

**The legal status  
of raw data:  
a guide for research practice**

# About this publication

The legal status of raw data: a guide for research practice

This legal guide was produced for SURFdirect, SURF's digital rights Expertise Community for higher education, by the Centre for Intellectual Property Law (CIER).

SURFdirect  
PO Box 2290  
NL 3500 GG UTRECHT  
The Netherlands  
T +31 (0)30 234 6600  
F +31 (0)30 233 2960  
E [info@surf.nl](mailto:info@surf.nl)  
W [www.surf.nl/surfdirect](http://www.surf.nl/surfdirect)

## Authors

Prof. Madeleine de Cock Buning, *CIER* (parts I, II and III)  
Allard Ringnalda, *CIER* (part I)  
Tina van der Linden, *CIER* (parts II and III)

## Final editing

Annemiek van der Kuil, *SURFdirect*

SURF is the Dutch higher education and research partnership for network services and information and communication technology ([www.surf.nl](http://www.surf.nl)). This publication can be downloaded from the SURFfoundation website: [www.surf.nl/publicaties](http://www.surf.nl/publicaties).

© SURFfoundation  
July 2009

This publication is published under Creative Commons Licence Attribution 3.0 Netherlands.



# Contents

- 1 Legal status of research data ..... 5
  - 1.1 Introduction ..... 5
  - 1.2 Definition of research data ..... 5
  - 1.3 Purpose ..... 6
  - 1.4 Plan of discussion ..... 6
  - 1.5 A brief guide to determining what consent is necessary to reuse someone else’s research data ..... 7
  - 1.6 Disclaimer ..... 8
- Part I –Intellectual Property ..... 11
- 2 Copyright law ..... 13
  - 2.1 Introduction ..... 13
    - 2.1.1 Owner or rightholder ..... 13
    - 2.1.2 Legal framework: the content is free ..... 13
  - 2.2 Structure of this section ..... 15
  - 2.3 When may research data be protected by copyright? ..... 15
    - 2.3.1 The form of the work ..... 15
    - 2.3.2 General criteria ..... 15
    - 2.3.3 Copyright-protected research data ..... 16
    - 2.3.4 Other kinds of data: photographs, maps, diagrams, etc. .... 17
    - 2.3.5 Copyright-protected databases ..... 18
    - 2.3.6 How copyright protection arises ..... 18
  - 2.4 What actions require consent? ..... 18
    - 2.4.1 Some components of the work are not subject to copyright ..... 18
    - 2.4.2 What use is subject to copyright? ..... 19
    - 2.4.3 Exception: works produced by the authorities ..... 19
    - 2.4.4 Exception: citations ..... 20
    - 2.4.5 Exception: private copy ..... 20
  - 2.5 Who is the rightholder? ..... 21
    - 2.5.1 Shared copyright or involvement of a superior ..... 21
    - 2.5.2 Cases in which the law allocates copyright ..... 21
    - 2.5.3 Cases in which contractual provisions may be relevant ..... 21
  - 2.6 Reuse of protected data: sometimes there is double copyright ..... 22
  - 2.7 Conclusion: the importance of copyright ..... 23
- 3 Database right ..... 25
  - 3.1 When is research data protected by database right? ..... 25
    - 3.1.1 Collection ..... 25
    - 3.1.2 Searchability ..... 26
    - 3.1.3 Substantial investment ..... 26
  - 3.2 What actions require consent? ..... 27
    - 3.2.1 Retrieving material ..... 28
    - 3.2.2 Reuse ..... 28
    - 3.2.3 Substantial portion ..... 28
    - 3.2.4 Derivation ..... 28
    - 3.2.5 Exceptions: scientific/scholarly research ..... 29
    - 3.2.6 Exception: government databases ..... 29
    - 3.2.7 Consent, consultation, and contractual arrangements ..... 29
    - 3.2.8 Borrowing from another database containing research data ..... 30
  - 3.3 Who is the rightholder? ..... 30
  - 3.4 Relationship to copyright ..... 30

3.5	Summary .....	31
4	Protection of non-original writings .....	33
4.1	When does protection of non-original writings apply? .....	33
4.2	What actions require consent? .....	33
4.3	Whose consent must be secured? .....	34
5	Internet .....	35
5.1	The Internet .....	35
5.2	Contractual arrangements .....	35
5.3	Schematic overview .....	36
Part II – Privacy .....		39
6	Privacy .....	41
6.1	Personal data .....	41
6.1.1	Anonymized information .....	41
6.1.2	Indirectly identifying data .....	41
6.1.3	Directly identifying data .....	41
6.1.4	Special personal data .....	42
6.2	Obtaining of personal data .....	42
6.3	Storage of personal data .....	44
6.4	Use of personal data .....	44
6.5	Sharing of personal data .....	45
6.6	Sanctions .....	45
Part III – Liability .....		47
7	Liability .....	49
7.1	Legal classification .....	49
7.2	Information obligations .....	49
7.3	Liability .....	50
7.4	General Terms and Conditions .....	51
7.5	Conclusion as regards liability .....	51

# 1 Legal status of research data

## 1.1 Introduction

The legal protection of raw research data is relevant for a number of reasons. Raw research data is primarily relevant to other researchers, who can use it for new research and new publications. It is therefore important to know what the legal status is of the data concerned, because under certain circumstances research data is protected by copyright. This means that certain action taken in respect of the data requires consent, for example from its “author” (i.e. the “maker” of the data). Not every use of protected research data requires the author’s consent, however; some actions involving data can be carried out without consent.

Conversely, it may be desirable in some cases for a researcher to protect his/her own data against unauthorised use by a third party. Can a third party simply go ahead and utilise research data, and what arrangements then need to be made? This legal guide explains three regimes for the protection of research data. First, however, we will give a more detailed definition of what we mean by “research data”.

## 1.2 Definition of research data

“Research data” (or “research information”) is a very broad term. In the humanities, one finds the following definition:

*All data collected in some way or another in the context of scientific/scholarly research. A distinction can be made between primary data (empirical, observed, measured data) and secondary data. Secondary data is data derived from sources created previously (figures published by the authorities, data assembled previously, archived data, texts, etc.).*

This definition is a formal and not a material one: it indicates only *when* data can be designated to be research data, but it does not comprise any actual definition of the term “data”. Given the wide-ranging nature of research data, that would not be an easy thing to do.

A somewhat different definition is current in the natural sciences:

*A datum is an element that has relevance and semantic value...Data is used to describe features of persons, things, actions, etc. taken from reality.*

We also have *metadata*. Metadata is data that describes the characteristic features of certain information; in other words, it is data about data. The metadata with a particular document may include such things as the name of the author, the number of pages, and the language in which the data is provided. The propose of including metadata with the data to which it relates is to make the latter easier to trace. Metadata is sometimes added automatically, but generally human intervention is required.

This legal guide is not written with a view to specific types of data. It can therefore basically be used by researchers in all disciplines. It does, however, present a number of criteria for determining the legally relevant application of a system of protection. It thus provides pointers for deciding, on a case-by-case basis, whether research data comes under a certain system of protection.

This guide is not concerned with data in any one particular form. Basically, any type of data – whether it be separate items of data, datasets, or databases – may qualify for some kind of protection. Whether it is in fact protected must be determined on a case-by-case basis. In deciding, we need to focus on research data in its most primary form, as it arises from measurements or

observations. This means data that has not yet been incorporated into a publication etc.; if it has been, one soon arrives in the domain of the copyright to which books, articles, etc. are subject. This guide will deal, however, with data that has been processed to a very limited extent, for example by being selected or filtered, and data that has been converted into an ordered database.

The guide will therefore consistently work on the basis of the *legally* relevant definitions of the term “data”. For some data, the law provides no protection. As will be explained below, facts are basically free; we then talk of “the bare facts”. Some data is in fact subject to legal protection. That may be the case if the facts concerned have been put into a certain form or if they constitute a protected database or data collection. In such cases, we are dealing with more than merely the bare facts themselves and we speak of “protected data”. We will use the term “research data” to refer both to the bare facts and to protected data. When reading this legal guide, the reader will do well to remember that the relevant legal terminology does not generally correspond very closely to ordinary usage. Whenever a term appears familiar, one needs to realise that it has a very specific meaning in legal language.

We can illustrate this with the term “database”. The law provides that, under certain conditions, databases may be subject to a certain amount of protection (see Section 3). This might make one think, in accordance with ordinary usage, that “databases” in the sense of large, indexed, digital collections of associated data enjoy protection. Legally speaking, however, that is not in fact the case. The legal definition of “database” is highly specific. On the one hand, it sets no requirements regarding the size of the database or the medium on which it is recorded. A telephone book may be a database, but so may a short table. On the other hand, the legal definition stipulates all sorts of other requirements for a data collection before it can be deemed to constitute a database. It must, for example, have been created by means of “substantial investment”. If there has been no substantial investment, then the collection of data concerned is not a database from the legal point of view, even though it may be considered to be one in normal usage.

### **1.3 Purpose**

Part I of this guide clarifies the legal protection afforded for research data. It envisages two different situations:

- The reader is a researcher who wishes to make use of other people’s data for a new study of his own. In that case, it is necessary to consider whether the research data concerned is subject to one of the protection regimes and whether the use to which the researcher wants to put the data means that consent is required.
- The reader is a researcher who wants to protect his own raw data. This may be relevant if the reader intends sharing the raw data (or files containing that data) with other researchers, for example by uploading it to the Internet.

This guide focuses on the first of these two situations. Its primary aim is to clarify matters for researchers who want to know what they can do with other people’s data. At the same time, it provides relevant information regarding the second situation by providing information for researchers who wish to know more about the status of their own research data.

### **1.4 Plan of discussion**

Part I outlines the scope of each protection regime. The reader can use the pointers provided to assess the legal status of actual research data. He can then check whether the use that he wishes to make of that data is subject to a protection regime and therefore whether consent is required. We then look at the practical side of things for each regime, namely whose consent must be secured and how that consent can be arranged. Each section concludes with a concise summary of the actions that require consent if the research data is subject to the regime concerned.

With a view to the reuse of other people’s research data, Section 5 presents a schematic overview to help determine whether a certain type of use requires consent. This is based on a number of

practical questions that will play a role in deciding whether certain research data is protected; it also gives references to the relevant sections of this guide.

We then go on to discuss the following three protection regimes:

- copyright (Section 2);
- database right (Section 3);
- protection of non-original writings (Section 4).

We have chosen to discuss the regimes in the order of the amount of protection that they provide. The most extensive – i.e. the “toughest” regime – is copyright, followed by database right, and finally protection of non-original writings. The more extensive the protection provided, the stronger the position of the author vis-à-vis third parties. On the other hand, the tougher regimes generally have a more restricted scope of application. If research data is not covered by copyright, it may well be protected by database right. If database right does not apply either, then it is possible that protection of non-original writings is still applicable. Copyright and database right may also apply to one and the same work simultaneously. Protection of non-original writings, however, is only relevant when copyright or database right do not apply.

## 1.5 A brief guide to determining what consent is necessary to reuse someone else’s research data

To help researchers quickly determine what consent they need in order to reuse someone else’s research data, we have produced a brief guide with references to the detailed explanations elsewhere in this legal guide. It is important to note that this brief guide cannot replace the full legal guide but is meant only as an aid to finding one’s way around this document. This brief guide is derived from the schematic overview in Section 5.3.

### **Do you want to make a copy for your own use?**

- You do not need to seek consent for this.

### **Do you want to input data into your own scientific/scholarly database, without sharing it with anyone other than your own team of researchers?**

- You do not need to seek consent for this.

*Explanation: Even if the data is protected by copyright or database right, this type of use is one of the permitted exceptions: making a copy for your own use (Section 2.4.5) of a copyright-protected work or retrieving substantial portions of a database in the context of scientific/scholarly research (Section 3.2.5; retrieving non-substantial portions does not require consent in any case: Section 3.2.3).*

### **Do you want to publish some of the data that you are borrowing or to share it with someone other than your own team of researchers?**

- Then you probably need to secure the author’s consent.

#### ***Check whether copyright applies:***

**Does the research data have an original character of its own and bear the personal stamp of the author?** (Tip: Is it conceivable that two different authors, working separately, could have arrived at precisely the same form?.)

- Copyright applies and you need to secure the consent of the author/authors.

*Explanation: The decisive point is therefore whether the author had freedom of choice in arranging the data in a particular form and whether he did this – deliberately or otherwise – on the basis of personal preference (Section 2.3.2). Even just a selection consisting solely of bare facts may be subject to copyright protection (Section 2.3.3).*

#### ***Check whether database right applies:***

**Are you dealing with a collection of independent materials (research data)?  
Is the data collection arranged systematically?**

**Has there been a substantial investment in obtaining, verifying, or presenting the materials?** The investment in carrying out the research that generates the data does not count.

- Database right applies and you need to secure the consent of the author/authors.

*Explanation: This is explained in detail in Sections 3.1.1., 3.1.2, and 3.1.3.*

**Do you intend publishing the protected data that you have borrowed in some way, or making it available to other people in some other way?**

- Copyright applies and you need to secure the consent of the author/authors.

*Explanation: in cases such as this, the law speaks of **publication** (Section 2.4.2). Other people's data can only be copied or utilised for the purpose of your own research without the author's consent if it is not published (Section 2.4.2).*

*The proposed use requires the author's consent (Section 4.3).*

**Are you dealing with the right to cite, works produced by a government body, the all-inclusive form, a selection of bare facts or other types of data (photos etc.)?**

- Then there is a copyright law exception and you do not need to secure consent.

*Explanation: These exceptions are explained in Sections 2.4, 2.1.2, 2.3.3, and 2.3.4.*

**Do you intend not only borrowing data from the database for your own research but also publishing it further, for example in a publication, or do you intend sharing your own material (which will then include the data you have borrowed) with other people?**

- Database right applies and you need to secure the consent of the author/authors.

*Explanation: In cases such as this, the law speaks of reuse.*

**Do you intend reusing a substantial portion of the database?**

- There are no clear criteria for what constitutes a "substantial" portion. It is necessary to consider the relationship between the reused portion and the database as a whole, as well as the technical and economic value of the reused portion.

*Explanation: See Section 3.2.3; see also 3.2.8.*

**Do you intend using a government database?**

- You do not need to seek consent for this.

*Explanation: Unless provided otherwise, government databases are basically not subject to database right (Section 3.2.6). You do not require the consent of the rightholder in order to retrieve substantial portions of a government database in the framework of scientific/scholarly research (Section 3.2.5)*

**Do you intend borrowing in full or publishing in full a text that has been published (or is intended for publication) but that is not protected by copyright or database right?**

- This is only permitted if you publish it in an altered version.

*Explanation: See Section 4.*

## 1.6 Disclaimer

It should be noted that research data may also be subject to other legal rules than the ones discussed here. Even if you can assume from the available information that certain research data can be used, it is still possible that other legal provisions restrict that use. When we speak here of data that is "free", we mean only that no consent for its use is required under copyright, database right, or the protection of non-original writings. There may still, however, be rights beyond these domains that will prevent the use of "free" data. Certain data may, for example, contain personal data, meaning that privacy rules apply. Under certain circumstances, one may also be liable as regards the inaccuracy of the research data. The various regimes may also be relevant if protected data is used with the author's consent. These topics are dealt with in Parts II and III.

Our intention is to provide an overview of the current situation on the basis of the most important legislation and case law. Our explanations will focus as far as possible on raw data. It is relevant,

however, that the legislation and case law devote hardly any attention to the legal status of research data. We have attempted to set out the main points of the existing law as clearly as possible so that the reader has some pointers for assessing the extent to which certain research data is protected. This guide does not, however, make any proposals for reinterpreting the existing legislation and case law, or suggestions for new or amended regulations. We have taken account of case law up to 1 December 2008.

It is important to note that this legal guide only gives indications for how current Dutch law applies. Whether a protection regime applies in a particular case must be determined on a case-by-case basis in the light of a large number of different circumstances. In cases of doubt, it is therefore advisable to secure the author's consent or to seek legal advice.



**Part I –Intellectual Property**



## 2 Copyright law

### 2.1 Introduction

Copyright gives the “author” (i.e. the maker or creator) of a work the exclusive right to decide on the publication and duplication of that work (or parts of it). This means that the author’s consent is required if another person wishes to “publish” or “duplicate” the work. The author is referred to as the *copyright holder*.<sup>1</sup>

Copyright law provides protection for works that display a certain creativity or originality. The law does not, however, explain what works enjoy copyright protection. It is the courts that decide this on a case-by-case basis; Section 2.3 considers the criteria that they apply in doing so. In many cases, it is clear that a work enjoys copyright protection: writing a book or article almost always involves personal creativity. In the case of research data, things are not so clear: data often consists of bare facts. In many cases, then, research data will not be protected by copyright.

It is important to distinguish between two different matters. Copyright is not intended to protect the researcher who makes a *discovery*, no matter how creative or original that discovery may be. Copyright cannot be used to protect newly discovered data. It can, however, protect the *form* in which the discoverer has written down the bare facts. That form must be the result of creative choices; if that is the case, others are only permitted to borrow the form if they secure consent to do so.

It is not necessary for the work to bear the “copyright notice” (©), which is simply an indication that the work is copyright-protected. It is important to note that works sometimes bear the copyright notice even though they do not meet the requirements and therefore do not in fact enjoy copyright protection.

#### 2.1.1 Owner or rightholder

Researchers sometimes wonder who the *owner* is of the data that they have collected or discovered. This is not really a relevant question within a legal context. As we have already seen, bare facts do not belong to anyone; information is not subject to being owned. Someone can, however, be the *rightholder* in respect of certain data. That is the case if the form in which the data has been put means that it qualifies for protection under the various regimes discussed in this guide. The law provides that the author of that form (or in some cases another party) enjoys certain *exclusive rights* to carry out certain actions. An exclusive right means that the party that enjoys that right – the rightholder – is the sole party that can perform any acts covered by the right; others must secure the consent of the rightholder.

There is generally no obligation to share research data with others. Regardless of whether the data is protected or whether the researcher is the rightholder, the researcher can always decide not to make his data available to others, so that they also cannot use it. Contractual agreements, for example with a client, may be relevant here. In some cases, the client stipulates that all the data should be sent to it. The contract may also contain provisions that oblige the researcher to keep the data secret. Such arrangements must be complied with, regardless of the regimes discussed here.

#### 2.1.2 Legal framework: the content is free

---

<sup>1</sup> In some cases, the law designates someone other than the author as the copyright holder; see Section 2.5.

Under certain circumstances, the law provides protection for certain products of the mind. This means that the author has the exclusive right to determine what is done with them. A third party that wishes to make use of the product concerned is then obliged to request the author's consent; the author can then stipulate that the third party pay to use the product.

When someone writes a book or article, for example, it is subject to copyright, meaning that the author enjoys certain exclusive rights in respect of the work. The author alone can decide whether and where the text may be published. If someone else wishes to republish the article or book (or part of it), for example in a collection or on the Internet, that person will need to secure the consent of the author. Not everything that third parties do with a work is relevant here, of course; they are always permitted, for example, to make a copy of the work for themselves, or to cite from it.

It is important to realise that the protection regimes that we are discussing never provide protection for the bare facts themselves: facts are free. There is a distinction between non-protected *content* (bare facts) and the *form* which the content takes. It is only the form that may qualify for protection under certain circumstances, meaning that third parties cannot simply adopt that form.

#### *Example*

If someone publishes a scientific/scholarly article on the basis of research that he has carried out, the article is protected by copyright. This means that a third party is basically not permitted to copy some or all of the article without the author's consent. However, copyright protects only the article as a *form*. The *content* – i.e. the bare facts that it contains – is free. It belongs to the domain of knowledge, which is a public good. The measurements that underlie the article also belong to the content, and are basically not protected by copyright. Third parties can write a new publication in their own words (i.e. a publication with a form of its own) on the basis of the measurements.

When determining whether raw research data is protected, it is important to bear this background in mind. Research data can generally be categorised as bare facts. After all, scientific/scholarly knowledge cannot be monopolised: a researcher does not have an exclusive right to utilise the data that he has discovered or collected. This means that research data as such will often be “free”.

In practice (i.e. in legal practice), however, it is not always clear where the boundary must be drawn between content and form: at what point do bare facts become protected data? After all, even bare facts have to be set out in a certain way, i.e. in a certain form. Measurements can be presented in tables or complex formulas; data of a more qualitative nature can often only be presented in words. Data also occurs in a variety of forms: it may consist not only of a few figures or words but sometimes of datasets or complete databases.

Before dealing with this in detail, we can already say in general that bare facts may become protected data in the following cases:

- the bare facts are written down in subjectively determined original wording (copyright, Section 2);
- the bare facts are presented in some other way in a subjectively determined original manner (copyright, Section 2);
- the bare facts have been selected from a larger set of data (copyright, Section 2);
- the bare facts form a database in the legal sense (database right, Section 3).

In all these cases, the bare facts have been arranged in a certain form, which may qualify for legal protection. In legal parlance, we then speak of a “(protected) work”. For the precise criteria and for the extent of the legal protection provided, see the sections referred to. In the cases not listed above, protection of non-original writings is relevant (see Section 4).

The distinction between form and content (bare facts) is illustrative for the way the law works. One needs to realise, however, that the distinction between form and content has no useful legal

relevance. After all, it is virtually impossible to determine legally which aspects of a work belong to the content and which to the form.

## 2.2 Structure of this section

This section tries to answer two important questions:

- When is research data protected by copyright?
- What kind of use requires consent when copyright-protected data is concerned?

Section 2.3 deals with the scope of copyright. It sets out criteria that can be used to determine whether certain research data may be protected by copyright. In doing so, it deals separately with copyright-protected databases. Section 2.4 explains what kind of use requires consent when copyright-protected data is concerned. Section 2.5 then deals with the question of whose consent needs to be sought and what arrangements should be made. Section 2.6 gives a brief and to the point summary of the copyright system.

## 2.3 When may research data be protected by copyright?

### 2.3.1 The form of the work

In legal terms, copyright covers a “work”. This means a perceptible manifestation of a certain “mental creation”. It is important that the work is in fact perceptible: thoughts alone are not covered by copyright. It is not necessary, however, for the work to have been put down on paper or given some other physical form. Oral statements or expressions such as lectures may also be protected by copyright. Research data will generally have been put down in writing or recorded in digital form but, under the circumstances referred to below, research data presented orally may also be covered.

### 2.3.2 General criteria

The legislation and case law on copyright set a number of requirements that must be met if a work is to enjoy copyright protection. As we already saw in the general introduction, copyright protection applies not to the content – i.e. the bare facts – but to the form that has been given to that content. However, not every form is in fact subject to copyright protection. The form must also show that the personal approach of the author has led to certain subjective choices being made.

In the relevant case law, there are two requirements for a work to enjoy copyright protection:<sup>2</sup>

- the work must have *an original character of its own*;
- the work must bear the *personal stamp* of the author.

Whether a work is actually protected will need to be determined on a case-by-case basis by the courts, with both these criteria being considered. Briefly, one can say the following. An “original character of its own” generally means that the work must not be borrowed from other works. This means that some or all of it must not be quoted verbatim, or with only a minimum of changes, from another, pre-existing work. Commonly utilised forms are in general not considered to be original. The “personal stamp” of the author means that the work must be the result of a person’s creative endeavour. In creating the work, the author must have made certain choices on the basis of personal preferences (whether or not those preferences are conscious or unconscious).

Simply presenting a series of measurements in a table, for example, will not easily meet these requirements; the person who produced the table has simply listed a number of figures one after

---

<sup>2</sup> Supreme Court of the Netherlands [*Hoge Raad, HR*], 30 May 2008, *LJN*: BC2153 (*Endstra tapes*), ground 4.5.1.

the other, probably in such a way that any other author could arrive at a similar table. That may be different if, for example, the figures are discussed in a text. If that is the case, then it is much more likely that one can speak of creative, personal endeavour which qualifies for copyright protection. Only then is there a copyright-protected “work” from the legal point of view.

For a work to be deemed to bear the personal stamp of its author, the latter must have had a certain freedom of choice when creating it; in other words, he could choose between various different forms for the work. If the author did not have that freedom, then his product will not usually enjoy copyright protection.

One practical rule of thumb is to ask oneself whether it is conceivable that two authors might have produced exactly the same work independently.<sup>3</sup> If that is in fact the case, then the product concerned will not have an original character of its own, nor will it bear the personal stamp of the author, because one can assume that its form is not based on any creativity. Rather, the form will be dictated by practical, objective considerations that would be the same for all authors. In such a case, it would be unfair for just one of those authors to have the exclusive right to that form.

### 2.3.3 Copyright-protected research data

The question that now arises is whether research data qualifies for copyright protection on the basis of the above criteria. There are in fact very few examples in case law.

In deciding whether research data is in fact protected, one needs to bear in mind the following general rules. When an author creates a work, he bases it on some objective data – the bare facts – regardless of whether he has discovered that data for himself. Those bare facts are not protected by copyright; after all, they are facts, and facts can be used by anyone (Section 2.1.2). When the author goes on to process those bare facts, for example by writing them down in a particular manner, he can make certain choices. If those choices are based on a personal preference, they may enjoy copyright protection. This then means that someone else cannot borrow them without consent (see Section 2.4).

What does this actually mean in practice as regards research data? One has to determine whether the author made subjective choices when processing the research data and, if so, what those choices were. When research data is presented, there will generally be little or no question of any subjectively motivated choices. The way in which research data is recorded is often determined solely by practical considerations. Categorising, structuring, and selecting the data, for example, will normally take place on the basis of scientific/scholarly insights or standards and considerations of a practical or scientific/scholarly nature (for example so that the data can be analysed using software such as SPSS). This means that there is only restricted scope for the researcher’s personal preferences. If one applies the rule of thumb mentioned above, one can indeed perfectly well imagine that two researchers would arrive independently at the same form for the data. In that case, there is no question of copyright protection.

Situations are conceivable, however, in which research data is given a particular form on the basis of subjective choices that are protected by copyright. According to rulings by the courts, that may be the case if the researcher makes a *selection* from a large number of bare facts and then *processes* that selection in some way, for example by placing the selection in a table or diagram. If the researcher makes subjective choices when doing so, then the selection and processing qualify for copyright protection. This may also be the case if some or all of his choices are made on the basis of scientific/scholarly or technical knowledge, insight, and experience.<sup>4</sup>

It remains difficult to determine in particular cases whether selecting and processing the bare facts has actually involved subjectively determined choices. How, for example, should one assess the

---

<sup>3</sup> J.H. Spoor, D.W.F. Verkade, and D.J.G. Visser, *Auteursrecht*, Kluwer 2005, Section 3.9 (infra: Spoor, Verkade, and Visser 2005).

<sup>4</sup> Supreme Court, 24 February 2006, *NJ* 2007, 37 (*Technip/Goossens*).

selection and processing of data with a view to statistical analysis? This will generally not involve any personal choices because both selection and processing are determined by statistical methodology. It has in any case been established in case law that if the selection and processing concerned were *necessary*, then no copyright protection applies; in such a case, there is after all no question of any real freedom of choice.<sup>5</sup>

Protection of selection and processing was considered possible, for example, in the case of a kinetic scheme that included chemical formulae describing the production process for a particular substance. Although chemical formulae are facts, it was considered possible for the scheme to enjoy copyright protection because the author of the scheme made a selection from a very large number of formulae and then incorporated them into the scheme. Even though they were based on scientific and technical knowledge, insight, and experience, his choices in doing so were subjectively determined; a different author might well have arrived at a different selection of formulae and a different scheme.

A selection of bare facts may therefore enjoy copyright protection, meaning that consent must be secured before someone can copy it. If someone copies it without securing consent, then he is infringing the copyright. This is only different if it can be shown that that person arrived independently – i.e. by means of his own labour – at the same (or virtually the same) selection. In that case, the selection has not been copied.

Qualitative data – for example descriptions, observations, etc. – that are in a form that displays an original character of its own and the personal stamp of the author may also enjoy copyright protection. That is not the case, however, if use is made only of short phrases or standard scientific terms and structures. Matters may be different if the data consists of lengthy descriptions that clearly demonstrate a creative choice on the part of the author, for example because the phrasing and vocabulary are unique.

#### **2.3.4 Other kinds of data: photographs, maps, diagrams, etc.**

So far, we have been talking about research data that has been put down in writing, i.e. in figures or words. There are many kinds of data, however, that are presented in a more visual form, for example photographs, videos, maps, and diagrams.

Photographs are more likely to enjoy copyright protection than research data that have been put down in writing. After all, taking a photograph generally involves a number of choices beforehand, for example selecting the subject, the perspective, the composition, etc. But not every photograph qualifies for protection: to be protected, a photograph needs to have an original character of its own and bear the personal stamp of the photographer. In the case of a photograph that has been taken by purely mechanical means, for example, no personal choices are involved: the photograph has simply been taken automatically without human intervention. Satellite photos are an example of this and they are not subject to copyright protection. *Processing* such photos, however, may indeed involve making personal choices. Adding colours or data to the photograph, for example, may bring it within the remit of copyright if doing so in fact involves personal choices. The same applies *mutatis mutandis* to video recordings.<sup>6</sup>

Maps are something of a special case. If a map represents an existing geographical situation, then creating it does not really involve any creative choices. But actually designing the map may of course require creative choices, and these may qualify for copyright protection. Adding research data to a map may also lead to it being protected if this is done in a way that involves personal, creative choices.

---

<sup>5</sup> Supreme Court, 16 June 2006, AMI 2006, 161 (*Kecofa/Lancôme*).

<sup>6</sup> It should be noted that databases containing photographs (for example a database comprising sorted satellite photographs) may be protected by database right (see Section 3).

For example, an updated map – for instance showing new roads – may use different colours and may (or may not) include viewpoints or other interesting features for tourists; it may consequently enjoy copyright protection.

The situation regarding diagrams is similar; they may enjoy copyright protection if personal choices were involved in compiling or formatting them.

### **2.3.5 Copyright-protected databases**

Certain databases may also be subject to copyright protection.<sup>7</sup> The following requirements apply:

- the database must consist of a collection of works, data, or other independent materials;
- it must be possible to trace those materials by means of a method or system.

A database is only protected by copyright if the *choice or arrangement of the materials* that it contains complies with the criterion that it should have an original character of its own and bear the personal stamp of the author. This may also be the case if the subjective choices made in creating the database – for example selection of the data – were determined by scientific/scholarly or technical knowledge, insight, or experience.<sup>8</sup>

### **2.3.6 How copyright protection arises**

Finally, it should be noted that copyright protection arises automatically if a work complies with the criteria explained above. No registration is required, nor does the work need to bear the “copyright notice” (©). Moreover, the work does not need to have been published in some way or another or be intended for publication. This means, for example, that a set of processed research data for internal use may enjoy copyright protection.

## **2.4 What actions require consent?**

If it has been determined that research data qualifies for copyright protection according to the above criteria, then certain types of use of that data require the consent of the author (or perhaps another party: see Section 2.5). This section deals with the actions for which consent must be secured and those for which consent is unnecessary.

### **2.4.1 Some components of the work are not subject to copyright**

One important point to note is that copyright does not apply to *all* elements and aspects of a work. It is only the subjectively determined elements or aspects, those determined by the personal creativity of the author, that are protected. This is important where research data is concerned. If research data is protected by copyright, then it is in general only the structuring, selection, or wording of the data that are covered (see above). The bare facts themselves can, however, be freely used by other people as long as those facts are presented in a structure of their own and in the user’s own words and if use of the facts does not constitute retrieval or reuse within the meaning of database law (see Section 3.2). Copying a copyright-protected work in unaltered form is therefore often not permitted, while copying the content – i.e. the bare facts – and presenting them in one’s own words or form is generally permissible.

Let us assume, for example, that a collection of research data in the form of figures is protected by copyright because the collection is structured in a special and original way that displays personal creativity on the part of the author. It is then only the structuring of the figures that is protected by copyright. This means that a person other than the author cannot copy that structure or publish it elsewhere without the author’s consent. The bare facts contained in the structure are not

---

<sup>7</sup> For databases subject to database right, see Section 3.

<sup>8</sup> Supreme Court, 24 February 2006, *NJ* 2007, 37 (*Technip/Goossens*)

protected, however, even if they were discovered by the maker. Another person may therefore freely copy the figures and use them for his own research, as long as he does not also copy the copyright-protected structure in which the figures are included. The same applies, *mutatis mutandis*, to research data continued in the form of a text which is protected.

For the sake of comparison: the copyright protection afforded to a scientific/scholarly publication (i.e. a book or article) is wide-ranging because the whole structuring and wording derive from personal choices made by the author. Nevertheless, such publications also contain objective aspects that fall outside the scope of copyright. Examples of this are such things as common – and therefore not original – phrases or expressions, but also the scientific or scholarly theory defended in the article. The theory is an objective, factual given, even if it has been conceived by the author himself; it is only the wording and the manner of treatment that have been chosen for the article – and which are therefore subjective – that are protected by copyright. Another person may, however, describe the theory in his own words and in a structure of his own without seeking the author's consent.

#### **2.4.2 What use is subject to copyright?**

If someone wishes to use not only the bare facts but also the copyright-protected form, it is necessary to determine whether the proposed use is such that the author's consent is required. The law applies two broad terms to indicate what is covered by the author's exclusive right. The author is the only person who can decide on *publication* and *duplication* of a copyright-protected work.<sup>9</sup> In so far as relevant, *duplication* here means the production of physical or digital copies of the work (or parts of it). This therefore involves making a copy or a print-out of a written work, scanning it, or downloading a digital copy. Duplication also includes adapting, translating, or imitating the work (or parts of it). In the latter case, we are dealing with non-literal copies that are nevertheless taken from the work so that the protected subjective elements are largely preserved.

*Publication* means making the work available to the public in some way or other. This may involve circulating prints of the work but also making it available digitally on the Internet, displaying it publicly, or reading it out.

##### *Example*

Let us assume that we are dealing with a copyright-protected set of research data because the data consists of observations expressed in personal and original wording. Someone other than the author wishes to write a new article on the basis of this data. That is permitted. The objective data from the observations can be used without the original author's consent if this is done in the second author's own words. However, if the second author intends incorporating the whole of the protected set of data into his article as an appendix, he must secure the consent of the original author. This is because verbatim inclusion of the protected data counts as publication and making the data available to the public; it also involves making copies (duplication).

When using protected data, it is always necessary to determine whether the proposed action constitutes *duplication* and/or *publication*.

Important: even if the author himself has made the protected data available to others, for example on the Internet, the work can still not simply be used by third parties (i.e. published or duplicated) in the manner described above. Doing so requires the consent of the rightholder.

#### **2.4.3 Exception: works produced by the authorities**

No consent is required, however, to use works produced by the authorities. Acts of parliament, regulations, and court rulings are never subject to copyright.<sup>10</sup> Other types of works published by the authorities are also basically not subject to copyright unless that it is explicitly stipulated in the

---

<sup>9</sup> Sections 1 and 13 of the Dutch Copyright Act [*Auteurswet*].

<sup>10</sup> Section 11 Copyright Act. This provision may also apply generally by law.

work concerned.<sup>11</sup> This applies, for example, to reports (including research reports), policy documents, data etc. published by a government body. This rule does not require the authorities to be the author of the publication: a government body may well publish a report or something of the kind that has been produced by a private research firm. Such works are also not protected by copyright unless that is explicitly stipulated.

Research data, for example statistics, published by the authorities can therefore basically be reused without consent needing to be sought, as long as the work does not include a stipulation to the contrary (for example "All rights reserved"). It is not always clear whether research data is in fact derived from the authorities; what bodies are actually concerned? Local and national authorities and ministries in any case. Publications by other government bodies are also subject to these rules. In cases of doubt, the best thing is to check whether the work includes a copyright statement (i.e. "All rights reserved", the © sign, etc.); otherwise, one should ask the body that published the work.

Statistics Netherlands [CBS] is an example of a government body; it is therefore covered by this system of rules. Data from the Statline database can therefore be freely used by others (even if the database itself is subject to copyright). According to the information given on its website, Statistics Netherlands does not reserve copyright but it does require that it should be designated as the source when using its data.<sup>12</sup> The Statline database is probably also subject to database right protection, meaning that that is also relevant (see Section 3).

#### **2.4.4 Exception: citations**

In a scientific/scholarly publication, one does not need consent to *cite* from a work.<sup>13</sup> This also applies to protected research data. A number of rules do need to be borne in mind, however; the most relevant of these are:

- As far as possible, the name of the author and the source of the citation must be given.
- The size of the passage cited must be in a reasonable relationship to the purpose. To take the example given in Section 2.4.2: one cannot simply "cite" a whole set of protected data as an appendix; nor can one cite a whole work.<sup>14</sup>
- Data which one cites from must have been published lawfully. This means that the author himself must have decided that the data could be made available to the public. The data is likely to have been published lawfully if it has been uploaded to the Internet by the author (or with the author's consent). That may be different if the data was only shared within a highly restricted network. If the data has been made known only personally to another person, there is often no question of lawful publication (for example if the author has sent the data on request). The data may therefore not be cited.
- The citation must also be justified in the context in which it is made. This means that there must be a substantive connection between what is cited and the context, for example the accompanying text.<sup>15</sup> There will be such a connection if, for example, the protected data that is cited is discussed in the text.

#### **2.4.5 Exception: private copy**

One is always permitted to make a duplicate of a copyright-protected work for one's own use or study if this is done without any commercial intention. This may, for example, involve making a physical copy or copying out or downloading the work. One cannot provide such copies to others,

---

<sup>11</sup> Section 15b Copyright Act.

<sup>12</sup> See [www.cbs.nl/nl-NL/menu/organisatie/website/copyright/default.htm](http://www.cbs.nl/nl-NL/menu/organisatie/website/copyright/default.htm).

<sup>13</sup> Section 15a Copyright Act.

<sup>14</sup> An exception applies, for example, to photographs, which may be cited in their entirety even if the article is not about the photographer but about the subject of the photograph, if all the requirements have been met. See Spoor, Verkade, and Visser 2005, p. 240.

<sup>15</sup> Spoor, Verkade, and Visser 2005, p. 239–240.

however, or publish the complete work in some other manner, whether or not as an appendix or as part of one's own publication. In the context of research, copyright-protected works can therefore be used by downloading them and placing them in one's own database, for example, but they may not be published in any way.

## **2.5 Who is the rightholder?**

If certain research data seems to qualify for copyright protection, it is important to determine who holds the right in respect of the work. That is, after all, the person who has the exclusive right of publication and duplication and whose consent is therefore required for use by third parties. So far, we have referred to the author as the rightholder. That is in fact usually the case: as a rule it is the author – i.e. the person who thought up the work – who is the rightholder.<sup>16</sup> The copyright may, however, belong to someone else.

### **2.5.1 Shared copyright or involvement of a superior**

If a copyright-protected work has been created by more than one person, it may not be clear who counts as the author and therefore as the rightholder. With research data, that situation may well arise, for example if the data has been assembled by a group of people. If all those people have had an equal, creative influence on the production of the work such that it is no longer possible to distinguish their individual contributions, then they all count as being authors. They consequently share the copyright and must decide jointly on the exercise of that right. If, however, a group is concerned that produces the work under the direction of a person who thought up the work and who instructs them accordingly, then that superior holds the copyright.<sup>17</sup>

When a research group processes data in such a way, for example, that a copyright-protected work is produced, it will be necessary to consider which situation is most applicable. If the study was set up by a superior who instructed a number of co-workers about how the data should be collected and processed, then that superior holds the copyright. One cannot then say that the co-workers themselves had a creative influence on the work. If, however, one is dealing with a group of equal researchers who came up with the concept jointly and implemented it jointly, then the copyright is likely to be vested in all of them.

### **2.5.2 Cases in which the law allocates copyright**

There are also a number of situations in which the law provides that the copyright is not vested in the actual author of a work. Two of those situations are relevant here. Firstly, the copyright may be vested in an employer if the author of the work concerned was working for that employer on the basis of a contract of employment.<sup>18</sup> The work must then have been created as part of the employee's employment. A researcher employed by a research firm will not therefore hold the rights to the research data that he collects and processes in the context of his work. The same also applies at universities, although there has been a lot of discussion of this matter.<sup>19</sup> Secondly, a legal entity – which includes business companies, foundations, and universities – may be deemed to be the rightholder. That is the case if a work has been published with only the legal entity being referred to, without any notification being given of the actual author (i.e. a natural person).<sup>20</sup>

### **2.5.3 Cases in which contractual provisions may be relevant**

---

<sup>16</sup> Section 1 Copyright Act; pursuant to Section 4 Copyright Act, one can basically base oneself on the statement regarding authorship shown on the work.

<sup>17</sup> Section 6 Copyright Act.

<sup>18</sup> Section 7 Copyright Act.

<sup>19</sup> See Spoor, Verkade, and Visser 2005, p. 45–46.

<sup>20</sup> Section 8 Copyright Act.

Finally, there are a number of situations in which contractual provisions may be relevant when deciding who is the rightholder. A client who commissions a piece of research, for example, may stipulate in the contract that the copyright will be transferred to it. In that case, the client becomes the rightholder and it is the client's consent that will be required for the work to be used by other parties. The contract may also stipulate that the researchers will retain the copyright but that they already give consent for the client to use the work (this is then referred to as a licence). The arrangement may also mean that the client is the sole party that is permitted to use the work (this is then an exclusive licence). In that case, the rightholder is not permitted to grant consent for other parties to use the work (at least not without the consent of the client).

For the sake of clarity, it should be noted that the author of research data that is not subject to copyright protection is not obliged to make his data available to third parties. He is basically at liberty not to share the data with third parties and thus prevent them using it.

An author may in fact grant consent in advance for certain types of use by any potential user. The author may attach a statement to that effect to the work. One popular way of doing this is by means of a Creative Commons licence (see Section 5).

## 2.6 Reuse of protected data: sometimes there is double copyright

Before someone can copy a whole set of copyright-protected data or include it in slightly altered form in his own collection of research data, he needs to secure the consent of the rightholder. The question may then arise as to the status of the new collection, which now contains the initial researcher's protected data. Who is the rightholder as regards the new collection?

That situation here is rather complicated because the author of the original data retains his rights even if his data is now contained in a new and larger collection of data. If the portion of the collection that has been included is utilised in a manner that requires consent to be given (see Section 2.4), the user will need to secure the consent of the original author. The same applies if the user is the author of the new collection of data.

### *Example*

Researcher A puts together a collection of data ( $c$ ). In it, he includes a table of data ( $t$ ) produced by researcher B. Table  $t$  is protected by copyright, so A asks B for consent to include it. A then wishes to upload collection  $c$  to the Internet. In order to do so, he will need to secure the consent of B because collection  $c$  now includes the copyright-protected table  $t$ . Uploading  $c$  to the Internet in fact also involves uploading  $t$ , and that requires the consent of  $t$ 's author.

If researcher C then downloads collection  $c$  from the Internet and wishes to take over table  $t$  from it and include that table in a publication, he will need to secure the consent of researcher B because B is the rightholder in respect of table  $t$ .

The new collection itself may also be subject to copyright protection. That right is then vested in the author of the new collection. The whole collection is then subject to two copyrights: one vested in the new author of the whole collection and one vested in the initial author in respect of that part of the data that has been copied. If someone wishes to use the new collection in a manner that requires consent to be given, he will need to secure the consent of both the author of the new collection and the author of the copied data.

If researcher C therefore wishes to copy and publish collection  $x$  in its entirety, he will need to secure the consent of both A and B. After all, the collection as a whole has been created by A, meaning that A is the rightholder, but collection  $x$  now also includes table  $t$ , in which B holds the copyright.

## 2.7 Conclusion: the importance of copyright

Research data may be copyright-protected if it has been processed and the author had freedom of choice which he exercised in a manner of his own. As a rule of thumb, one can ask whether it is conceivable that two different authors, working separately, could have arrived at precisely the same form. If that is in fact conceivable, then the work concerned will not usually enjoy copyright protection.

Even though data may be copyright-protected, the following actions can still generally be carried out without consent:

- incorporation of the factual data in one's own words and in a structure of one's own;
- making a copy (including a digital copy) and utilising that copy for one's own research, as long as the original data is not made available to others;
- citing from the research data;
- works produced by the authorities are not protected by copyright unless that is explicitly stipulated.

Consent is therefore necessary for other types of use of copyright-protected data (or data collections). This may involve such things as:

- inclusion of the research data in a publication;
- sharing the research data with other people;
- including the whole of the research data in a database of one's own that is also shared with other people.

In such cases, one must secure the consent of the author (or under certain circumstances of another party). The form in which consent can be arranged is basically unrestricted, but it is a good idea to put it in writing. Consent can also be arranged by e-mail. When arranging consent, one must agree on exactly what use of the work is permitted.



## 3 Database right

A collection of data may also be subject to protection by database right.<sup>21</sup> Database right is an entirely different matter to copyright. The question of an “original character of its own” and the “personal stamp of the author” are not relevant here. Quite the contrary: where the protection of databases is concerned, it is not the creativity or originality of the database or its contents that counts but protection of the investment made in order to assemble the collection of data. It follows that database right is specifically intended to ensure that someone who has invested in the database can recoup his investment by exploiting the database; the database right protects the investor against third parties that wish to take over large portions of the database or to use it without consent or without paying. Database right applies just as much to non-commercial databases as long as the requirements explained below have been complied with.

Database right may therefore be relevant to collections of research data. If a collection of research data is deemed to be a protected database, then third parties must request consent before they can *retrieve* or *reuse* substantial portions of the database. Consent is also required for systematically retrieving non-substantial portions of the database (see Section 3.2).

### 3.1 When is research data protected by database right?

The law assigns a different meaning to the term “database” than in normal parlance: not every collection of data is a database in the legal sense. The legal definition of a database specifies when database right applies. It will therefore be necessary to determine whether the collection of data concerned is in fact a database within the meaning of the law.

Dutch law defines a database as follows:

*A collection of works, data, or other independent items arranged in a systematic or methodical way and individually accessible by electronic or other means and for which the obtaining, verification or presentation of the contents, evaluated qualitatively or quantitatively, bears witness to a substantial investment.*<sup>22</sup>

This definition comprises three essential elements:

- the database must consist of independent items;
- the database must be searchable or systematically arranged so that the individual items can be traced;
- there must have been a substantial investment in the database.

From the legal perspective, a collection of data is only considered to be a database if it meets these three requirements. A searchable collection of data is therefore only a database in the legal sense if there has been a substantial investment.

#### 3.1.1 Collection

A database in the legal sense must in the first place be a collection containing works, data, or other independent items. That definition is a broad one, and it certainly includes research data (both bare facts and copyright-protected data). This may involve data such as figures, codes, words, or descriptions, but also photographs etc. The materials in the collection do, however, need to be independent; in other words, it must be possible to separate them without their losing their substantive value or meaning.

---

<sup>21</sup> Database right is provided for in the Dutch Databases (Legal Protection) Act [*Databankenwet*]. This act is based on EU legislation, namely Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases.

<sup>22</sup> Section 1(1) Databases (Legal Protection) Act.

A collection of measurements, for example, will often consist of independent items, each having its own substantive meaning which is not lost if that particular item of data is viewed separately from other data in the collection, because the value of that item of data remains unchanged or just as significant. This may be different, for example, in the case of a set of cohesive data (for example a series of figures) that only have a substantive meaning when viewed as a whole and whose components cannot be viewed separately without their significance changing. In that case, there is no question of a protected database.

The law does not impose any requirements regarding the size of the collection. This means that a small collection can also be a database as long as it complies with all the components of the definition.

A database does not have to be in electronic form. It does, however, have to be recorded on a data medium, but this can be a paper medium as well as a digital one. The database may be distributed between a number of different media; in other words, it does not need to consist of only a single computer file or a single book.

### 3.1.2 Searchability

A database must also comprise a method or system that makes it possible to access each of the items in the database individually. Such a system can consist of an arrangement or indexation that allows each item to be traced. If the database is not arranged in that way, then a search programme can ensure that each item can be traced. In the case of research data, however, it is more likely that each item of data can be traced because the data is arranged in the form of tables or matrices.

### 3.1.3 Substantial investment

Finally, a substantial investment must have been involved in obtaining, verifying, or presenting the data in the database if it is to qualify for database right protection. Not every investment qualifies for protection: it must be a substantial investment. Consideration needs to be given here to the time, effort, and money that went into compiling the database.

The law does not give any detailed indications of when an investment should be considered to be *substantial*. The investment concerned must in any case be both qualitatively and quantitatively substantial. It involves investment in:

- *obtaining* the contents of the database;
- *presenting* the contents;
- *verifying* the contents.

Obtaining means collecting the contents of the database. Verifying means checking the correctness of the data; it can take place either as the database is being assembled or at a later stage when it already exists. Presentation means setting up the database, organising it, etc.<sup>23</sup>

A database comprising research data does not necessarily involve a substantial investment within the meaning of database law. Although the data will often have been collected in the course of lengthy and expensive research, protection under database law requires a substantial investment in the database itself. From the legal point of view, a distinction needs to be made between *creating* the data on the one hand and *obtaining, presenting, or verifying* it on the other.<sup>24</sup> Creating means producing or discovering the data that goes into the database. It is important to note that creating is not the same from the legal point of view as obtaining; the latter refers mainly to collecting existing data and not to "discovering" new data.

---

<sup>23</sup> C. Gielen (ed.), *Kort Begrip van het Intellectuele Eigendomsrecht*, Kluwer 2007, p. 550–553 (referred to below as Gielen 2007).

<sup>24</sup> Gielen 2007, p. 552.

Investment in the creation of data does not count when deciding whether there has been a substantial investment in the database.<sup>25</sup> That fact is important as regards databases comprising research data. With such databases, the main investment will have been in the research itself and thus in “creating” the data; that investment does not count, however, when deciding whether there has been a substantial investment in the database *as such*. The same applies to investment in verifying the data: that investment takes place, after all, in the context of research and not for the sake of the database.

Deciding whether there has been a substantial investment in the database is based solely on whether – regardless of the time, money, and effort that went into the research that produced the data – there has been a substantial investment in obtaining, presenting, and verifying the data. This means that a collection of research data will not always be deemed to be a protected database, although that is not impossible. If there has been a substantial investment in the database itself – in addition to the investment in the research – then the database will indeed qualify for legal protection. The investment may, for example, involve systematically or methodically arranging the research data for the database.

#### *Example*

Let us assume that a qualitative field study has led to the storage of the measurements taken in a large matrix, with a large number of variables being filled in for each case investigated. The variables themselves are simple, consisting only of a single word or a simple code. The structure of the database is also straightforward. There is therefore no question of any originality or a personal stamp, meaning that neither the database as a whole (because of its structure) nor the materials as such (the measurements) are protected by copyright. The database may, however, be deemed to be a protected database. There is, after all, a collection of separate data: each item has its own meaning. Moreover, the matrix has been structured systematically: the data has been arranged so that each item of data can be accessed separately. The crucial point, however, is whether there has been a substantial investment in obtaining, presenting, or verifying the contents of the database. That does not appear to be the case here. It is true that a lot of time, money, and effort has been invested in the research and in “creating” the data, but collecting the data in the database was merely a by-product of that research. There has been no investment in obtaining the data because the data had already been obtained in the course of the research. Nor has there been any investment in verifying the data; that took place in the course of the research. All that therefore remains is the presentation of the database. This may have involved a substantial investment if, for example, a lot of time, money, and effort went into formatting and programming the database; in this particular case, there is no question of that.

Sometimes, a database comprising research data may well have involved a substantial investment in obtaining (i.e. collecting) the contents. This may be the case, for example, if someone has gone to a great deal of trouble to compile a database comprising an overview of all existing research data. In such a case, a great deal of time, money, and effort may have gone into collecting and perhaps verifying the data, with that investment being solely in creating the database. The database does therefore qualify for protection and can therefore be covered by database right.

### **3.2 What actions require consent?**

If the database meets the criteria explained in the previous subsection, its producer enjoys an exclusive right. The producer is then the sole person who can grant consent for *retrieving* a substantial portion of the database, repeatedly systematically retrieving non-substantial portions, or *reusing* a substantial portion.<sup>26</sup> A person who wishes to perform these actions therefore requires consent, otherwise he will be infringing the right held by the producer of the database.

---

<sup>25</sup> European Court of Justice (ECJ), 9 November 2004, case C-203/02 (*British Horseracing Board/William Hill*).

<sup>26</sup> Section 2(1) Databases (Legal Protection) Act.

### 3.2.1 Retrieving material

Retrieving material involves appropriating or transferring some or all of the content of a database.<sup>27</sup> This may, for example, involve making a copy – either written or digital – or (temporarily) downloading material from the database. However, borrowing data from the database also counts as retrieval, for example when the data is copied out or typed into another database. That is also the case if the data (or a large part of it) is copied by hand after a critical selection has been made.<sup>28</sup> Taking data from a protected database for one's own research purposes, for example, also counts as retrieving material. It is irrelevant whether the research data is then used to create a new database. Consent is only required in two types of case:<sup>29</sup>

- a *substantial* portion of the database is retrieved (see Section 3.2.3);
- *non-substantial portions of the database are retrieved systematically and repeatedly.*

### 3.2.2 Reuse

The term “reuse” as applied in Dutch law<sup>30</sup> does not mean the same as in normal parlance. By reuse, the law means any action whereby the database is made available to the public. This therefore means providing *access* to the database. It may involve an intermediary who makes other parties' databases available to third parties via the Internet. This is also not permitted except with the consent of the producer of the database concerned.

Reuse will not always be relevant, of course, as regards the use of databases containing research data. In actual practice, consulting the data from a database containing research data will be the main action involved, something that is not covered by the legal use of the term “retrieval”. Reuse is relevant, however, if other people are given access to a database containing research data. This may be the case, for example, if that database is included in a new one. In such a case, one is dealing first of all with data being retrieved but if the new database is made available to other people, then reuse is also being made of the original database. In all cases, substantial portions of the database must be involved (see Section 3.2.3).

### 3.2.3 Substantial portion

The law specifies what it means by the term “substantial portion”. The portion concerned may be substantial both qualitatively and quantitatively. The qualitative aspect refers to the economic or technical value of the portion that is retrieved or reused. From the quantitative point of view, it is probably necessary to consider the size of the retrieved or reused portion in relation to the size of the database as a whole. It can sometimes be difficult to determine what is to be viewed as the whole database; a large database may, after all, consist of component databases. A component database is also deemed to exist if a small portion of the database also complies independently with the requirements explained above. In that case, the component database is also deemed to be an “independent unit”.

Consent is also required for repeatedly and systematically retrieving non-substantial portions. The thinking here is that constantly retrieving (in the sense of borrowing) small portions of the database can ultimately lead to reconstruction of the whole database or a substantial part of it; this is also harmful for the producer.

### 3.2.4 Derivation

If someone compiles a database that is wholly or to a substantial extent identical to an existing database, he does not automatically infringe the relevant database right. He only does so if the

---

<sup>27</sup> Section 1(1)(c) Databases (Legal Protection) Act.

<sup>28</sup> ECJ, 9 October 2008, case C-304/07 (*Directmedia/Albert-Ludwigs-Universität Freiburg*).

<sup>29</sup> Section 2(1)(b) Databases (Legal Protection) Act.

<sup>30</sup> Section 1(1)(d) Databases (Legal Protection) Act.

new database is entirely or to a large extent *derived* from the original database.<sup>31</sup> What this means is that the author of the new database has taken the data from the existing one; this requires consent. If, however, someone compiles a database by means of his own efforts that is wholly or to a large extent the same as an existing database, he is not infringing the relevant database right. He will, however, need to be able to demonstrate that the new database is not derived from the existing one. If someone compiles a new database by examining materials in an existing database one by one and making a selection from them for the new database, then there will be deemed to be a case of derivation.

### **3.2.5 Exceptions: scientific/scholarly research**

The law provides that retrieving substantial portions of a database for the purpose of scientific/scholarly research does not require consent.<sup>32</sup> This exception is restricted, however, to retrieval. This covers actions such as inspecting the database, downloading it, and making printouts. Copying the database into one's own database also constitutes retrieval.

One cannot, however, reuse substantial retrieved portions of a database in the framework of scientific/scholarly research without securing consent. Reuse basically means any action whereby the data is made available to persons other than the research team. The retrieved data cannot therefore be shared with persons other than one's own team of researchers. One is also not permitted to include the substantial portions of the database that one has retrieved in a publication. If such portions are placed in a researcher's own database, he must not share that database with others because it will then contain a substantial portion of the original database, and that portion may not be shared. All of these actions do require consent.

### **3.2.6 Exception: government databases**

Government databases have a special position. Databases comprising legislation, regulations, and court rulings are not subject to database right.<sup>33</sup> The same applies to other types of databases produced by the authorities, unless there is an explicit provision to the contrary.<sup>34</sup> In general, therefore, no consent is required to reuse or retrieve material from a government database. It is necessary, however, to determine whether the database right in the particular database concerned has been reserved. A government database is one that has been produced by a government body (see also Section 2.4.3). Important: it is also possible that a different producer (i.e. a commercial organisation) compiles a database containing solely government material; in that case, this exception does not apply.

### **3.2.7 Consent, consultation, and contractual arrangements**

Where the consent requirement is concerned, it is irrelevant whether the database is already available (including generally available) to third parties. Even if the database can be consulted by other people, consent is in principle still required to retrieve (i.e. borrow) substantial portions of the database or reuse them (i.e. make them accessible). It is also irrelevant whether such retrieval or reuse serves a commercial purpose: the producer's consent must also be acquired for non-commercial use (but see Section 3.2.4 regarding the exception that applies to scientific/scholarly research).

Consent is not necessary to inspect a database, nor for borrowing small quantities of information from it. It is important to remember, however, that this rule does not prevent the producer of the database making different arrangements. The producer of the database may, for example, stipulate in the relevant contract that access to the database depends on a payment being made. However,

---

<sup>31</sup> Spoor, Verkade, and Visser 2005, p. 627.

<sup>32</sup> Section 5(b) Databases (Legal Protection) Act.

<sup>33</sup> Section 8(1) Databases (Legal Protection) Act.

<sup>34</sup> Section 8(2) Databases (Legal Protection) Act.

someone who has secured legitimate access to the database can always retrieve and reuse non-substantial portions of it without the consent of the producer; that right cannot be excluded contractually.<sup>35</sup> In this regard, one should consult the user agreement concluded with the producer of the database. If no other arrangements have been made, then basically the database right regime that we have discussed will apply.

Finally, it should be noted that some commercially available databases are programmed in such a way that certain actions are technically impossible. It may not be possible, for example, to print out or download the database. The law provides that such “technical protection measures” must not be circumvented, even if the regime provides that making copies or downloads is in fact permitted.

### **3.2.8 Borrowing from another database containing research data**

Where assembling research data for scientific/scholarly research is concerned, it is important to know whether one can borrow data from a protected database containing research data (or borrow the whole database) without securing the consent of the producer of that database. As we have already seen, bare facts are free; the bare facts contained in a database can therefore be used for new research. However, one must always remember that substantial portions of the database may not be copied. If substantial portions are retrieved in the context of scientific/scholarly research, however, the exception dealt with in Section 3.2.4 will often apply.

Things are more problematical when it comes to borrowing and publishing an existing database (or a large part of it) in a new database in a new research publication. In that case, one is retrieving and reusing a substantial portion of the existing database. That is also the case if a large part of the data in the pre-existing database is copied out by hand after a critical selection has been made.<sup>36</sup> This also requires the consent of the producer of the database concerned. Important: this only applies to databases for which a substantial investment has been made in obtaining, arranging, and verifying the data, something that is by no means always the case with databases containing research information. In the case of government databases, borrowing data only requires consent if the database right has been explicitly reserved.

## **3.3 Who is the rightholder?**

We have already noted that the producer is deemed to be the rightholder in respect of the database; he is the party that can grant consent for other people to retrieve and reuse substantial portions or to repeatedly and systematically retrieve non-substantial portions. The relevant question here is who counts as the producer of the database; this is by no means always the actual author (i.e. the “maker” of the database). According to the law, the producer of the database is the party that bears the risk in respect of the investment in the database etc. This may be the actual author, but it may also be the party that instructed that the database be created. Who bears the risk in this relationship must be determined on the basis of the contractual arrangements that apply between them. It is advisable to reach agreement on this matter. In many cases, the rightholder may therefore be a legal entity.

## **3.4 Relationship to copyright**

In conclusion, we will give a brief explanation of the relationship between copyright and database right. These are two separate areas of the law which by no means exclude one another. If it is sufficiently original, a database itself can therefore also be a copyright-protected work (see Section 2.3.4). The various materials in a database can also themselves be a copyright-protected

---

<sup>35</sup> Section 3 Databases (Legal Protection) Act. See Spoor, Verkade, and Visser 2005, p. 631–632.

<sup>36</sup> ECJ, 9 October 2008, case C-304/07 (*Directmedia/Albert-Ludwigs-Universität Freiburg*).

work. In that case, borrowing those materials requires the consent of the copyright holder in accordance with the way the copyright has been arranged (see Section 2).

### **3.5 Summary**

A collection of separate items of data may constitute a protected database if those items are systematically arranged or are traceable in some other manner. It is also necessary for creation of the database to have involved a substantial investment; the investment made in researching and creating the data that is included in the database does not count towards that substantial investment.

If there is in fact a protected database, the producer's consent is required for the following actions:

- retrieving (i.e. copying or downloading) substantial portions of the database;
- repeatedly and systematically retrieving non-substantial portions of the database;
- reusing (i.e. publishing) substantial portions of the database.

The producer's consent is not necessary for:

- using a database for scientific/scholarly research if no substantial portions of the database are published (reused);
- using a government database (unless there is an explicit stipulation to the contrary).



## 4 Protection of non-original writings

If neither copyright nor database right applies, then “protection of non-original writings” [*onpersoonlijke geschriftenbescherming*] may be relevant. This involves a very “narrow” type of protection for works that do not meet the criteria for copyright protection. The requirements that a work have an original character of its own and bear the personal stamp of the author are irrelevant where this type of protection is concerned. Protection of non-original writings may also apply to collections of data for which no substantial investment has been made and that are consequently not subject to database right protection.<sup>37</sup>

Important: protection of non-original writings only applies if neither copyright (see Section 2) nor database right (see Section 3) applies.

### 4.1 When does protection of non-original writings apply?

Any “writing”, whether analogue or digital, can basically qualify for protection of non-original writings; no detailed criteria can be given.<sup>38</sup> Given the cases in which this type of protection has been assumed, it is quite conceivable that lists or collections of research data are also covered by this regime.

There is in fact one specific requirement for a non-original writing to qualify for protection, namely that it must have been published or be intended for publication. This means that it must have been provided for inspection by the public, or at least that must be the intention. This restricts the number of cases in which research data is protected as being non-original writing. It is not uncommon, after all, for unprocessed data to be intended only for a researcher’s own use, as the basis for a publication. When the raw data is distributed within a research community, for example on the Internet, then we are in fact dealing with a published piece of writing.

One example of data that qualifies for protection as non-original writing but not for copyright protection is metadata. Metadata is data that describes the characteristic features of certain information; it is insufficiently original to qualify for copyright protection. Important: metadata only qualifies for protection as non-original writing if it has been brought to the attention of the public or if the intention is that that should be done. One can argue that this is not the case when we are dealing with metadata that has been recorded solely in machine-readable form. That may be different if it is a simple matter to make the metadata visible, as in the case of the source code for web pages.

Moreover, it is only a *collection* of independent metadata that is systematically arranged and in which the obtaining, verifying, or presentation of the materials has involved a substantial investment that qualifies for database right protection and not the single set of metadata for a particular document.

### 4.2 What actions require consent?

Protection of non-original writings covers a broad scope in the sense that it applies to many different types of works, but the protection that it affords is much narrower than in the case of copyright or database right. Here too, the individual items of data that have been brought together in a non-original piece of writing are not each individually protected. Moreover, consent is only required for copying the piece of writing verbatim or with limited changes. That is therefore a significantly narrower type of protection than that afforded by copyright, because the latter also requires that a work must also not be *adapted* without the consent of the rightholder.

---

<sup>37</sup> For protection of non-original writings in general, see Spoor, Verkade, and Visser 2005, p. 81–92.

<sup>38</sup> Protection of non-original writings follows from Section 10(1)(1) of the Copyright Act.

### **4.3 Whose consent must be secured?**

The person that can grant consent is the same as in the case of copyright. In the first place, that means the author or authors of the material concerned, i.e. the person/persons who put down the data in writing. Depending on the circumstances, a superior, employer, or legal entity may also be the rightholder and thus the party that can grant consent. Contractual arrangements may also be relevant. (For all these matters, see Section 2.5.)

## 5 Internet

### 5.1 The Internet

The Internet plays an important role in research. Not only is it a source of other people's research data, it also allows researchers to share their own data. Copyright, database right, and protection of non-original writings apply just as much to the Internet as they do otherwise. The fact that data is in digital form is irrelevant: copyright, database right, and the protection of non-original writings apply to digital products in just the same way as they do to analogue products. It is also irrelevant whether the data originates in the Netherlands or has been produced elsewhere. Within the Netherlands and for websites directed to the Netherlands, it is Dutch law as described above that applies. If someone uploads their research data to the Internet, they need to remember that different rules may apply in other countries. Even if Dutch law provides that the data is protected by one of the regimes, different rules apply in other countries and these may mean that the data is not protected there, or that in fact fewer actions are permissible without the author's consent.

In some cases, the author of a protected work indicates in advance that he already grants consent for certain actions. He may do this by attaching a statement to that effect to the work, indicating that he already issues a licence to any potential user to use the work in a certain way. The *Creative Commons* licences are one popular standard. Authors can choose from a number of different licences that comprise particular conditions. Other people can then see at a glance what use they are permitted to make of the work and no longer need to contact the author to negotiate about consent.

Creative Commons licences allow authors to make all kinds of works available – texts, photos, music, films, etc. – on licence conditions that they determine independently to suit their particular situation. The licences are available in a number of different versions and allow copyright-protected works to be made available on various different conditions: attribution required and/or no commercial use and/or no adaptations permitted ("share-a-like") or adaptations permitted but with further distribution subject to the same conditions. In addition, the author can require payment to be made for commercial use of his work. The licence conditions can be easily drawn up using the licence tool provided by the Creative Commons organisation; this was adapted last year to bring it in line with Dutch law. A wide range of different licences are available on the websites [www.creativecommons.org](http://www.creativecommons.org) and [www.creativecommons.nl](http://www.creativecommons.nl).<sup>39</sup>

SURF has expressed an emphatic preference for the most liberal of the Creative Commons licences, namely "Creative Commons Attribution 3.0 Netherlands Licence". Using this licence creates the fewest possible obstacles to the future use of repositories. The Creative Commons Attribution 3.0 Netherlands Licence allows users to copy a work, distribute it and pass it on, and produce derivative works and distribute them on condition that the work is attributed to its author.<sup>40</sup>

### 5.2 Contractual arrangements

When sharing data with others, the author can make arrangements about what use they can make of it. He may stipulate, for example, that they can only inspect it but not use it in any other way, or that they must not publish it. Such arrangements can also be made in respect of research data that is not protected by copyright or database right or the protection afforded to non-original writings.

---

<sup>39</sup> M. de Cock Buning, "Creative Commons-licenties in het auteursrecht: common sense?", in: J.J. Brinkhof, M. de Cock Buning and E.H. Hondius (eds.), *Contracteren internationaal*, Boom Juridische uitgevers, 2006.

<sup>40</sup> [SURFdirect, P. Keller and W. Mossink \*Hergebruik van materiaal in onderwijs- en onderzoekomgevingen\*, Utrecht 2009.](#)

We also need to refer to the exchange of research data via a third party that acts as an intermediary. The Netherlands, for example, now has “DANS” (Data Archiving and Networked Services). This organisation has a website where researchers can upload data so that other researchers can then download it. Certain arrangements can be made when supplying research data to the intermediary, for example that the intermediary may make the data available to third parties regardless of whether it is subject to copyright or database right. It can also be stipulated that the data must not be made available to third parties, or only on certain conditions. These, however, are arrangements that apply between the supplier of the data and the intermediary; basically, they do not apply to third parties.

Anyone who receives data from such an intermediary must observe the rules applying to copyright or database right. The fact that the data is made available via an intermediary does not therefore mean that the author is also waiving the copyright or database right that may apply to the data. If consent is required for the data to be used, it will be necessary to contact the author. This is only different if the author has already given consent in advance for certain actions, for example by including a Creative Commons licence.

### 5.3 Schematic overview

This subsection gives a schematic overview of the necessary procedure; it is intended to clarify matters for researchers who wish to determine the status of other people’s research data. This overview therefore assumes that research data is to be reused. It is important to note that this overview cannot replace the full legal guide but is meant only as a practical step-by-step plan for finding one’s way around this guide. This overview provides the basis for the brief guide in Section 1.5.

#### 1. Determine whether consent is required for use of the data

Do you wish to make a copy of the data for your own use or do you wish to input data into your own scientific/scholarly database, without sharing it with anyone other than your own team of researchers?

In that case, you probably do not need to request consent; even if the data is protected by copyright or database right, this type of use is one of the permitted exceptions: making a copy for your own use (Section 2.4.5) of a copyright-protected work or retrieving substantial portions of a database in the context of scientific/scholarly research (Section 3.2.5; retrieving non-substantial portions does not require consent in any case: Section 3.2.3).

Does the intended use go further than this, for example because you want to publish some of the data that you are borrowing or share it with someone other than your own team of researchers?

In these cases, consent may well be required; you therefore need to find answers to the following questions:

- Does **copyright** apply?  
That is the case if the research data has an original character of its own and bears the personal stamp of the author. The decisive point is therefore whether the author had freedom of choice in arranging the data in a particular form and whether he did this – deliberately or otherwise – on the basis of personal preference (Section 2.3.2). Even just a selection consisting solely of bare facts may be subject to copyright protection (Section 2.3.3).

As a rule of thumb, one can ask whether it is conceivable that two different authors, working separately, could have arrived at precisely the same form (Section 2.3.2).

If copyright applies, you should also answer question 2 and check whether database right also applies (see below). See also Section 2.3.5 regarding copyright-protected databases.

- Does **database right** apply? There are a number of requirements for database right to apply:
  - Are you dealing with a collection of independent materials (research data) (Section 3.1.1)?
  - Is the data collection arranged systematically (Section 3.1.2)?
  - Has there been a substantial investment in obtaining, verifying, or presenting the materials (Section 3.1.3)? The investment in carrying out the research that generates the data does not count.

If database right applies, go on to question 3.

## 2. If copyright applies

Copyright requires that consent be given if you intend making the protected data publicly available in some way, for example by publishing it or providing it to other people in some other way (such as sharing it with other researchers). In cases such as this, the law speaks of publication (Section 2.4.2). Other people's data can only be copied or utilised for the purpose of your own research without the author's consent if it is not published (Section 2.4.2).

Copyright probably applies and the proposed use requires the author's consent (Section 4.3).

You should, however, check whether an exception applies:

- the right to cite (Section 2.4.4);
- works produced by government bodies (Section 2.4.3);
- the all-inclusive form (Section 2.4 and 2.1.2);
- a selection of bare facts (Section 2.3.3);
- other kinds of data (for example photos) (Section 2.3.4).

Check whether database right applies (question 3); if that is not the case, you probably do not need consent to borrow bare facts in this way if they do not constitute a protected selection (Section 2.3.2).

It is only the bare facts that can be taken over, in your own words and with a structure that you have created yourself (Section 2.4.1).

## 3. If database right applies

Database right requires that consent be given if you intend not only borrowing data from the database for your own research but also publishing it further, for example in a publication, or by sharing your own material (which will then include the data you have taken over) with other people. In cases such as this, the law speaks of reuse.

- Do you intend reusing a substantial portion of the database?

There are no clear criteria for what constitutes a "substantial" portion. It is necessary to consider the relationship between the reused portion and the database as a whole, as well as the technical and economic value of the reused portion (Section 3.2.3; see also Section 3.2.8).

- Check whether an exception applies:

Unless provided otherwise, government databases are basically not subject to database right (Section 3.2.6). You do not require the consent of the rightholder in order to retrieve substantial portions of a database in the framework of scientific/scholarly research (Section 3.2.5).

- Does copyright apply to the various items making up the database? (See question 2).

When borrowing those items, you need to observe the copyright; you should therefore check whether consent is required (question 2).

#### **4. Protection of non-original writings**

You are not permitted to borrow in full or publish in full a text that has been published (or is intended for publication) but that is not protected by copyright or database right. You are, however, permitted to borrow and publish it in slightly altered form (see Section 4).

**Part II – Privacy**



## 6 Privacy

### 6.1 Personal data

Privacy problems may arise if the research data concerned contains information about living persons. Briefly, the right to privacy of information means that every person must have the maximum influence on what people know about him or her. This is important because personal data says something about who someone is. If it is known that someone drives a big car and frequently eats at expensive restaurants, then the image of that person will be different to if he lives in a deprived area and receives supplementary benefit. Personal data must therefore be dealt with carefully because it says something about the person to whom it relates.

The rules discussed below do not apply to deceased persons or to legal entities. However, it is important to note that information about a deceased person may also comprise information about living persons (for example a surviving relative) and that information about a legal entity may well relate to a single identifiable individual (for example in the case of a one-person business); in that case, the rules regarding processing personal data continue to apply. We will refer to the person to whom certain information relates as “the person concerned”.

#### 6.1.1 Anonymized information

Information regarding living persons may be anonymized, i.e. it would take a disproportionate effort to trace the person to whom the information refers. Anonymized data is not covered by the Personal Data Protection Act [*Wet bescherming persoonsgegevens, Wbp*] and can be used or shared with other people without concerns about privacy. It is therefore a good idea to work with anonymized data, if possible. It is important to note the following, however. Combining anonymized data with other data – which may also be anonymized – may in fact produce data that can be used to trace the identity of the person concerned. This means that anonymized data can become indirectly identifying data. One familiar example is an uncommon occupation which when combined with someone’s age or the place where he lives can make clear who the person concerned is.

#### 6.1.2 Indirectly identifying data

Indirectly identifying data is data that cannot be traced to someone directly but that still makes it possible – without a disproportionate amount of time and effort – to determine the identity of the person concerned (as in the case of the combination of an uncommon occupation and the place where someone lives). The criterion of “without disproportionate time and effort” must be interpreted broadly: if it is possible to connect data with an individual by means of the tools available to a researcher, then the data concerned counts as indirectly identifying data. Practically speaking, the data may still appear to be “anonymized” for the researcher but processing the data must be done in accordance with the provisions of the Personal Data Protection Act.

#### 6.1.3 Directly identifying data

From the privacy perspective, the most sensitive data is directly identifying data, i.e. collections of data which immediately make clear the identity of the person to whom they relate. This may be because, for example, the person’s name and address are given, or a telephone number, bank account number, etc.

Both directly and indirectly identifying data count as personal data and are subject to the provisions of the Personal Data Protection Act.

#### 6.1.4 Special personal data

“Special personal data” requires extra protection because it counts as “sensitive”: data concerning someone’s religion or convictions about life, race, political affiliation, health, sexual orientation, membership of professional associations, criminal record, and details regarding illegal actions or misbehaviour in connection with a prohibition imposed in the light of that misbehaviour.<sup>41</sup> Financial information, for example regarding someone’s income or assets, is therefore not included. Processing of special personal data is subject to a stricter regime than “ordinary” personal data. Even when special personal data has been anonymized, special care is necessary because of the risk that it may later be traced to an individual.

## 6.2 Obtaining of personal data

The purpose of a scientific/scholarly study may require use to be made of data about living persons. As we have already seen, the data should if possible be anonymized. The purpose for which the data is necessary must in any case be clearly specified, and no more data may be collected than is necessary to achieve that purpose.<sup>42</sup>

The data required to carry out a study can be obtained in a number of different ways:

- a) by collecting the data from the persons concerned themselves;
- b) by collecting data about the persons concerned from third parties;
- c) by reusing a file that has already been created.

### Re (a)

When data is collected from the persons concerned themselves – for example by means of surveys or interviews (conducted either by the researcher himself or by someone else on his behalf) – it is important that the persons concerned give their consent.<sup>43</sup> That consent must be unambiguous. The person concerned must know what the purpose is of the research, who commissioned it, who is carrying it out, and where any additional information about the research can be found,<sup>44</sup> with the person concerned then granting consent for his details to be used for the research. Data collected with the individual’s consent in this way can also be used later, on certain conditions, for other research (see below). The consent of the legal representative is required for the collection of personal data from persons younger than 16, and from persons who are subject to a guardianship order or who have been made the object of a mentorship.<sup>45</sup>

With a view to being able to provide evidence of consent, it is a good idea to get the person concerned to sign a consent form.

---

<sup>41</sup> Section 16 Personal Data Protection Act.

<sup>42</sup> Section 7 Personal Data Protection Act.

<sup>43</sup> Section Art. 8(a) Personal Data Protection Act. Consent means any freely given and specific indication of intent, based on information, with which the person concerned accepts that personal data concerning him will be processed (Section 1(i) Personal Data Protection Act). If special personal data is concerned, explicit consent is required (i.e. not tacit or implicit consent): the person concerned must then have expressed, in words, writing, or behaviour, his will to give consent for processing of the data concerning him.

<sup>44</sup> Section 33 Personal Data Protection Act.

<sup>45</sup> Section 5 Personal Data Protection Act.

## Consent

The undersigned, <name>, born on <date of birth> in <place of birth>, hereby declares that he/she is aware of the purpose of <name of research study>, the commissioning party, and the party carrying out <name of research study>, and of where additional information about <name of research study> is available.

On the basis of this information, the undersigned gives consent for the use of <his/her> personal data in <name of research study>.

<place>, <date>,

<signature>

If data is to be collected without the consent or knowledge of the persons concerned, for example by means of observation, a request must be submitted beforehand to the Dutch Data Protection Authority [*College Bescherming Persoonsgegevens, CBP*] so that it can carry out an investigation. Personal data may only be collected in this way if the CBP notifies the researcher that it does not consider an investigation to be necessary or if it decides that the proposed study is legally permissible.

It is perhaps unnecessary to point out that data may only be used for scientific/scholarly research if it is obtained in a legal manner.

### Re (b)

Collecting data about someone from someone else (for example “tell me what you know about...”) is only permissible with the consent of the person concerned or, in the case of a longer study during which information is regularly collected by phone, if there are no indications that the person concerned wishes to withdraw from the study.

### Re (c)

One important legal provision as regards science and scholarship is Section 9(3) of the Personal Data Protection Act. The basic rule of the Act is that data may not be processed further in a manner incompatible with the purposes for which it has been obtained; Section 9(3) provides, however, that processing data for historical, statistical, or scientific/scholarly purposes is not deemed to be incompatible. This means that reusing a file for scientific/scholarly research that has already been created is permitted, even if the file was created for a different purpose. As always, there are exceptions, namely for special data (see above) or if the data concerned is subject to professional confidentiality, as is the case with medical information. In such cases, the relevant rules of professional confidentiality must be observed.<sup>46</sup>

Reuse of special details requires the explicit consent of the person concerned or, if securing his explicit consent would be impossible or involve disproportionate effort, if all three of the following conditions are met:

- a) the research must serve the public interest;
- b) it must not be possible to carry out the research without these details and sufficient guarantees are provided during the research to ensure that no undue harm will be done to the privacy of the person concerned;
- c) consultation takes place regarding the research with the person who created the file.

---

<sup>46</sup> See in particular the *Code of Conduct for Health Research, Code of Conduct for the Dutch Biomedical Research Community* [*Gedragcode Gezondheidsonderzoek, Gedragcode van de Nederlandse biomedische onderzoeksgemeenschap*], approved by the Data Protection Authority in 2004 (available online at <http://www.giantt.nl/gedragcode%20gezondheidsonderzoek.pdf>).

Such reuse is again deemed to be the processing of personal data, even if the data is derived from public or publicly accessible sources (for example the Internet). The provisions of the Personal Data Protection Act therefore still apply.

### **6.3 Storage of personal data**

As we have already seen, it is preferable for the data collected to be anonymized. Indirectly identifying data is second best, and collecting directly identifying data for research purposes should be avoided as far as possible. Sometimes, however, it may be necessary after analysing certain data to contact one of the persons concerned to request additional data. It is then advisable, if and as far as possible, to separate the contact details and the research data from one another and to store them in separate files. The two types of data can then be linked by means of an “empty” record number, for example a serial number.

If it is absolutely necessary to work with directly identifying data, processing of personal data in this way must be reported to the CBP (or to the data protection officer if the institution where the researcher works has appointed one). Notification to the CBP must be made by the “party responsible”; in the case of a university, this will be the Executive Board.

Notification is unnecessary if a file containing directly identifying data is separated out into research data and contact details and the contact details are also not kept for more than six months after they have been obtained from the person concerned. It is important to keep track of this six-month period and to notify the CBP of processing if it is necessary to keep the contact details for longer than six months. As the “party responsible”, the Executive Board of the institution where the researcher works must be aware of studies involving the use of directly identifying data and of what directly identifying personal data has been provided to third parties.

Files containing personal data must be properly secured, by means of both technical and organisational measures. This will involve, in the first place, a username and password to access the data. Depending on the nature of the data, the system of access control may be more stringent, fewer people will have access to the data, and the data may perhaps be encrypted.

Anyone involved in processing personal data must sign a confidentiality statement in respect of personal data;<sup>47</sup> this includes temporary employment agency workers (“temps”), student assistants, etc. who help process the data.

Normally speaking, the person concerned has a number of rights, for example the right to be informed that his details are being processed and the right to submit a request to inspect his details.<sup>48</sup> If, however, the data is processed by an institution or service for scientific/scholarly research or statistics, and if the necessary arrangements have been made to ensure that the personal data can only be used for statistical or scientific/scholarly purposes, then these rights do not apply.

### **6.4 Use of personal data**

The personal data that has been collected is then used for the purpose for which it was obtained, namely scientific/scholarly research. It is important that the results of the research are published in such a way that it is absolutely impossible to trace them to the persons concerned, unless those persons give consent for that to be done. That consent may only be requested if it is not possible

---

<sup>47</sup> Important: a confidentiality statement is not the same thing as professional confidentiality. Professional confidentiality is required by law, for example in the relationship between a doctor and his patient or a lawyer and his client. A confidentiality statement is a promise not to pass on personal data which one is made aware of.

<sup>48</sup> Sections 34 and 35 Personal Data Protection Act.

for the research results to be published without the risk of their being traced to the persons concerned. Only the minimum of information may be published, i.e. no superfluous details that do not contribute to the publication but that do facilitate identification.

## 6.5 Sharing of personal data

Making data available to fellow researchers – not only the research results but also the raw data on which those results are based – is permitted under certain conditions, even if personal data within the meaning of the Personal Data Protection Act is concerned. The most important condition here is that the fellow researcher who receives the data is bound by the same rules for careful use, as described in the various codes of conduct.<sup>49</sup> It is the responsibility of the researcher who provides the data to record this clearly.

Provision of data to fellow researchers outside the European Union is more problematical: it is only permitted if the country concerned guarantees an appropriate level of protection.<sup>50</sup> The United States, in particular, has no legislation regarding the protection of personal data. The European Commission has stipulated that an appropriate level of protection can only be deemed to exist in the case of organisations that have undertaken to comply with the “Safe Harbor Principles”. These principles are intended to guarantee that organisations in the United States and European Union that process personal data do not act contrary to the guarantees set forth in the EU directive on the protection of personal data (Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data). Data may, however, be provided if the person concerned has granted unambiguous consent for this to be done.<sup>51</sup> It is therefore advisable when collecting personal data for scientific/scholarly purposes to immediately request consent for sharing the data with colleagues, including perhaps colleagues abroad.

As we have seen, it is the responsibility of the researcher who provides the data to ensure that a colleague who receives it is bound by the same rules for careful use; it is therefore advisable not to place files containing personal data in the publicly accessible part of a repository. What the researcher can perhaps do, however, is indicate that certain files are available and then only provide them to certain fellow researchers for certain studies on individual request and under specified conditions.

There is of course no objection to including publications (see above) and anonymized data in a repository.

## 6.6 Sanctions

The CBP sees to it that personal data is processed in accordance with the law. To that end, the CBP may carry out an investigation. The CBP can require that processing be terminated or only continue if it is in accordance with the law. In extreme cases – for example unwillingness or obstruction –

---

<sup>49</sup> Code of Conduct for the use of Personal Data in Scientific/Scholarly Research [*Gedragscode voor gebruik van persoonsgegevens in wetenschappelijk onderzoek*], Advisory Report by the Social Sciences Council, Royal Netherlands Academy of Arts and Sciences, Amsterdam, April 2003 (available online at <http://www.knaw.nl/publicaties/pdf/20031019.pdf>) and *Code of Conduct for Health Research, Code of Conduct for the Dutch Biomedical Research Community*, approved by the Data Protection Authority in 2004 (available online at <http://www.giantt.nl/gedragscode%20gezondheidsonderzoek.pdf>).

<sup>50</sup> Section 76 Personal Data Protection Act. Such countries include Argentina, Canada, and Switzerland.

<sup>51</sup> Section Art. 77(1)(a) Personal Data Protection Act.

the CBP can impose penalties. In addition, acting contrary to the Personal Data Protection Act is a criminal offence subject to the imposition of a penalty by the courts.<sup>52</sup>

---

<sup>52</sup> The right to prosecute lapses, however, if the CBP has already imposed a penalty (Section 75(3) Personal Data Protection Act).

**Part III – Liability**



## 7 Liability

This part of the guide deals with the legal position of repositories. A “repository” here means a storage location for research data. The question is whether – and if so when – a repository is liable if, for example, material that is sensitive from the point of view of privacy is published or dealt with in some other manner that is contrary to legislation and regulations. This part of the guide restricts itself to liability on the part of repositories for acts carried out by others (more specifically: researchers who deposit their research data in a repository).<sup>53</sup>

### 7.1 Legal classification

The service provided by a repository consists of:

- storing (on request) research data provided by researchers;
- making research data accessible and providing it to other researchers (on request);
- making metadata accessible, i.e. data about who has what research data so that another researcher who is interested can contact that person or body.

Because a repository is accessible electronically,<sup>54</sup> the service that it provides is deemed to be a “service of the information society”. Section 3:15(d)(3) of the Dutch Civil Code [*Burgerlijk Wetboek*] defines “service of information society” as:

*Any service that is provided, normally in return for payment, by electronic means, at a distance, and by individual request of the user of the service without parties being simultaneously present at the same location.*

The element “normally in return for payment” is assigned a broad significance, so broad, in fact, that it also covers free services, as long as an economic activity is involved.<sup>55</sup> The element “by electronic means” is explained in the same section as:

*A service is provided by electronic means if it is sent, transmitted, and received entirely by wire, by radio, or by means of optical or other electromagnetic means with the aid of electronic equipment for the processing – including digital compression – and storage of data.*

It is clear that the Internet, but also restricted-access intranets and “virtual private networks” operate “by electronic means”.

### 7.2 Information obligations

A party that provides a service of the information society is required by law to make certain information available – easily, directly, and permanently – to potential users. That information comprises:<sup>56</sup>

---

<sup>53</sup> If a repository itself acts wrongly, for example by infringing copyright, then it is naturally the repository that is liable.

<sup>54</sup> If a researcher, in order to access research data, is required to physically visit a repository and to take the data with him in physical form – either on paper, a USB stick, or in some other physical form – then the service is not provided electronically and the repository does not therefore provide a “service of the information society”; the liability regime that we are discussing here consequently does not apply.

<sup>55</sup> See recital 18 of the Directive on Electronic Commerce (2000/31/EC), available online at <http://eur-lex.europa.eu/Notice.do?val=236968:cs&lang=en&list=275398:cs,236968:cs,236200:cs,237247:cs.&pos=2&page=1&nbl=4&pgs=10&hwords=&checktexte=checkbox&visu=texte>.

<sup>56</sup> Section 3:15(d)(1) Dutch Civil Code.

- a) his/its identity and the address where he/it is located;
- b) details making possible rapid contact and direct and effective communication with him/it, including his/its e-mail address;<sup>57</sup>
- c) if registered in a trade register or comparable public register: the name of the register where he/it is registered, his/its registration number, or comparable means of identification in that register;
- d) if an activity is subject to a permits system: details of the competent supervisory authority;
- e) if he is a member of a regulated profession:
  - the professional association or organisation with which he is registered;
  - the professional title and the Member State of the European Union or other state which is party to the Agreement on the European Economic Area where that professional title has been awarded;
  - a reference to the professional rules that apply in the Netherlands and the means of access to them;
- f) insofar as he/it carries out an activity subject to VAT: the VAT identification number within the meaning of Section 2(a)(1)(g) of the 1968 Turnover Tax Act [*Wet op de Omzetbelasting 1968*].

It is not too complicated a matter for repositories to comply with these information obligations.

### 7.3 Liability

Being classed as a supplier of a service of the information society brings with it not only the above-mentioned information obligations but also a significant benefit. If the service provided consists of “storing, on request, information derived from another” – which is the case with repositories – then the service provider is not liable for the information stored if he/it complies with two requirements:

1. He/it must not know that the activity or information is of an unlawful nature, and can also not reasonably be expected to know.
2. As soon as he/it knows, or can reasonably be expected to know, that the information is of an unlawful nature, the service provider must immediately remove it or make access to it impossible.<sup>58</sup>

This provision was drawn up with a view to Internet service providers that offer hosting services. The purpose is to prevent hosting providers needing to screen material that they publish in advance. It basically means that a repository does not have any obligation to investigate material in advance but that as soon as it becomes clear that certain data involves an infringement or is unlawful in some other way then that data may no longer be made available.

How can a repository know whether information is of an unlawful nature? This may be, for example, because the repository has been informed by means of a “notice-and-take-down” request, for example from an interested party. Let us assume, for example, that someone sees that material has been made available in a repository for which he holds the copyright. He can then inform the repository to that effect. Needless to say, he will need to substantiate his claim; he will need to show that he does in fact hold the copyright and that it is unlawful for the repository to make that information available. The repository will only itself be liable if it fails to take action in response to that request.

It is therefore advisable to have a good “notice-and-take-down” procedure in place. This might involve, for example, having a Web form that makes it easy for someone to submit a complete

---

<sup>57</sup> The European Court of Justice recently ruled that under certain circumstances – namely if contact by e-mail or a Web form does not produce the desired result – a telephone number should be included also. European Court of Justice, 16 October 2008 C-298/07 (*Bundesverband der Verbraucherzentralen*).

<sup>58</sup> Pursuant to Section 6:196(c)(4) Dutch Civil Code.

notice-and-take-down request, and to provide organisational guarantees that such requests will be dealt with carefully.<sup>59</sup>

Something similar applies to information that constitutes a criminal offence (for example child pornography). Section 54(a) of the Dutch Criminal Code [*Wetboek van Strafrecht*] provides as follows:

*An intermediary that provides a telecommunications service consisting of passing on or storing data deriving from another shall as such not be prosecuted if he complies with an order by the public prosecutor, after written authorisation has been provided by the delegated judge at the request of the public prosecutor, to take all measures that can reasonably be required of him to make that data inaccessible.*

In other words, only if an order by the public prosecutor to remove information is ignored can the repository itself be criminally liable (unless the repository knew itself that the information concerned was subject to criminal prosecution).

## 7.4 General Terms and Conditions

Repositories receive the material that they make available from researchers. Proper arrangements need to be made with those researchers regarding what information will be included in the repository and made available to fellow researchers, and on what conditions. Those arrangements will normally be comprised in a set of "general terms and conditions", "conditions of use", or something similar.

It is important to stipulate in that set of rules that only data that does not constitute an infringement or that is not unlawful in some other way will be included in the repository. It should also be stipulated that the researcher concerned is at all times liable himself should something be amiss in this regard.<sup>60</sup>

In this way, the repository can always correctly assert that it has done everything possible to ensure that no infringing or otherwise unlawful information is made available by means of its services.

The general terms and conditions must also state that if any discussion arises as to the lawfulness of certain data, the complainant will be informed of the identity of the researcher that provided that data and of the institution where he or she works. It is perhaps a good idea in any case to always give the origin of data included in a repository; this is also good scientific/scholarly procedure.

## 7.5 Conclusion as regards liability

---

<sup>59</sup> See, for example, the Notice-and-take-down Code of Conduct at [http://isoc.nl/info/nieuws/NTD\\_Gedragcode.pdf](http://isoc.nl/info/nieuws/NTD_Gedragcode.pdf).

<sup>60</sup> See, for example, provision 2 of the DANS (Data Archiving and Networked Services) Licence Agreement: "2. The Depositor

- a. The Depositor declares that he is a holder of rights to the dataset, or the only holder of rights to the dataset, under the Databases act (Databankenwet) and where relevant the Copyright Act (Auteurswet) or otherwise, and/or is entitled to act in the present matter with the permission of other parties that hold rights.
- b. The Depositor indemnifies the Repository against all claims made by other parties against the Repository with regard to the dataset, the transfer of the dataset, and the form and/or content of the dataset."

A repository may be liable if, for example, material that is sensitive from the point of view of privacy is published or dealt with in some other manner that is contrary to legislation or regulations if

- it knows, or can reasonably be expected to know, that the information is of an unlawful nature but fails to remove it or make access to it impossible;
- when ordered by a public prosecutor,<sup>61</sup> it fails to take all measures that can reasonably be required of it to make that data inaccessible.

A repository does not therefore need to check material provided to it for possible irregularities; responding properly to complaints is sufficient.

It is advisable to record in the agreement with researchers who provide material

- that researchers who provide material are at all times liable in respect of that material; and
- that if any discussion arises as to the lawfulness of certain data, the complainant will be informed of the identity of the researcher who provided that data and of the institution where he or she works.

---

<sup>61</sup> with authorisation from the delegated judge.