

Chapter 3

Reuse of Excess Research Data for New Researches

Amiram Porath
College for Academic Studies, Israel

ABSTRACT

The chapter presents a specific niche dilemma regarding the ethical aspects of utilization of data gathered for one specific research for another research, which was not specified to the data suppliers (sample) at the time of gathering, or that the data suppliers were not even informed of the possibility of such occurrence. That dilemma of the reuse of research data stems from motives that are also (among others) rooted in the aim to increase the public good, and the dilemma is between potential benefits and potential harm. As the forces that create the dilemma are growing in concurrence with current trends in research and research financing, the dilemma commands some attention, even if it seems at first glance to be minor compared to issues related to business data mining and governmental data bases. The discussion ends with a possible solution, but the reader is encouraged to think about the dilemma and understand it rather than “solve” it. The novice can regard this as an introduction to the dilemma while the experienced researcher will view it as a summary. However, this sort of research need to be supported to uphold the ethical aspects of data mining and their various applications in the socio-economic development processes of a country, such as generic or specific researches, entrepreneurship development through innovation, and trading, commerce or e-governance through the utilization of innovative technologies.

INTRODUCTION

The dilemma described in this chapter occurs between two public interests, the first is the moral obligation to maximize the benefit for the public derived from the utilization of public funds; the second is the moral obligation to utilize

private information for the purpose presented as the background to which it was supplied. That second moral obligation is rooted not only in the moral obligation to use information gathered for research just for that specific reason and not for any other. Additionally there is another driving force for maximizing the research efficiency and

DOI: 10.4018/978-1-4666-4078-8.ch003

publication of its results. It is expressed in the academic world as the saying, “Publish or perish,” which is a well-known cultural proverb (Porath, 2012). It is understood not as a recommendation but as a way of life for academics. It signifies the increasing demand on researcher to publish increasing empirical or empirically supported research which is leading to a higher pressure to gather information, enlist bigger samples and to improve the statistical analysis. The pressure to use larger samples is often equated with stronger statistical significance. Gathering information is therefore, becoming more costly all the time due to the need for bigger samples. Added to that, is the need to perform multiple disciplinary researches that has led to a situation where more complex issues are researched in social sciences.

The factors mentioned above, the need for larger samples, more complex issues and the public lack of will to supply evidence is making the research performance more complex and more expansive. The increasingly expansive cost is measured both in effort and time as well as in actual expenditure.

This study illustrates through four specific examples, such as Research in Psychology, Academy Industry Links, R&D Financing and the Mutual Card the finer points of an ethical dilemma, discussed in it. The topic of government data bases where there are issues of personal as well as national security balanced versus individual and groups rights, is not discussed, but only mentioned here as it is felt that it deserves special care. See Example 1.

Example 1. Research in psychology

Assume a research in psychology requiring interviews which later on need to be transcribed and analyzed. An interviewer cost is 35\$ per hour (Master student) and each transcription of such an interview is another 20\$. If the research deals with a small sample presented as case studies (up to 5 cases) requiring each just one hour interview to deal with the specific research question, the cost would be under 300\$. However, if the research would like to present a statistically significance and the calculated sample would be over 75 cases the cost would raise to over 4,000 \$ and that is before additional cost raising requirements, such as administrating the entire operation including the coordination of interviewers and interviewees and the additional cost of statistical analysis.

While this situation is not research prohibitive, it is enough to make the researcher wish to extract the maximum possible benefit from each information gathering event.

For our discussion here, we can define a Research Gathering Event as both the distribution and collection of a questionnaire or the setting up of interviews and performing them. It includes also the effort invested in preparing for them, such as the validation of the questionnaires, the identification of the sample group and the calculation of the size of it. The relevant cost would include also the cost of the statistical analysis required in order to extract the answer to the research questions mentioned.

It is claimed that the increased transaction cost has two effects, one is the preference of case studies to large statistical researches; and the second is the aspiration to extract the maximum amount of information from each information gathering event. See Example 2.

Therefore, in order to make the most of the effort of gathering information we tend to gather more information than is strictly envisioned as required by the research questions and hypotheses. We create data collections tools (questionnaires, interviews etc.) that are more comprehensive than strictly required for the original research.

The creation of research data bases and placing them to the researchers benefit has been on-going for some time. The EU since FP5 has invested significant funds (EU, 2003; 2006; 2009), towards the creation and then toward assuring access to research infrastructure which includes

11 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/reuse-excess-research-data-new/76256

Related Content

Chinese Text Sentiment Analysis Utilizing Emotion Degree Lexicon and Fuzzy Semantic Model

Xing Wu and Shaojian Zhuo (2016). *Big Data: Concepts, Methodologies, Tools, and Applications* (pp. 1077-1090).

www.irma-international.org/chapter/chinese-text-sentiment-analysis-utilizing-emotion-degree-lexicon-and-fuzzy-semantic-model/150206

Clustering of COVID-19 Multi-Time Series-Based K-Means and PCA With Forecasting

Sundus Naji Alaziz, Bakr Albayati, Abd al-Aziz H. El-Bagoury and Wasswa Shafik (2023). *International Journal of Data Warehousing and Mining* (pp. 1-25).

www.irma-international.org/article/clustering-of-covid-19-multi-time-series-based-k-means-and-pca-with-forecasting/317374

Cluster-Based Input Selection for Transparent Fuzzy Modeling

Can Yang, Jun Meng and Shanan Zhu (2006). *International Journal of Data Warehousing and Mining* (pp. 57-75).

www.irma-international.org/article/cluster-based-input-selection-transparent/1771

Extraction of Medical Pathways from Electronic Patient Records

Dario Antonelli, Elena Baralis, Giulia Bruno, Silvia Chiusano, Naeem A. Mahoto and Caterina Petrigni (2013). *Data Mining: Concepts, Methodologies, Tools, and Applications* (pp. 1004-1018).

www.irma-international.org/chapter/extraction-medical-pathways-electronic-patient/73481

Towards Spatial Decision Support System for Animals Traceability

Marcos Visoli, Sandro Bimonte, Sônia Ternes, François Pinet and Jean-Pierre Chanet (2013). *Data Mining: Concepts, Methodologies, Tools, and Applications* (pp. 2094-2116).

www.irma-international.org/chapter/towards-spatial-decision-support-system/73536