- Behrens, C. A., & Read, D. W. (1993). Anthropology: Moving from task-driven to science-driven computing. *Social Science Computer Review*, 11(4), 429–451.
- Beuving, J. J. (2020). Ethnography's future in the big data era. *Information, Communication & Society*, 23(11), 1625–1639. <https://doi.org/10.1080/1369118X.2019.1602664>
- Bharwani, S. (2006). Understanding Complex Behavior and Decision Making Using Ethnographic Knowledge Elicitation Tools (KnETs). *Social Science Computer Review*, 24(1), 78–105. <https://doi.org/10.1177/0894439305282346>
- Brooker, P. (2022). Computational ethnography: A view from sociology. *Big Data & Society*, 9(1), 1–6. <https://doi.org/10.1177/20539517211069892>
- Burton, M. L. (1973). Recent computer applications in cultural anthropology. *Computers and the Humanities*, 7(6), 337–341. <https://doi.org/10.1007/BF02395108>
- Carlsen, H. B., & Ralund, S. (2022). Computational grounded theory revisited: From computer-led to computer-assisted text analysis. *Big Data & Society*, 9(1), 1–16. <https://doi.org/10.1177/20539517221080146>
- Carlson, S., & Anderson, B. (2007). What Are Data? The Many Kinds of Data and Their Implications for Data Re-Use. *Journal of Computer-Mediated Communication*, 12(2), 635–651. <https://doi.org/10.1111/j.1083-6101.2007.00342.x>
- Chantas, G., Karavarsamis, S., Nikolopoulos, S., & Kompatsiaris, I. (2018). A Probabilistic, Ontological Framework for Safeguarding the Intangible Cultural Heritage. *Journal on Computing and Cultural Heritage*, 11(3), 12:1-12:29. <https://doi.org/10.1145/3131610>
- Chi, Y.-L., Sung, H.-Y., & Lien, Y.-Y. (2020). Towards the Ethnic Understanding of Taiwanese Indigenous Peoples: A Mashup Based on Semantic Web and Open Data. In P.-L. P. Rau (Ed.), *Cross-Cultural Design. User Experience of Products, Services, and Intelligent Environments* (pp. 287–297). Springer International Publishing. https://doi.org/10.1007/978-3-030-49788-0_21
- Chui, C., Grüninger, M., & Wong, J. (2020). An Ontology for Formal Models of Kinship. *Formal Ontology in Information Systems*, 92–106. <https://doi.org/10.3233/FAIA200663>
- Cioffi-Revilla, C. (2016). *Bigger Computational Social Science: Data, Theories, Models, and Simulations -- Not Just Big Data* (SSRN Scholarly Paper ID 2784278; pp. 1–5). Social Science Research Network. <http://ssrn.com/abstract=2784278>
- Combi, M. (1992). The imaginary, the computer, artificial intelligence: A cultural anthropological approach. *AI & Society*, 6(1), 41–49. <https://doi.org/10.1007/BF02472768>
- Doerr, M., & Iorizzo, D. (2008). The dream of a global knowledge network: A new approach. *Journal on Computing and Cultural Heritage*, 1(1), 5:1-5:23. <https://doi.org/10.1145/1367080.1367085>

- Fauconnier, G., & Turner, M. (2003). *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. Basic Books.
- Fischer, M. (2004). Integrating Anthropological Approaches to the Study of Culture: The “hard” and the “soft.” *Cybernetics and Systems*, 35(2–3), 147–162.
<https://doi.org/10.1080/01969720490426830>
- Fischer, M. D. (1994). *Applications in Computing for Social Anthropologists* (edition published in the Taylor&Francis e-Library, 2005). Routledge.
- Fischer, M. D. (2006). Introduction: Configuring Anthropology. *Social Science Computer Review*, 24(1), 3–14. <https://doi.org/10.1177/0894439305282575>
- Fischer, M. D., & Ember, C. (2018). Big Data and Research Opportunities Using HRAF Databases. In C. Shu-Heng (Ed.), *Big Data in Computational Social Science and Humanities*. Springer. http://dx.doi.org/10.1007/978-3-319-95465-3_17
- Fischer, M. D., & Finkelstein, A. (1991). Social knowledge representation: A case study. In N. Fielding & R. Lee (Eds.), *Using computers in qualitative research* (pp. 119–135). SAGE.
- Fischer, M. D., Lyon, S. M., Sosna, D., & Henig, D. (2013). Harmonizing Diversity Tuning Anthropological Research to Complexity. *Social Science Computer Review*, 31(1), 3–15. <https://doi.org/10.1177/0894439312455311>
- Fortun, M., Fortun, K., & Marcus, G. E. (2017). Computers in/and anthropology: The poetics and politics of digitization. In L. Hjorth, H. A. Horst, A. Galloway, & G. Bell (Eds.), *The Routledge companion to digital ethnography* (pp. 11–20). Routledge, Taylor & Francis Group. <https://www.routledge.com/The-Routledge-Companion-to-Digital-Ethnography-1st-Edition/Hjorth-Horst-Galloway-Bell/p/book/9781138940918>
- Fuentes, A., & Wiessner, P. (2016). Reintegrating Anthropology: From Inside Out: An Introduction to Supplement 13. *Current Anthropology*, 57(S13), S3–S12.
<https://doi.org/10.1086/685694>
- Geertz, C. (1973). *The Interpretation of Cultures: Selected Essays*. Basic Books.
- 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.
<https://doi.org/10.1006/ijhc.1995.1081>
- Gruber, T. R. (2007). Ontology of folksonomy: A mash-up of apples and oranges. *International Journal on Semantic Web and Information Systems*, 3(1), 1–11.
- Hakken, D. (1993). Computing and Social Change: New Technology and Workplace Transformation, 1980-1990. *Annual Review of Anthropology*, 22, 107–132.
- Haron, H., & Hamiz, M. (2014a). An Ontological Framework to Preserve Malay Indigenous Health Knowledge. *Advanced Science Letters*, 20(1), 226–230.
<https://doi.org/10.1166/asl.2014.5261>
- Haron, H., & Hamiz, M. (2014b). An Ontological Model for Indigenous Knowledge of Malay Confinement Dietary. *Journal of Software*, 9(5), 1302–1312.
<https://doi.org/10.4304/jsw.9.5.1302-1312>
- Hjorth, L., Horst, H. A., Galloway, A., & Bell, G. (Eds.). (2017). *The Routledge companion to digital ethnography*. Routledge, Taylor & Francis Group.
- Knox, H., & Nafus, D. (Eds.). (2018). *Ethnography for a data-saturated world*. Manchester University Press.
- Kohne, J. (2014). Ontology, Its Origins and Its Meaning in Information Science. In R. Hagengruber & U. Riss (Eds.), *Philosophy, computing and information science* (pp. 85–89). Pickering & Chatto.

- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Alstynne, M. V. (2009). Computational Social Science. *Science*, 323(5915), 721–723. <https://doi.org/10.1126/science.1167742>
- Leetaru, K. H., Perkins, T., & Rewerts, C. (2014). Cultural Computing at Literature Scale: Encoding the Cultural Knowledge of Tens of Billions of Words of Academic Literature. *D-Lib Magazine*, 20(9/10). <https://doi.org/10.1045/september2014-leetaru>
- L’Homme, M.-C., & Bernier-Colborne, G. (2012). Terms as labels for concepts, terms as lexical units: A comparative analysis in ontologies and specialized dictionaries. *Applied Ontology*, 7(4), 387–400. <https://doi.org/10.3233/AO-2012-0116>
- Lyon, S. M. (2013). Networks and Kinship Formal Models of Alliance, Descent, and Inheritance in a Pakistani Punjabi Village. *Social Science Computer Review*, 31(1), 45–55. <https://doi.org/10.1177/0894439312453275>
- Lyon, S. M., & Magliveras, S. S. (2006). Kinship, Computing, and Anthropology. *Social Science Computer Review*, 24(1), 30–42. <https://doi.org/10.1177/0894439305281494>
- Marcus, G. E. (1998). The once and future ethnographic archive. *History of the Human Sciences*, 11(4), 49–63. <https://doi.org/10.1177/095269519801100404>
- Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., & Aiden, E. L. (2011). Quantitative Analysis of Culture Using Millions of Digitized Books. *Science*, 331(6014), 176–182. <https://doi.org/10.1126/science.1199644>
- Moretti, F. (2013). *Distant Reading*. Verso Books.
- Mueller, A. (2016). Beyond ethnographic scriptocentrism: Modelling multi-scalar processes, networks, and relationships. *Anthropological Theory*, 16(1), 98–130. <https://doi.org/10.1177/1463499615626621>
- Munk, A. K., & Winthereik, B. R. (2022). Computational Ethnography: A Case of COVID-19’s Methodological Consequences. In M. H. Bruun, A. Wahlberg, R. Douglas-Jones, C. Hasse, K. Hoeyer, D. B. Kristensen, & B. R. Winthereik (Eds.), *The Palgrave Handbook of the Anthropology of Technology* (pp. 201–214). Springer. https://doi.org/10.1007/978-981-16-7084-8_10
- Nelson, L. K. (2017). Computational Grounded Theory: A Methodological Framework. *Sociological Methods & Research*. <https://doi.org/10.1177/0049124117729703>
- Paff, S. (2022). Anthropology by Data Science. *Annals of Anthropological Practice*, 46(1), 7–18. <https://doi.org/10.1111/napa.12169>
- Papakyriakopoulos, O., & Mboya, A. M. (2022). Beyond Algorithmic Bias: A Socio-Computational Interrogation of the Google Search by Image Algorithm. *Social Science Computer Review*, 08944393211073169. <https://doi.org/10.1177/08944393211073169>
- Pels, P., Boog, I., Florusbosch, J. H., Kripe, Z., Minter, T., Postma, M., Sleeboom-Faulkner, M., Simpson, B., Dilger, H., Schönhuth, M., Poser, A. von, Castillo, R. C. A., Lederman, R., & Richards-Rissetto, H. (2018). Data management in anthropology: The next phase in ethics governance? *Social Anthropology*, 26(3), 391–413. <https://doi.org/10.1111/1469-8676.12526>
- Phefo, O. S. D., Kefitiley, N., & Hlomani, H. (2015). Towards the Cultural Knowledge Ontology. *2015 IEEE International Conference on Information Reuse and Integration (IRI)*, 526–533. <https://doi.org/10.1109/IRI.2015.85>

- Pokraev, S., Reichert, M., Steen, M. W. A., & Wieringa, R. J. (2005). *Semantic and pragmatic interoperability: A model for understanding*. 160. Scopus. <http://ceur-ws.org/Vol-160/paper21.pdf>
- Pool, R. (2017). The verification of ethnographic data. *Ethnography*, 18(3), 281–286. <https://doi.org/10.1177/1466138117723936>
- Purzycki, B. G., & Jamieson-Lane, A. (2017). AnthroTools: An R Package for Cross-Cultural Ethnographic Data Analysis. *Cross-Cultural Research*, 51(1), 51–74. <https://doi.org/10.1177/1069397116680352>
- Read, D., Fischer, M., & Leaf, M. (2013). What Are Kinship Terminologies, and Why Do We Care? A Computational Approach to Analyzing Symbolic Domains. *Social Science Computer Review*, 31(1), 16–44. <https://doi.org/10.1177/0894439312455914>
- Read, D. W. (2006). Kinship Algebra Expert System (KAES) A Software Implementation of a Cultural Theory. *Social Science Computer Review*, 24(1), 43–67. <https://doi.org/10.1177/0894439305282372>
- Read, D. W., & Behrens, C. A. (1990). KAES: An expert system for the algebraic analysis of kinship terminologies. *Journal of Quantitative Anthropology*, 2(4), 353–393.
- Ribes, D., & Bowker, G. C. (2009). Between meaning and machine: Learning to represent the knowledge of communities. *Information and Organization*, 19(4), 199–217. <https://doi.org/10.1016/j.infoandorg.2009.04.001>
- Romney, A. K., Weller, S. C., & Batchelder, W. H. (1986). Culture as Consensus: A Theory of Culture and Informant Accuracy. *American Anthropologist*, 88(2), 313–338. <https://doi.org/10.1525/aa.1986.88.2.02a00020>
- Sahlins, M. (2002). *Waiting for Foucault, Still*. Prickly Paradigm Press.
- Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, 4(2), 1–12. <https://doi.org/10.1177/2053951717738104>
- Seo, J., Moon, J., Choi, G. W., & Do, J. (2022). A Scoping Review of Three Computational Approaches to Ethnographic Research in Digital Learning Environments. *TechTrends*, 66(1), 102–111. <https://doi.org/10.1007/s11528-021-00689-3>
- Sosna, D., Galeta, P., Šmejda, L., Sladek, V., & Bruzek, J. (2013). Burials and Graphs Relational Approach to Mortuary Analysis. *Social Science Computer Review*, 31(1), 56–70. <https://doi.org/10.1177/0894439312453277>
- Stubbersfield, J., & Tehrani, J. (2013). Expect the Unexpected? Testing for Minimally Counterintuitive (MCI) Bias in the Transmission of Contemporary Legends A Computational Phylogenetic Approach. *Social Science Computer Review*, 31(1), 90–102. <https://doi.org/10.1177/0894439312453567>
- Tallyn, E., Fried, H., Gianni, R., Isard, A., & Speed, C. (2018). The Ethnobot: Gathering Ethnographies in the Age of IoT. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 604:1-604:13. <https://doi.org/10.1145/3173574.3174178>
- TEDx Talks (Director). (2011, July 14). *A picture is worth 500 billion words: Erez Lieberman Aiden and Jean-Baptiste Michel at TEDxBoston*. <https://youtu.be/WtJ50v7qByE>
- Totaro, P., & Ninno, D. (2014). The Concept of Algorithm as an Interpretative Key of Modern Rationality. *Theory, Culture & Society*, 31(4), 29–49. <https://doi.org/10.1177/0263276413510051>
- Van Der Leeuw, S. E. (2004). Why Model? *Cybernetics and Systems*, 35(2–3), 117–128. <https://doi.org/10.1080/01969720490426803>

- Wagner, R. A. (1989). The Rise of Computing in Anthropology: Hammers and Nails. *Social Science Computer Review*, 7(4), 418–430.
<https://doi.org/10.1177/089443938900700403>
- Walsh, D., & Downe, S. (2005). Meta-synthesis method for qualitative research: A literature review. *Journal of Advanced Nursing*, 50(2), 204–211.
<https://doi.org/10.1111/j.1365-2648.2005.03380.x>
- Weingart, S., & Jorgensen, J. (2013). Computational analysis of the body in European fairy tales. *Literary and Linguistic Computing*, 28(3), 404–416.
<https://doi.org/10.1093/lc/fqs015>
- Wilf, E. (2013). Toward an Anthropology of Computer-Mediated, Algorithmic Forms of Sociality. *Current Anthropology*, 54(6), 716–739. <https://doi.org/10.1086/673321>
- Wing, J. M. (2006). Computational Thinking. *Communications of the ACM*, 49(3), 33–35.
<https://doi.org/10.1145/1118178.1118215>
- Zheng, K., Hanauer, D. A., Weibel, N., & Agha, Z. (2015). Computational Ethnography: Automated and Unobtrusive Means for Collecting Data In Situ for Human–Computer Interaction Evaluation Studies. In V. Patel, T. Kannampallil, & D. Kaufman (Eds.), *Cognitive Informatics for Biomedicine* (pp. 111–140). Springer, Cham.
https://doi.org/10.1007/978-3-319-17272-9_6