INTERFACES: Getting Data, Using API's, or Stopping the Same

By: Paul H. Arne[1,2,3]

Introduction

Is it legal for a company to extract data from another company's website? Can a company use the other company's interface information[4] to extract or upload data or invoke the functionality of the other company's software? What if the other company objects?

Interactions between systems are largely governed by the same laws that are used to protect software: copyright law, trade secret/confidential information law, patent law, and contract rights. However, the interactions between computing systems have additional considerations— related to rights associated with interoperability, the "ownership" of data, and the invoking of functionality of another company's software or systems. There is also a new legal player in town: the Computer Fraud and Abuse Act ("CFAA").

This article addresses the following basic fact pattern:

- A company (the "Source") has certain data that it makes available on the Web.
- Another company (the "Recipient") wants to aggregate data from the Source for various business reasons.
- If the data identifies or is related to an individual person (the "Data Owner"), such as a member of a social media website or the owner of a bank account that has online access, the individual person has consented for the Recipient (by providing the person's ID and password, for example) to obtain and use the data available from the Source's website.
- If the Source's website is for subscribers only, but is otherwise available to a subscriber without charge, a subscriber has authorized the Recipient to access the Source's website on behalf of the subscriber.
- The data itself is not subject to copyright protection. No pictures, movies, etc.

---

[1] Copyright © Paul H. Arne, 2017. All rights reserved. Special thanks to Lindsey Gearhart and Joseph Wallace, Jr. for their research in preparing this article.

[2] Paul is the chair of the Technology Transactions Group of Morris, Manning & Martin, L.L.P.

[3] This article does not create an attorney/client relationship with you and does not provide specific legal advice to you or your company. Certain legal concepts have not been fully developed and certain legal issues have been stated as fact for which arguments can be made to the contrary, due to space constraints. It is provided for educational purposes only.

[4] This interface information is frequently called an "API" or "Application Programming Interface." Both in terms of case law and the general practice of law in this area, the author has found that different people ascribe different meanings to this term. Accordingly, in this article the information used to communicate with or extract data from a computing system will be called "interface information." Interface information can include programming commands, the syntax and parameters of programming commands and information received, as well as documentation that explains how to use other interface information to communicate with or extract data from a computer system.

- The Source does not want the Recipient to obtain the data.

This article shows that the type of technology used to obtain the data from the Source can impact the legal result. As with many technology law issues, the law is not well-developed. While addressing these issues from basic IP law principles and foundational case law is critical, recent case law adds important additional considerations. Regarding the CFAA, the case law is not yet clear, but the CFAA poses additional legal risks to the Recipient and another possible arrow in the legal quiver of the Source.

These issues arise fairly regularly in a technology law practice. Clients frequently want to interface with, retrieve data from, or provide data to computing systems of another party. Other clients want to prevent other parties from exchanging data or interfacing with their systems.

In practice, analyzing these issues frequently requires the analysis of interfaces at a highly technical level combined with the application of legal principles with ambiguous statutory law and limited case law available as guidance.

Data Extraction

How data is extracted from the Source can make a difference in the legal analysis. Because of this, there may be a need to understand the technical means by which the data will be obtained. This section describes the general ways that data is obtained from a web interface.[5]

Screen Scraping

Since the days of dumb terminals connected to mainframe computers, "screen scraping" has been used to gain information from a database. In many situations, data that is displayed on a computer screen is placed in a specific position on the screen. Because a computer screen consists of a number of pixels horizontally and vertically, it may be possible to extract the data by copying the text that occupies a specific place on a screen. This was easier when there were fewer choices in terms of screen resolution, when screens didn't scroll, and when the screen interface was textual rather than graphical. Nevertheless, screen scraping remains a well-established means of extracting data, although frequently an inexact way of doing so.

Screens can also be printed. Using PDF or other electronic document format, an electronic document can be created that reflects what is on a screen. A variant of screen scraping can then be used on the PDF document itself for data extraction purposes.

Screen scraping does not require the Source's web server to do anything special. All of the extraction of data occurs after the server has provided information sufficient to display that information in a web browser.

Web Scraping

With the advent of the World Wide Web, other means of obtaining data from a screen display—more specifically the coding that is used to display data on a screen—became available. Taking a simple example, data streams that are used by a web browser to display data on a screen use HTML, and possibly XML. Both HTML and XML use pairs of tags inserted into the data stream that identify the formatting, placement,[6] and in the case of XML, content of the data that exists between the two tags. "Web scraping," where the HTML/XML data streams are analyzed to extract data from them, is now

[5] Extraction methods typically utilized within a single company's computing environment are not addressed.

[6] For example, <bold> text </bold> uses an HTML tag that causes "**text**" to be bolded.

another means of extracting data. Use of HTML/XML tags is only one of many ways to use screen display data to obtain content.[7]

Similar to screen scraping, Web scraping technologies also do not depend on causing a web server to do anything special.

Using Interface Information

Another way of extracting data is to ask the web server for it. A Recipient uses interface information to issue commands (typically called an "API call") to the web server, which then delivers the requested information to the Recipient. Frequently, using API calls causes the Source's web server to provide information that is different from, or formatted differently than, a normal webpage request. It may therefore be considered to be a more invasive means of obtaining data than screen scraping or web scraping.

The Law Associated with Data Extraction[8]

Copyright Law

Extracting data from a database, or using interface information, without permission can touch on three basic concepts in copyright law.

The Idea-Expression Dichotomy

Section 102(b) of the US Copyright Act codifies the idea-expression dichotomy.

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

By virtue of this section, copyright protection does not seem to extend to ideas, procedures, processes, systems, methods of operation, etc.

From this doctrine, courts have held that facts[9] themselves and short phrases[10] are not protected by copyright law.

Legal Protection for Derivative Works

Section 103(b) of the US Copyright Act describes the scope of copyright protection for derivative works.

> The copyright in a compilation or derivative work extends only to the material contributed by the author of such work, as distinguished from the pre-existing material employed in the work, and does not imply any exclusive right in the pre-existing material.

---

[7] *See* generally, https://ahrefs.com/blog/web-scraping-for-marketers/.

[8] The CFAA will be addressed in later sections of this article.

[9] *See* Computer Associates Intern. Inc. v. Altai, Inc., 982 F.2d 693 (2d Cir. 1992), at 711.

[10] *See* Oracle America, Inc. v. Google Inc., 750 F.3d 1339 (Fed. Cir. 2014), at 1362.

If the interface information that is used for extraction was obtained by the Source from a third party, then use of that interface information does not give the Source the right to sue the Recipient for copyright infringement. Only the owner of the copyright has the right to sue for copyright infringement.

Fair Use

Section 107 of the US Copyright Act places limits on the exclusive rights granted to a copyright holder.

> Notwithstanding the provisions of sections 106[11] and 106A[12], the fair use of a copyrighted work … for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright. In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall include—
>
> (1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;
> (2) the nature of the copyrighted work;
> (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
> (4) the effect of the use upon the potential market for or value of the copyrighted work.

Importantly, Section 107 contains no explicit reference to the use of a copyrighted work for the purposes of extracting non-copyrighted material or for the purposes of allowing for interoperability among computing systems.

Patent Law

Patent law rarely applies in these situations, so this article does not address patent issues.

Trade Secrets and Confidential Information

All states provide protection for the misappropriation of trade secrets. Basically, a trade secret is information that is both valuable because it is not generally known and is the subject of reasonable efforts under the circumstances to maintain its secrecy.[13]

Confidential information protection is the contract counterpart to trade secret law. The parameters of scope and protection of confidential information is defined and mitigated through contract provisions.

Contract Law

One can agree to a contract that goes beyond the rights and responsibilities that exist under copyright and trade secret law. If one agrees to online terms of use, then those terms of use may be another source of rights that may be asserted by the Source.

---

[11] Section 106 sets forth the exclusive rights of a copyright holder.

[12] Section 106A sets forth the exclusive rights of a holder of moral rights.

[13] *See* Uniform Trade Secrets Act, Section 1. Found at http://www.uniformlaws.org/shared/docs/trade%20secrets/utsa_final_85.pdf.

Legal IP Analysis

Screen Scraping

Our hypothetical fact pattern assumes that the data itself is not subject to copyright protection. If the data is not copyrighted, what else is being taken from the Source's website when a Recipient uses screen scraping?

- The choices surrounding the actual location of the data on a screen?
- The choices of what data to provide on a screen?

As you can see, it is difficult to identify anything expressive, and therefore copyrighted, when data is taken using screen scraping. "Screen" scraping using the placement of data on a PDF does not seem to lend itself to copyright protection, either.

A Source could argue that the selection or order of the data elements is protected under copyright law. However, the selection and order of data elements is typically not important to a Recipient. The Recipient wants the data. Therefore, even if the selection or order of data elements are protected under copyright, the Recipient has a good argument that it is only using the copyrighted elements of the screen display as a means to get to the non-copyrighted data.[14]

Screen scraping publicly available information or where the Recipient has the permission of the Data Owner to access the information, is rarely impacted by trade secret law. Social media sites, sharing economy sites (such as, Uber), e-commerce sites (such as, Amazon or Ticketmaster), or bank websites, to name a few, typically do not have confidential information protection built into their terms of use.[15] Accordingly, trade secret law would not seem to apply, either.

Because screen scraping does not require a web server to do anything different from a normal request for a webpage, screen scraping is not particularly invasive of the Source's server.

Based on the analysis above, traditional screen scraping or reading data from PDFs is not likely to result in a copyright infringement. Trade secret protection is also unlikely, unless sufficient protections are taken to protect the screen displays themselves.

Not surprisingly, screen scraping of data seems to be a time-honored and legal way of extracting data from a Source's website,[16] at least from a copyright and trade secret perspective. Until recently, screen scraping seemed to be a relatively safe way of extracting data.[17]

Web Scraping

The use of web scraping methods and the use of interface information are potentially more problematical than the use of screen scraping to obtain data. Web scraping methods generally depend on information delivered to a browser to enable the browser screen to display properly. It can be readily seen in all modern browsers. To the extent that web scraping utilizes an XML schema in capturing the data, and if the particular XML schema is protected under copyright, trade secret, or other intellectual

---

[14] *See generally*, Sega v. Accolade, Inc., 977 F.2d 1510 (9th Cir. 1992).

[15] Protection of the data itself as confidential or subject to various privacy laws is a different issue that is not addressed in this article.

[16] *See, e.g.*, Ticketmaster Corp. v. Tickets.Com, Inc., No. 2000 WL 525390, at *2 (C.D. Cal. Mar. 27, 2000).

[17] *But see* the discussion of the CFAA, below.

property right, the copying of the XML schema as a part of web scraping may result in a copyright infringement or trade secret misappropriation.

However, while the risk of copyright protection for such information is greater than with screen scraping, under normal circumstances, the sequence of the HTML and XML coding, or other metadata,[18] is still not particularly likely to result in a successful copyright claim. Because it is readily viewable, its trade secret status is normally equivalent to the trade secret risk of using screen scraping methods.

Use of Interface Information

Interface information is not necessarily available in connection with analyzing screen displays. In addition, using interface information, such as API calls, can result in causing the web server to provide information that is different from what is available for a normal screen output of the web server. Generally speaking, the risk of trade secret protection for interface information is therefore greater than using screen scraping methods, especially if the interface information is not published by the Source. Depending on the original source of the interface information and the complexity of the interface information, there may also be a greater exposure to copyright infringement claims.

Similar to using XML schemas above, API calls that are sufficiently detailed may be protected by copyright law. Interface information may also be the subject of trade secret protection.

Because API calls are somewhat standardized, and because they are harder to keep confidential, the Source may still have considerable problems in successfully asserting a copyright infringement or trade secret misappropriation claim.

Unlike screen scraping and web scraping, the use of API calls can result in the Source's servers doing things that are different from a normal request for a webpage. Generally speaking, using API calls is therefore more invasive of the property of the Source that is screen scraping or web scraping.

General Intellectual Property Risk Conclusions

As can be seen from the above, screen scraping, Web scraping, and use of interface information form a continuum of increasing risk of IP claims. In each instance, the actual facts are critically important. At least from an intellectual property right standpoint, it may be perfectly acceptable to obtain data from a server using API calls, even if the potential legal risks are higher than the other means of data extraction.

The CFAA

The Computer Fraud and Abuse Act[19] was enacted in 1986, well-before the rise of the internet and the World Wide Web as mainstream phenomena. "Whoever… intentionally accesses a computer *without authorization* or *exceeds authorized access*, and thereby obtains… Information from any protected computer…" violates the CFAA. A "protected computer" includes any computer that is used in or affecting interstate commerce in the United States.

The CFAA is a criminal statute that also provides for civil remedies. Different courts have more broadly or more narrowly construed "without authorization" and "exceeds authorized access." Relatively recently, courts have started to apply the CFAA to activities that are closely associated with the normal activities that a user may do with a website. The most important case law developments seem to be in the 9th Circuit.

[18] Metadata is information about data. An example of metadata is that the text at the bottom of the pages of this article are footnotes. "Footnote" is a description of what the data is, not the data itself.

[19] 18 U.S.C. §1030.

U.S. v. Nosal[20]

U.S. v. Nosal was a criminal proceeding. This case is nevertheless important because the rationale used in this case has been followed, or not, in subsequent civil cases. After leaving his employment from Korn/Ferry International, an executive search firm, Nosal convinced three then-current employees of Korn/Ferry to extract confidential and proprietary information from Korn/Ferry's computing systems, apparently in preparation for competing against Korn/Ferry. Among other charges, Nosal was charged with violating the CFAA. Nosal defended, claiming that the employees who actually extracted the confidential information were authorized users of the Korn/Ferry systems, and that the CFAA "does not cover employees who misappropriate information or who violate contractual confidentiality agreements by using employer-own information in a manner inconsistent with those agreements."

The legal question became one of whether the employee accomplices of Nosal exceeded "authorized access" by gaining access for purposes that were not authorized. The court held that "'exceeds authorized access" in the CFAA is limited to violations of restrictions on *access* to information, and not restrictions on its *use*."[21] Accordingly, the access of data by the Korn/Ferry employees for the purposes of using it to provide such information to Nosal in violation of their obligations to Korn/Ferry was not a violation of the CFAA.

Craigslist v. 3Taps[22]

Hopefully, everyone who reads this article is familiar with Craigslist, a website that posts hundreds of millions of classified ads every year. An ID and password is not necessary to access the Craigslist website. 3Tap apparently used screen scraping to obtain data on the Craigslist website. Upon learning of this activity, Craigslist sent a cease-and-desist letter to 3Tap that 3Tap was no longer authorized to access the Craigslist website. In addition, Craigslist configured its website in a manner that blocked access to the site from 3Tap IP addresses. Through various means, 3Tap continued to access the Craigslist website and copy the Craigslist data. Craigslist sued.

In holding for Craigslist, the court reasoned that after receipt of the cease-and-desist letter, 3Tap was no longer *authorized* to access the Craigslist website. The court distinguished the ruling in *Nosal* by noting that the issue was not one of exceeding authorized access, as was the case in *Nosal*, but 3Tap was not authorized to access the Craigslist website for any purpose. In holding that Craigslist had the right to selectively deny access of 3Tap to a website that is otherwise freely accessible to the public, the court stated: "[A] meaningful distinction exists between restricting uses of a website for a certain purpose and selectively restricting access to a website altogether."[23]

Craigslist stands for the proposition that a website can deny the use of screen scraping by merely sending a cease-and-desist letter, revoking all access rights to the otherwise public website.

Facebook v. Power Ventures[24]

Power Ventures ("Power") had a business model of aggregating information on social media websites. In order to aggregate this information, powers received access information to Facebook and other social media sites, which allowed Power to aggregate the data. Power offered users of its services an opportunity to possibly win $100. For certain Power customers who agreed, Power utilized the functionality of Facebook itself to send messages to its customer's friends using Facebook's functionality.

[20] United States v. Nosal, 676 F.3d 854 (2012).

[21] Id. at 864.

[22] Craigslist, Inc. v. 3Taps Inc., 964 F.Supp.2d 1178 (2013).

[23] Id. at 1184.

[24] Facebook, Inc. v. Power Venture, Inc., 828 F.3d 1068 (9th Cir. 2016).

The "from" line of these email messages stated that it was "from Facebook," and the emails were signed "The Facebook Team."[25]

Upon learning of Power's activities, Facebook sent Power a cease-and-desist letter. Power's activities also violated various provisions of Facebook's Terms of Use. Facebook also attempted to block Power's IP addresses. Despite these actions, Power continued to access the Facebook website.

This case differs from *Craigslist* in at least two important ways. First, Power was granted permission by Facebook users to access the Facebook account on their behalf. Second, Power was not screen scraping. Power actually utilized the functionality of Facebook to send email messages to other Facebook members.

The Ninth Circuit held that obtaining Facebook user consent was not sufficient. Facebook's consent was also required.

> An analogy from the physical world may help to illustrate why this is so. Suppose that a person wants to borrow a friend's jewelry that is held in a safe deposit box at a bank. The friend gives permission for the person to access the safe deposit box and lends him a key. Upon receiving the key, though, the person decides to visit the bank while carrying a shotgun. The bank ejects the person from its premises and bans his reentry. The gun-toting jewelry borrower could not then reenter the bank, claiming that access to the safe deposit box gave him authority to stride about the bank's property while armed. In other words, to access the safe deposit box, the person needs permission *both* from his friend (who controls access to the safe) and from the bank (which controls access to its premises). (Emphasis supplied) [26]

While the Ninth Circuit did not address this issue, would the ruling have changed if the Facebook users information was what Power extracted rather than Power's using of Facebook's functionality? The reasoning of the case suggests that this would not make a difference. However, it is certainly possible that a Facebook users authorizing a third party to extract the user's information was not a factual scenario that the court considered in its decision.

Power's application for a *writ of certiorari* to the Supreme Court was denied a week before this article was completed.

hiQ Labs v. LinkedIn[27]

The most recent case in this group was decided in August of this year. hiQ is in the business of providing information to companies about the prospects of their employees based upon a statistical analysis of publicly available information. Much of the information obtained by hiQ came from users of LinkedIn. hiQ only accessed those LinkedIn customers who set their privacy settings in a way that the entire public had access to that information.

LinkedIn sent hiQ a cease-and-desist letter, demanding that hiQ cease unauthorized data scraping, which was prohibited by LinkedIn's terms of use, as well as other violations of those terms of use. hiQ then filed suit against LinkedIn, seeking to enjoin LinkedIn from preventing hiQ's access to the website.

---

[25] Id. at 1073.

[26] Id. at 1078.

[27] Case: 3:17-cv-03301-EMC (filed 8/14/17).

The District Court's decision was in connection with a motion for a plenary injunction brought by hiQ. The court identified the key issue regarding the CFAA to be "whether, by continuing to access public LinkedIn profiles after LinkedIn has explicitly revoked permission to do so, hiQ has 'accesse[d] a computer without authorization' within the meaning of the CFAA." In granting the primary injunction, the District Court found that it was likely that the CFAA was not violated by hiQ's actions. While the court was aware of the decision in *Craigslist* it apparently did not consider that ruling dispositive. The court held that the CFAA "was not intended to police traffic to publicly available websites on the Internet – the Internet did not exist in 1984. The CFAA was intended instead to deal with 'hacking' or 'trespass' onto private, often password-protected mainframe computers."

Analysis

The facts in both the *Craigslist* and *hiQ* cases involved:

- access to websites that were not password-protected;
- the sending of a cease-and-desist letter revoking authorization to access the website;
- the purported violation of the Terms of Use of the website; and
- screen scraping.

Yet the opposite results were reached in these cases.

Based upon these cases, it is not clear what the effect of the CFAA has on data collection from websites.

Conclusion

In the battle between sources of data and those who would like to otherwise exploit that data, the CFAA is a relatively new tactical weapon. Not surprisingly, many sources of data have modified their terms of use and attempted to prevent companies from accessing "their" data, even when screen scraping is used, even when the "owner" of the data has authorized its collection and use, and even when a traditional IP analysis of the facts would not suggest a right to preclude the Recipient's use.

The factual parameters of when the CFAA is or is not violated are currently unclear. The rise of the CFAA as a litigation tool has resulted in the sources of data being in a better position to argue that their data are in fact "owned" by those data aggregators, even when made publicly available.

Businesses are better able to compete when the rules are clear. Hopefully, the applicability of the CFAA to these fact patterns, especially given that it is a criminal statute, will become clearer in the next few years.