ÉTICA E PROPOSTA DE APROVEITAMENTO DE DADOS NÃO UTILIZADOS DO TRABALHO CIENTÍFICO
ETHICS AND A PROPOSAL TO USE UNUSED DATA IN SCIENTIFIC WORK
ÉTICA Y PROPUESTA DE APROVECHAMIENTO DE DATOS NO UTILIZADOS DEL TRABAJO CIENTÍFICO

Isabel Camalhão¹
Serafim Camalhão²

¹ Centro de Estudos Interdisciplinares em Educação e Desenvolvimento (CeIEd) Instituto de Educação – Universidade Lusófona, Lisboa, Portugal.  
² CIES-IUL, Lisboa, Portugal

Isabel Camalhão - isabelaferreira66@hotmail.com | Serafim Camalhão - serafimleopoldo@hotmail.com

Corresponding Author
Serafim Camalhão
CIES-IUL
Edifício ISCTE, Av. das Forças Armadas
1649-026 Lisboa – Portugal
serafimleopoldo@hotmail.com

RECEIVED: 29th November 2017
ACCEPTED: 18th January 2018
RESUMO
Introdução: Todo o trabalho científico nas áreas das Ciências Sociais e Humanas não utiliza todos os dados que recolhe e exclui elementos que não vão integrar os resultados finais. Estes dados podem ser úteis para outras investigações, mas colocam-se problemas éticos que dificultam o seu aproveitamento.
Objetivos: Desenvolver uma metodologia para usar os dados não utilizados e propor uma interpretação ética que facilite a utilização destes dados.
Métodos: Utilizaram-se os procedimentos e fases de construção de um projeto de investigação, tendo por centro uma área e um problema metodológico.
Resultados: A criação uma proposta de uma metodologia especificamente destinada à utilização de dados não utilizados, aberta e flexível.
Conclusões: A utilidade de aproveitar não só os dados não utilizados, assim como, todos os elementos de uma investigação social, que vem beneficiar todo o conhecimento. No campo da ética, centrou-se na questão de proteger as comunidades, os participantes e o investigador como critério central no aproveitamento de dados. As questões éticas são demasiado complexas para caber num único artigo.
Palavras-chave: Dados não utilizados; Metodologia; Conhecimento; Ética.

ABSTRACT
Introduction: All scientific work in the social sciences don’t use all collected data and exclude elements that won’t be part of the result. This data can be useful to other researches, but it has ethic problems that difficult their use.
Objectives: Develop a methodology to use the data not analysed and propose an ethical interpretation that facilitates its use.
Methods: It was used the procedures and steps of a construction of an research project, having as a centre one methodological area and problem.
Results: The creation of a proposal of one methodology specifically destined to the use of unused data, open and flexible.
Conclusions: The utility of the use, not only of unused data, but also, all elements of a social research, that will benefit all knowledge. In the field of ethics, the focus is in the question of protecting communities, participants and the researcher. The ethical questions are too complex to be treated in one only article.
Keywords: Unused data; Methodology; Knowledge; Ethics.

RESUMEN
Introducción: Todo lo trabajo científico en las áreas de las Ciencias Sociales e Humanas no utiliza todos los datos que recoge y excluye elementos que no tendrán parte de los resultados finales. Estos datos pueden ser útiles para otras investigaciones, pero se colocan problemas éticos que dificulten su utilización.
Objetivos: Desarrollar una metodología para utilizar los datos no analizados y proponer una interpretación ética que facilite su uso.
Métodos: En esto artículo se utilizarán los procedimientos y fases de la construcción de uno proyecto de investigación, tenido por centro un área y un problema metodológico.
Resultados: La creación de una propuesta de una metodología destinada à la utilización de datos no utilizados, abierta y flexible.
Conclusiones: La utilidad de aprovechar, no solo los datos no utilizados, así como, de todos los elementos de una investigación social, que viene beneficiar todo lo concomiento. En lo campo de la ética, se cetro en la cuestión de proteger las comunidades, participantes y lo investigador como criterio central no aprovechamiento de datos. Las cuestiones éticas demasiado complexas, pero caber en un único artículo.
Palabras Clave: Datos no utilizados; Metodología; Conocimiento; Ética.

INTRODUCTION
This article was born the confluence of an troubled PHD Thesis with many excluded data that result in the intention to use this unused data, present in the article “Os “despojos do trabalho científico em Ciências Sociais: O potencial do conhecimento dos “restos” (Camalhão & Camalhão), in this article it was proposed a way to use this unused data and the last his public presentation in the Iberic-American Congress in Qualitative research.
The article published in the proceedings of CIAIQ 2017 is structured in a substantive form, it is used Glasser (1992) term concerned with two aspects, becoming more didactic, assessible to all and present a scientific problem inherent to an area substantive and concrete. Starting from this observation, it becomes evident that was needed give a more formal form to the article.

The oral presentation of this article revelled that would be important to add an ethical dimension. It is an old problem that we have seen personally in the academic environment, the creation of walls and barriers to knowledge. What surprised us was the extensions of the problem presented by the Brazilian researchers, they couldn’t use the unused data, for scientific proposals. The explanation that was given than, was that the Ethics Committees forbid and become impossible the use of unused data through one highly restrictive interpretations of ethical rules.

In our point of view, it more important propose one solution, rather than just criticise. We put ourselves in the place of Ethics Committees and they have reason in the sense that Hammersley & Trainou (2012) remember us that doing social research have rules and risks for the researchers and participants. Our proposal implies doing a just interpretation of ethical aspects, allowing the use of unused data, and in this broaden this idea to reuse all information present in social research.

The content of this article it was born from a common problem of two researchers, the end of a PHD Thesis with a lot of unused information and the other a restart of a PHD Thesis where is needed review all the work that has been done. This work it was born from an collaborative ethnography (Chang, Ngunjiri & Hernandes, 2013), where all observations and academic experience of both researchers become important to the construction of this work. Because of that many statements are this field, that easily could be justified in the literature, but they don’t belong to any author, just to ourselves. Nevertheless, in the essential, our concern, just as in Grounded Theory, we look for formalize and transform one problem in a coherent proposal with a relevant literature.

In our experience as students in academic world, we observe, that whenever end an academic work, que do one evaluation of what have produced. Also, we have seen that our colleges did the same. In our case the evaluation was made comparing what we pretended to do and what we had accomplished, this reveal flaws, limitations and what we could do better. The literature has an abundant in production of this indicators, that we recommend the reading of Patton (2015, p. 21) in which the topics are the proposal of a research; to whom is addressed, research question, answer to the question, resources and quality of results. In the case of ended PHD Thesis, the evaluation revealed that from 126 interviews done just 78 were used. When we spoke with our colleges, it was revealed that is normal, in the function criteria applied to academic work that formally limit their dimensions, the boundaries of a research project and exclude that don’t fit in that limits. This unused data can be utilized.

In methodological terms, we use the same procedures used in any research design (Quivy & Campenhoud, 2011) with a the choosing of a topic or theme for a research, a research problem, a framework and methodology. In the article that originate these, it has the concrete data to give them structure. Here are just the abstract aspects with a higher formalization, the result is the one final proposal to treat this data.

What the reader of this article will find in the next pages is that is presented isn’t new, there are several methodologies that reuse data. What we propose it is more one proposal to be debated. In a realistic way, we assume that is impossible to expose all arguments e references about these theme, but only the ones that we consider more representative.

In our academic experience, we verify that, when we start one research process we can define previously what will be included and excluded in the presentations of the results. In this process, we will add and remove elements, in which is important to keep all the data an know why we have taken this or that decision.

The type of elements collected and excluded, are the same of any academic work. They are essentially, bibliographic, documents, empirical data collected in the field and reflexive with aspect related to the researcher. The main point is in seeing if the information collected is complete or incomplete. The complete information is the one in which all elements or components, in his whole has one coherent and logical form. The information is incomplete if all the elements aren’t clearly enough, the data are inconsistent, contradictory, it is needed to do more research to reach a minimal state of coherence. They are just clues. We, as authors of this text, defend that there no complete information, just coherent and logical enough.

The unused data, are data or excluded information from a final work in Social and Human Sciences. The quality and quantity of this unused data, are an important criterion in what we can do with them. We defend that don’t exist perfect research process, and these leave always unused data.

This evidence, raises a criticism about the waste of knowledge generated by the inevitable destruction of the unused data. Knowledge exists to be shared, and this is the centre of this article.

1. THE STARTING POINT

Our starting point, was a popular image, that it doesn’t has author or owner, it belongs to the practices of any family with scarce resources or spared. The popular wisdom, tells us that in a world where are people with hunger, it is a crime to throw away food. What is normal is to use the scratches in two ways by gathering all the chunks or joining more elements to what they have. In these practices no one knows previously, what will be left over from a meal.
Using this argument, we changed the term scratches to unused data, but the principle is the same. Any research, leave unused data, that point the existence of any information excluded for one or another reason. There is many information that can happen in a different way in a research. Thigs happen that change what was initially planned, namely the idea of Rossman and Rallis (2012), which pointed the objective of a researcher is to learn in the field, be flexible, to become able to listen the other. This may lead to the exclusion of gathered information. In other situations, Quivy & Campenhoud (2011), warn for the risk of research design with conceptions problems, too much ambitious to the skills and researcher resources.

In a practical point of view, that we have seen in our academic life made us to adapt some designations. Flick (2005) presents a distinction between linear and circular research. To this author, linear means quantitative where all aspect are previously defined. In our proposal, linear means only that the researcher has an equilibrated research design, in the access to the field didn’t had problems, everything went well, it didn’t have nothing to change, but to add. The term circular means that the researcher along the research process is going to adapt the research design to the concrete research. In our conception circular, we add a term troubled and the successions of problems that lead to the research be necessarily circular. The unused data are present in both situations. In the first the field are so rich that something has to be excluded, but in the most part was used. In the second, the succession of problems leads a number of changes in a research that leave a lot of unused information. The notion that we want to tell is that this data can either be used on new research process or complete other academic works.

2. WHAT ARE UNUSED DATA?

In his simplest form, we define unused data or excluded information as the data or excluded information from the final results in Human and Social Sciences. It is not enough to present this definition, it becomes useful to search in the literature, to what can correspond. In the literature we found several terms, in which we highlight three: data reduction, traces and corpus, that is near academic experience. As we can observe:

- Miles & Huberman (1994) uses the term "Data Reduction" means that in any investigation there is a need to reduce or condense data, abstracts inevitably imply deleting or losing data.
- Ghiglione & Matalon (1992) uses the term traces to designate documents, the remains leave records showing signs of a phenomenon.
- Paillé & Mucchielli (2009) uses the term "corpus" to designate the set of elements that constitutes an investigation.

2.1. The use of unused in the Social Sciences

What is presented is not new, we find that researchers and students publish articles, conference proceedings, seminars, books and in multiple forms, throughout a process of investigation extending beyond this. These are evaluated by their scientific production. The provisional results of an investigation are remains very similar to remains, because provisional and the accomplishment of articles after an academic work, a reutilization of the data because it implies the deepening of the same ones.

This article has origin and basis in the work of Rodriguez (2011) Analysis of Incomplete Data in Social Sciences. The author addresses a recurrent problem in quantitative methodologies databases are incomplete in the surveys with non-responses and the resulting voids. The problem was solved with some observations: (1) he has cheeked that some data that should be present are absent, we look for in the matrix or structure elements that allow us to find the lost or missing elements; (2) when trying to recover information it is enough to reorganize the elements around a question; (3) contrasting the various variables by looking for correlations, it is possible to estimate the lost value. These three solutions indicate that the desired information can be found in a structure, simply by reorganizing the data or contrasting the various variables. This first reading warned us to the possibility of taking advantage of the remains of any academic work.

The above referred work is quantitative in nature but its principles can be applied to any research. The next challenge was to do the same qualitatively in a more comprehensive way, with the possibility of looking for information inside and outside the structure to complete the missing elements.

In the final version of a Doctoral Thesis, the data recorded on any paper or electronic medium are documents with different types of information. The literature confirms this statement to be coupled with some examples that allow the use of data: Bernard & Ryan (2010) indicate the term secondary analysis of data, to designate the use of data of other investigations the problem implies lack of information; Wästefors, Åkerström & Jacobsson (2014) speak in reanalysis of qualitative data is the analysis of the data of an investigation by another investigator different from that had done it.

We could very easily go to one of these ways of reusing and seizing data, but the philosophy of this article indicates that before making any decision one has to have the notion of the type of information produced. The image that we reproduce more than theoretical is practical, in this case the main concern was solving problems, there is no care to record what is left out, the question of re-use of the data appears at the end and knowing what we had implied making an assessment of what we have.

In a qualitative approach, contextualization is fundamental (Rossman & Rallis, 2012), the challenge is to understand the data without its context with the maximum information. After recovering the information taken to re-analyze the same, regardless if
they were the researcher who collected the data or another. All methods and techniques can be used. There are, however, many remains that are primary data, collected by the researcher which have not been yet analyzed, distinguishing themselves from the reanalysis or even from a secondary analysis. Among the options you can still return to the field to complete what is missing. From this point of view, we consider that whenever the data are used in the original record, for example interview transcriptions, the distinction is useless, they are data of the same nature.

Both the unused data and the reanalysis of the data imply an attitude of openness to all forms of information and methodologies, because before making an evaluation of what one has, it is not possible to take any decision.

3. PROPOSAL OF A METHODOLOGY FOR THE USE OF REMAINS

The use of unused data aims to treat the unused data in the form of scientific publication or participation in academic events. We do not predefine the way because this depends on the potentiality of the remains and the solutions to the complementary ones. The proposed methodology is comprehensive and open, leaving room to each one’s to adapt it to its concrete situation. We define four phases: hermeneutics, analysis and taking decision, complementary analysis of data and formalization of results.

2.1 The hermeneutic phase

The hermeneutic phase is not really a stage, its main activity is to evaluate the data that were not used (Rodrigues, 2011). This work goes beyond a mere description and exploration of its contents including the understanding, relationships, structure, meaning and meaning of the data. Using a simple definition is the moment when it is necessary to identify what is left of from academic or scientific work. In this evaluation, it is necessary to avoid any preconceptions (Glasser, 2013), this would lead to sub or overestimate one aspect over another and in the expression “all is data” (Glasser, 2001, p. 145) in all unused data regardless the form or the source they are potentially useful sources of information. This is not a stage or a phase, because it isn’t the moment where we take decisions, the objective. The objective is simply knowing what we have without other preoccupations. In the case that we had to select or decide what to do, like in one PhD Thesis, we would trough out data. Than what it is? It is a moment in which we will return recurrently, to seek the elements we need.

You will not make decisions at this stage, this is only possible once the evaluation has been completed. Hermeneutics arises in a context where it is necessary to go beyond the mere description or relation among data. In the first place, the scientific works in Social Sciences take a written form and of discourse. These elements can discover meaning, reference and meaning (Ricour, 2013). Wernet (2014) presents four principles: Exclude the context; Take the literal meaning of the text seriously; Sequentially and Extensiveness. This contribution is accepted partially, the context is part of the data, without this it cannot understand the unused data, but the notion is that the first reading goes through the interpretation only of what is written, sequentially comes to expose the structure that is in the text and the extensively indicates that all the elements must be analyzed in detail, depth, the hypotheses are referring to small parts of the whole, so it was realized, the extensively arises from the analysis of the set of the various parts. This first analysis does not need to be so detailed, but needs to be understood and its structure exposed. The initial process can be summed up to the three stages of the circle of hermeneutics which, according to Carpenter (2013) are the simple reading, the structural analysis and the interpretation of the whole. As it is proposed, at this initial moment the intention is to become familiar with the data revealing its structure and understanding the potential of the presented unused data.

3.2. First phase: Analysis and decision making

There is enough information at this stage to start a research process. The decisions to be made depend on the resources of the researcher (Quivy & Campenhoud, 2011), of his ambition and of the data that he possesses, at least a small article at most a new research funded.

The basis of this and the next phase are based on Saldana (2013), in this phase is the first cycle of coding, with the first contact, treatment and approach to data. The first challenge is in choosing a coding process. The investigator may wish to conduct an investigation from the data, revealing its structure, often using an in vivo coding can have an attitude similar to an analytic induction, with the use of hypotheses and a provisional coding scheme previously created. Finally, it is possible to start from a previous encoding scheme where a protocol with procedures and classifications may be involved. There is an openness here to all kinds of qualitative methodologies and if the researcher wants mixed or even quantitative, it depends on the objective, formation and intention of those who study. The unused data, which however, always present failures and by nature incomplete, were excluded from the structure from which they emerged. The objective of this phase is to reveal the structure and consistency of the data selected for the use of the unused data.

We use the language of Grounded Theory to exemplify this phase, the procedure goes through an open coding having only the basis of the data (Glasser, 1992). The goal is always to achieve a substantive theory (Glasser & Strauss, 1967), but at this stage it is neither possible nor supposed to reach this level. In this methodological approach, both the objectives and the concern of the researcher, is to systematize and maximize what the data has to offer in order to identify the inevitable failures.
3.3 Second phase: Complementary analysis of data

The complementary analysis of data refers to the discovered failures or limitations. It is important to note that the unused data, based on the Grounded Theory, reveal a structure, a logic and even limited hypotheses.

We start with this approach based on the principle that data (Rodriguez, 2011), are not lost, are unused data and information is simply missing and needs to be completed. The strategies used consist of the crossing of data, which here being numerical take the form of tables. At a more qualitative level we work in a similar way, but attention is given to the text, to the letter then the crossing is done through matrices (Miles & Huberman, 1994).

In the use of unused data, we proceed as if it were a new investigation, where failures to complement are detected. In our perspective, the unused data are part of a whole (Paillé & Mucchielli, 2009) Irrespective of whether they have been used, it is in this whole body the unused data and traces that one must look for first what is needed to complete what is missing. When the original research capability is exhausted, if the researcher has resources and the process of recovery of remains is included in a scientific project, then there is the option to go back to the field as Strauss & Corbin (2008) do to improve and correct failures until the end of the investigation. But it can happen that the researcher does not have the resources, you can also look for another study about the area or thematic and ask permission to use the data collected in the field before being analyzed can still join forces with the colleague with the unused data of both doing a fabulous job on something that would go to waste. In the latter case, there is always the modesty of accepting what the research data allows, the researcher was happy because he did not throw away something that might be useful to others.

The question that remains to be solved is how will we join elements of both the research itself and external to it? In all investigations there is always a theme, a question of departure and objectives, this is the guiding thread, from which it will be necessary ton all investigations there is always a theme, a question of departure and objectives, this is the guiding thread, from which it will be necessary to use the comparative method, in which Glasser & Strauss (1967) presents a table where the comparison between groups is made by crossing information by similarity and difference in the columns and the minimum and maximum degree in the lines, the result is that it allows finding pertinent information for what is missing. It can not be said that it is something that can be done with two, three or more moments, the only rule that defends itself, in order to avoid confusion, is that one must first begin with the data of the investigation and after this goes to the external data, for prudence the analysis must be done by groups or types of information and of these not everything is usable.

In a reference to Saldaña (2013) this phase would correspond to the second cycle of coding methods which is no more than translating, "reorganization and reanalysis of the coded data through the first cycle of methods" is the search for patterns, structures, hypotheses. There is a difference here, these elements necessarily have to be discovered in the previous phase. Here, we only try to complete what is missing, which does not prevent what has already been done and rethink the analysis.

3.4 Third phase: Formalization of results

The third phase consists of the passage of all work in written form with the formalization of results. One of the tasks present in this task is to re-consult the bibliographical references again in this process to discover data that need the confrontation with the literature. The reality is that many problems are only discovered when one begins to write formally.

Writing is not easy, (Wolcott, 2009) imply reconstruct all the research process, and present it in a coherent form to a given public. Although it isn’t doing Classic Grounded Theory (Glasser, 2007), in ascertain way we look for the core category presented in the data, and then compare it with the relevant literature. This stage can be longer or shorter depending of the unused data, body of presented data and type of research.

This research process needs passing through one instrument for validation, verification and generalization of results. The test is done by confronting the literature, comparing the data and complementing failures naturally through the formalization of results. In the case of the unused data it is necessary to be humble and modest both in relation to the results and to their generalization. We recommend doing what Miles & Huberman (1994) refer to as the design and verification of conclusions, as in any investigation to present and discuss the results of the analysis of the unused data.

4. HOW TO PROCEED TO THE USE OF UNUSED DATA.

In this section, the objective, is in specify how can we proceed in a more technical way. The hermeneutic moment is the base of all research process, where we put the information in graphic representation and create one data storage system accessible to both data treatment or identification. To this end we use matrices (Miles & Huberman, 1994) where gather data by resemblance and difference. In this matrix, we can put either numbers or text, as it can be observed in figure 1.
5. ETHICAL ISSUES

Data sharing in social research is not free of ethical issues, we could start with copyright in our point of view is not a problem, because access to primary data depends exclusively on universities and researchers give them access. Usually, as we have seen, only the finished work and the data used are accessible, to those the question arises, as above depends on authorization.

The main ethical problem must be put in the use of these and in the protection of the participants. There is a lot of literature on ethics in research, in the case of a qualitative approach the question that is in question is the text and the contextualization (Rossman & Rallis, 2002; Miles & Huberman, 1994), in practice we will not reduce everything to numbers and abstractions, but we reflect on the subject and conclude that this exposes people.

How can we do it? We start from our practical experience, what we did in our theses. It’s a question of common sense, but you can find reference in the literature, we consider three levels of protection, the first is to change the name of the participants, omit the name of the entities, data as addresses, small geographical areas, the second implies rewriting parts of the original leaving only the content and the third pure and simple elimination either because the participant requested or because the content harms the own, and is not relevant. Not everything is profited or revealed.

In our proposed recovery, based on Wästerfors, Åkerström & Jacobsson, (2014) we replace the term deidentification and relocation instead of decontextualization. In protecting the participants, it is enough that we can not identify them and not locate them in space, the rest is loss of data.

It is important to legitimize the reuse of data as well as the use with rule which we consider the following:

1. All data reuse must be presented as a project and be approved either by the scientific committee or other competent entity to verify the rules below.
2. All data can be reused once properly de-identified and delocalized.
3. Sharing data can only be used exclusively for scientific purposes, and this is the purpose of every research.
4. There is a hierarchy of data access, some of which are reserved access.
5. Sharing the data of several researchers depends on their prior authorization to do so.
6. When the data are sensitive, the original documents should not be exposed, leaving only the results without ethical problems.
7. Access to data is provided mainly to disabled people with no possibility of doing field work, students without resources who wish to finish their courses and teaching with practical cases.
8. A researcher with a lot of empirical material, can reuse all his data without restrictions, only needing to present a project for ethical approval.

What we present is a starting point for discussion, the important thing is that we share the information we have, creating a useful knowledge and for the benefit of all.
CONCLUSIONS

This article presents one proposal to the use of amused data. His authors made bricolage to produce one methodological propose that allow the use of this data. The problem exceeds the methodological question, if we do a rigid interpretation of ethics in research, then we lose knowledge, if there are no etic, all produced knowledge became invalid. It is important the use of good sense, all data can be reused since the participants and researchers be protected with well-balanced rules. Do not pretend the reader see in this simple article completely treated in one article. We only pretend to lunch the discussion and invite all to criticize and develop this simple ide, the use of unused data in an ethical way.

REFERENCES