

## Old Dogs and New Tricks: Law and Digital Curation

**Mags McGinley**

Digital Curation Centre, University of Edinburgh

Email: [mags.mcginley@ed.ac.uk](mailto:mags.mcginley@ed.ac.uk)

Scientists, researchers and scholars across the world generate increasingly vast amounts of digital data. Together with institutional investment in digitisation and licensing of digital content, this activity results in what has been termed 'the data deluge'. The scientific record and the documentary heritage created in digital form are at risk from technology obsolescence, from the fragility of digital media, and from lack of the basics of good practice, such as adequate documentation for the data.

The Digital Curation Centre (the DCC) was launched just over two years ago in order to address these issues. It provides a national focus for research and development into digital curation issues and promotes expertise and good practice for the management of all research outputs in digital format. To carry out this project the DCC took on staff from some of the more obvious disciplines such as computer science, database theory and records management. In addition to this, the project creators had the foresight to recognise that digital curation involves a number of legal elements. On that basis they teamed up with the AHRC Research Centre for Studies in Intellectual Property and Technology Law (the "AHRC Research Centre") and appointed a full time member of staff to address the legal aspects of the project.

It is crucial that such an important topic is brought to the attention of the legal profession so that the legal issues involved in this developing area can be fully debated from the outset. The areas of law of most significance to digital curation are intellectual property (in particular copyright and the database right) and information governance (in particular data protection, freedom of information and re-use of public sector information). However, when you consider that the process may involve creating, re-using, preserving, annotating or amending all types of digital data or documents, whether they are legal case reports or databases containing petabytes of astro-physics results, many more areas of law become relevant. Examples include: defamation, contract, evidence, confidentiality, legal deposit, accessibility and human rights.

From the perspective of the sole legal personnel at the DCC the author would like to draw the topic of digital curation to the attention of other lawyers. The paper will explain the purpose of digital curation and describe its importance in an increasingly sophisticated technological environment. This will be followed by an overview of the DCC and its work. The main part of the paper will take a more in-depth look at the areas of law that have most relevance for digital curation (intellectual property and information governance) and the live and largely unresolved issues in these areas.

### WHAT IS DIGITAL CURATION?

This is not a simple question to answer. At a symposium at the 1<sup>st</sup> International Digital Curation Conference this was the only question discussed – for two and a half hours. The DCC website gives the following description:

*"Digital curation, broadly interpreted, is about maintaining and adding value to a trusted body of digital information for current and future use. The digital archiving and preservation community now looks beyond the preservation, cataloguing and cross referencing of static digital objects such as documents. The scientific community has data characterised by*

*structure, volatility and scale. These require us to extend our notions of curation. We must also investigate the principles that underlie appraisal, and lessons learnt about the economics of preservation.*"<sup>1</sup>

It is worth clarifying here that *curation* and *preservation* are not the same things. This is a common misconception. Curation may include preservation but it is not limited to that. This is highlighted in the following passage:

*"The term 'digital curation' is increasingly being used for the actions needed to maintain and utilise digital data and research results over their entire life-cycle for current and future generations of users.... Implicit in this are the processes of digital archiving and digital preservation but it also includes all the processes needed for good data creation and management, and the capacity to add value to data to generate new sources of information and knowledge. In most research fields, capturing 'knowledge' is more than just the archiving and preservation of source data and associated metadata. It generally involves interaction between creators and providers of data, the archivers of data, and most importantly the consumers of data. Successful curation of data requires users to be able to utilise the data using their current tools and methodologies"*<sup>2</sup>

This is an important thing to remember when identifying the areas of law that this relatively nascent topic has implications for.

## **WHY IS THERE A NEED FOR CURATION?**

### **Digital assets are fragile**

Many people are surprised to discover this. In many digital assets are not as robust as their non-digital counterparts. Media degradation is a big problem as is semantic loss and systems that are frozen or subject to incomplete migration processes. Lynne Brindley, Chief Executive of the British Library confirms:

*"In many ways digital material is more fragile than physical material and if we don't manage it effectively it won't survive for future generations."*<sup>3</sup>

### **Software and format obsolescence make things unusable**

A simple domestic example of this is that people have their favourite films stored on old VHS videotape (or even Betamax!) However, nowadays most people watch films on DVD players or computers or portable devices. The films on these occasions are not on VHS videos but on newer formats. Digital curation deals with this format obsolescence but on a grander scale. Consider the Microsoft Word application that this paper was written on. Newer versions of this are frequently being released. Over time the standard Microsoft Word document of today may not be accessible on the Microsoft format of the time. What happens if the code behind these applications is not available and people no longer understand how the older version worked? Consider finding an important digital document in 40 years time and not being able to access it because the software required is incompatible with your current hardware. In that case, digital curators could use a variety of techniques to allow access to the document. However, if the information that allows someone to do that is lost (or was never available) access may prove impossible.

### **Data sets can be badly labelled or organised and missing back up information that verifies their authenticity**

---

<sup>1</sup> <http://www.dcc.ac.uk/about/what/>

<sup>2</sup> Joint Information Systems Committee Circular 6/03 Revised. <http://www.dcc.ac.uk/docs/6-03Circular.pdf>

<sup>3</sup> Press Release: 'British Library predicts switch to digital by 2020.' <http://www.bl.uk/news/2005/pressrelease20050629.html>

Badly organised or lost descriptive information may sound like carelessness on the part of the creator. However, this happens even at the highest levels. Take for example this recent news story:

*“Houston, we have a problem. There is probably no artefact in the history of space exploration more precious than the first television images of the Moon captured by Neil Armstrong and his fellow astronauts as they disembarked from their lunar module in July 1969.*

*Unfortunately, the magnetic tapes of those images have gone missing. Worse still, they appear to have been missing for at least 30 years - and nobody, until now, even noticed...Their hope is to track the tapes down before they deteriorate so far as to be unreadable, then transfer them to digital format so they can be preserved for ever.”<sup>4</sup>*

A smaller scale domestic example that many people can relate to is home photograph collections. Households all over the country have drawers or albums full of disorganised photographs. The owners themselves may not be able to identify where the picture was taken, or when, or even who the subject is. Perhaps no information is given. At the time the picture was taken people thought that they would remember the details. As little as 10 years on that may not be the case. Consider then all those wonderful photos from 100 years ago. How much information do we have about them now? These principles equally apply to all sorts of digital data. In order for it to be useful it is not only the object that is required but the descriptive data that accompanies it (the metadata<sup>5</sup>). This is often missing or inadequate.

### **Data deluge**

Because of advancements in technology the generation of data has rocketed. In science alone the amount of information produced is reckoned to double every couple of years. Instead of gigabytes, we now use terabytes<sup>6</sup> as the standard volume of measurement. And the use of petabytes as a measurement is now not unusual. This presents us with a massive amount of data of a wider variety of types which need to be evaluated, organised, stored, and preserved and much more. No simple task.

### **WHY IS CURATION CONSIDERED TO BE IMPORTANT?**

The reasons that call for curation have been identified, but why are they valuable enough to be prioritised? Digital data is increasingly important as evidence. One example of this is email. Nowadays many things are agreed by email that would previously have been done by a more formal document. Emails become important as records. However, people’s email accounts are frequently a mess! E-discovery is becoming increasingly important in legal cases, particularly in the United States. In some cases the outcome rests on an item of digital data. Other times digital data gained from observations and experiments are unrepeatable. This is particularly true in the case of environmental observations and experiments. Preservation of cultural heritage is also a big factor. People care about the documents or artefacts that authenticate the occurrences of history. Lastly continuing access to digital data is vital to ensure contemporary scholarship is reproducible and verifiable and that further progress can be made by building on what has gone before. All of these things, to one extent or another, are evidence. For that evidence to have strong evidential value, data must be curated.

### **THE DIGITAL CURATION CENTRE**

The DCC is a consortium project made up of four partner institutions: The University of Edinburgh<sup>7</sup>, HATII at the University of Glasgow<sup>8</sup>, UKOLN at the University of Bath<sup>9</sup> and the

<sup>4</sup> ‘Houston, our tapes have gone missing’, Andrew Gumbel, The Independent on Sunday, 13 August 2006

<sup>5</sup> For a description of the term ‘metadata’ see <http://en.wikipedia.org/wiki/Metadata>

<sup>6</sup> One terabyte = 1000 gigabytes and roughly 100,000 times the capacity of the average home computer

<sup>7</sup> <http://www.ed.ac.uk/>

<sup>8</sup> <http://www.hatii.arts.gla.ac.uk/>

Council for the Central Laboratory of Research Councils (“CCLRC”)<sup>10</sup>. Edinburgh is the lead site and provides input from the Database Research Group within the School of Informatics, EDINA (a JISC-funded national data centre) and from the AHRC Research Centre at the School of Law. HATII provides expertise in the use of information and communication technology to enhance research and teaching in the arts and humanities. UKOLN is a key contributor of expertise in digital information management. Lastly, CCLRC’s experience is crucially important as it is one of Europe’s largest multidisciplinary research organisations supporting scientists and engineers worldwide. It operates large scale ‘big science’ research facilities so it provides really important input at that end of the scale.

Working with other practitioners, the DCC supports UK institutions that store, manage and preserve digital data to help ensure their enhancement and their continuing long-term use. In the first phase of the DCC there were four teams (Research, Development, Services and Outreach).

During Phase One of the DCC the Research team addressed: data integration and publishing; annotation; metadata extraction; provenance and data quality; and legal issues. The Development team was involved in: monitoring international standards; developing tools and methods; developing a representation information registry; and creating testbeds. The Services team began compiling a digital curation manual written by world experts; manned a helpdesk system; created a variety of themed resources; developed audit and certification guidelines; provided support for LOCKSS users;<sup>11</sup> and presented information days and training throughout the UK. The Outreach team created and developed the DCC web portal; organised two international conferences; created an associates network of individuals and organisations interested in curation; and launched the International Journal of Digital Curation.

In Phase Two of the DCC, which began recently, there is a re-profiled organisational structure resulting in the four original teams being amended to six work packages. These are: management, community development, tools and infrastructure, resources, events and curation services.

The objectives of the DCC for phase 2 are:

- Provide strategic leadership in digital curation and preservation for the UK research community, with particular emphasis on science data
- Influence and inform national and international policy
- Provide advocacy and expert advice and guidance to practitioners and funding bodies
- Create, manage and develop an outstanding suite of resources and tools
- Raise the level of awareness and expertise amongst data creators and curators, and other individuals with a curation role
- Strengthen community curation networks and collaborative partnerships
- Continue the strong association with the research programme<sup>12</sup>

## **WHAT AREAS OF LAW ARE MOST RELEVANT AND WHY?**

As already discussed, curation is a very wide-ranging discipline. When considering creating, re-using, preserving or amending all types of digital data or documents, from a legal journal article to a database containing petabytes of astrophysics results, many areas of law become relevant. These include: defamation; evidence; contract; confidentiality; legal deposit; accessibility; and human rights. The ones that the DCC has found to be of major relevance and which this paper will discuss are:

- Copyright
- The Database Right

---

<sup>9</sup> <http://www.ukoln.ac.uk>

<sup>10</sup> <http://www.cclrc.ac.uk/>

<sup>11</sup> LOCKSS is a project that delivers software that provides institutions with a way to collect, store and preserve access to their own local copy of licensed content. <http://www.lockss.org/lockss/Home>

<sup>12</sup> <http://www.dcc.ac.uk/about/#objectives>

- Data Protection
- Freedom of Information
- Re-Use of Public Sector Information

## COPYRIGHT

Copyright is governed by the Copyright, Designs and Patents Act 1988 (“the CDPA”).<sup>13</sup> It is the branch of intellectual property law that protects the expression of ideas.<sup>14</sup> The categories of work that copyright applies to are: literary, dramatic, artistic and musical works, databases, typographical arrangements, sound recordings, films and broadcasts. These categories cover things such as: books, films, songs, websites, computer programmes, scientific databases, emails, diagrams, maps and much more. Preservation and curation of digital materials is dependent on a range of strategies that require the making of copies and modifications. It can therefore be appreciated that copyright’s scope for affecting the curation of digital data is vast.

Copyright comes under a lot of criticism; especially from those involved in digital curation where the rights of the copyright holder to prevent copying and modification of a work<sup>15</sup> can be seen as restricting vital curation activities. However, copyright is not without its redeeming qualities. There are a number of reasons that justify its existence. It is argued that the monopoly it imposes provides an incentive for creators to create. As one of the objectives of curation is to support and encourage future progress and creativity, the existence of an incentive to create (in the form of copyright) is a plus point. The issues with copyright arise around trying to hit the right balance between meeting the interests of right holders by providing incentives to create and protections for creative works with the interests of wider society to use the work in order to encourage further creativity and progress and (in relation to the topic of this paper) to allow ease of curation activity.

One of the more significant aspects of copyright for curation purposes is the fact that a user may not carry out an act restricted by copyright without the permission of the copyright owner.<sup>16</sup> Examples of the types of activities that a person may want to carry out include:

- Making a copy of a work as part of a digitisation programme
- Undertaking preservation activity such as emulation<sup>17</sup> or migration<sup>18</sup>
- Incorporating the work, or part of the work into a new work of their own

Depending on the circumstances exceptions to copyright may allow these but in many circumstances they will not. There are certain exceptions to this principle, including what have been termed the ‘fair dealing’ provisions.<sup>19</sup> These are aimed at research for non-commercial purposes, private study, criticism or review. They may be of use to digital curators. It is important to note that at present the provisions in section 29 (applying to research and private study) do not apply to copyright in films, sound recordings, broadcasts or typographical arrangements (only to literary, dramatic, musical and artistic works). However, this was addressed in the recent Gowers Review of Intellectual Property (the Gowers Review) where it

---

<sup>13</sup> As this paper is intended for a legal audience, a basic knowledge of copyright is assumed. For further information please see L. Bently and B. Sherman, Intellectual Property Law, (2004) Part I

<sup>14</sup> It is important to recognise that copyright does not protect the *information* but the *expression* of that information.

<sup>15</sup> Section 16 CDPA

<sup>16</sup> Section 16 CDPA

<sup>17</sup> “Emulation : A means of overcoming technical obsolescence of hardware and software by developing techniques for imitating obsolete systems on future generations of computers”, Jones and Beagrie, Preservation Management of Digital Materials, (2003) p105

<sup>18</sup> “Migration: A means of overcoming technical obsolescence by transferring digital resources from one hardware/software generation to the next. The purpose of migration is to preserve the intellectual content of digital objects and to retain the ability for clients to retrieve, display, and otherwise use them in the face of constantly changing technology.” Ibid p106

<sup>19</sup> Sections 29 and 30 CDPA

was recommended that private copying for research be allowed to cover all forms of content.<sup>20</sup> On this basis, changes may be afoot.

It is worth noting that fair dealing is unlikely to apply where the whole of a work has been copied. So, if a situation arises where a user comes within a copyright exception but would like to copy an entire work that will still be infringement. Unfortunately that is likely to be the case in a number of curation scenarios where the purpose is to preserve the resource as a whole.

This would seem a suitable time to segue into the difficulties presented by digital rights management technology (DRM). DRM is an umbrella term referring to any of several technologies used to enforce pre-defined policies controlling access to and use of protected works. There is huge concern that DRM technology to protect content could lead to 'digital lock out' and the diminishment of the material freely available for use existing in the public domain. DRM can prevent a person from using a legitimate exception to copyright or the database right because it doesn't recognise that a user fits into one of the relevant categories. This was also a topic of discussion in the recent Gowers Review. This is a huge issue for curation and preservation initiatives which may often rely on exceptions and who also hope to be the beneficiaries of future expansion of the exceptions such as the Gowers Review recommendation to enable libraries to format-shift archival copies to ensure records do not become obsolete.<sup>21</sup>

The potential for DRM technologies to render fair dealing and other exceptions unavailable is recognised in the legislation. Section 296ZE of the CDPA provides that in the case of this happening the Government can issue civilly enforceable directions to rights holders. However, it stipulates that no complaint may be made where the copyright work in question has been made available on contractual terms in a way that members of the public may access it at a place and a time chosen by them. This seems to suggest that the only complaint likely to succeed is where access is completely blocked and that a party making material available for a price, however exorbitant, is likely to be safe.

Concern has also been expressed that DRM won't recognise the expiry of the copyright period. After a statutorily defined period of time any copyright work becomes part of the public domain for anyone to use freely. Public domain works are much easier to curate, preserve and re-use. In its evidence to the All Parliamentary Internet Group Inquiry into Digital Rights Management the Libraries and Archives Copyright Alliance expressed concern that restrictions will be allowed to continue after copyright expires:

*"The effect of DRMS, if conditions are not applied to these measures, is to make copyright perpetual, which goes against the long-standing principles of all existing intellectual property laws. By the time copyright expires the rightholder company may have gone out of business or merged one or more times with other companies. The ownership of the rights may be impossible to trace rendering the product orphaned. It is probable that no key would still exist to unlock the DRMS. For libraries this is serious. As custodians of human memory, a number would keep digital works in perpetuity and may need to be able to transfer them to other formats in order to preserve them and make the content fully accessible and usable once out of copyright."*<sup>22</sup>

These difficulties aside, if a use or user does not fall within an exception, the potential user<sup>23</sup> will need to do one of a number of things:

---

<sup>20</sup> Recommendation 9, The Gowers Review of Intellectual Property. [http://www.hm-treasury.gov.uk/independent\\_reviews/gowers\\_review\\_intellectual\\_property/gowersreview\\_index.cfm](http://www.hm-treasury.gov.uk/independent_reviews/gowers_review_intellectual_property/gowersreview_index.cfm)

You can read the report of the Review at : [http://www.hm-treasury.gov.uk/media/53F/C8/pbr06\\_gowers\\_report\\_755.pdf](http://www.hm-treasury.gov.uk/media/53F/C8/pbr06_gowers_report_755.pdf)

<sup>21</sup> The Report of the Gowers Review of Intellectual Property, Recommendation 8 : [http://www.hm-treasury.gov.uk/media/53F/C8/pbr06\\_gowers\\_report\\_755.pdf](http://www.hm-treasury.gov.uk/media/53F/C8/pbr06_gowers_report_755.pdf)

<sup>22</sup> Para 1.2 LACA response: APiG inquiry into DRMS (December 2005) <http://www.apig.org.uk/current-activities/apig-inquiry-into-digital-rights-management/apig-drm-written-evidence.html>

<sup>23</sup> The term 'user' is used to cover an individual or organisation wishing to undertake all types of curation activity or to benefit from access to curated data

- Obtain an assignment or an assignation<sup>24</sup> from the copyright holder. This means that the user then owns the copyright and all the rights that accompany it. This is the traditional scenario with publishers.
- Obtain a licence from the copyright holder. This licence may be exclusive to the user or non-exclusive (others will also be allowed to use the work). It will cover the specific acts that may be carried out and can be a very flexible document. It is important to remember here that the licensor remains the copyright owner.<sup>25</sup>
- Obtain a licence from a collective licensing society (such as the Copyright Licensing Agency or the Publishers Licensing Society) to whom authors have licensed specific uses of their material. This is especially prevalent in the education sector.

The difficulty here is that specific permission can be challenging and time and resource consuming to obtain. This is especially acute with digital items that are orphan works.<sup>26</sup>

Another important element to consider in relation to copyright is that a single digital object may embody more than one copyright work, each with a different owner. For example a television programme may have literary, dramatic and musical copyrights for the script, screenplay and score respectively. There will then be further protection for the recording of the programme. Many web pages contain multiple copyrights, with text (and indeed the html coding and metadata) protected as literary work, graphics and photographs as artistic works, and sound files as musical work. In such cases permission will be even more burdensome to obtain.

A final area that can cause difficulty in obtaining permissions is the complex copyright ownership rules, especially in the cases of employment relationships and multiple authorship. Multiple authorship is not only common but the norm in the case of database authorship, to which we now turn.

## **DATABASE RIGHT**

Much digital curation activity, especially in the scientific arena, involves databases. Nowadays much information is stored and made accessible through a database. For legal purposes, 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.”<sup>27</sup>*

Unsurprisingly, the principal intellectual property right to look at when considering databases is the database right. This sui generis right was enacted relatively recently (in comparison to copyright's 400 year existence) as a result of Directive 96/9/EC on the legal protection of databases (the Database Directive)<sup>28</sup> and then implemented in the UK by the Copyright and Rights in Databases Regulations 1997 (the Database Regulations).<sup>29</sup> A key concern in relation to the Database Directive has been a perception that the sui generis right seems close to the grant of an intellectual property right in data and information per se, allowing only limited extractions for the purposes of non-commercial research. The result is that scientists may suffer restrictions on access to, and ability to re-use the raw data necessary for scientific progress.<sup>30</sup>

<sup>24</sup> 'Assignment' is the term in the law of England and Wales and 'assignation' is the law of Scotland.

<sup>25</sup> A lengthy paper could be written on solely the implications of licensing for digital curation. Especially interesting is the prevalence of Creative Commons licences and the affect this may have on digital curation.

<sup>26</sup> Broadly speaking orphan works are copyrighted works where the copyright owner is difficult or impossible to find

<sup>27</sup> Section 3A (1) CDPA

<sup>28</sup> <http://europa.eu.int/ISPO/infosoc/legreg/docs/969ec.html>

<sup>29</sup> <http://www.opsi.gov.uk/si/si1997/19973032.htm>

<sup>30</sup> International Council for Science, Scientific Data and Information – A Report of the CSPR Assessment Panel, December 2004. Available @:

Before moving on, it is important to clarify the relationship between the database right and copyright when the work in question is a database. A database can attract copyright protection, database right protection, both or neither. This is because there are separate qualifying criteria for the two rights. Copyright may exist in the structure of a database if, by reason of the selection and arrangement, it constitutes the authors own intellectual creation. In addition the contents of database, depending on what they are, may attract their own copyright protection (a simple example might be a database of poems).

On the other hand the database right protects the investment in the contents of a database (although it is hoped not the data per se). A database will attract database right protection if the criteria discussed below are met.

As this right was created quite recently it merits further description. First of all it is very important to note that the database right doesn't apply to all databases. The right subsists in a database only if there has been substantial investment in at least one of three things. These are *obtaining*, *verifying* or *presenting* the contents of the database. Investment in the creation of the data or other activity pre-inclusion does not count. If a database qualifies for the protection afforded by the database right this enables the owner to prevent others from unauthorised extraction and/or reutilisation of all or a substantial part of the contents of the database.

A few definitions of the terms involved may be helpful:

- Investment is defined as the deployment of financial resources; and/or the expending of time, effort and energy.<sup>31</sup>
- Obtaining very specifically relates to the collection or collation and not creation of data. In referring to the investment needed in obtaining the contents of the database the ECJ said that this "*must be understood to refer to the resources used to seek out existing independent materials and collect them in the database and not to resources used for the creation as such of independent materials.*"<sup>32</sup>
- Verifying covers verification of the accuracy of the data *once the materials are in the database.*
- Presentation concerns the presentation of *the contents* of the database as opposed to the database itself.
- Extraction is defined as the permanent or temporary transfer of the contents of the database.
- Re-utilisation is making the contents available to others.

The fair dealing exceptions for the database right are not as wide as those of copyright. Extraction of a substantial part of the contents is allowed by a lawful user for the purposes of illustration for teaching or scientific research provided that the source is indicated and it is to the extent justified by the non-commercial purpose to be achieved.<sup>33</sup> Reutilisation of the contents of the database is not permitted.

The Royal Society commented on this in their 2003 report, Keeping Science Open: the effects of intellectual property policy on the conduct of science:

*"The fair dealing exception under UK law, in line with the EC Directive permits only extraction and not re-utilisation. Re-utilisation is an essential part of scientific endeavour, and so this limitation does not address the scientific community's needs. The effects of these limitations.. are difficult to assess quantitatively.. but in our view they will in the longer term if vigorously*

---

[http://www.icsu.org/Gestion/img/ICSU\\_DOC\\_DOWNLOAD/551\\_DD\\_FILE\\_PAA\\_Data\\_and\\_Information.pdf](http://www.icsu.org/Gestion/img/ICSU_DOC_DOWNLOAD/551_DD_FILE_PAA_Data_and_Information.pdf).

<sup>31</sup> Database Directive, Recital 40

<sup>32</sup> Fixtures Marketing Ltd. v Organismoa Prognostikon Agnon Podosfairou (OPAP) C-444/02

<sup>33</sup> Database Directive, Article 9

*enforced become a serious impediment to scientific research and hence to the national interest.*<sup>34</sup>

The term of the database right is much shorter than that of copyright at only fifteen years.<sup>35</sup> From the perspective of things falling into the public domain and the ensuing potential for greater re-use this is a good thing. However, there is some uncertainty surrounding this. Article 10(3) of the Database Directive states that there is renewal of the term of the right each time there is a “substantial change” to the contents. Many databases are dynamic. If a database is continually changed and updated could the database right in any one database last indefinitely? This is presently unresolved by the courts or the legislators.

Regardless of the term of the right, a user may still extract and re-utilise an insubstantial part of the contents of the database without permission and substantial parts without permission for the purposes of non-commercial research<sup>36</sup> (assuming always that access can be gained). However, it is not yet clear (from neither the legislation or case law) what is meant by “a substantial part”.<sup>37</sup>

As can be seen, the ambit of the database right is still rather uncertain which creates an unwillingness to act in this area which in turn reduces the re-use people are willing to make. Unfortunately this affects the free flow of digital information and the productivity it encourages. Scepticism about the existence of this right is certainly not limited to the digital curation community. Criticism has been widespread. The recent British Academy Review of Copyright and Research in the Humanities and Social Sciences expressed concern that “*the Database Directive is at once vague and wide-ranging*”<sup>38</sup> and included the following recommendation:

*“Both public authorities and academic communities should monitor carefully the assertion of database rights and charges made for access to database contents in order to ensure that this right does not become a growing impediment to scholarship.”*<sup>39</sup>

## **DATA PROTECTION**

Data Protection is covered by the Data Protection Act 1998 (the DPA)<sup>40</sup> and governs the handling of a living individual’s personal data. It seeks to strike a balance between the interests of an individual in maintaining privacy over their personal details with the sometimes competing interest of those with legitimate reasons for using personal information. It gives individuals certain rights regarding information held about them, whilst placing obligations on those who process that data.

Data protection is relevant to digital curation from two different perspectives. Firstly, the DPA impacts the way those who have control over personal data can use it. If someone is curating or reusing data that is covered by the DPA (perhaps a researcher using medical or social science data) they will need to be aware of the legislation and take steps to comply with the Act. Although researchers are exempted from certain of the data protection principles, an awareness of the constraints imposed by the legislation and how it impacts the way the data can be used is crucial. This becomes even more significant if the personal data is of the type defined by the DPA as ‘sensitive personal data’.<sup>41</sup> In such cases more stringent rules apply.

---

<sup>34</sup> Keeping Science Open: the effects of intellectual property policy on the conduct of science. Available @ <http://www.royalsoc.ac.uk/displaypagedoc.asp?id=11403> at 23.

<sup>35</sup> Database Directive, Article 10

<sup>36</sup> Database Directive Article 9(b).

<sup>37</sup> C Waelde and M McGinley, "Public Domain; Public Interest; Public Funding: focussing on the 'three Ps' in scientific research", (2005) 2:1 *SCRIPT-ed* 71 @: <http://www.law.ed.ac.uk/ahrc/script-ed/vol2-1/3ps.asp>

<sup>38</sup> British Academy Review of Copyright and Research in the Humanities and Social Sciences, September 2006, para 46

<sup>39</sup> Above, Recommendation 9

<sup>40</sup> <http://www.opsi.gov.uk/ACTS/acts1998/19980029.htm>

<sup>41</sup> Section 2, the DPA

From the other perspective, implementing robust curation practices in relation to digital data will be of great assistance in not only not falling foul of the legislation but facilitating faster and more efficient compliance, thus requiring fewer resources in the long term.

## **FREEDOM OF INFORMATION**

The Acts that introduced freedom of information in the UK were the Freedom of Information Act 2000 and the Freedom of Information (Scotland) Act 2002 (the FOI Acts).<sup>42</sup>

In broad terms the FOI Acts set out a right of access by the general public to all information held by public authorities.<sup>43</sup> Information may be accessed by two means: via the public authority's publication scheme and via a right to request information. The FOI Acts and associated secondary legislation proscribe in further detail the time limits and exemptions relating to these two rights.

Freedom of information's relevance to digital curation lies in a few areas. Firstly, the FOI Acts apply to electronic records and paper records equally. The fact that any recorded information can be requested means that all applicable data has to be stored and retrieved effectively. Systems for organisation and retrieval are well established in the paper environment, but most records are now in digital format. A public authority's ability to comply with the legislation will be facilitated by effective electronic records management which is very closely linked to digital data curation in the humanities and social sciences.

In addition to this a record has not only to be retrievable but also needs to be readable or useable over the long term. For this reason it needs to be curated. In some cases the Acts have forced public authorities to tackle the question of longevity of data and allocate funds to support curation and preservation initiatives where they may not have otherwise done so.

The FOI Acts may also cause public authorities to consider their disposal schedule. Under the provisions, unless records are disposed of (preferably in accordance with an agreed disposal schedule) they can be requested by members of the public. As discussed earlier, curation is not just about keeping and preserving everything. It involves decisions about what is still useful. Evaluation of data is a key part of the process to decide whether it merits preservation or should be disposed of. In addition, disposal of a digital record is a whole different procedure from disposal of a paper record. Many versions of a digital record may exist, some of which the average-user may not even be aware of. Further renditions may end up being preserved even though the 'original' has been destroyed in good faith. Well curated data will reduce the risk of such a 'deleted' document turning out not to be so and will improve the ease with which the Acts can be complied with and lead to overall improvements in efficiency.

Lastly, and simply, freedom of information is about access. So is curation.

## **RE-USE OF PUBLIC SECTOR INFORMATION**

The law relating to re-use of public sector information was introduced in 2003 by a European Directive.<sup>44</sup> This Directive was implemented in the UK by the Re-use of Public Sector Information Regulations 2005 (the Regulations).<sup>45</sup>

The aim of the Regulations is to encourage re-use of public sector information by removing obstacles that stand in the way of such action. The intention is that this will stimulate the development of innovative information products and services across Europe thereby boosting the information industry, in a similar way to the United States. Re-use of information means the reproduction of documents in a way that was not originally intended when they were created. This includes copying, adapting, developing, adding value, broadcasting,

---

<sup>42</sup> <http://www.opsi.gov.uk/acts/acts2000/20000036.htm> and

<http://www.opsi.gov.uk/legislation/scotland/acts2002/20020013.htm>

<sup>43</sup> The majority of higher and further education institutions fall within the definition of public authority.

<sup>44</sup> Directive 2003/98/EC on the Re-use of Public Sector Information. [http://europa.eu.int/eur-lex/pri/en/oj/dat/2003/l\\_345/l\\_34520031231en00900096.pdf](http://europa.eu.int/eur-lex/pri/en/oj/dat/2003/l_345/l_34520031231en00900096.pdf)

<sup>45</sup> S.I. 2005 No. 1515 <http://www.opsi.gov.uk/si/si2005/20051515.htm>

downloading and other actions. 'Documents' covers written texts, databases, audio files and film fragments. Once again, it can be recognised that the scope of the legislation is very broad.

It is important to note that at present the Regulations do not apply to educational and research establishments. They apply to all public sector organisations other than public service broadcasters, educational and research establishments, archives, libraries and museums and cultural establishments. However, this has been criticised and there is a strong chance that when the Regulations are reviewed in 2008 the bodies that they apply to will be extended.

It is still important for those involved in curation within education and research to know about the Regulations because educational and research establishments are not only creators of intellectual property but great re-users of intellectual property. Therefore, although the Regulations do not require them to make their information available, they may wish to avail themselves of the Regulations to utilise the information created or gathered by others

The Regulations are incredibly relevant to digital curation. They are about access and re-use.<sup>46</sup> These are integral parts of curation. As emphasised a number of times, curation is not just preservation.

Effective curation efforts involve knowing what information held digitally is re-useable or useful to others and making that available to promote innovation downstream.

Again, as with other information governance topics, the legislation's relationship with digital curation is reciprocal. Having robust curation practices facilitates compliance. For example, one of the requirements of the Regulations is to produce an Information Asset List which details what is available and how the information can be obtained. This will be significantly easier to achieve in an organisation with good curation practices.

## **WHERE DO LAWYERS COME IN?**

It is evident that digital material equals new challenges in many areas of the law. Digital curation is not optional. Long term neglect will lead to large scale losses. Although this may be unfamiliar territory for many lawyers, there is a lot of work to be done in investigating the interplay between digital curation and the law. In addition there will also be much work researching solutions to any difficulties that become apparent. This will be ongoing as technology advances and curation methods change. The legal team at the DCC will continue dissemination work on this topic. It is hoped that other legal professionals will consider these matters also. Both can then work together towards the ultimate goal of raising awareness of the importance of integrating law with science in decisions about digital curation activity.

---

<sup>46</sup> For an explanation of the interplay between FOI's 'access' and Re-use of PSI's 're-use' see <http://www.opsi.gov.uk/advice/psi-regulations/advice-and-guidance/psi-guidance-notes/links-between-access-and-reuse.htm>