



COLUMN

Copyright in the networked world: using facts

Copyright in the networked world

463

Michael Seadle
MSU Libraries, Michigan State University, East Lansing, Michigan, USA

Received 14 June 2006
Revised 16 June 2006
Accepted 18 June 2006

Abstract

Purpose – The purpose of this column is to look at how US copyright law deals with facts and what can reasonably be considered as a fact.

Design/methodology/approach – The US statutes and case law are examined, as well as standard practices within academic settings for the use and analysis of facts.

Findings – Using facts may require a risk assessment under certain circumstances. Although facts cannot be protected under US law, compilations of facts can have protection under the European Database Directive, and certain apparent facts, such as normal human body temperature, represent judgments based on extensive research and analysis. It is not always clear when a research result achieves the status of a fact and is therefore exempt from copyright protection.

Originality/value – The use of facts is an essential part of scholarly work and factual compilations are important tools for many scholarly communities. Understanding how far facts may be used and manipulated helps researchers and teachers who have grown anxious about copyright infringement.

Keywords Copyright law, United States of America

Paper type Case study

Introduction

[My mother] always says, my Lord, that facts are like cows. If you look them in the face hard enough they generally run away – Bunter to Lord Peter (Sayers, 1927).

One of the most common copyright questions I get in my role as copyright advisor is whether a graphical representation of a statistical analysis of data may be copied and used online. In other words, can the data and the resulting graph be considered facts? Or do the data points on the graph represent an analysis with enough originality to meet the requirement of the copyright law? And do the data underlying the analysis have the legal status of “facts” or did the work that went into creating or assembling them suffice for copyright protection? Facts are rarely simple when looked at carefully enough.

The goal of this column is to help answer these questions. Only a court can give a definitive answer. I will present some of the factors that a US court might consider, including:

- the US copyright statutes;
- the 1991 Feist decision that discussed both facts and compilations of facts; and
- the 1999 Bridgeman Art Library decision, where the court considered the effect of a mechanical process on public domain materials.



I will also talk briefly about how the European Database directive might affect the handling of facts within the European Union.

Facts in the statutes

The subject matter of US copyright law lies in “original works of authorship fixed in any tangible medium of expression” and:

In no case does copyright protection for an original work of authorship extend to any idea, procedure, process, system, method of operation, concept, principle, or discovery, regardless of the form in which it is described, explained, illustrated, or embodied in such work (17 USC 102).

Ideas are, in other words, excluded from copyright protection. Facts are too, but not quite so explicitly. The exclusion of facts rests largely on the meaning of the word “original.” Originality is a quality that facts are not supposed to have.

The US copyright law has a section with definitions (17 USC 101), but that section contains no definition of the words “fact”. As far as the US copyright statute is concerned, the word has no special legal meaning, unlike, for example, “publication” – an equally ordinary word whose legal meaning is broader than its conventional usage.

Even in conventional usage the precise meaning of the word “fact” can be difficult to pin down. The Merriam-Webster Online Dictionary has five definitions for the word, the most specific of which is: “a piece of information presented as having objective reality” (Merriam-Webster Online Dictionary, 2006). Objective reality can, however, be a subjective matter.

For example, 37°C (98.6 Fahrenheit) is widely quoted as the normal temperature for humans and is often treated as a fact when teaching school children. It is actually an average that researchers determined in the nineteenth century from a composite of readings. That average is seen today as possibly somewhat high and as more of a point within a range of reasonable healthy human temperatures than as a single expression of objective reality.

Nonetheless few people outside the medical profession would dispute 37°C (98.6 Fahrenheit) as a fact. Does that mean it is one? Or is it not a fact because of the intellectual originality and physical labor that went into establishing and, more recently, reconsidering the average? This is the kind of uncertainty that faculty and students face in knowing whether something is a “fact” for online publications or teaching.

Facts as defined in Feist

The 1991 Supreme Court ruling in *Feist Publications v. Rural Telephone Inc.* is a landmark case that defines how US copyright law views facts and compilations of facts.

Rural Telephone was a utility that served parts of Kansas. Feist was a publishing company that reused information from Rural Telephone’s white pages listings in its own directory without the utility’s consent. Rural Telephone sued for infringement, and the District Court upheld their infringement complaint. The Appeals Court did too, but the Supreme Court reversed that decision:

... it was held that (1) the names, towns, and telephone numbers listed in the white pages were not protected by the telephone company’s copyright in its combined white and yellow pages directory, because the listings in the white pages were not original to the telephone company, since (a) the listings, rather than owing their origin to the telephone company, were uncopyrightable facts, and (b) the telephone company has not selected, coordinated, or

arranged these uncopyrightable facts in an original way sufficient to satisfy the minimum standards for copyright protection (US Supreme Court, 1991).

The Court explained that there were circumstances under which a factual compilation might have copyright protection as well as the limits of that protection:

Although a factual compilation is eligible for copyright if it features an original selection or arrangement of facts, and although a compilation author who clothes facts with an original collocation of words may be able to claim a copyright in this written expression, others may copy from the publication the underlying facts but not the precise words used to present them, because (1) the copyright is limited to the particular selection or arrangement, and (2) in no event may copyright extend to the facts themselves, since, no matter how original the format of the selection or arrangement, the facts do not become original through association; thus, notwithstanding a valid copyright, a subsequent compiler remains free to use the facts contained in another's publication to aid in preparing a competing work, so long as the competing work does not feature the same selection and arrangement.

The Court's reasoning rested on the constitutional basis underlying US Copyright law:

... the primary objective of copyright is not to reward the labor of authors, but to promote the progress of science and useful arts, where to this end, copyright, under the principle known as the idea/expression or fact/expression dichotomy, which applies to all works of authorship, assures authors the right to their original expression, but encourages others to build freely upon the ideas and information conveyed by a work.

The Court rejected the "sweat of the brow" doctrine because it distorts "basic copyright principles":

Although protection for facts and ideas that are the fruits of an author's research may in certain circumstances be available under a theory of unfair competition, to accord copyright protection on the basis of the doctrine known as "sweat of the brow" or "industrious collection" – the underlying notion of which is that copyright is a reward for the hard work that goes into compiling facts – distorts basic copyright principles, in that it creates a monopoly in public domain materials without the necessary justification of protecting and encouraging the creation of writings by authors, where (1) the most fundamental axiom of basic copyright principles is that no one may copyright facts or ideas, and (2) it is just such wasted effort that would result from relying on facts contained in prior works that the proscription against the copyright of ideas and facts is designed to prevent.

The Court also interpreted the definition of "compilation" in the Copyright Act as meaning that collections of facts are not copyrightable *per se*, and emphasized the originality requirement for any legitimate copyright of a compilation:

The purpose of the definition of "compilation" in 101 of the Copyright Act of 1976 (17 USCS 101) – which defines a compilation in the copyright sense as a work formed by the collection and assembly of pre-existing materials or of data that are selected, coordinated, or arranged in such a way that the resulting work as a whole constitutes an original work of authorship – is to emphasize that collections of facts are not copyrightable *per se*; this message is conveyed through 101's structure identifying three distinct elements required to be met for a work to qualify as a copyrightable compilation – (1) collection and assembly of pre-existing material, (2) selection, coordination, or arrangement of the material, and (3) originality based on selection, coordination, or arrangement; the plain language of 101 indicates that not every collection of facts receives copyright protection, since otherwise there would be a period in 101 after the word "data".

In this decision the Supreme Court does not define facts explicitly, but it does emphasize one key characteristic: that facts are never original. At a common-sense level this makes sense, much as did the dictionary definition about facts having “objective reality.” But in the realm of scientific predictions, originality and reality can suddenly switch places.

For example, in 1909 when Pickering and Lowell predicted the existence of Pluto, the planet was essentially just the copyrightable expression of their original ideas. In 1930 after astronomers at the Lowell Observatory saw Pluto, the planet suddenly transformed into a pre-existing unoriginal objective reality.

This happened even though the very act of seeing something in a telescope some millions kilometers away is itself a creative act that requires human judgment and technical skill – a level of originality that long sufficed for photographs to have copyright protection even when the subject was (like facts) in the public domain.

Bridgeman and slavish transformation

The issue in the case of *Bridgeman Art Library vs Corel* was whether an exact photographic reproduction of a two-dimensional public domain work of art was itself in the public domain, or whether the new image had sufficient originality for copyright protection. The court reviewed cases where prior courts had granted copyright protection to photographs, but decided that an attempt at slavish copying lacked the originality necessary for copyright protection:

In this case, plaintiff by its own admission has labored to create “slavish copies” of public domain works of art. While it may be assumed that this required both skill and effort, there was no spark of originality – indeed, the point of the exercise was to reproduce the underlying works with absolute fidelity. Copyright is not available in these circumstances (US District Court for the Southern District of New York, 1991).

The court also considered British law, since the *Bridgeman Art Library* is a British institution. The decision quoted a work by Hugh Laddie, Peter Prescott, and Mary Vitoria, *The Modern Law of Copyright and Design*. Laddie is now a British judge:

Under the 1988 Act the author is the person who made the original contribution and it will be evident that this person need not be he who pressed the trigger, who might be a mere assistant. Originality presupposes the exercise of substantial independent skill, labour, judgment and so forth (US District Court for the Southern District of New York, 1991).

The court also disagreed with plaintiff claims based on the skill need to take a photograph and the change in medium from oil painting to a transparent slide:

Plaintiff nevertheless argues that the photocopier analogy is inapt because taking a photograph requires greater skill than making a photocopy and because these transparencies involved a change in medium. But the argument is as unpersuasive under British as under U.S. law (US District Court for the Southern District of New York, 1991).

What makes *Bridgeman* relevant to a discussion about facts and data is the degree to which the comments about slavish copying carry over to other mechanical processes. Statistics is a branch of mathematics that clearly requires significant understanding and originality to create new tests or get any meaning out of a test. But contemporary spreadsheet software will run a wide variety of statistical tests on any data that is typed into it. The process of doing (and graphing) the calculation is as mechanical as

the process of holding a camera in front of a painting, and, as the court said in the *Bridgeman* decision, the person who “pressed the trigger” or clicked on the calculate button has not necessarily made an original contribution. Whether the resulting numbers or graphs have any meaning certainly requires as much originality as interpreting a painting. The raw process of mechanical creation may not.

Database protection

Statistical analysis involves a compilation of data. If those data are plainly facts, the *Feist* ruling protects their public domain status in the US. The same is not necessarily true in Europe.

The European Database Directive of 1996 required member states to enact legislation protecting “a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means” (European Union, 1996). Database protection is somewhat different than copyright protection and is more limited in time.

A British court decision suggests that copying only part of a database could be an infringement, depending on the importance of the material, not just the amount:

On February 9, 2001, the High Court settled a case dealing with William Hill’s use of information listed in the databases of the British Horseracing Board (BHB). William Hill had published information about horses, jockeys and race lists for upcoming races on its website without BHB’s consent. In the High Court’s decision, Justice Laddie based the court’s ruling on the 1997 Copyright and Rights in Database Regulation in a novel way. The database legislation protects the rights of owners if a substantial portion of the database is copied. The Court held that William Hill had infringed on BHB’s copyrighted material protected by the Database legislation, based on the importance of the material and not on the specific amount of the copied information (Askanazi *et al.*, 2001).

This suggests that use of someone else’s compilation of data, even if that data is purely factual, could be a problem even if the process for analyzing it were considered as mechanical as the slavish copying in the *Bridgeman* decision.

Conclusion

In copyright decisions the details matter. Those who wonder whether they can post someone else’s graph of a statistical analysis on the web need to consider the following points:

- *Data* – how closely do the data resemble something that most people would consider to be an unoriginal fact that represented some form of objective reality? It seems likely that most people would consider an average body temperature a fact, even if on closer examination it turned out to be the result of prior scientific analysis. Most people might also consider a set of astronomical observations to be facts, despite the skill needed to get and interpret them. More people might be skeptical about a set of composite measures like the index of consumer confidence. The more unoriginal the data is, the safer its status as fact.
- *Analysis* – how mechanical was the analysis? Could multiple software packages produce the same result with a minimum of human intervention? It seems possible that the courts would regard a purely mechanical analytical process as lacking the necessary originality for protection. But the amount of thought that

had to go into the selection of data could be a factor in favor of its originality. For example, an graph showing the percentage of the world's production of cars that each country has might involve less originality than a *t*-test showing a confidence interval for a political poll. The former involved significant sweat of the brow for someone to assemble, but is making no choices about what to include or exclude. The latter involves careful choices about the nature of the sample and the population as a whole.

- *Venue* – a US venue (either in the form of a US-based server or a US publisher) would have less risk for the use of someone else's compilation of factual data than a venue within the European Union, where compilations of data have more protection.

Risk assessment has become a fact of life for many individuals and institutions making judgments about whether to use materials when the rights may belong to others. Those who want certainty need either to find the rights owner and get permission, or to create their own data and analysis.

Taking that low-risk route may please royalty-collecting organizations, but it also contributes to undermining that key principle in US law that the point of copyright is to promote the progress of science and the useful arts. Facts are supposed to be free for all to use and to manipulate. That right should not be neglected.

References

- 17 USC 101 (n.d.), United States Code, Title 17, Chapter 1, section 101, available at: www.copyright.gov/title17/92chap1.html#101 (accessed March 2006).
- 17 USC 102 (n.d.), United States Code, Title 17, Chapter 1, section 102, available at: www.copyright.gov/title17/92chap1.html#102 (accessed March 2006).
- Askanazi, J., Caplan, G., Descoteaux, D., Donohue, K., Glasser, D., Johnson, A. and Mena, E. (2001), "The future of database protection in US copyright law", in *Duke Law & Technology Review*, 22 May, available at: www.law.duke.edu/journals/dltr/articles/2001dltr0017.html
- European Union (1996), Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases. available at: http://europa.eu.int/ISPO/ecommerce/legal/documents/396L0009/396L0009_EN.doc
- Merriam-Webster Online Dictionary (2006), "Fact", available at: www.m-w.com/dictionary/fact
- Sayers, D. (1927), *Clouds of Witness*, Harper & Row, New York, NY, p. 68.
- US Supreme Court (1991), *Feist Publications Inc v. Rural Telephone Company*, March, available at: <http://supreme.justia.com/us/499/340/>
- US District Court for the Southern District of New York (1991), *Bridgeman Art Library v. Corel Corporation*, available at: www.law.cornell.edu/copyright/cases/36_FSupp2d_191.htm

About the author

Michael Seadle is Editor of *Library Hi Tech*. He is also Copyright Librarian and Assistant Director for Information Technology at Michigan State University. He is not a lawyer, and nothing in this column should be considered legal advice. Michael Seadle can be contacted at: Seadle@msu.edu

To purchase reprints of this article please e-mail: reprints@emeraldinsight.com
Or visit our web site for further details: www.emeraldinsight.com/reprints