Factors Influencing Research Data Reuse in the Social Sciences: An Exploratory Study

Renata Gonçalves Curty
Universidade Estadual de Londrina, Brazil

Abstract

The development of e-Research infrastructure has enabled data to be shared and accessed more openly. Policy mandates for data sharing have contributed to the increasing availability of research data through data repositories, which create favourable conditions for the re-use of data for purposes not always anticipated by original collectors. Despite the current efforts to promote transparency and reproducibility in science, data re-use cannot be assumed, nor merely considered a 'thrifting' activity where scientists shop around in data repositories considering only the ease of access to data. The lack of an integrated view of individual, social and technological influential factors to intentional and actual data re-use behaviour was the key motivator for this study. Interviews with 13 social scientists produced 25 factors that were found to influence their perceptions and experiences, including both their unsuccessful and successful attempts to re-use data. These factors were grouped into six theoretical variables: perceived benefits, perceived risks, perceived effort, social influence, facilitating conditions, and perceived re-usability. These research findings provide an in-depth understanding about the re-use of research data in the context of open science, which can be valuable in terms of theory and practice to help leverage data re-use and make publicly available data more actionable.

Introduction

Primary data has become the prime currency of science (Davis and Vickery, 2007). The broader availability and accessibility of research data is a fundamental item of the open science agenda, which aims to maximize the cost-effectiveness of socio-economic resources, enhance the utility and application of data beyond the focus or time constraints of the original data collectors, and promote better scrutiny, reproducibility, and transparency in science (Fienberg, Martin, and Straf, 1985).

In the quest to expand the availability of research data and comply with new governmental directives, a number of funding agencies, journal publishers, academic institutions, and research organizations started implementing mandates and deploying data repositories, initiating a call for research data sharing. As new data repositories are created to house research data, and more data accumulate in their servers, attention shifts to find ways to sustain the value of these research outputs and maximize their re-use within and across disciplines (Faniel and Zimmerman, 2011). The benefits of data sharing can only be reaped through data re-use (Niu, 2009b) because the value of data increases when scientists can make more use of it. In this sense, the sustainability of open science's life cycle depends on finding ways to maximize data re-use, rather than merely stock-pilling data assets to sit idle in data repositories.

Time, money and effort saving are widely acknowledged as key motivators for scientists to re-use research data (e.g. Castle, 2003; Hyman, 1972; Kiecolt and Nathan, 1985; Law, 2005). While recognizing that frugality is a quintessential driver for scientists to consider re-using research data, this study argues that data re-use cannot be seen simply as a 'thrifting' activity, where scientists shop around in data repositories for 'cheaper' second-hand data. The resource savings associated with data re-use accounts for only one aspect of it. The re-use of research data is a more complex process which requires scientists to have the ability to discover and access intelligible, trustworthy, and relevant data (Thessen and Patterson, 2011). Furthermore, and perhaps more importantly, it requires re-users to be capable of translating and re-contextualizing primary data collected by others in order to apply to their own purposes, without misinterpreting or misusing them.

Scientists seeking to re-use publicly available research data face the duality between the convenience of having ready data and the effort of dealing with data produced by someone else. On one hand, working with existing data has the advantage of significantly minimizing costs and time associated with data collection (Castle, 2003; Law, 2005). On the other hand, the re-use of already available data faces the constraints of dealing with data which were created under particular circumstances, following specific data collection procedures and techniques, in order to answer specific research questions (Boslaugh, 2007; Devine, 2003).

Still, very little is known about how scientists perceive the process of data re-use and the different factors that motivate and/or discourage them to make use of data collected by others (Wallis, Rolando and Borgman, 2013; Zimmerman, 2007; 2008). Even less is understood about how these factors affect not only scientists' intentions to re-use data, but also how they impact on the actual re-use of data. In spite of the wide recognition of the importance of data re-use in science, this issue has been addressed by the literature in a more peripheral way, as a desirable and expected outcome of data sharing practices, rather than a research phenomenon itself. The lack of a more

integrated view of how these factors collectively affect data re-use motivated the study reported in this paper, which aims to offer a clearer picture of the different influential factors (individual, social and technological) that may encourage or discourage scientists to re-use publicly available research data.

Literature Review

There are only a few empirical studies in the literature that investigated scientists' behaviours or perceptions towards data re-use as the central topic of interest. These studies can be grouped into two categories based on their research approach. One examines data re-use behaviour through citation analysis, and the other investigates data re-use behaviour or perceptions among scientists.

The first group of studies attempt to explain scientists' data re-use behaviour through bibliometric analysis. Examples for this approach include Piwowar (2008; 2010a; 2010b), Piwowar and Vision (2013), and Chao (2011, 2012), in which the authors tracked citations to datasets in Biomedical and Earth Sciences research publications. In short, these studies consider that citations and attributions to datasets are good measures for re-use and can be traced to demonstrate some patterns of scientists' data re-use behaviour. For example, Piwowar and Vision (2013) found that scientists who use openly available microarray data in their papers tend to be cited more than those who publish papers based on their own datasets. The citation analysis for data use focuses on the re-use outcomes and is suitable for finding out what data have been used, identify individual scientists' actual re-use behaviours, and in what disciplinary groups the re-use of data is more common. This approach, however, cannot explain how scientists decided to re-use the data, or capture the different nuances involved in this process.

Another approach investigates scientists' perceptions, experiences, and attitudes towards the re-use of research data. A common thread across the studies in this category is that data re-use varies according to the type of data and disciplinary community under question. Authors agree that there is no one-size-fits-all model to understand data re-use (Carlson and Anderson, 2007; Faniel and Jacobsen, 2010; Howard et al., 2010; Zimmerman, 2003, 2007). This is aligned with the fact that studies on this topic have predominantly focused on specific disciplinary fields or scientific communities, such as engineering (Howard et al., 2010), earthquake engineering (Faniel and Jacobsen, 2010), astronomy (Sands, Borgman, Wynholds and Traweek, 2012), social sciences (Faniel, Kriesberg and Yakel, 2012; Faniel et al., 2013b), meteorology (Kelder, 2005), and ecology (Zimmerman, 2003, 2007, 2008). Because these studies address very specific communities of scientists and consider particular tasks or contexts involving data re-use, it is difficult to compare the findings from this body of literature.

Not all studies in this group chose a single community to study data re-use practices; some examined a range of research projects in science disciplines, including Carlson and Anderson (2007), Davis, Alston, and D'Ignazio (2011), Borgman et al. (2012), Faniel et al. (2013a) and Kriesberg et al. (2013).

It also should be pointed out that previous empirical studies on data re-use behaviour are fundamentally atheoretical, which indicates that the problem presented by Zimmerman still persists: "little direct research or theory exists on the sharing and re-use of data, and this makes it difficult to identify variables or to state research hypotheses" (Zimmerman, 2003). Due to the novelty of the topic, and the relatively

incipient current literature about this emerging phenomenon, an exploratory study was necessary to better understand and elucidate factors related to data re-use among scientists. By exploring scientists' experiences about the re-use of research data, this study aimed to examine how scientists assess the re-usability of data collected by others and what factors they perceive as determinants when deciding whether to re-use the data or not.

Methodology

The target population consisted of social scientists. This decision was made based on the possibility to draw out more nuance and variation in the process of data re-use than in other disciplines. Social sciences include a wide range of sub-disciplines, and this diversity is particularly useful to provide examples of data re-use practices with reference to heterogeneous and different types of research data, originated by a variety of scientific methods and grounded in a rich spectrum of disciplinary traditions. In addition, research data in the social sciences are generally intensive, contextual and time-dependent, which means they are expected to require extra effort from re-users to preserve the data interconnectedness and reflexivity necessary to guarantee their understandability and informative value (Friedhoff et al., 2013; Jacoby, 2010).

The recruitment process required participants to be knowledgeable about the topic and familiar with the process of re-using research data. In this sense, the rule was that potential subjects had to have at least attempted to re-use third parties' primary research data once for the purpose of their own research; regardless if the final outcome resulted in concrete re-use of data or not.

A non-probabilistic purposive sampling (Tashakkori and Teddlie, 1998) technique was applied in order to recruit potential participants with characteristics relevant to the study and who would be the most informative. The strategy adopted for the recruitment process was to send out participation calls to registered users of social sciences data repositories. This recruitment phase was supported and mediated by two data repositories with one of the largest collection of social sciences data hosted in the United States: Interuniversity Consortium for Political and Social Research (ICPSR) and Harvard Dataverse Network. Three rounds of participation requests were submitted by each of the two data repositories facilitators. A total of 13 social scientists (seven men and six women) were interviewed: 11 that were reached out via ICPSR and two via Dataverse. However, during interviews some participants mentioned have used both data repositories.

Interviewees were affiliated to different academic and research institutions located in different states of the United States, including New York, Oregon, Massachusetts, Pennsylvania and Mississippi. Participants belong to different academic sub-disciplines within the social sciences, including economics, political sciences, sociology, communication and social media, child and family studies, clinical psychology, and public administration and international affairs.

An interview protocol was developed to guide some initial questions which were complemented by follow-ups and probes during the interviewing process. Questions followed a funnel interviewing fashion, where participants were initially asked to talk more broadly about their research agendas and areas of study, as well as about their general understanding about data re-use in science and in their discipline. Then, the interview moved to more specific questions about their own experiences of re-using or attempting to re-use research data, allowing for a better understanding and interpretation of their narratives.

All interviews were audio-recorded for the convenience of transcription, with a total of seven hours and 55 minutes of audio recording. The average length of an interview was 40 minutes. Different interview modes were used to accommodate geographical barriers, as well as participants' preferences and availability. Five interviews were conducted face-to-face, five via Skype video calls, and three by phone.

Interviews were transcribed using the free transcription software Express Scribe. Interviews' transcripts were uploaded to QSR NVivo 10 for the convenience of data organization and coding scheme development. Interviews were coded respecting a data driven and inductive bottom-up approach. Therefore, the coding process did not follow any preconceived scheme of coding categories, which were developed solely by the researcher. Initially, the coding was focused on some patterns in the responses according to more obvious and general topics related to some of the questions, such as factors that discourage social scientists from re-using someone else's data or motivators for doing so. Different rounds of close scrutiny of the transcripts were performed to both identify emerging themes and group similar coding occurrences within and across interviews.

Findings

Interviews revealed a collection of factors that were found to influence social scientists' practices regarding data re-use. Six major categories (theoretical variables) were created to represent the different emerging themes identified in interviews about social scientists' data re-use experiences: a) perceived benefits, b) perceived risks, c) perceived effort, d) re-usability assessment and judgment, e) enabling factors, and f) social factors. Data analysis revealed a total of 25 codes and 430 utterances of these codes across the 13 interviews (see Table 1, Appendix).

Perceived Benefits

Perceived benefits represent factors that interviewees mentioned as personal motivations and/or motivations that they believe are significant for scientists to attempt re-using data collected by others in their own research.

Knowledge expansion

Social scientists find the re-use of data beneficial to yield new discoveries and contribute to the development of a particular field through the re-use of data. This aspect converges with the idea of 'benefits for theory and substantive knowledge' presented by Hyman (1972) who describes the ability of scientists to widen intellectual horizons through secondary analysis. Hyman postulates that the examination of the wide array of materials through the course of re-using data expands the intellectual horizons of researchers, and consequently their field of study. Researchers are stimulated to think about otherwise forgotten problems and to think in a direction of higher level of abstraction (Hyman, 1972). A similar idea was expressed by Frank:

> '…someone else can look at it [data] some other time, some other place, maybe with completely different names and objectives or tools than someone else, and get new information, you know, this is data, it's old data. But we analyse it in a different way and we get new information… you know, something applicable, a new idea from using this old data' (Frank,

Social Sciences).

Frugality

Social scientists perceive data re-use as a way to circumvent problems associated with primary data collection and gathering, including the reduction of time and effort needed for obtaining data, as well as the notion of minimization of duplicated efforts and necessary skills to perform data collection. Social scientists believe that data re-use is beneficial because it is an opportunity to obtain existing data that would have been difficult to obtain through a new primary data collection endeavour. The notion of frugality as a benefit and driver associated with data re-use has been substantially addressed in the literature. Some examples are Kiecolt and Nathan (1985) who articulate on the notion of secondary data analysis as a resource saving activity. Hyman (1972) emphasizes that re-use of existing data economizes money, time and personnel. Law (2005) associates the re-use of data to parsimony and Castle (2003) elaborates on the possibility of re-users to count on data collection skills from more experienced researchers. Likewise, this was one of the most recurrent factors across interviews:

> 'It takes a lot of effort to collect data, and as an experimental researcher I know that, so certainly the availability of already existent data can make the whole process of studying something quicker' (Cindy, Mass Communication).

Pre-endorsement

Social scientists perceive data re-use as beneficial because data available for re-use are considered to some extent credible and reliable, otherwise they would not be shared and available to the public and to be subject to scrutiny and verification. This perception of endorsement is found in Ellen's comment:

> 'Most of the data that gets upload to ICPSR or other repositories are collected in some way or the other, you know...probably they got grants, which were evaluated by their peers, and they are weighed to some extent' (Ellen, Child and Family Studies).

Perceived Risks

Perceived risks are considered as foreseeable, harmful consequences associated with the re-use of research data. Four types of risks were found in interviewees' narratives.

Fear of being undervalued

When re-using other people's data social scientists might fear that their work would receive less credit in comparison with scientists who conducted primary data collection and used original data in their research. Goodwin (2012) elaborates on this matter, indicating that social scientists, especially with a qualitative approach to inquiry, hold a general belief of undervalue towards data re-use. Martin (1995) and Fahs, Morgan and Kalman (2003) indicate that especially in cases of replication studies there is a sense

that data re-use might not be as highly regarded as a research activity. This issue was presented in Denise's opinion:

> 'There was a value issue going on, yeah, there was actually a value issue and it was not as respected as collecting my own data, doing that [re-using existing data]' (Denise, Child and Family Studies).

Fear of infringing ethical codes

Social scientists might hesitate to re-use existing data generated by other researchers if they perceive risks associated with the consent and approval for conducting the study, which was granted only to the original data collectors. Where sensitive data is involved, informed consent cannot be simply presumed by re-users and there is a need to verify whether the re-use of data violates the contract established between subjects and primary investigators (Heaton, 2004). Additionally, copyright and confidentiality issues might be unclear for data re-users (Heaton, 2004). Grinyer (2009) and Law (2005) indicate that this might be a result of a lack of clarity with regard to codes of ethical conduct, especially for qualitative data archived for future re-use. The distinction between 'once-and-for-all' consent and the need for renewed consent for re-use are not always well defined and apparent to re-users.

> 'There are some datasets that require, you know...confidentiality I would say, that have confidentiality requirements. So, but you need to find that re-using those datasets require confidentiality arrangements' (Ellen, Child and Family Studies).

Slippage

Social scientists are concerned with potential misinterpretation, incorrect or unintentional misuse that might result from re-using someone else's data. Kuula (2010) describes the misuse as one major concern of social scientists. Corti and Thompson (2004) explain that 'concerns about misinterpretation of data may arise from fear of selective and opportunistic interpretation in reanalysis.' In other words, the process of trying to explore data in different ways may cause re-users to make wrong assumptions based on data, because data is pulled out from its original framework. Tenopir et al.'s (2011) study on data sharing practices among scientists identified the issue of misinterpretation and misuse of data as a concern of scientists when asked about their views on the use of data across their research field. This particular issue was raised by Cindy:

> 'I kind of have the same concerns about someone using my data as I would have for that data I am using... I would certainly want to know if they are following certain standards and things like that, and that they are not misusing it' (Cindy, Mass Communication).

Vulnerability to hidden errors

Data collected by others might contain hidden errors that are not easily identifiable by re-users. The idea of hidden errors is articulated by Kiecolt and Nathan (1985) as a potential risk for re-users. Similarly, Hyman (1972) asserts that while conducting

analysis of existing research data, re-users face difficulties in detecting errors. Castle (2003) comments on this matter, emphasizing that data re-users experience a lack of control over data quality.

Perceived Efforts

Perceived efforts refer to the amount of work that social scientists estimate that they have to face in dealing with data they did not produce or collect themselves.

Being innovative with old data

When social scientists consider re-using data, they take into account that they have to invest effort in identifying new ways to approach old data that would differentiate their work from the original research and/or subsequent re-uses of that particular work. Zimmerman (2008) discusses the idea of new knowledge from old data. She found that scientists not only devote attention to understanding the data they have at hand, but that they also have to look for ways to expand science from old data. This notion was explicit in Michael's comment:

> 'If it's a sort of publicly used dataset sometimes I get the feeling that, well, [it is] someone else's big work, they've already asked these questions, somebody must have had more time than me…so there's at least this perception that I'm competing with a bunch of people over the same data' (Michael, Social Media and Communications).

Obtain access to data

The access to datasets varies depending on who owns and control access, where data is held and in which format (Heaton, 2004). The process of obtaining access to data is recognized by social scientists as an expenditure of effort in the re-use of data. Faniel et al. (2013b) found that the easiness of access to data was the strongest predictor of data re-use satisfaction. Tenopir et al. (2011) also found that scientists indicate strong interest in using datasets from other researchers, if the data were easy to access. This particular factor can be illustrated with an excerpt of Denise's interview, in which she describes the efforts of gaining access to data. In particular, she mentions two distinct circumstances in this passage, one positive and other negative, with the latter concerning a dataset with some restricted data:

> 'I expected that to access and use restricted data would be not a complicated process...and it took six months…So you have to apply for the data. So, that process was six months. That was horrible!' (Denise, Child and Family Studies).

Data discovery process

Social scientists perceive that there is effort required to discover data for potential re-use. Faniel and Majchrzak's (2002) found that the effort associated with search to be one of predictors for re-use. Darby et al. (2012) also reveal that the ease of discovery plays an important role in scientists' willingness to re-use data or not. Adam explained that sometimes data discovery involves a set of activities, and thus, more effort to discover data:

'It involves making phone calls and looking around in the web and figuring it out where I can find the information that I need to answer that particular question' (Adam, Economics).

Dealing with mismatches

Social scientists recognize that the re-use of data implies a devotion of some efforts to deal with mismatches between the data they have at hand and the data they wish to have in order to answer their research questions. The primary data was collected under particular circumstances, in a given context, time-frame, and in order to investigate particular issues, meaning it rarely captures all the elements that re-users would collect if they had the chance. Kiecolt and Nathan (1985) indicate that re-users usually experience a mismatch between primary and secondary research objectives. Dale (2004) describes how re-users have to adapt to the available data: 'because the data have been collected by another researcher, the secondary analyst will have had no opportunity to influence the questions asked or the coding frames used, and this important factor must be borne in mind at all stages of analysis.' Ivan mentioned this idea of mismatch:

'One of the limitations with working with secondary data is that...you know, you didn't ask the questions yourself, you didn't write the questions yourself, some of the wording might be, you know, not what you would have asked...or it is similar to what you would like to know, but not exactly not what you want to know' (Ivan, Sociology).

Preparation for re-use

Social scientists recognize that the re-use of data often requires additional work prior to the analysis. This includes screening data, formatting it in a particular way, deciding how to manage missing data as well as complement data in cases where they combine different data sources. Faniel et al. (2013b) study, found that the ease of operation plays an important role on ones' willingness to re-use existing data. In this sense, the recognition of the additional efforts associated with re-use might inhibit scientists from re-using data. On this matter, Jen commented:

'I would have to do individual manual recoding of the observations from natural disasters, into natural disasters in the state and specific year, and that was just overwhelming. Even like with shortcuts, it is just too much' (Jen, Political Sciences).

Understanding the original study

Social scientists recognize that the process of making sense of data produced by others requires extra effort to gain a thorough comprehension of the original study; that is, the study from which primary data is derived. Scientists invest a great deal of their time to analyze the study in order to avoid the potential of inadvertent slippage. Faniel and Jacobsen (2010) found that scientists invest a significant amount of effort seeking confidence that data is fully understood. Boh (2008) found that asset complexity is an important factor for one's intention to re-use a knowledge asset. She particularly emphasizes the time required for re-users to understand an asset in order to determine

'how the ideas can be adapted to meet the problem at hand.' This particular type of effort was brought up by Michael:

> 'I recall it was a little bit of a task [in] itself to interpret what questions had been asked at what times. You know I had to kind of comb through some of the actual files of the survey' (Michael, Social Media and Communication).

Reusability Assessment

When asked about important factors they consider prior to re-using someone else's data and during the description of the events of data re-use they experienced, social scientists disclosed different attributes of data they consider when deciding to re-use data produced by other researchers. Data re-usability means the condition of being re-usable, as appraised by the potential re-user, who relies on their best judgment about the attributes of the data. In other words, data has to possess certain characteristics to be considered re-usable. Such attributes were found to be important to social scientists and are expressed in the different factors described below.

Data documentation

When considering re-using data, social scientists tend to judge whether data documentation is of good quality; that is, if the documentation is sufficiently complete and clear. Data documentation includes a variety of supplementary materials that have the function of supporting the understanding and re-use of the data. Data documentation may vary depending on the type of data, and can include: codebooks or data dictionaries, reports about the data collection process, data collection instruments, previous publications based on the data, user guides or handbooks, statistical manuals, data extraction software, and institutional review board (IRB) documents. David (1991) emphasizes the role of documentation in data re-use and the risks of failure or induced avoidable error in secondary analysis caused by poor data documentation. Many scholars have addressed the importance of documentation quality as a condition of research data re-use, both conceptually and empirically (e.g. Faniel et al., 2013b; Zimmerman, 2003; 2007; 2008; Niu, 2009a; 2009b; Niu and Hedstrom, 2009). Ivan expressed this reusability concern:

> 'It was really kind of confusing and kind of hard to find what you are looking for in terms of the explanations and all the variables...working with [a] dataset that you know has 40 years of collection is really difficult, so there is no perfect way of doing it... but the codebook was kind of a mess as far as I am concerned' (Ivan, Sociology).

Data fitness

When social scientists consider re-using existing data, they examine different factors such as the topic, the level of analysis, and the type of data in order to help them to judge whether data is suitable or not to their purpose (Hinds, Vogel and Clarke-Steffen, 1997). Palmer, Weber and Crager (2011) assimilate fit for purpose and the attribute of utility of data. Faniel et al. (2013b) use the attribute of data relevance to represent data fitness. The need of data to fit some or all the criteria mentioned above was consistent in the interviewees' narratives, as illustrated by Adam's comment:

'If they have the variables that you are looking for, the level of observations it is probably the best what you are gonna [sic] do' (Adam, Economics).

Data producer trustworthiness and credibility

Before considering re-using data, social scientists tend to evaluate how trustable and credible data producers are. There is an extensive body of literature available about information credibility from communication studies on information credibility. Carlson and Anderson (2007) consider data trustworthiness and the credibility of data provenance as critical factors for data re-use and Darby et al. (2012) underscore the importance that re-users 'feel safe' about second-hand data. Assessing the credibility of a source might be important for the initial stage of re-use, but not so much for the evaluation of the re-use outcome (post-fact). With regard to trustworthiness and credibility, some interviewees not only underscored the importance of trusting in the data producers, but also expressed their preference for datasets institutions produced by institutional groups rather from individual researchers, as illustrated below:

'It would just be easier to convince reviewers that the data that came from the big research institutions. […] The individual research you know there's a certain amount of trust there and I can't know for sure what processes are clinical and followed by that researcher. […] I usually would prefer somebody from an established institution' (Michael, Social Media and Communication).

Data quality

There is also a sense among social scientists about how consistent data are perceived to be. Data documentation overlaps with data and the distinctions between them are not always easy to capture (Niu, 2009a; 2009b). Nonetheless, the data quality factor grouped interviewees' perceptions when they described attributes more directly to the dataset level, rather than the supplementary materials. Data quality represents the attributes of data in terms of consistency and completeness. While consistency refers to how accurate data is perceived to be, completeness refers to no or minimal missing data (Hinds, Vogel and Clarke-Steffen, 1997). Faniel et al.'s (2013b) study on data completeness, found that after data accessibility, data quality is the strongest contributor to data re-use satisfaction among quantitative social scientists. As observed in Beth's comment, there is often an association between missing data and quality, and these factors directly contribute to the reusability judgment.

'So, first thing I will look at missing data. If they have many dots, then I know data are not available or not applicable...means that it is not a very good dataset because there are only couple variable cases' (Beth, Political Sciences).

Study rigor

When re-using someone else's data, social scientists also consider the original study design and execution. Hinds, Vogel and Clarke-Steffen (1997) indicate that a prime question that should be asked by re-users is about how well the study was designed and executed. The

first apparent indicator of study rigor might be expressed in the dataset itself, and data documentation is the prime source for re-users to evaluate and understand the study. The factor "study rigor" groups interviewee's perceptions outside the artifacts they have in hand (data and supplementary materials) and the evaluation of these in terms of quality. These are the materialized forms through which re-users can understand and interpret the study. On the other hand, the assessment of how rigorous the study is, represents their overall judgment of appropriateness of methods and procedures, as well as the transition between goals/objectives, methods, and outcomes. Adam highlighted this issue:

> 'Another would be, you know, if I think [the way] they collected the data was done in a rigorous way' (Adam, Economics).

Enabling Factors

Enablers are facilitators which provide some of the necessary conditions and infrastructure that, according to interviewees, facilitate the re-use of data.

Data documentation availability

In the re-usability assessment category, data documentation is evaluated in terms of completeness, organization and clarity. A number of authors have addressed the essential role of data documentation to enable re-use (e.g. David, 1991; Pigott, Hobs and Gammack, 2001; Markus, 2001; Niu, 2009a; 2009b; Niu and Hedstrom, 2009; Zimmerman, 2003; 2007; 2008). For example, Markus (2001) elaborates on the importance of dissimilar others. In her view, documenting for future re-use is a challenging task because applications cannot be fully anticipated. However, without data documentation providing the rationale for the digital assets, chances of re-use are heavily compromised. Comments about the importance of the existence of data documentation were considered enabling factors. When data is openly available for re-use, it can be subject to countless applications and re-uses. Thus, some interviewees emphasized the importance of having access or means for accessing an updated list of studies which have re-used a particular dataset. Not only they believe it is helpful to know what was already done in terms of research beyond the original study with that particular dataset, but also they see it as an opportunity to identify gaps or potential opportunities for new research. Similarly, Faniel et al. (2013a) found in a study with quantitative social scientists that they tend to look for related literature written by data producers as well as for articles written by other re-users to see how the dataset was critiqued and re-used in different ways. As indicated above, the notion of re-users competing over the same data is a part of social scientists' concern about how to be innovative with old data. Some interviewees' emphasized the importance of having access or means to access different studies that have re-used datasets. Niu (2009a, 2009b) states that related bibliographies should be provided along with data documentation in order to optimize the re-use process. ICPSR offers the users of the platform the possibility of accessing a list or related studies with this purpose. This enabler was indicated by Denise:

> 'I did read a lot and I read dissertations that were done with the dataset just to see kind of the scope of how people used the dataset, so this was really helpful to me... the research that was done with the dataset was incredibly important to me' (Denise, Child and Family Studies).

Data repositories availability

Data repositories are vessels of data for potential re-use. Considering the current state of digital scholarship, relying on informal ad-hoc mechanisms for data sharing and re-use is not effective. The importance of data repositories as a central technological infrastructure for data sharing and future re-use of data assets is well-recognized by academics (e.g. Markus, 2001; Borgman, 2007; Marcial and Hemminger, 2010; Tenopir et al., 2011). Markus (2001) discusses the role of repositories, particularly in the business realm. However, she recognizes that repositories are determinants of failure or success of re-use in different contexts. Cindy, who said she never had the opportunity to find data in her area of research, highlighted the importance of data repositories:

> '[A] data repository of data that could be shared in some way, even from experiments, that you know, more people could ask certain questions and I could look at other items there, that would be really valuable' (Cindy, Mass Communication).

Primary investigators reach

Social scientists recognize the importance of having the technological infrastructure of data repositories to facilitate the re-use of data, but they also highlight the importance of establishing communication with primary investigators (data collectors) in order to gather a better understanding of the nuances behind the study, which cannot be always easily attainable from the analysis of the data and the provided documentation. Boh (2008) highlights that complementary person-to-person interactions between authors/collectors of data assets and data re-users are desirable and facilitate re-use, especially in circumstances of high-complexity data. Similarly, Faniel and Zimmerman (2011) recognize the value of social exchange between data producers and re-users, but indicate that this social exchange is difficult to accomplish on large scale. In spite of that, some interviewees described the process of contacting primary investigators to clarify or request additional information in order to re-use data. An example is illustrated in Adam's narrative:

> 'There were some institutional details that I wanted to know about how bails worked in Philadelphia in the sense that these people weren't talking that much and we weren't able to find some of their files, so I asked them about some of the files that weren't included [in the study files in the repository] […] and they gave me a lot of the institutional details that I needed' (Adam, Economics).

Support and assistance availability

Markus (2001) and Behboudi and Hart (2008) discuss the role of human intermediaries for re-use and their particular importance in re-use success. In a comparative case study between archeologists and social scientists, Faniel et al. (2013a) found that both disciplines rely on human intermediaries to re-use data. Data producers can be considered human intermediaries, but here support and assistance aggregates the more formal type of support provided by the institution the re-user is affiliated with and the data repository from which he/she obtained data. This support was expressed in Nathan's comment:

'I really needed a lot of external support for the data preparation and data
analysis process. So, you know, went to like research and stats camps, I
asked for a lot of help from statisticians from the thesis department, so they
have staff support there that I actually had' (Nathan, Clinical Psychology).

Training and expertise

The re-use of data minimizes data collection skills, but demands ability and
experience in data analysis. Hyman (1972) emphasizes that skills should be built
through methodological training. Hyman focuses primarily on the importance of
statistical knowledge, considering that his arguments are centered on the secondary
analysis of survey data. Corti and Bishop (2005), on the other hand, explore the need for
techniques and skills development for the re-use of qualitative data and the importance
of training programs to both build awareness of the general opportunity for data re-use
and prepare more scientists to re-use data. Similarly, Kriesberg et al. (2013) articulate
on the importance of formal training on data re-use, especially for novice scholars. With
regard to the importance of training, Denise provided a retrospect of her learning
process after conducting data re-use without previous training. Frank also spoke about
the set of skills required for the re-use data.

'I think [secondary data analysis] is a huge skill set that is different and it is
your own skill set and there are a lot of strengths to it and there are also
challenges like there would be in any research' (Denise, Child and Family
Studies).

Social Factors

Social factors correspond to elements of the scientists' social environment which can
influence their intention to re-use data. When asked about their opinion regarding data
re-use as a scientific practice, interviewees disclosed two aspects related to their social
environment they consider important when deciding whether to re-use data or not: their
discipline and their peers.

Disciplinary receptiveness

There is a general assumption that some fields and disciplines are keener or more
receptive to re-use second-hand data than others (Borgman, 2007; Faniel and Jacobsen,
2010; Thesen and Patternson, 2011). Even though this small-scale exploratory study did
not aim to conduct any sort of comparative analysis between different disciplines within
the social sciences, some questions asked about their general views on the re-use of data
in science revealed how open they perceive their discipline to be towards the re-use of
data.

'Sociology is always trying to defend itself as 'a' Social Sciences, you know
it is really changing to more positivistic scientific approaches, which is
fine... you know I am all for it... I mean… quantitative is generally... data is
generalizable to the agenda... and obviously the results are certainly more
valid... so that is why I really want to explore secondary sources...specially,
you know, public available data' (Ivan, Sociology).

Peer encouragement

The idea of support from peers was represented in some interviews. Most of the comments came from PhD student or candidate interviewees while describing a specific scenario where a professor or senior researcher recommended he/she look at a particular dataset and consider it to the study.

> 'Somebody else in my lab had been using the data from the MIT, Harvard, MIT data center and so my professor introduced me to the center and this really rich dataset' (Nathan, Clinical Psychology).

This notion of peer encouragement relates to some extent to the discussion about the role of senior researchers in modelling general re-use practices for novice scholars (Kriesberg et al., 2013) and the idea that research peers can influence social scientists' decisions to re-use data or not.

Conclusions

Empirical findings support the assumption that scientists take into account a combination of factors beyond frugal motivations when considering whether to re-use other people's research data. These preliminary findings suggest that frugality accounts for only one of the aspects identified amidst other dimensions of benefits and factors social scientists associate with the re-use of research data. The results from this preliminary exploratory study allow us to infer that more than merely 'thrifting' for available research data, by considering advantages such as the time, resources, and money saving when working with existing data, social scientists also weigh different conditions that they judge as relevant before re-using data collected or generated by others. These conditions include other benefits and potential harms associated with the re-use of data, the perceived re-usability of data, the effort required to deal with data they have not collected themselves, the availability of technical and personnel support to facilitate the data re-use process, and how receptive their peers and research field are with regard to research based on secondary/existing data. Findings from this exploratory study offer an important initial conceptual foreground for future research on research data re-use behaviour.

References

Behboudi, M. & Hart, D. (2008). Human knowledge intermediaries for successful knowledge re-use and sharing. Paper presented at the 5th International Conference on Intellectual Capital, Knowledge Management and Organizational Learning (ICICKM), New York.

Boh, W.F. (2008). Re-use of knowledge assets from repositories: A mixed methods study. *Information and Management, 45*(6), 365-375. doi:10.1016/j.im.2008.06.001

Borgman, C.L. (2007). *Scholarship in the digital age.* Cambridge: The MIT Press.

Borgman, C.L., Wynholds, L.A., Wallis, J.C., Sands, A. & Traweek, S. (2012). Data, data use, and scientific inquiry: Two case studies of data practices. *ACM Press, 19*. doi:10.1145/2232817.2232822

Boslaugh, S. (2007). An introduction to secondary data analysis. Secondary Data Sources for Public Health: A Practical Guide, 2-10.

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. doi:10.1111/j.1083-6101.2007.00342.x

Castle, J.E. (2003). Maximizing research opportunities: Secondary data analysis. *Journal of Neuroscience Nursing, 35*(5), 287-290. Retrieved from http://journals.lww.com/jnnonline/Citation/2003/10000/Maximizing_Research_Opportunities__Secondary_Data.8.aspx

Chao, T.C. (2011). Disciplinary reach: Investigating the impact of dataset re-use in the earth sciences. *Proceedings of the American Society for Information Science and Technology, 48*(1), 1-8. doi:10.1002/meet.2011.14504801125

Chao, T.C. (2012). Exploring the rhythms of scientific data use. Paper presented at the Proceedings of the 2012 iConference.

Corti, L. & Bishop, L. (2005). Strategies in teaching secondary analysis of qualitative data. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research, 6*(1). Retrieved from http://nbn-resolving.de/urn:nbn:de:0114-fqs0501470

Corti, L. & Thompson, P. (2004). Secondary analysis of archived data. Qualitative Research Practice. Thousand Oaks, CA: SAGE Publications, 297-314.

Curty, R.G. (2015). Beyond "data thrifting": An investigation of factors influencing research data re-use in the social sciences (Doctoral thesis, Syracuse University, Syracuse, New York). Retrieved from http://surface.syr.edu/etd/266/

Dale, A. (2004). Secondary analysis of quantitative data. In M. Lewis-Beck, A.E. Bryman & T.F. Liao (Eds.), The Sage Encyclopedia of Social Science Research Methods (Vol. 1). Thousand Oaks, CA: SAGE Publications.

Darby, R., Lambert, S., Matthews, B., Wilson, M., Gitmans, K., Dallmeier-Tiessen, S., … Suhonen, J. (2012). Enabling scientific data sharing and re-use. Paper presented at the IEEE 8th International Conference on e-Science.

David, M. (1991). The science of data sharing: documentation. In J.E. Sieber (Ed.), Sharing social science data: Advantages and challenges (pp. 91-115): Thousand Oaks, CA: SAGE Publications.

Davis, L.K., Alston, P., & D'Ignazio, J. (2011). Repurposing data across disciplines: a study of data reuse issues between climate science and social science. In Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries (pp. 433-434). doi:10.1145/1998076.1998171

Davis, H.M. & Vickery, J.N. (2007). Datasets, a shift in the currency of scholarly communication: Implications for library collections and acquisitions. *Serials Review, 33*(1), 26-32. doi:10.1016/j.serrev.2006.11.004

Devine, P. (Ed.) (2003) The AZ of social research. Thousand Oaks, CA: SAGE Publications.

Fahs, P.S., Morgan, L.L. & Kalman, M. (2003). A call for replication. *Journal of Nursing Scholarship, 35*(1), 67-72. doi:10.1111/j.1547-5069.2003.00067.x

Faniel, I. & Majchrzak, A. (2002). An exploratory study of the factors associated with successful re-use of others knowledge. Paper presented at the Americas Conference on Information Systems (AMCIS).

Faniel, I.M. & Jacobsen, T.E. (2010). Re-using scientific data: How earthquake engineering researchers assess the re-usability of colleagues' data. *Computer Supported Cooperative Work (CSCW), 19*(3), 355-375. doi:10.1007/s10606-010-9117-8

Faniel, I.M. & Zimmerman, A. (2011). Beyond the data deluge: A research agenda for large-scale data sharing and re-use. *International Journal of Digital Curation, 6*(1), 58-69. doi:10.2218/ijdc.v6i1.172

Faniel, I.M., Barrera-Gomez, J., Kriesberg, A. & Yakel, E. (2013a). A comparative study of data re-use among quantitative social scientists and archaeologists. Paper presented at the iConference Fort Worth.

Faniel, I.M., Kriesberg, A. & Yakel, E. (2012). Data re-use and sensemaking among novice social scientists. Paper presented at the ASIST 2012 Annual Meeting, Baltimore, MD.

Faniel, I.M., Yakel, E., Kriesberg, A. & Daniels, M. (2013b). Can quantitative social scientists get data re-use satisfaction? Panel presented at the Research Data Access and Preservation Summit, Baltimore, MD.

Fienberg, S.E., Martin, M.E. & Straf, M.L. (Eds.). (1985). Sharing research data. Washington, DC.: National Academy Press.

Friedhoff, S., Meier zu Verl, C., Pietsch, C., Meyer, C., Vompras, J,. & Liebig, S. (2013). Replicability and comprehensibility of social research and its technical implementation. Working Paper Series of the German Data Forum (RatSWD) (Vol 219, pp. 1-27): German Data Forum (RatSWD).

Goodwin, J. (2012). The secondary analysis of qualitative data. Thousand Oaks, CA: SAGE Publications.

Grinyer, A. (2009). The ethics of the secondary analysis and further use of qualitative data. *Social Research Update, 56*(4). Retrieved from http://sru.soc.surrey.ac.uk/SRU56.pdf

Heaton, J. (2004). Reworking qualitative data. Thousand Oaks, CA: SAGE Publications.

Hinds, P.S., Vogel, R.J. & Clarke-Steffen, L. (1997). The possibilities and pitfalls of doing a secondary analysis of a qualitative data set. *Qualitative Health Research, 7*(3), 408-424. doi:10.1177/104973239700700306

Howard, T., Darlington, M., Ball, A., Culley, S. & McMahon, C. (2010). Opportunities for and barriers to engineering research data re-use. University of Bath. Retrieved from http://opus.bath.ac.uk/21166/

Hyman, H.H. (1972). Secondary analysis of sample surveys: Principles, procedures and potentialities. New York: John Wiley & Sons.

Jacoby, J. (2010). Share and share alike? Data sharing practices in different disciplinary domains. In S.W. Witt & L.M. Rudasill (Eds.), Social Science Libraries: Interdisciplinary Collections, Services, Networks (pp. 79-94). The Hague: IFLA Publications.

Kankanhalli, A., Lee, O.-K.D. & Lim, K.H. (2011). Knowledge re-use through electronic repositories: A study in the context of customer service support. *Information and Management, 48*(2), 106-113. doi:10.1016/j.im.2011.02.002

Kelder, J.A. (2005). Using someone else's data: Problems, pragmatics and provisions. *Forum: Qualitative Social Research, 6*(1). Retrieved from http://nbn-resolving.de/urn:nbn:de:0114-fqs0501396

Kiecolt, K.J. & Nathan, L.E. (1985). Secondary analysis of survey data. Beverly Hills: SAGE Publications.

Klein, G., Phillips, J.K., Rall, E.L. & Peluso, D.A. (2007). A data-frame theory of sensemaking. New York: Lawrence Erlbaum Associates.

Kriesberg, A., Frank, R., Faniel, I. & Yakel, E. (2013). The role of data re-use in the apprenticeship process. Paper presented at the 76th Annual Meeting of the Association for Information Science and Technology, Montreal.

Kuula, A. (2010). Methodological and ethical dilemmas of archiving qualitative data. IASSIST Quarterly, 34(3–4/1–2), pp.12-17. Retrieved from http://iassistdata.org/sites/default/files/iq/iqvol34_35_kuula.pdf

Law, M. (2005). Reduce, re-use, recycle: Issues in the secondary use of research data. *IASSIST Quarterly, 29*(1), 5-10. Retrieved from http://www.iassistdata.org/sites/default/files/iqvol291law.pdf

Marcial, L.H. & Hemminger, B.M. (2010). Scientific data repositories on the Web: An initial survey. *Journal of the American Society for Information Science and Technology, 61*(10), 2029-2048. doi:10.1002/asi.21339

Markus, M.L. (2001). Toward a theory of knowledge re-use: Types of knowledge re-use situations and factors in re-use success. *Journal of management information systems, 18*(1), 57-94. Retrieved from http://dsi.esade.edu/theorybuilding/papers/01-Markus%20Knowledge%20Reuse.pdf

Martin, P. (1995). More replication studies needed. *Applied nursing research: ANR, 8*(2), 102. doi:10.1016/S0897-1897(95)80574-5

Niu, J. & Hedstrom, M. (2009). Documentation evaluation model for social science data: An empirical test. *Proceedings of DigCCurr2009 Digital Curation: Practice, Promise and Prospects, 125*. Chapel Hill, NC. doi:10.1002/meet.2008.1450450223

Niu, J. (2009a). Overcoming inadequate documentation. *Proceedings of the American Society for Information Science and Technology, 46*(1), 1-14. doi:10.1002/meet.2009.145046024

Niu, J. (2009b). Perceived documentation quality of social science Data. (Doctor of Philosophy Dissertation), University of Michigan, Ann Arbor, MI. Retrieved from http://deepblue.lib.umich.edu/bitstream/handle/2027.42/63871/niujf_1.pdf?sequence=1

Palmer, C.L., Weber, N.M. & Cragin, M.H. (2011). The analytic potential of scientific data: Understanding re–use value. *Proceedings of the American Society for Information Science and Technology, 48*(1), 1-10. doi:10.1002/meet.2011.14504801174

Pigott, D., Hobbs, V. & Gammack, J. (2001). An approach to managing repurposing of digitised knowledge assets. *Australasian Journal of Information Systems, 9*(1). doi:10.3127/ajis.v9i1.217

Piwowar, H. & Vision, T.J. (2013). Data re-use and the open data citation advantage. 1(e175). doi:10.7717/peerj.175

Piwowar, H.A. (2008). Proposed foundations for evaluating data sharing and re-use in the biomedical literature. *Bulletin of IEEE Technical Committee on Digital Libraries, 4*(2). Retrieved from http://www.ieee-tcdl.org/Bulletin/v4n2/piwowar/piwowar.html

Piwowar, H.A. (2010a). Foundational studies for measuring the impact, prevalence, and patterns of publicly sharing biomedical research data. (Doctor of Philosophy Dissertation), University of Pittsburgh, Pittsburgh.

Piwowar, H.A. (2010b). A method to track dataset re-use in biomedicine: Filtered GEO accession numbers in PubMed Central. *Proceedings of the American Society for Information Science and Technology, 47*(1), 1-2. doi:10.1002/meet.14504701450

Sands, A., Borgman, C.L., Wynholds, L. & Traweek, S. (2012). Follow the data: How astronomers use and re-use data. Paper presented at the ASIST 2012 Annual Meeting. Baltimore, MD.

Tashakkori, A. & Teddlie, C. (1998). Mixed methodology: Combining qualitative and quantitative approaches. Thousand Oaks, CA: SAGE Publications.

Tenopir, C., Allard, S., Douglass, K.L., Aydinoglu, A U., Wu, L., Read, E., ... Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLoS One, 6*(6). doi:10.1371/journal.pone.0021101

Thessen, A.E. & Patterson, D.J. (2011). Data issues in the life sciences. *ZooKeys(150), 15*. doi:10.3897/zookeys.150.1766

Wallis, J.C., Rolando, E. & Borgman, C.L. (2013). If we share data, will anyone use them? Data sharing and re-use in the long tail of science and technology. *PloS One, 8*(7), e67332. doi:10.1371/journal.pone.0067332

Zimmerman, A. (2007). Not by metadata alone: The use of diverse forms of knowledge to locate data for re-use. *International Journal on Digital Libraries, 7*(1-2), 5-16. doi:10.1007/s00799-007-0015-8

Zimmerman, A.S. (2003). Data sharing and secondary use of scientific data: Experiences of ecologists. (Doctor of Philosophy in Information and Library Studies Dissertation), The University of Michigan.

Zimmerman, A.S. (2008). New knowledge from old data the role of standards in the sharing and re-use of ecological data. *Science, Technology and Human Values, 33*(5), 631-652. doi:10.1177/0162243907306704

Appendix

Table 1. Summary of coding results. Column sources (S) represent the number of individual interviews where the factor occurred. References (R) represent the total utterances of each factor across the different interviews.

Category/Factors		S	R	Description	Related Literature
Perceived Benefits					
(PB1)	Knowledge Expansion	9	17	Data can be re-used to answer different questions other than the ones covered by primary studies or for replication/validation	Darby et al. (2012) Hyman (1972) Kankanhalli, Lee and Lim (2011)
(PB2)	Frugality	13	63	Ways to circumvent data collection problems associated with time and cost (money) to minimize duplicated efforts or the need to develop data collection skills	Castle (2003) Hyman (1972) Kiecolt and Nathan (1985) Law (2005)
(PB3)	Pre-Endorsement	5	6	Data available for re-use are considered to some extent credible and reliable, otherwise they would not be shared and available to the public	N/A
Perceived Risks					
(PR1)	Fear of being Undervalued	2	2	Re-using other people's data in research can be perceived as less valuable, and thus have fewer pay-offs than conducting research based on new data	Goodwin (2012) Fahs, Morgan and Kalman (2003) Martin (1995)
(PR2)	Fear of Infringing Ethical Codes	1	1	Hesitation to re-use data which was obtained through consent to a particular study and/or unwary violating aspects of confidentiality, copyright and data protection	Grinyer (2009) Heaton (2004) Law (2005)
(PR3)	Slippage	3	7	Misinterpretation, incorrect association, or misuse that might occur while re-using other people's data	Corti and Thompson (2004) Kuula (2010) Tenopir, et al. (2011)
(PR4)	Vulnerability to Hidden Errors	8	10	The susceptibility to faulty data given the difficulty of identifying potential errors on data collected by others	Castle (2003) Hyman (1972) Kiecolt and Nathan (1985)
Perceived Effort					
(PE1)	Be Innovative with Old Data	5	7	The effort of identifying original contributions from second-hand data and exploring different issues not yet explored or overlooked by primary researchers, as well as other re-users	Zimmerman (2008)
(PE2)	Obtain Data Access	3	7	Refers to data accessibility. The effort associated with obtaining access and retrieving data.	Faniel et al. (2013b) Heaton (2004) Tenopir, et al. (2011)
(PE3)	Data Discovery Process	5	6	Refers to data discoverability. The effort associated with data discovery and the identification of relevant and potentially re-usable datasets	Darby et al. (2012) Faniel and Majchrzak (2002)
(PE4)	Dealing with Mismatches	12	75	The effort of working with data that was generated based on different research questions and/or hypotheses, using particular instruments or techniques for data collection, in a particular context and time-frame, and having specific variables, constructs, and measurements. It also accounts for the effort associated with resigning initial ideas and reframing the study design and goals in order to accommodate the existing data	Dale (2004) Heaton (2004) Kiecolt and Nathan (1985) Klein et al. (2007)

(PE5)	Preparation for Re-use	9	35	Refers to the effort to get data ready for re-use and manipulation, including: screening and cleaning processes, dealing with missing data, adding/complementing data, and putting it in an appropriate format, sorting, recoding etc.	Faniel et al. (2013b) Heaton (2004)
(PE6)	Understanding the Original Study	8	29	The effort associated with making sense of the data and thoroughly comprehending the original study	Boh (2008) Faniel and Jacobsen (2010)
Re-usability Assessment						
(RA1)	Data Documentation	6	24	Whether the supplementary documentation provided along with the data is sufficient, easy to understand and clearly explains the methodology, the rationale of the study, etc. to support re-use.	David (1991) Faniel et al. (2013) Niu and Hedstrom (2009) Niu (2009ab) Zimmerman (2003, 2007, 2008)
(RA2)	Data Fitness	9	19	Whether the topic, level of analysis, and type of data are compatible with the purpose of re-use	Faniel et al. (2013b) Hinds, Vogel and Clarke-Steffen (1997) Palmer, Weber and Cragin (2011)
(RA3)	Data Producer Trustworthiness & Credibility	7	9	How trustful and credible data producers (institutions or individual authors/contributors) are	Boh (2008) Carlson & Anderson (2007) Darby et al. (2012) Faniel et al. (2013b)
(RA4)	Data Quality	7	14	How consistent and complete data are perceived to be	Faniel et al (2013b) Hinds, Vogel and Clarke-Steffen (1997)
(RA5)	Study Rigor	3	8	How well-designed and executed the study was	Hinds, Vogel and Clarke-Steffen (1997)
Enabling Factors						
(EF1)	Data Documentation Availability	6	19	The availability of comprehensive and detailed data documentation improves chances for data re-use	David (1991); Faniel and Majchrzak (2002); Markus (2001); Pigott, Hobs & Gammack (2001); Niu & Hedstrom (2009); Niu (2009ab); Zimmerman (2003, 2007, 2008)
(EF2)	Data Repositories Availability	5	9	The existence of repositories and their capability to organize, self-guard, and facilitate access to re-usable data improves conditions for re-use	Markus (2001); Palmer, Weber and Cragin (2012); Tenopir, et al (2011)
(EF3)	Primary Investigators Reach	5	10	Communication with primary investigators helps re-users to obtain additional information about the data and the study	Boh (2008) Faniel and Zimmerman (2011)
(EF4)	Support & Assistance Availability	5	16	Having institutional support and assistance from the data repository personnel or at the university level (e.g. statistical center, library, IT center, advisors)	Behboudi and Hart (2008) Faniel et al. (2013a) Markus (2001)

(EF5)	Training & Expertise	8	21	Importance of training on secondary analysis for skill development. Expertise in secondary analysis will lead to more re-use of data	Corti and Bishop (2005) Hyman (1972) Kriesberg et al. (2013)
Social Factors					
(SF1)	Disciplinary Receptiveness	10	12	Disciplinary tradition or perceived acceptance of the re-use of data. Some disciplines are more prone to data re-use than others	Borgman (2007) Faniel and Jacobsen (2010) Thessen and Patterson (2011)
(SF2)	Peer Encouragement	4	4	The acceptance or habitual practice of data re-use among colleagues and peer recommendations to re-use particular datasets	Kriesberg et al. (2013)