

## Hand out 1

### Background

This session assumes that you already have some background in open data and there isn't time in this short session to go into legal details. This sheet contains a few headline legal points.

- Most of this session will relate to the law deriving from the Database Directive
- We will be discussing two rights:
  - database right
  - database copyright (or copyright in a database)
- This applies to databases made on or after **1 January 1998** (for older databases, the law is more complicated)

A **database** is defined as:

*a collection of independent works, data or other materials which—*

*(a) are arranged in a systematic or methodical way, and*

*(b) are individually accessible by electronic or other means.*

In short: almost anything you would think of as a database is a database for legal purposes along with other things such as poetry anthologies and libraries which you might not think of as a “database”.

### Contents v database

All we are going to care about is the database itself. The contents of the database may be protected in other ways, in particular by copyright law. The only copyright law we will look at today will be copyright in the database itself and not the contents of the database.

## Hand out 2

### IP rights in databases

#### Database Right

- Protects *investment*.
- Requires *substantial investment* in at least one of the following:
  - **obtaining** the contents
  - **verifying** the contents
  - **presenting** the contents
- “substantial” could be:
  - qualitative (eg using highly skilled volunteers)
  - quantitative (eg investing lots of money)
- 15 years since database was last updated

Database right is infringed (amongst other things) by:

- extraction (“the permanent or temporary transfer of all or a substantial part of the contents of a database to another medium by any means or in any form”)
- re-utilization (“any form of making available to the public all or a substantial part of the contents of a database by the distribution of copies, by renting, by on-line or other forms of transmission”)

#### Database Copyright

- Requires one or more *authors*.
- Protects creative authorship.
- Requires that the author(s) has invested their *own intellectual creation* in one of the following:
  - **selecting** the contents
  - **arranging** the contents
- Life of author + 70 years from date database was created

Copyright is infringed by (amongst other things):

- Copying
- Making an adaptation

## Hand out 3

### Exercise - Football fixtures

#### Team A: **Football Fixtures**

The English Premier League has 20 teams, each playing 38 matches each season, giving a total of 380 matches. The three other leagues have 24 teams and a correspondingly larger number of matches. The Scottish League is smaller.

The matches have to be organised subject to lots of constraints such as general rules (eg no 3 consecutive home or away matches); requests from clubs (eg Chelsea playing away during the Notting Hill Carnival); or other competitions (such as the FA Cup).

Mr Thompson, employed by the English and Scottish Leagues, spends a significant amount of time and skill preparing the list each year with quite a bit of help from a computer.

Do the leagues have database right or database copyright over the list of fixtures?

#### Team B: “**Football Live**”

- Football analysts attend football games
- Send in reports by mobile phone
- Collect data such as:
  - goals, times of goals, scorers, fouls, substitutions
  - dominant player for 10 minute periods, man of the match, severity of a foul
- Costs £600,000 per season.

Do the leagues have database right or database copyright over Football Live?

## **Hand out 4**

### **Football – what the courts decided**

#### **Football Fixtures Lists**

No Database Right – Fixtures Marketing v Organismos prognostikon agonon podosfairou C-444/02 (European Court of Justice)

No Database Copyright – Football Dataco v Yahoo! UK C-604/10 (European Court of Justice)

- database copyright is about the structure not contents
- intellectual effort and skill in creating data irrelevant to database copyright

#### **Football Live Data**

Database Right – Football Dataco v Stan James (Court of Appeal of England and Wales)

- Facts (such as the scoring of a goal) are pre-existing data which is “obtained” by the football analysis on behalf of Football Dataco
- Directive clearly intended that database right would protect this kind of data, otherwise it would be a “mouse of a right”
- Unmeasured information (eg a temperature) is an “existing independent material”

Database copyright assumed not to exist.

## Hand out 5

### Exercise - German poems

Dataset: the “1,100 most important poems in German literature between 1730 and 1900”.

A German university (Albrecht-Ludwigs-Universität Freiburg) created this list of poems in two stages:

- A German academic created a list of titles of German poems that (in his view) were important.
- A representative corpus of German-language poetry anthologies was selected and all 20,000 poems in those anthologies had their titles and first lines standardised.
- A statistical analysis was run to determine how frequently each poem was referenced, only those more frequently mentioned were selected
- A publication was created listing the poems in order of frequency of mention and giving their title, author, opening line and year of publication
- The whole process took 2.5 years and cost 34,900 Euro

Directmedia has their own collection of 1000 poems. Directmedia staff consulted the 1,100 poem list for inspiration and ended up with 856 poems in common.

Team A: Has there been an infringement of database copyright?

Team B: Has there been an infringement of database right?

Hint: make sure you consult Hand out 2.

## **Hand out 6:**

### **German poems – what the courts decided**

#### **Database Right**

Yes. The Database right was infringed by *extraction* – Directmedia Publishing v Albert-Ludwigs-Universitt Freiburg C-304/07 (European Court of Justice)

- extraction can be done by hand (no computer is needed)
- it doesn't matter that the data is presented in a different form

#### **Database Copyright**

Yes. Database copyright was also infringed. (German Federal Court of Justice)

## Hand out 7

### Virality and Criminality

#### Viral Licences

*A viral licence on a work forces others to openly share works they create from that work.*

Example: CC-BY-SA (Creative Commons Attribution Share-Alike)

If I write a book and license it under CC-BY-SA, anyone else can freely copy it and publish it. However, if they publish it they must also license it under a compatible open licence. If they make a derived work from it, they must also license that under a compatible open licence – hence the name “viral” because the “openness” is catching.

Application: a viral licence allows wide dissemination and re-use of your work amongst the community, eg of developers, but anyone wanting to commercially exploit it may want to be able to impose a non-open licence on their product. To do so they would need to obtain another (possibly non open) licence from you.

#### Criminality of Access

*Scraping and other access to a dataset may be criminal in the UK if it is contrary to the site owner's permission.*

Section 1 of the Computer Misuse Act 1990. It is an offence to access a computer if:

- the access is unauthorised
- you know the access is unauthorised

Example: police officer uses police computer to find out whose car is parked outside his ex-wife's house.

Application:

If your data is:

- Only useful if it is fresh (eg bus locations)
- Supplied via API using a key or some similar process
- You can control use, by controlling access

## Hand out 8

### Summary of open licences

You should already have encountered some well-known open licences. This sheet summarises features of the main open data licences.

<b>Unrestricted</b>	
<b>CC0</b>	
<b>PDDL</b>	Applies only to the database and not its contents.
<b>Attribution</b>	
<b>OGL</b> (Open Government Licence)	Revocable. Very short and clean licence. Not restricted to the public sector (despite the name).
<b>CC-BY</b> (Creative Commons Attribution)	Irrevocable. Applies to contents and database.
<b>ODC-BY</b> (Open Data Commons Attribution)	Irrevocable. Applies to database only.
<b>Viral</b>	
<b>CC-BY-SA (from version 4.0)</b>	Contents and database.
<b>ODbL</b>	Database only. Works as a contract as well as licence so may have limited effect in the US. More viral than CC-BY-SA: Virality attaches to derived works (such as infographics) as well as other databases.



## Hand out 9

### OpenCorporates data

Opencorporates is an open data startup that collects data from numerous national registries of companies. The data is obtained in different ways – often by scraping the source but by no means always. Each registry makes available different information but it may include:

- Name and registration number of the company
- Name and address of directors
- Documents filed by the company (eg annual reports)
- Accounting information

Opencorporates wants to:

- encourage re-use, particularly for people doing research or making other open data websites and services, but also commercially
- ensure that no-one can simply re-sell their data (and therefore compete with them)
- ensure that significant uses of their computer systems are paid for

Discuss how Opencorporates might achieve this by reference to:

- the technical methods they use to publish the data
- the licence(s) used