

## RESEARCH PAPER

# The Definition of Reuse

Stephanie van de Sandt<sup>1,2</sup>, Sünje Dallmeier-Tiessen<sup>1</sup>, Artemis Lavasa<sup>1</sup> and Vivien Petras<sup>2</sup>

<sup>1</sup> CERN, Geneva, CH

<sup>2</sup> Humboldt-Universität zu Berlin, DE

Corresponding author: Stephanie van de Sandt ([stephanie.van.de.sandt@cern.ch](mailto:stephanie.van.de.sandt@cern.ch))

The ability to reuse research data is now considered a key benefit for the wider research community. Researchers of all disciplines are confronted with the pressure to share their research data so that it can be reused. The demand for data use and reuse has implications on how we document, publish and share research in the first place, and, perhaps most importantly, it affects how we measure the impact of research, which is commonly a measurement of its use and reuse. It is surprising that research communities, policy makers, etc. have not clearly defined what use and reuse is yet.

We postulate that a clear definition of use and reuse is needed to establish better metrics for a comprehensive scholarly record of individuals, institutions, organizations, etc. Hence, this article presents a first definition of reuse of research data. Characteristics of reuse are identified by examining the etymology of the term and the analysis of the current discourse, leading to a range of reuse scenarios that show the complexity of today's research landscape, which has been moving towards a data-driven approach. The analysis underlines that there is no reason to distinguish use and reuse. We discuss what that means for possible new metrics that attempt to cover Open Science practices more comprehensively. We hope that the resulting definition will enable a better and more refined strategy for Open Science.

**Keywords:** reuse; research data; open science; definition; characteristics

## Introduction

During recent years, we have observed a strong trend towards Open Science across different stakeholders and disciplines (Pampel & Dallmeier-Tiessen (2014)). Researchers must now submit their research data as supplementary information in order to be in compliance with the data storing requirements of major funding agencies, high profile journals and data journals (Molloy (2011)).

One of the driving aspects behind such requirements is the idea that by being able to reuse research data, it is possible to save costs and research efforts and validate or evaluate or reproduce the research that is reported. Reuse is not only a key concept of the overall research data life cycle that is affecting research data management plans. The term recently gained attention as many funding agencies and expert organizations started demanding data management plans to ensure the later reusability of funded research data.

The term reuse is a complicated concept and the exact meaning of "reuse of research data" does not seem to be fixed yet. The understanding of reuse varies between disciplines and even individuals and no common standard seems to be applied yet. But finding a definition of the term is crucial, because reuse is starting to get more attention now as it has been recognized as an important topic for all research communities across disciplines and institutions (DCC (2018); NSF (2017); EC (2016); OSTP (2013); Wellcome Trust (2010)).

Having a commonly agreed definition can clear misunderstandings in science communication. As it is of relevance for individuals and institutions to make their data count, it is also crucial to make reuse measurable and thus measure the impact of research resources. The desire that a certain impact or even the value of resources can be measured by its reuse count can only be fulfilled if reuse can easily be broken down into attributes or characteristics. Attributes are distinguishable and comparable, which is important as value

implicites that some events are “less valuable” than others. But attributes as a basis for developing better data metrics can only be achieved by sharing a clear and commonly-understandable definition.

## Methodology

The confusion and the lack of practical steps towards reusability of research data in guidelines and policies urge for a clarification of what is meant by reuse in a science policy context.

In general, there are two methods that lead to a definition. One way is empirical. Inductive reasoning is based on the generalization of observed premises, which serve as evidence for a theory. This approach can be considered a bottom-up strategy. The other way starts with the development of the theoretical concept. By deductive reasoning, a theory leads to predictions on the outcome of real world relations. This approach can be considered a top-down strategy. As reuse is assumed to be a concept used to describe real world phenomena, deductive reasoning seems to be a promising approach. The analysis of the word reuse and its etymology may already reveal the meaning of the concept, as the term itself differs from related concepts such as replication, reanalysis, reproduction, subsequent research and reinterpretation. In order to understand the meaning and etymology of the word, definitions in encyclopedias were examined and compared to definitions of related terms. The results from the decluttering of the terminology were then contextualized in relation to the current discourse on the reuse of research data.

For the purposes of this article, all scholarly works with a combination of reuse (or re-use or secondary analysis) and data in their title were selected. In total, 65 works were identified by querying the information resources Google Scholar, Scopus and LISA. The references part of all works were also checked with the same criteria to identify further candidates until no further new literature was found. Even though the resulting list may not be complete, it covers a large part of the discussion on this topic. If there was already a broadly used and commonly accepted definition of reuse, it would certainly be reflected in the literature sample.

The resulting sample consists of works from different authors, different publication years and disciplinary backgrounds. Although many works focus on the social sciences, there are also STEM related publications. All works were processed for definitions. The definitions were usually given in the introduction or in the beginning sections. Most definitions were clearly identifiable as they were formalized and clearly expressed as a definition. E.g. Zimmerman (2008) starts her definition with the introductory words ‘Thus, I define reuse as [...]’, which makes the identification of the definition easy. Simply using the term in a certain context, reflecting on previous work on reuse and stating the advantages of reuse is not considered a proper definition. Although a definition can be guessed, it is not clearly expressed in the author’s own words. Moreover, a definition that is not explicitly conveyed is open to interpretation, which may distort the result. For these reasons, only clearly identifiable definitions were collected.

The definitions were then categorized and analyzed for specific characteristics of reuse, which may differentiate reuse from simple use. Each characteristic was tested by comparing it to models that represent certain aspects of the current research landscape in order to find out if the characteristics resulting from theoretical definitions would also play a role in practice and if the difference between reuse and use is really measurable.

In summary, meaning of reuse is examined by the general etymology of the term. We use definitions from dictionaries to distinguish the term from related concepts. The understanding of the term in the concrete context of research data reuse is analyzed by comparing definitions of reuse which were extracted from a selection of publications related to research data reuse. Out of these definitions, characteristics of reuse are extracted. In a last step, we compare the extracted characteristics of research data reuse to observable real-life research scenarios in order to evaluate their validity. The goal of this research approach is to evaluate criteria that distinguish reuse from all other related actions and to find a definition that makes reuse measurable and understandable to a broader audience.

## *Etymology of Reuse*

The Oxford Dictionary defines the term reuse as ‘to use again or more than once’. The nature of use is not further explained. This broad explanation can be compared to other concepts that indicate the usage of research data. It may be possible to draw conclusions on the nature of reuse from the delimitation of related terms. First of all, replication means here: ‘The action of copying or reproducing something’. This explanation already includes the term reproduction, which is defined as: ‘The action or process of copying something’. At this point, this dictionary does not contribute to an overall understanding of reuse, as it gives a circular reasoning to distinguish two concepts.

The definition of other related terms is not helpful either. The meaning of Reanalysis is to 'conduct a further analysis, or, analyze again' and reinterpretation stands for 'the action of interpreting something in a new or different light'. A definition for subsequent research is not included, but the dictionary does offer a definition for the verb restudy, which means to 'study (something) again'.

The action of using becomes more clear through the definition of related terms (replication, reproduction, reanalysis, restudy). The Oxford Dictionary only indicates that the object has to be used several times, which is true for all of the terms above. According to this definition, replication, reanalysis, reproduction, reinterpretation and subsequent research can be forms of reuse.

Wikipedia gives a more detailed explanation of the term: 'Reuse is the action or practice of using something again, whether for its original purpose (conventional reuse) or to fulfill a different function (creative reuse or repurposing)'. In this definition, two usage scenarios are differentiated. Conventional reuse is related to concepts like replication, reproduction or reanalysis. The purpose for using the object is the same as it was the first time. Creative reuse differs from all other concepts, as the the intention must be distinguishable from the original purpose.

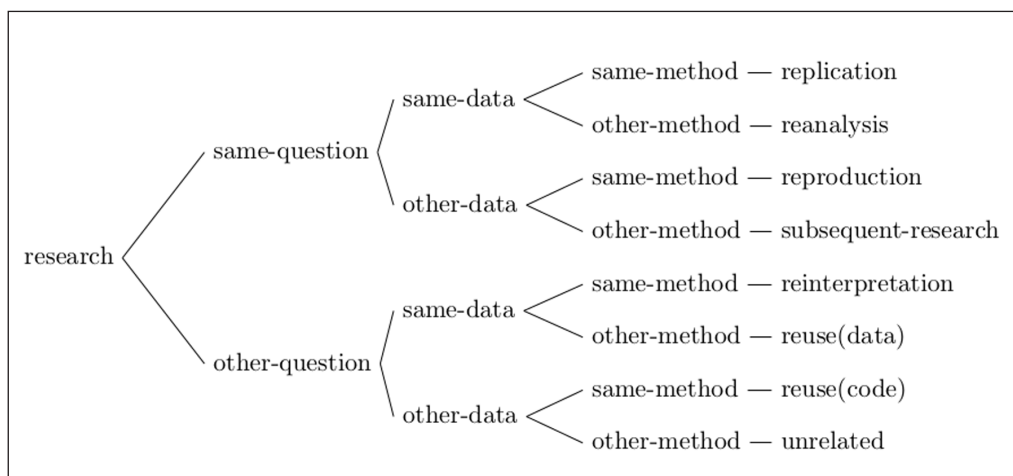
Both definitions have in common that an object has to be used somehow before it can be reused. The first usage of a resource for its original purpose cannot be counted as reuse or any other related concept. They all contain the prefix "re", which supports the differentiation of a first and a second usage. However, reuse cannot be clearly distinguished from related concepts in the context of research data reuse at this point, as the examined definitions in encyclopedias/dictionaries of the term are not precise enough yet. This may be problematic as the classic research data life period is usually drawn as a cycle. After the first use of research resources, data is made publicly available for reuse purposes. The end of one's research can be the beginning of a third party's research, using the same data. This ideal cycle is related to Wikipedia's creative reuse concept, as the data fulfills a different function when it is used again.

A differentiation of reuse and related terms is still difficult at this moment. Schöch (2017) demonstrates a method to distinguish reuse from other concepts (see **Figure 1**).

He distinguishes the terms by three variables: research question, research data and research method. This enables a clear differentiation between reuse and related concepts, as the reuse of data enforces the usage of the same data with a different method and a different research question in mind. Code is not counted as a research resource, although code and software are used and produced for research processes. At this point, it makes sense to consider both data and code. The code is handled as the method of the research project, which stays the same while every other component changes. Still, both reuse cases (data and code) could be renamed as repurpose or creative reuse according to the definition from Wikipedia, as the need of a different method and question can be understood as research while using the same data for another purpose – repurposed resources.

### **Detailed literature review**

Out of 65 identified works related to the reuse of research data, 20 provided a definition. In 45 papers no definition of the term was found, although it was used in the title.



**Figure 1:** The concept of reusable research, adapted from Schöch (2017).

Nearly all gathered definitions share the repurposing thought as a common element. Authors often refer to a definition given in another paper and authors with several papers often used the same or a similar definition. This limits the diversity of different angles.

The following definition of Zimmerman was often cited and may have the highest influence in the recent discourse:

Zimmerman (2003): 'In this study, I define secondary use as the use of data **collected for one purpose to study a new problem.**'

Zimmerman developed this definition in her doctoral theses and did not change it in her later publications:

Zimmerman (2008): 'Thus, I define reuse as the use of data **collected for one purpose to study a new problem.** I also use the phrase secondary use, which I intend to be synonymous with the term reuse.'

Although Zimmerman did not change her core definition through the years, it is noteworthy that she changed the term slightly. It is not clearly expressed what is meant by 'a new problem', e.g. if both the question and method must be different as Schöch assumed. However, it is clear that some detail of the research objective must be different compared to the first usage. She also considers that secondary use happens after a first use of the data. So, two core components can be identified: purpose and time.

The following definitions are built on Zimmerman's definition and reflect the repurposing thought:

Faniel & Jacobsen (2010): 'Few studies on scientific data reuse formally define reuse but generally agree that it includes the secondary use of data for a **purpose other than originally intended** (Karasti and Baker 2008; Zimmerman 2008), which presents one of the major challenges in providing reusable data.'

Faniel et al. (2016): 'By data reuse, we mean the secondary use of data for a **purpose other than the original** intention of the data producer (Karasti & Baker, 2008; Zimmerman, 2008).'

Daniels (2014): 'From Zimmerman I borrow the terms secondary data use and data reuse to describe, "the use of data **collected for one purpose to study a new problem**" (Zimmerman 2003, p. 7).'

Curty & Quin (2014): 'Data reuse represents the re-analysis of a dataset or a combination of different datasets for the purpose of answering the **original research questions** with a **new method** of analysis, or answering new questions based on old data that was not necessarily the focus of the original data collection (Law, 2005; Zimmerman, 2003).'

Curty's and Quin's definition is more explicit about what exactly must be different to fulfill another purpose. The research question must be differentiable from the original when combining different datasets. Thus, the process can be distinct. She also names a new component: the character of the reused object, as she considers the combination of several datasets.

Fear (2013): 'Data reuse is use of data one which did not collect oneself, for example to answer a **new question** from **existing data** (Zimmerman, 2008), to combine with other existing or newly collected data, or to **reproduce or replicate** the results of a **prior study** (King, 1995). **Data producers repurposing their own data is not considered data reuse** for the purposes of this study. In the social sciences, the terms secondary analysis (Gleit & Graham, 1989; Hinds, Vogel, & Clarke – Steffen, 1997) or reanalysis (King, 2003; Weber & Chao, 2011) are often used synonymously with my definition of data reuse.'

Fear agrees with Zimmerman on repurposed data. She also agrees with Curty & Quin that several datasets can be combined. But she widens the scope as she also considers the replication or reproduction of a research result as reuse. At this point she contrasts the findings of Schöch with the definitions related to Zimmerman. It is noteworthy that she excludes self-reuse. These definitions underline that the user of the data is another considerable component of the reuse concept.

The following definitions do not refer to Zimmerman's definition, but also support the repurposing thought:

Francis (2017): 'Data are reused when they are collected for one use and then **used a second time**. They are repurposed when the second use has a **different aim** than the first.'

Sun (2017): 'This paper takes the position that as long as data is **used for purposes other than that for which they were originally collected**, data are reused. In other communities, the term "secondary use of research data" is employed to refer to this activity.'

Yoon (2015)/Yoon (2017): 'Data reuse refers to the **secondary use** of data – not for its original purpose but for studying **new problems**.'

Law (2005): 'Secondary research refers to the use of research data to study a **problem that was not the focus of the original data collection**. This may be data collected for administrative, health or educational purposes, census data, or data collected as part of a previous study. This secondary analysis may involve the combination of one data set with another, address new questions or use new analytical methods for evaluation (Szabo & Strang, 1997).'

Hinds et al. (1997): 'Secondary analysis is the use of an existing data set to find answers to research questions that **differs from the question asked in the original or primary study** (Lobo, 1986; McArt & McDougal, 1985; McCall & Applebaum, 1991).'

Rolland & Lee (2013): 'Data reuse, then, is the work **done by the recipient of those shared data**. It involves identification of a dataset of interest, receipt of the dataset and appropriate use of the data for analysis.'

The involvement of a third-party is controversial. E.g. Fear and Rolland explicitly do not consider self-reuse as reuse. The following two definitions also include the user aspect, but include self-reuse and give a different picture:

Heaton (1998): 'Secondary analysis involves the use of existing data, collected for the purposes of a prior study, in order to pursue a **research interest which is distinct** from that of the original work; this may be a **new research question** or an **alternative perspective** on the original question (Hinds, Vogel and Clarke-Steffen 1997, Szabo and Strang 1997). In this respect, secondary analysis differs from systematic reviews and meta-analyses of qualitative studies which aim instead to compile and assess the evidence relating to a common concern or area of practice (Popay, Rogers and Williams 1998). As will be shown below, secondary analysis can involve the use of **single or multiple qualitative data sets**, as well as **mixed qualitative and quantitative data sets**. In addition, the approach may either be employed by researchers to re-use their own data or by independent analysts using previously established qualitative data sets.'

Szabo & Strang (1997): 'Secondary analysis involves the analysis of data that was gathered for a previous research study. The analysis is done **either by the original researcher or another researcher** and addresses **new questions** or looks at the same questions with **different analysis methods**.'

Then, there are definitions that are very discipline-specific, such as those by Meystre et al. and Safran et al.:

Meystre et al. (2017): 'Secondary use (or reuse) of clinical data is defined as "non-direct care use of personal health information including but not limited to analysis, research, quality/safety measurement, public health, payment, provider certification or accreditation, and marketing and other business including strictly commercial activities."

Safran et al. (2007): 'For purposes of this meeting, secondary use of data was defined as non-direct care use of personal health information (PHI) including but not limited to analysis, research,

quality/safety measurement, public health, payment, provider certification or accreditation, and marketing and other business including strictly commercial activities.'

While other definitions are so broad that they could fit any need:

Castle (2003): 'Simply defined, SDA [secondary data analysis] refers to a research process, or a set of endeavors, that uses existing data to answer research questions (Kiecolt & Nathan, 1985).'

Summarizing, there is no commonly agreed understanding of reuse or a de-facto standard definition. However, what is clear from the definitions in the sample is that they were based on the following components:

- **character of data**
- **user**
- **purpose**
- **time**

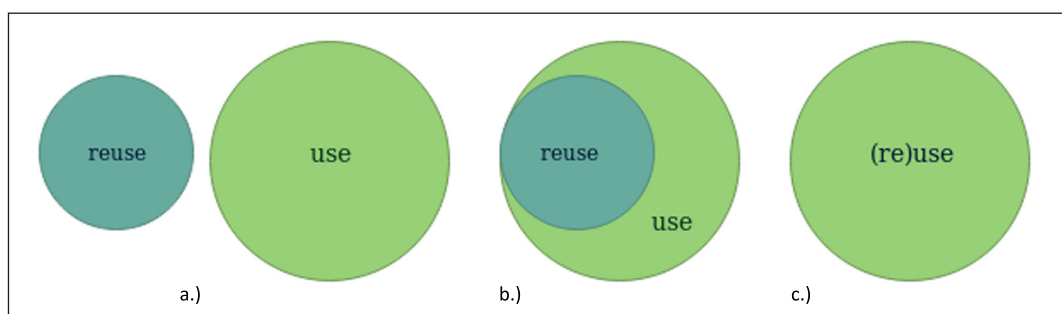
### Is "Reuse" = "Use"?

The question that arises now is if reuse is distinguishable at all from using data generally. As reflected in Pasquetto's definition, differentiating between these two concepts is the main issue:

Pasquetto (2017): 'The most fundamental problem in understanding data reuse is to **distinguish between a "use" and a "reuse."** In the simplest situation, data are collected by one individual, for a specific research project, and the first "use" is by that individual to ask a specific research question. If that same individual returns to that same dataset later, whether for the same or a later project, that usually would be considered a "use." When that dataset is contributed to a repository, retrieved by someone else, and deployed for another project, it usually would be considered a "reuse." In the common parlance of data practices, reuse usually implies the usage of a dataset by someone other than the originator.'

It is clear that reusers use the data somehow, but what exactly must be different from simple cases of use? Is there any difference at all between using own datasets in a different research context or using other's data for a very similar purpose. And after all, is it possible to measure the difference to prove that a usage scenario is clearly reuse?

**Figure 2** illustrates the problem. Reuse and use could be disjoint sets, as shown in subfigure a. In this case, using an object is a self-contained process and unrelated to its reuse. Both processes share no common elements. Reuse could also be considered as subset of use, as demonstrated in subfigure b. In this case, use is the generic term and reuse a specification. Both concepts would share common ground elements, but reuse also implies particularities that distinguishes it from usage. E.g. both the first use of data and the second use of data for another purpose could then be considered as usage. But the second use is more specifically expressed as reuse, as it meets certain criteria, like a different purpose. One last scenario is that there is no difference between use and reuse (subfigure c). Here, the two sets do not only perfectly overlap, but there is only one set to consider. This would deny the existence of a specific reuse.



**Figure 2:** Reuse in relation to use.



### The Linear Model of Reuse

All the previously given definitions from the discourse refer to a rather linear model of usage. They all assume that research data was created for a specific purpose, analyzed and processed to serve this purpose and finally published after leading to certain research results. In this model, reuse only happens after the data was used for its original purpose. A graphic depiction of this perspective could look like **Figure 3**. Researcher 1 (R1) has a certain research interest and raises a research question (Q1). Therefore, R1 gathers data. The analysis of the data leads to a final publication (P1). The data is uploaded to a public research data repository, where another third-party user (R2) seeks suitable data for their research question (Q2). R2 may transform the data set ( $D \rightarrow D'$ ) to suit their question and method, and publishes the results in a paper (P2) some time after the first publication (P1). This model demonstrates a reuse scenario that fits all criteria discussed so far (character of data, user, purpose, time).

However, research reality is more complex. There are plenty of real world scenarios that do not fit into this linear-shaped model. The characteristics identified so far from the discourse are promising candidates for the differentiation between the two terms. Throughout this paper, they are tested and validated in order to prove or reject the assumptions so far.

### Characteristic 1: Character of Data

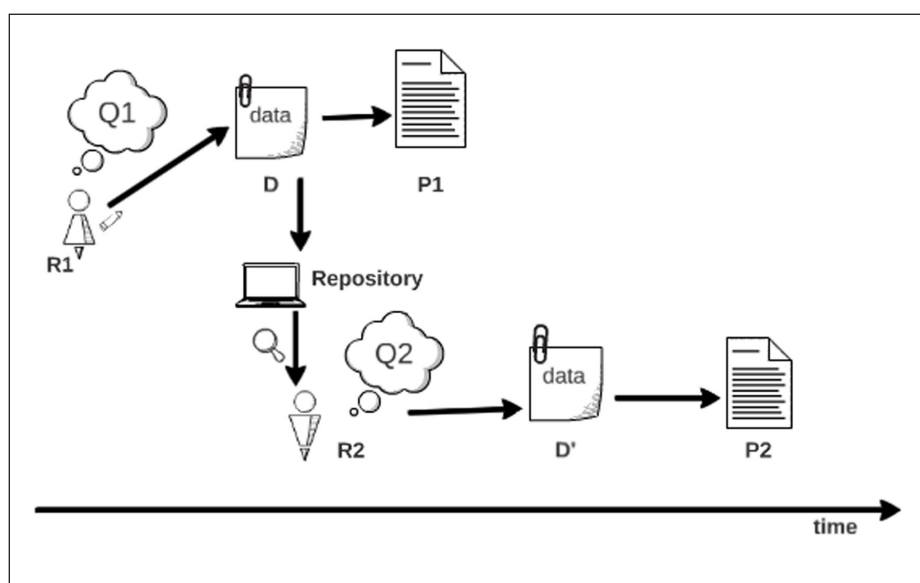
One of the identified characteristics include the “character of data”. This characteristic was raised, for example, by Curty, as she contrasts the use of multiple datasets to using one single dataset in her definition. Now the question arises if the character of data has any effect on reuse.

**Figure 4** demonstrates the usage of multiple datasets. Here, researcher A produces several datasets (DA-1 and DA-2) to answer their research question (Q1). Researcher B creates several datasets (DB-1 and DB-2) to answer research question Q2 and finally writes publication P1. Both researchers publish their datasets on a public research data repository where researcher C selects three datasets (DA-1, DA-2 and DB-1) to answer research question Q3. After finishing the analysis, publication P2 is written.

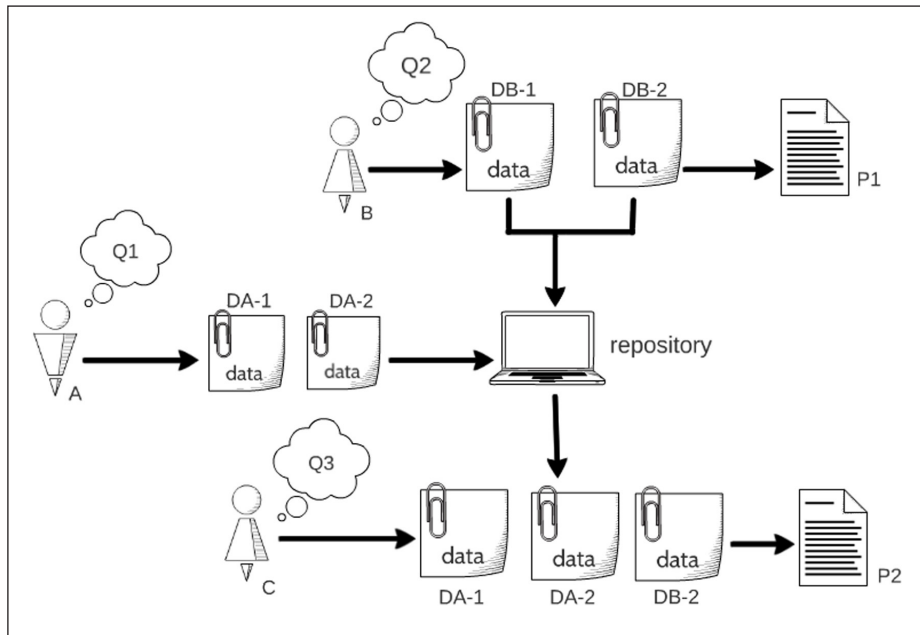
In this model, multiple datasets from different sources were selected to answer one question (Q3). The resulting publication is undoubtedly based on several sources of data that were used by a third-party researcher to fulfill a different research interest. It would probably even be possible to trace this usage, though it would be slightly more complicated.

**Thus, both models show how data of different character and composition is properly used which strengthens the conclusion that reuse does not only refer to single dataset usage.**

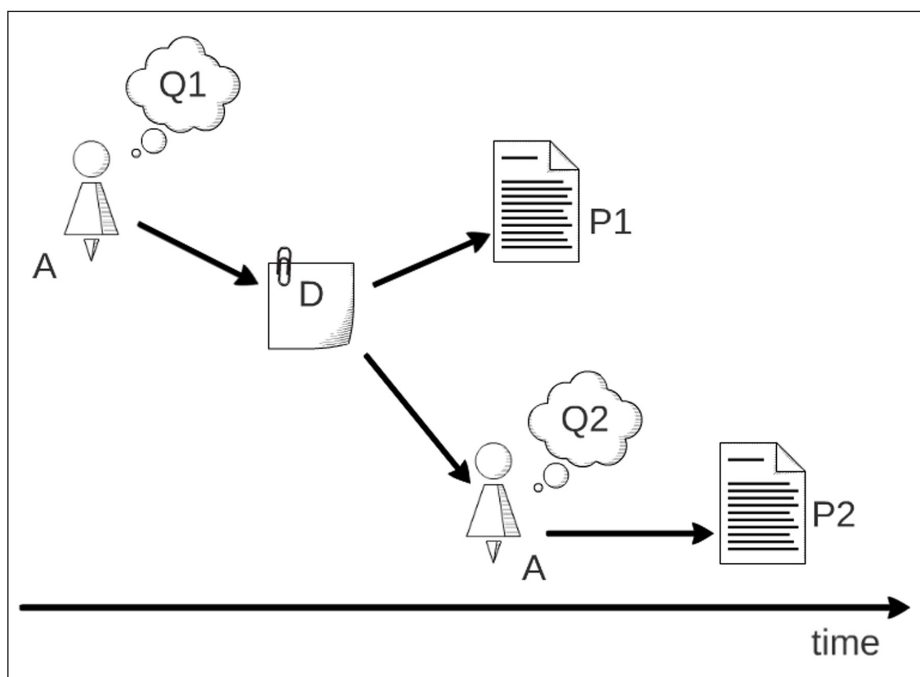
The character of data does not only concern the amount of datasets but also their state. Previous literature on data trajectories, e.g. Missier (2016), suggests the transformation of the resource after each use case. **Figure 5** demonstrates the usage of a data resource without changing it.



**Figure 3:** Linear reuse model.



**Figure 4:** (Re)using multiple datasets.



**Figure 5:** Data characteristics of (re)use.

In this fictional example, researcher A raises a research question (Q1) and collects data (D). The results of the analysis get published in publication 1 (P1). Years later, researcher A may be more experienced and analyses the same data again, but for a slightly different question (Q2) or maybe with a more advanced tool. The results are published in a second paper (P2) based on the same old data. It is not important who uses the data in this case. The important aspect of this model is that the data was not transformed as it was used again. The resulting data were just different, as the question or method differed, which means that there are two publications based on the same data.

Coming back to the question of the relation between reuse and use, which is a key question raised by Pasquetto et al. (2017), the issue here is whether this model demonstrates use or reuse and if the difference between the two concepts can be measured. The ability to be measured or counted in some way is an important consideration for a theory based definition that should be applicable in practice.



In the above described model, it would be very hard to prove the lack of transformation of the data or to measure the degree of transformation. The dataset in that example was not published on a repository and as papers usually deal with the discussion of the resulting data, it would be very hard to trace the transformation of the data based on information in the scholarly work. Only the author could upload both datasets as versions on a research data repository and document the transformation in high quality metadata, which would also need to contain a link to both publications. If the dataset resulting from new research on old data doesn't get published, the transformation of a dataset for repurposed research cannot be measured in practice.

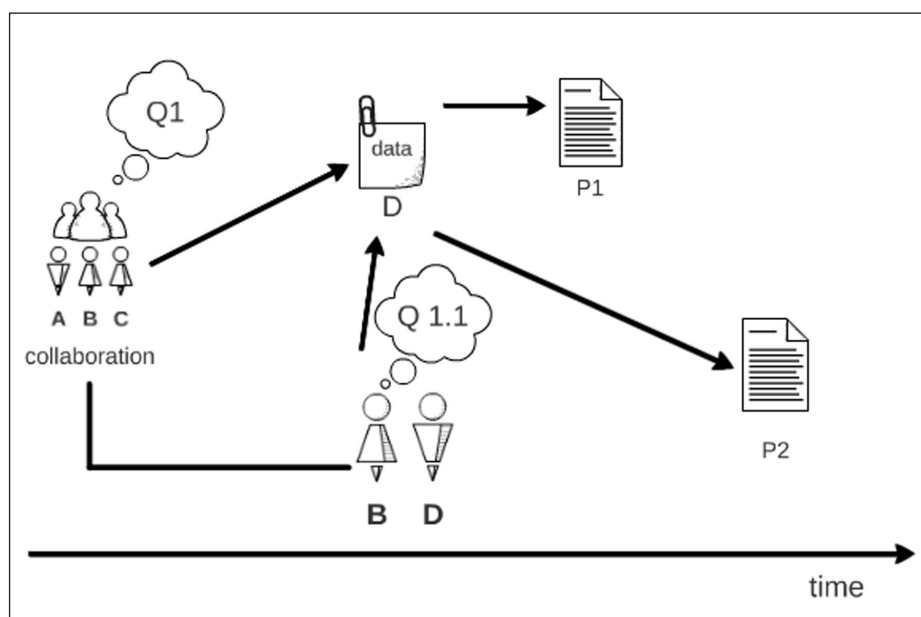
**Since there is no evidence that the transformation degree can be proven or that the degree affects the use of the data set, this factor cannot be a criteria for reuse.**

**Summarizing, the “character of data”, as described in the discourse, is not a reliable characteristic of reuse.**

### **Characteristic 2: User**

The linear reuse model implicates a third-party user. This may be easy to measure when there is just one person involved in the research process. However, today's data driven science is often executed by collaborations consisting of a large number of people. For example, for experiments in high-energy physics or other large-scale collaborative research areas, it is typical to have publications authored by thousands of people. The researchers within these collaborations change on a very regular basis, which makes it extremely difficult to define third-party usage. What if one member of the collaboration uses the data again for another purpose? What if the same collaboration uses the data, but with slightly changed members? After all, even if just one person is involved in the research process, it seems natural that a researcher may want to use their data again after some time when circumstances change or they become more experienced. Are there levels of reuse? Does using one's own data several times then count as “extended use”?

**Figure 6** demonstrates the problem in a model. Here, a collaboration of scientists consisting of hundreds of people, including researcher A, B and C, uses data for a research question (Q1), which leads to a publication (P1). In this scenario, an individual from the collaboration (researcher B) collaborates with an external researcher (D). They have a slightly different research question in mind (Q 1.1) and write a second paper (P2). Now, researcher B is an author of both P1 and P2. It can neither be considered as third-party reuse nor self-reuse. The problem remains even if the same collaboration writes another paper based on the same data because the researchers in that collaboration may have changed several times during the writing period. Maybe a new researcher joins the collaboration or another leaves meaning that the collaboration does not consist of the exact same people. Also, researchers may work for several collaborations around the world and write papers independently from the collaboration. A collaboration is the sum of individuals that come and go, which makes it impossible to determine and measure if a certain use case is third-party reuse for sure.



**Figure 6:** Involvement of collaborations in data reuse.

The scenario in **Figure 7** is less complex. Here, the same researcher (A) uses the same data for two different questions (Q1 and Q2), which leads to two different publications (P1 and P2). As the second use may happen several years after the first use, the question arises if A can still be considered to be the first person. Maybe A changed workplaces or is more experienced and interprets the results in a different way than before. Researcher A may no longer possess any insight on the data collection process after several years and may now depend on documentation just like any other third-party user would.

**These models demonstrate the difficulties of measuring third-party use in practice. The user of a research resource is hard to define and even harder to verify in practice. Thus, reuse cannot be differentiated from use by the “user” characteristic.**

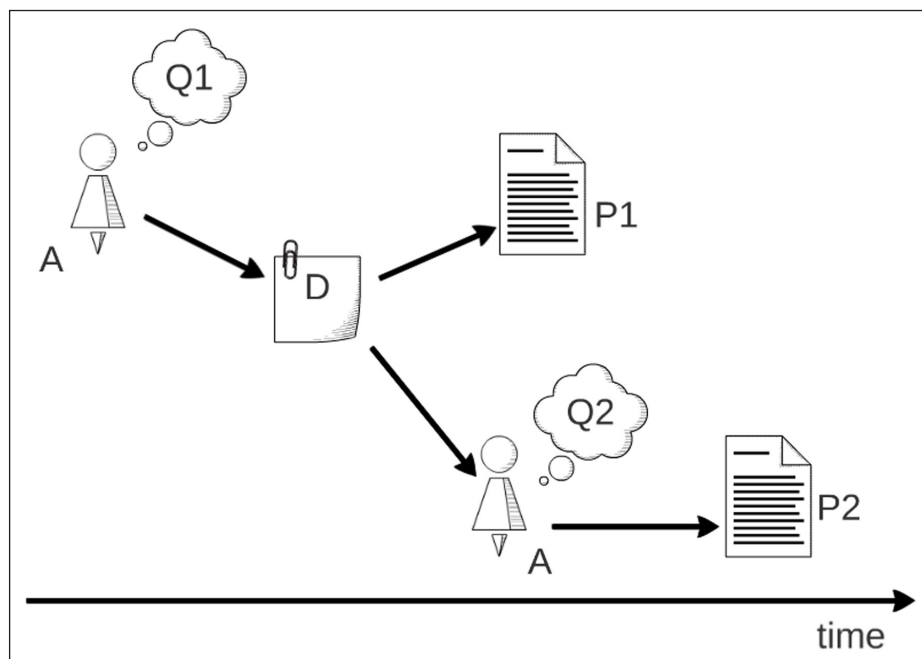
### **Characteristic 3: Purpose**

There was some agreement in the literature review that data has to be repurposed to be considered reuse. This can be the case when there is a different research question from the original and/or by using a different method, as demonstrated by Schöch (2017). However, as discussed before, purpose is quite discipline-specific and varies. Research question and method are only two components to consider. A distinguishable research objective is relatively easy to imagine in a linear model. But this simple perspective on the research workflow does not match with what can be observed nowadays in data-driven science where purpose is not that easy to determine.

What if data was used for a different purpose but in the end was dismissed? This case would be extremely difficult to measure as publications on failed analyses are very rare. Yet it can be argued that the data was used, just without any visible outcome.

This case is demonstrated in **Figure 8**. After the data was collected by a collaboration for a certain research interest (Q1) and uploaded on a research data repository, researcher B uses the data for another purpose (Q2) but fails and no publication is written in the end. This usage of data is not measurable. Thus, it cannot contribute to the differentiation between use and reuse.

Another scenario is the usage of data for validation purposes. In this case, the data does not support one single research question, as the research process is not linear. This is demonstrated in **Figure 9**. A computer scientist may be interested in machine learning and needs data to test their code and to demonstrate the effectiveness of it. Therefore, the researcher may use well-known datasets from a public repository and indicates the usage of said dataset in the resulting publication by giving a data reference. However, the content of the data is not the actual focus of his research interest, which is rather the development of software, which means that the reused dataset does not serve the linear research purpose. This scenario indicates data use, but is it also reuse?



**Figure 7:** Extended use.

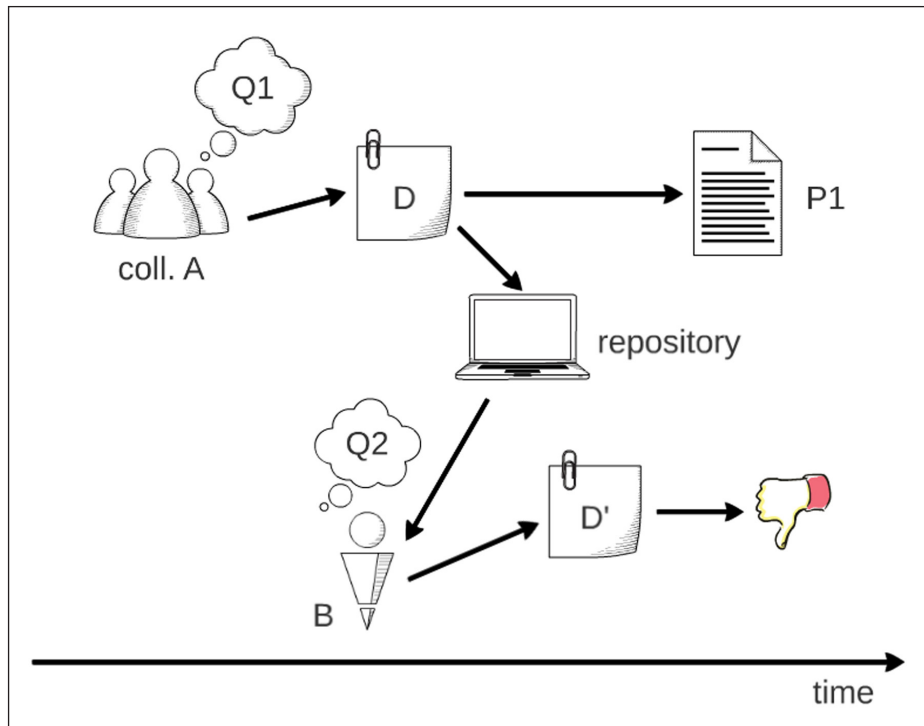


Figure 8: Failed analyses.

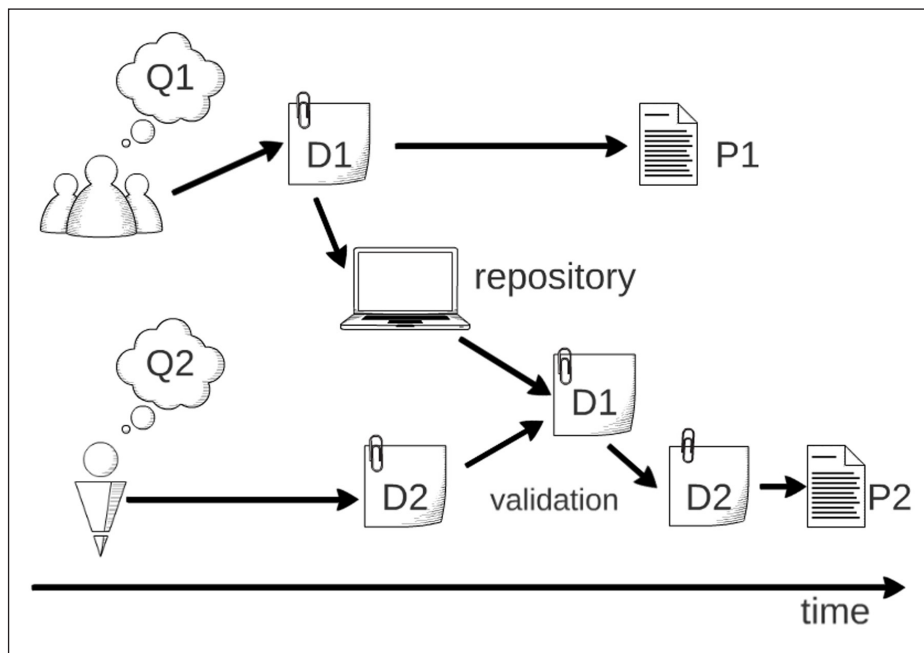


Figure 9: Usage for validation purposes only.

The “linearity” of data use for a research purpose is extremely hard to measure, which is why the “purpose” characteristic cannot be used to differentiate reuse from use.

**Characteristic 4: Time**

The linear model is linear with regard to time. But does original primary use of data always visibly happen before the second use of data? What if data is produced on a regular basis, e.g. on a research vessel. These datasets get published on a regular basis in a time series. What if this data attracts another researcher? And what if this other researcher produces a publication out of data that is clearly connected to another research question before the results of the original analysis get published? Which publication should be considered as reuse?

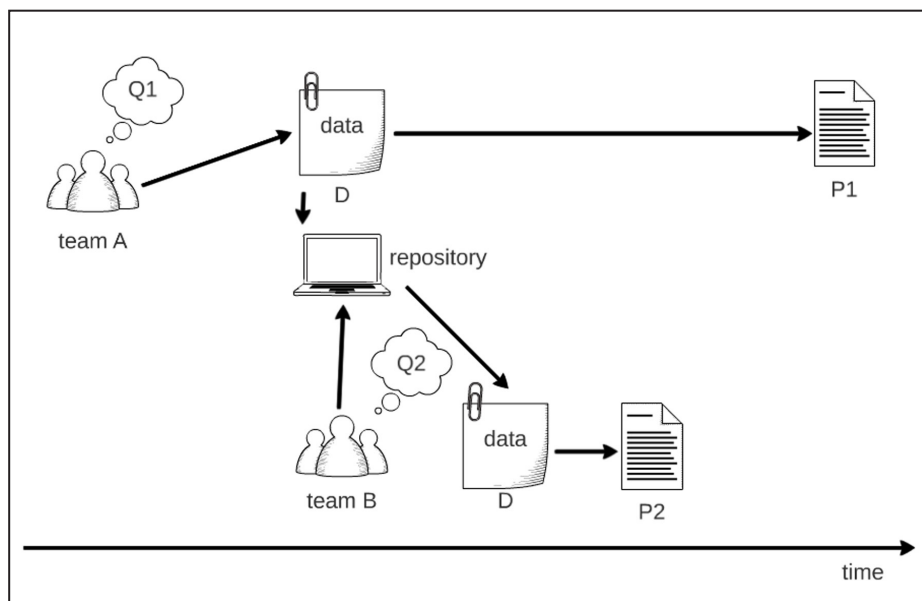
The usage of data does not always happen linearly in time as demonstrated in **Figure 10**. The publication (P2) of team B is submitted earlier than publication (P1) of team A, which answers the original research question (Q1). Therefore, the publication date does not necessarily indicate, which publication was the result of reused data.

A related problem is the usage of data produced in relation to several research interests simultaneously, as shown in **Figure 11**. For example, in astrophysics the cosmos is observed on a regular basis. A certain area of the cosmos is often observed over a long period of time. Upon completion of the observational phase, the data is published on a shared platform so that researchers can use specific datasets that fit their research interest. Thus, the data production is not connected to one specific research interest. It was produced for several possible research questions and by a large collaboration of people. In these cases, it is tremendously hard to say, which the original purpose for the use of the data is. And if there is no original purpose, data can hardly be repurposed as the original purpose is not fixed. But still, these data are undoubtedly used several times. Is then the publication that is submitted first automatically the original purpose?

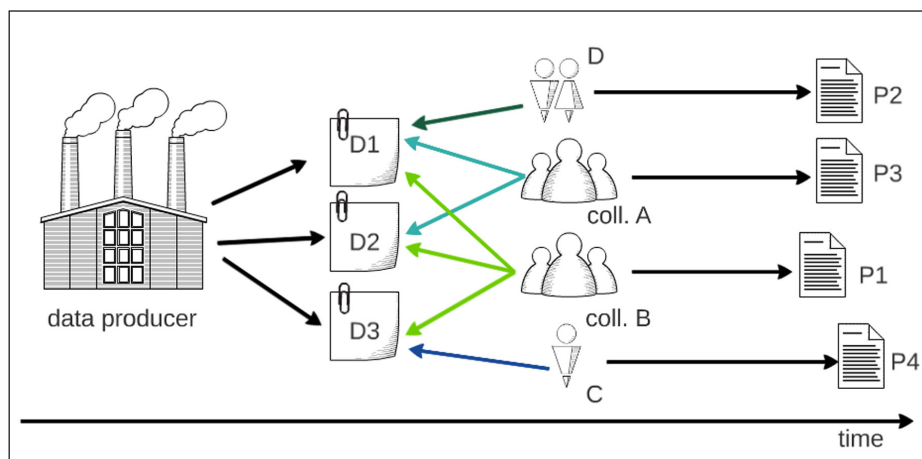
**As demonstrated, time is not a reliable characteristic of reuse, because the first published paper does not necessarily represent the first or original purpose of the data.**

### Summary of the Reuse Models

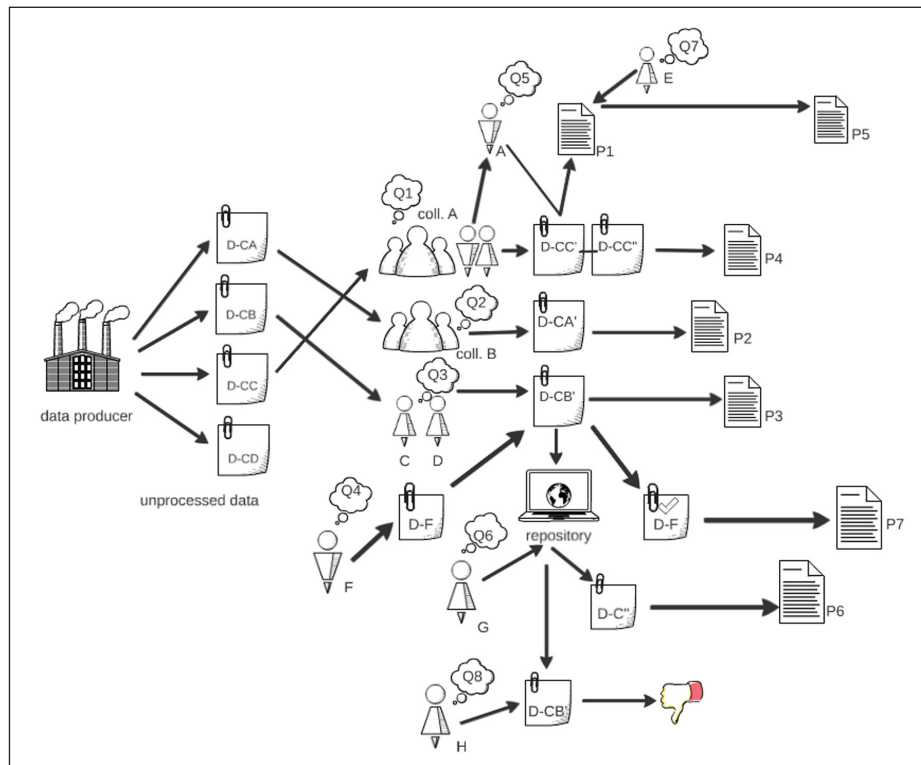
To sum up all observations, the landscape of research data usage looks like the model demonstrated in **Figure 12**. This model is a combination of all previously discussed research scenarios. Here, several datasets are produced for several research purposes. All datasets share similar elements, as they were produced



**Figure 10:** Non-linear usage of data with regard to time.



**Figure 11:** Data produced for several research purposes.



**Figure 12:** Complex (re)use scenario.

within the same context. This is expressed by the similar dataset names (D-CA – D-CD). Collaboration A, collaboration B, researcher C and D use the produced datasets for different research purposes (Q1, Q2 and Q3). Collaboration A transforms dataset D-CC several times in order to answer research question Q1. They publish the results in paper P4. One individual from collaboration A, researcher A, has a new question in mind (Q5), uses some aspects of the previously used datasets again (only D-CC') and writes a publication (P1) about his findings. Although researcher A benefits from their work in the collaboration, P1 is published before P4. For researcher E, publications are the basis for answering research question 7. Some disciplines may consider written works as research data or that publications are the foundation for bibliometrics. Therefore, researcher E only uses dataset D-CC' indirectly. Researcher C and D publish their transformed dataset D-DB' on a repository. This is where researcher G finds the data and uses them for answering research question Q6. Researcher H is also interested in the data and tests it for their research interest, but disregards it in the end. Researcher F has developed software (D-F) and needs data to produce noise for robustness testing issues. Therefore, F uses dataset D.CB' from researcher C and D. After validation, publication P7 is submitted.

This model is very complex and confusing. It is not always easy to tell what is happening and what the life cycle of a specific data set really is. But that reflects the reality of research. The linear model can still be found in these complex connections and is undoubtedly a real life data use case.

## Discussion

Based on the conclusions drawn from the comparison with the research landscape models in the previous chapters, none of the identified characteristics from the discourse proved a difference between reuse and use. It can now be stated that:

- the character of data is not necessarily an indicator of (re)use
- the user is not necessarily an indicator of (re)use
- purpose is not necessarily an indicator of (re)use
- time is not necessarily an indicator of (re)use

In all models, data resources were used. Researchers took action and interacted with the resources. The linear model was considered as reuse in the beginning. This model shows that the assumption of use and reuse as disjoint sets must be rejected. Reuse and use share some common mechanics such as the interaction with resources.

(Re)use is not the sum of certain characteristics that differentiate it from use, neither is it the denial of special characteristics. (Re)use does not necessarily depend on any other existing interactions, is not necessarily linked to other actions, but can happen in its own dynamic context. The identified specifications are still characteristics of reuse. In fact, (re)use cases can show one or several characteristics. There may even be (re)use cases where all of the characteristics occur. Yet there can always be a case found that buck the rule. Some data use cases may not even imply one of the identified characteristics of reuse. Still, they were proven to be valid (re)use cases. No unbreakable rule can be applied and no case can be excluded.

**Thus, we define (re)use as the use of any research resource regardless of when it is used, the purpose, the characteristics of the data and its user.**

The former use of the term reuse may be a relic from the traditional theory-driven research landscape. The research process in a paper-centered environment may have been more linear and straightforward, as less irregularities, complexities and dynamics may have occurred. In a paper-oriented research environment, it is rather easy to define purpose and to assume that the reuse of resources includes a new research objective. The variety of research designs were more limited and did not include complex computer simulations, for example. Hermeneutic, empirical and conceptual approaches require less research resources and can be performed by smaller research collaborations. An illustration of this research process may look like the linear use model, which includes some characteristics of the former theory-driven landscape.

However, the research landscape evolved and, as expressed by Anderson (2008), research is nowadays also data-driven. There is much more computational power and the amount of data dramatically increased in the ‘fourth paradigm of data-intensive discoveries’ (Hey (2009)). The increasing possibilities in the ‘petabyte age’ enabled new research designs, such as computer simulations and complex experiments, which are data-driven approaches. So, research itself changed. What was once linear and straightforward has now expanded to a more complex and dynamic landscape where characteristics like “character of data”, “the user”, “purpose” and “time” are no longer solid pillars. In this new landscape, theory-driven approaches exist next to data-driven science, where the question is rather “what” instead of “why”.

Maybe the reason for the confusion over the exact meaning of (re)use results from trying to use the same paper oriented schemas on an evolved research landscape and thinking of research data as just another form of paper.

The deconstruction of the reuse concept may not only affect the discourse, but it may also affect the way we approach data citation counts as a base for the evaluation of the impact or value of research data resources. If there was a difference between the two concepts, data citations would be categorized in such a way that would allow us to learn about their contextualized impact. For example, reuse cases, which imply for example repurposing from third-parties, could have a greater impact than self-citations. But as there is only use, citation counts must be considered with another significance:

- The occurrence of one (re)use case citation rate ( $n = 1$ ) equals proof of existence of the research resource.
- The occurrence of multiple (re)use cases citation rate ( $n \geq 2$ ) equals multiple uses of the research resource.

This simple differentiation could be used for displaying a contextualized citation count and help researchers, bibliometricians and others to understand what is meant by this impact expressed in an abstract number. A simple system that is understandable regardless of discipline or expertise in bibliometrics would contribute to the transparency of research impact counts. Services and publishers would both benefit from adopting more concrete and understandable metrics for their products.

The proposed significance of data and software citations could be extended by attributes of (re)use to serve the community needs of a more granular and differentiable system that can be associated with certain levels of value. There are various ways to use research data resources. These concrete actions have the potential to be categorized into attributes of research data usage. Attributes of usage have the potential to be measurable, be distinguishable and may add the granularity to the use of research data that is needed for the transmission of usage counts into impact metrics. These attributes must be derived from real-life observations and should be generic enough to be applicable to all disciplines. Certainly, this aim deserves research on its own and should be discussed separately from the general definition of (re)use. But a good starting point for deriving usage attributes could be (data) citation purpose classifications. These classifications (e.g. research use, data sharing, attribution, misc.) could be used to differentiate between the integration of research resources in research settings and non-research related actions like the featuring of specifics



or methodological aspects of data/software. But a more refined classification of attributes must be subject of another publication separated from the general discussion about reuse and use.

## Conclusion

The literature on (re)use could benefit from accepting a more dynamic, chaotic and less categorizable research world. The statement that the (re)use of research resources equals their usage is a clarification that could support a terminology that matches the complex and evolving research landscape. The new terminology could lead to a common understanding regardless of disciplines or research approaches, which in turn could help us shaping future Open Science strategies.

A simplification may also contribute towards an increasing engagement in the Open Science movement. Researchers may feel less enthusiastic to document and publish high quality research resources for a goal (reuse) that is not entirely clear and that they might not fully understand.

Solving the confusion about the vague concept of reuse may also invite a broader audience to the discussion and generate an overall feeling of involvement. Recommendations and reports by expert organizations and funding agencies, resulting in funding principles and guidelines, affect the research community directly. All in all, communication is always easier if the message is clearly understandable. Guidelines should include realistic and concrete steps towards a golden standard in research. But this is only achievable when there is a common interpretation of the guidelines, based on a deep understanding of the subject. Only then can individuals be empowered to implement top-down approach policies. Without concrete actions to follow, policies are nothing more than just paper carrying nicely-written words.

## Appendix

**Table 1:** Existence of a definition of reuse of research data in scholarly works.

References	Definition given?
Castle (2003); Curty & Qin (2014); Daniels (2014); Faniel & Jacobsen (2010); Faniel et al. (2016); Fear (2013); Francis & Francis (2017); Heaton (1998); Hinds et al. (1997); Law (2005); Meystre et al. (2017); Pasquetto et al. (2017); Rolland & Lee (2013); Safran et al. (2007); Sun & Khoo (2017); Szabo & Strang (1997); Yoon (2015a); Yoon (2017a); Zimmerman (2003); Zimmerman (2008)	yes
Boslaugh (2007); Brakewood & Poldrack (2013); Carlson & Anderson (2007); Chao (2011); Corti & Backhouse (2005); Corti & Bishop (2005); Corti (2007); Curty (2016); Curty et al. (2016); Darby et al. (2012); Davis et al. (2011); Dwork et al. (2017); Faniel & Zimmerman (2011); Faniel et al. (2012); Faniel, Barrera-Gomez, Kriesberg & Yakel (2013); Faniel, Kansa, Whitcher Kansa, Barrera-Gomez & Yakel (2013); Farquhar & Brase (2014); Federer et al. (2015); Garnett & Edmond (2014); Gleit & Graham (1989); Howard et al. (2010); Kansa (2015); Kim & Yoon (2017); Kriesberg et al. (2013); Li et al. (2016); Missier (2016); Mooney & Newton (2012); Moore (2007); Murillo (2014); Palmer et al. (2012); Piwowar & Vision (2013); Richards (1997); Rung & Brazma (2013); Sands et al. (2013); Scaffidi et al. (2006); Shen (2016); Tenopir et al. (2015); Ullman (2017); Wallis et al. (2013); Wallis (2014); Wicherts et al. (2006); Yoon (2015b); Yoon (2016); Yoon (2017b); Zimmerman (2007)	no

## Acknowledgements

This work has been sponsored by the Wolfgang Gentner Programme of the German Federal Ministry of Education and Research (grant no. 05E15CHA). The paper is also inspired by discussions within the EC funded project FREYA (grant agreement no. 777523).

## Competing Interests

The authors have no competing interests to declare.

## References

- Anderson, C.** 2008. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete, *Wired*. URL: <https://www.wired.com/2008/06/pb-theory/>.
- Boslaugh, S.** 2007. An Introduction to Secondary Data Analysis, in *Secondary Data Sources for Public Health: A Practical Guide*, Cambridge University Press, pp. 2–10. DOI: <https://doi.org/10.1017/CBO9780511618802>

- Brakewood, B** and **Poldrack, RA**. 2013. The ethics of secondary data analysis: Considering the application of Belmont principles to the sharing of neuroimaging data. *NeuroImage*, 82: 671–676. DOI: <https://doi.org/10.1016/j.neuroimage.2013.02.040>
- Carlson, S** and **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: <https://doi.org/10.1111/j.1083-6101.2007.00342.x>
- Castle, JE**. 2003. Maximizing research opportunities: Secondary data analysis. *Journal of Neuroscience Nursing*, 35(5): 287–90. DOI: <https://doi.org/10.1097/01376517-200310000-00008>
- Chao, TC**. 2011. Disciplinary reach: Investigating the impact of dataset reuse in the earth sciences. *Proceedings of the American Society for Information Science and Technology*, 48(1): 1–8. DOI: <https://doi.org/10.1002/meet.2011.14504801125>
- Corti, L**. 2007. Re-using archived qualitative data – where, how, why? *Archival Science*, 7(1): 37–54. DOI: <https://doi.org/10.1007/s10502-006-9038-y>
- Corti, L** and **Backhouse, G**. 2005. Acquiring Qualitative Data for Secondary Analysis. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 6(2). DOI: <https://doi.org/10.17169/fqs-6.2.459>
- Corti, L** and **Bishop, L**. 2005. Strategies in Teaching Secondary Analysis of Qualitative Data. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 6(1). DOI: <https://doi.org/10.17169/fqs-6.1.509>
- Curry, RG**. 2016. Factors Influencing Research Data Reuse in the Social Sciences: An Exploratory Study. *International Journal of Digital Curation*, 11(1): 96–117. DOI: <https://doi.org/10.2218/ijdc.v11i1.401>
- Curry, RG** and **Qin, J**. 2014. Towards a model for research data reuse behavior. *Proceedings of the American Society for Information Science and Technology*, 51(1): 1–4. DOI: <https://doi.org/10.1002/meet.2014.14505101072>
- Curry, R**, **Yoon, A**, **Jeng, W** and **Qin, J**. 2016. Untangling data sharing and reuse in social sciences. *Proceedings of the Association for Information Science and Technology*, 53(1): 1–5. DOI: <https://doi.org/10.1002/pra2.2016.14505301025>
- Daniels, MG**. 2014. Data Reuse in Museum Contexts: Experiences of Archaeologists and Botanists, Dissertation, University of Michigan, Michigan. URL: <http://hdl.handle.net/2027.42/108953>.
- Darby, R**, **Lambert, S**, **Matthews, B**, **Wilson, M**, **Gitmans, K**, **Dallmeier-Tiessen, S**, **Mele, S** and **Suhonen, J**. 2012. Enabling scientific data sharing and re-use. In *2012 IEEE 8th International Conference on E-Science*, pp. 1–8. DOI: <https://doi.org/10.1109/eScience.2012.6404475>
- Davis, LK**, **Alston, P** and **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, JCDL '11*, New York, NY, USA: ACM, pp. 433–434. DOI: <https://doi.org/10.1145/1998076.1998171>
- DCC – Digital Curation Center**. 2018. Overview of funders' data policies. URL: <http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>.
- Dwork, C**, **Feldman, V**, **Hardt, M**, **Pitassi, T**, **Reingold, O** and **Roth, A**. 2017. Guilt-free Data Reuse, *Commun. ACM*, 60(4): 86–93. DOI: <https://doi.org/10.1145/3051088>
- EC – European Commission**. 2016. Guidelines on open access to publications and research data in Horizon 2020. URL: [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf).
- Faniel, I**, **Kansa, E**, **Whitcher Kansa, S**, **Barrera-Gomez, J** and **Yakel, E**. 2013. The Challenges of Digging Data: A Study of Context in Archaeological Data Reuse. In *'Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries', JCDL '13*, New York, NY, USA: ACM, pp. 295–304. DOI: <https://doi.org/10.1145/2467696.2467712>
- Faniel, IM**, **Barrera-Gomez, J**, **Kriesberg, A** and **Yakel, E**. 2013. A comparative study of data reuse among quantitative social scientists and archaeologists. URL: <https://www.ideals.illinois.edu/handle/2142/42099>.
- Faniel, IM** and **Jacobsen, TE**. 2010. Reusing Scientific Data: How Earthquake Engineering Researchers Assess the Reusability of Colleagues' Data. *Computer Supported Cooperative Work (CSCW)*, 19(3–4): 355–375. DOI: <https://doi.org/10.1007/s10606-010-9117-8>
- Faniel, IM**, **Kriesberg, A** and **Yakel, E**. 2012. Data reuse and sensemaking among novice social scientists. *Proceedings of the American Society for Information Science and Technology*, 49(1): 1–10. DOI: <https://doi.org/10.1002/meet.14504901068>

- Faniel, IM, Kriesberg, A and Yakel, E.** 2016. Social scientists' satisfaction with data reuse. *Journal of the Association for Information Science and Technology*, 67(6): 1404– 1416. DOI: <https://doi.org/10.1002/asi.23480>
- Faniel, IM and Zimmerman, A.** 2011. Beyond the Data Deluge: A Research Agenda for Large-Scale Data Sharing and Reuse. *International Journal of Digital Curation*, 6(1): 58–69. DOI: <https://doi.org/10.2218/ijdc.v6i1.172>
- Farquhar, A and Brase, J.** 2014. Data Identification and Citation – The Key to Unlocking the Promise of Data Sharing and Reuse. *D-Lib Magazine*, 20(1/2). DOI: <https://doi.org/10.1045/january2014-farquhar>
- Fear, KM.** 2013. Measuring and anticipating the impact of data reuse, PhD thesis, University of Michigan. URL: <https://deepblue.lib.umich.edu/handle/2027.42/102481>.
- Federer, LM, Lu, Y-L, Joubert, DJ, Welsh, J and Brandys, B.** 2015. Biomedical Data Sharing and Reuse: Attitudes and Practices of Clinical and Scientific Research Staff. *PLOS ONE*, 10(6): e0129506. DOI: <https://doi.org/10.1371/journal.pone.0129506>
- Francis, LP and Francis, JG.** 2017. Data Reuse and the Problem of Group Identity. In *Studies in Law, Politics, and Society*. Emerald Publishing Limited, pp. 141–164. DOI: <https://doi.org/10.1108/S1059-433720170000073004>
- Garnett, V and Edmond, J.** 2014. Building an API is not enough! Investigating Reuse of Cultural Heritage Data. URL: <http://eprints.lse.ac.uk/71611/>.
- Gleit, C and Graham, B.** 1989. Secondary Data Analysis. *A Valuable Resource, Nursing Research*, 38(6): 380–381. URL: [https://journals.lww.com/nursingresearchonline/Citation/1989/11000/Secondary\\_Data\\_Analysis\\_\\_A\\_Valuable\\_Resource.18.aspx](https://journals.lww.com/nursingresearchonline/Citation/1989/11000/Secondary_Data_Analysis__A_Valuable_Resource.18.aspx). DOI: <https://doi.org/10.1097/00006199-198911000-00018>
- Heaton, J.** 1998. Social Research Update 22: Secondary analysis of qualitative data. *social research UPDATE*, 22. URL: <http://sru.soc.surrey.ac.uk/SRU22.html>.
- Hey, T, Tansley, S and Tolle, KM.** 2009. *The fourth paradigm: data-intensive scientific discovery*, Vol. 1. Microsoft Research Redmond, WA
- Hinds, PS, Vogel, RJ and 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: <https://doi.org/10.1177/104973239700700306>
- Howard, T, Darlington, M, Ball, A, Culley, S and McMahon, C.** 2010. Opportunities for and Barriers to Engineering Research Data Re-use. URL: <http://opus.bath.ac.uk/21166/>.
- Kansa, SW.** 2015. Using Linked Open Data to Improve Data Reuse in Zooarchaeology. *Ethnobiology Letters*, 6(2): 224–231. DOI: <https://doi.org/10.14237/eb1.6.2.2015.467>
- Kim, Y and Yoon, A.** 2017. Scientists' data reuse behaviors: A multilevel analysis. *Journal of the Association for Information Science and Technology*. DOI: <https://doi.org/10.1002/asi.23892>
- Kriesberg, A, Frank, RD, Faniel, IM and Yakel, E.** 2013. The role of data reuse in the apprenticeship process. *Proceedings of the American Society for Information Science and Technology*, 50(1): 1–10. DOI: <https://doi.org/10.1002/meet.14505001051>
- Law, M.** 2005. Reduce, Reuse, Recycle: Issues in the Secondary Use of Research Data. *IASSIST Quarterly*, 29(1): 5–10. URL: <http://www.iassistdata.org/sites/default/files/iqvol291law.pdf>. DOI: <https://doi.org/10.29173/iq599>
- Li, K, Lin, X and Greenberg, J.** 2016. Software citation, reuse and metadata considerations: An exploratory study examining LAMMPS. *Proceedings of the Association for Information Science and Technology*, 53(1): 1–10. DOI: <https://doi.org/10.1002/pr2.2016.14505301072>
- Meystre, SM, Lovis, C, Bürkle, T, Tognola, G, Budrionis, A and Lehmann, CU.** 2017. Clinical Data Reuse or Secondary Use: Current Status and Potential Future Progress. *IMIA Yearbook*. DOI: <https://doi.org/10.15265/IY-2017-007>
- Missier, P.** 2016. Data trajectories: tracking reuse of published data for transitive credit attribution. *International Journal of Digital Curation*, 11(1): 1–16. DOI: <https://doi.org/10.2218/ijdc.v11i1.425>
- Molloy, J.** 2011. The Open Knowledge Foundation: Open Data Means Better Science. *PLoS Biol*, 9(12): e1001195 DOI: <https://doi.org/10.1371/journal.pbio.1001195>
- Mooney, H and Newton, M.** 2012. The Anatomy of a Data Citation: Discovery, Reuse, and Credit. *Journal of Librarianship and Scholarly Communication*, 1(1). DOI: <https://doi.org/10.7710/2162-3309.1035>
- Moore, N.** 2007. (Re)Using Qualitative Data? *Sociological Research Online*, 12(3): 1–13. DOI: <https://doi.org/10.5153/sro.1496>

- Murillo, AP.** 2014. Examining data sharing and data reuse in the dataone environment. *Proceedings of the American Society for Information Science and Technology*, 51(1): 1–5. DOI: <https://doi.org/10.1002/meet.2014.14505101155>
- NSF – National Science Foundation.** 2017. Dissemination and sharing of research results. URL: <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>.
- OSTP – Office of Science and Technology Policy.** 2013. Increasing access to the results of federally funded scientific research. URL: [https://www.usaid.gov/sites/default/files/documents/1865/NW2-CCBY-HO2-Public\\_Access\\_Memo\\_2013.pdf](https://www.usaid.gov/sites/default/files/documents/1865/NW2-CCBY-HO2-Public_Access_Memo_2013.pdf).
- Palmer, CL, Weber, NM and Cragin, MH.** 2012. The analytic potential of scientific data: Understanding reuse value. *Proceedings of the American Society for Information Science and Technology*, 48(1): 1–10. DOI: <https://doi.org/10.1002/meet.2011.14504801174>
- Pampel, H and Dallmeier-Tiessen, S.** 2014. Open Research Data. From Vision to Practice. In: Bartling, S and Friesike, S. (eds.), *Opening Science*. Cham: Springer. DOI: [https://doi.org/10.1007/978-3-319-00026-8\\_14](https://doi.org/10.1007/978-3-319-00026-8_14)
- Pasquetto, I, Randles, B and Borgman, C.** 2017. On the Reuse of Scientific Data. *Data Science Journal*, 16. DOI: <https://doi.org/10.5334/dsj-2017-008>
- Piwowar, HA and Vision, TJ.** 2013. Data reuse and the open data citation advantage. *PeerJ*, 1: e175. DOI: <https://doi.org/10.7717/peerj.175>
- Richards, JD.** 1997. Preservation and re-use of digital data: the role of the Archaeology Data Service. *Antiquity*, 71(274): 1057–1059. DOI: <https://doi.org/10.1017/S0003598X00086014>
- Rolland, B and Lee, CP.** 2013. Beyond Trust and Reliability: Reusing Data in Collaborative Cancer Epidemiology Research. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work, CSCW '13*, New York, NY, USA: ACM, pp. 435–444. DOI: <https://doi.org/10.1145/2441776.2441826>
- Rung, J and Brazma, A.** 2013. Reuse of public genome-wide gene expression data. *Nature Reviews Genetics*, 14(2): 89–99. URL: <https://www.nature.com/nrg/journal/v14/n2/abs/nrg3394.html>. DOI: <https://doi.org/10.1038/nrg3394>
- Safran, C, Bloomrosen, M, Hammond, WE, Labkoff, S, Markel-Fox, S, Tang, PC and Detmer, DE.** 2007. Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper. *Journal of the American Medical Informatics Association*, 14(1): 1–9. DOI: <https://doi.org/10.1197/jamia.M2273>
- Sands, A, Borgman, CL, Wynholds, L and Traweek, S.** 2013. Follow the data: How astronomers use and reuse data. *Proceedings of the American Society for Information Science and Technology*, 49(1): 1–3. DOI: <https://doi.org/10.1002/meet.14504901341>
- Scaffidi, C, Shaw, M and Myers, B.** 2006. Games Programs Play: Obstacles to Data Reuse, Montreal. URL: <https://pdfs.semanticscholar.org/6af5/225449d78b11de69b705ec0ed64217e2a760.pdf>.
- Schöch, C.** 2017. Wiederholende Forschung in den digitalen Geisteswissenschaften. URL: <https://christofs.github.io/wiederholende-forschung-dhd/>.
- Shen, Y.** 2016. Research Data Sharing and Reuse Practices of Academic Faculty Researchers: A Study of the Virginia Tech Data Landscape. *International Journal of Digital Curation*, 10(2): 157–175. DOI: <https://doi.org/10.2218/ijdc.v10i2.359>
- Sun, G and Khoo, CSG.** 2017. Social science research data curation: issues of reuse. *Libellarium: journal for the research of writing, books, and cultural heritage institutions*, 9(2). DOI: <https://doi.org/10.15291/libellarium.v9i2.291>
- Szabo, V and Strang, VR.** 1997. Secondary Analysis of Qualitative Data. *Advances in Nursing Science*, 20(2): 66. URL: <https://insights.ovid.com/pubmed?pmid=9398940>. DOI: <https://doi.org/10.1097/00012272-199712000-00008>
- Tenopir, C, Dalton, ED, Allard, S, Frame, M, Pjesivac, I, Birch, B, Pollock, D and Dorsett, K.** 2015. Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. *PLOS ONE*, 10(8): e0134826. DOI: <https://doi.org/10.1371/journal.pone.0134826>
- Ullman, J.** 2017. Technical Perspective: Building a Safety Net for Data Reuse, Commun. *ACM*, 60(4): 85–85. DOI: <https://doi.org/10.1145/3051086>
- Wallis, J.** 2014. Data Producers Courting Data Reusers: Two Cases from Modeling Communities. *International Journal of Digital Curation*, 9(1): 98–109. DOI: <https://doi.org/10.2218/ijdc.v9i1.304>
- Wallis, JC, Rolando, E and Borgman, CL.** 2013. If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. *PLOS ONE*, 8(7): e67332. DOI: <https://doi.org/10.1371/journal.pone.0067332>




- Wellcome Trust.** 2010. Policy on data management and sharing. URL: <http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm>.
- Wicherts, JM, Borsboom, D, Kats, J and Molenaar, D.** 2006. The poor availability of psychological research data for reanalysis. *American Psychologist*, 61(7): 726–728. DOI: <https://doi.org/10.1037/0003-066X.61.7.726>
- Yoon, A.** 2015a. Data Reuse and Users' Trust Judgments: Toward Trusted Data Curation, Dissertation, University of North Carolina at Chapel Hill Graduate School, Chapel Hill, NC. URL: <https://cdr.lib.unc.edu/record/uuid:2c2268b3-88cf-4397-b038-b39e88f80d83>.
- Yoon, A.** 2015b. "Making a square fit into a circle": Researchers' experiences reusing qualitative data. *Proceedings of the American Society for Information Science and Technology*, 51(1): 1–4. DOI: <https://doi.org/10.1002/meet.2014.14505101140>
- Yoon, A.** 2016. Red flags in data: Learning from failed data reuse experiences. *Proceedings of the Association for Information Science and Technology*, 53(1): 1–6. DOI: <https://doi.org/10.1002/pr2.2016.14505301126>
- Yoon, A.** 2017a. Data reusers' trust development. *Journal of the Association for Information Science and Technology*, 68(4): 946–956. DOI: <https://doi.org/10.1002/asi.23730>
- Yoon, A.** 2017b. Role of communication in data reuse. *Proceedings of the Association for Information Science and Technology*, 54(1): 463–471. DOI: <https://doi.org/10.1002/pr2.2017.14505401050>
- Zimmerman, A.** 2007. Not by metadata alone: the use of diverse forms of knowledge to locate data for reuse. *International Journal on Digital Libraries*, 7(1–2): 5–16. DOI: <https://doi.org/10.1007/s00799-007-0015-8>
- Zimmerman, AS.** 2003. Data Sharing and Secondary Use of Scientific Data: Experiences of Ecologists, Dissertation, The University of Michigan, Michigan. URL: <http://hdl.handle.net/2027.42/39373>.
- Zimmerman, AS.** 2008. New Knowledge from Old Data: The Role of Standards in the Sharing and Reuse of Ecological Data. *Science, Technology, & Human Values*, 33(5), 631–652. DOI: <https://doi.org/10.1177/0162243907306704>

**How to cite this article:** van de Sandt, S, Dallmeier-Tiessen, S, Lavasa, A and Petras, V. 2019. The Definition of Reuse. *Data Science Journal*, 18: 22, pp. 1–19. DOI: <https://doi.org/10.5334/dsj-2019-022>

**Submitted:** 30 November 2018    **Accepted:** 24 May 2019    **Published:** 20 June 2019

**Copyright:** © 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Data Science Journal* is a peer-reviewed open access journal published by Ubiquity Press.

**OPEN ACCESS** 