Cutting Edge Issues in IP Ownership: Big Data and New Technologies

Sponsored by BakerHostetler LLP as part of the ACC High Tech Summit

June 1, 2016
Big Data: Who owns what?

• Where did we get the data?
• What can we do with the data?
• How will we control ownership of the analytics?
• (Should we control ownership?)
What Are The Concerns?

- Licenses
- Contracts
- Customer expectations
- Shareholder expectations
- Regulatory environment
- Media environment
- WILL WE GET SUED?
What is Big Data?

- Massive amounts of data that are difficult to analyze and handle using common database management tools
- Big data includes:
  - Raw data points
  - Database
  - Analytic tools
  - Output / analysis
What is the IP?

- Is there copyright in the raw data?
- What rights in the databases?
- Who owns the tools used to analyze the data?
- Who owns the analytic reports?
- Who owns the insights gained from the reports?
Why Is This Important?

• Big Data spending to grow from $28.6 B to $66.7 B by 2021.

• Use of Big Data to drive innovation and “reduce the barriers between man and machine” has been dubbed the “Fourth Industrial Revolution.”
Where Does Raw Data Come From?

- Music, books, conversations, photos, videos, social media
- Trains, planes, automobiles, office buildings
- Smart watches, smart houses (microwave, thermostat, etc.), smart cars
What are Uses of Big Data?

• Telecom – improve network performance and customer experience; subscriber analytics and segmentation
• Banking – fraud detection, increase personalized offerings
• Media – clickstream analysis, social/customer profile analysis, loyalty programs
• Manufacturing – inventory management, preventive maintenance and repairs, transportation efficiency
• Healthcare – monitoring for diseases, improving patient experience, stocking supplies
Hadoop and MapReduce

- Distributed file system (HDFS)
- Data processing tool (MapReduce)
- Open source
- Proprietary add-ons
MapReduce

The overall MapReduce word count process

Input
- Deer Bear River
- Car Car River
- Deer Car Bear

Splitting
- Deer Bear River
- Car Car River
- Deer Car Bear

Mapping
- Deer, 1
- Bear, 1
- River, 1
- Car, 1
- Car, 1
- Car, 1
- Deer, 1
- Deer, 1
- Car, 1
- Bear, 1
- River, 1
- River, 1

Shuffling
- Bear, 1
- Bear, 1
- Car, 1
- Car, 1
- Car, 1
- Deer, 1
- Deer, 1
- Car, 1
- Bear, 1
- River, 1
- River, 1

Reducing
- Bear, 2
- Car, 3
- Deer, 2
- River, 2

Final result
- Bear, 2
- Car, 3
- Deer, 2
- River, 2
Many pieces to the puzzle
Many approaches to data “ownership”

- US vs. EU
- Life Sciences vs. Technology
- Public vs. Private
- Open source vs. proprietary
Hold On To What You Got

• Interest in retaining rights to content and process have led to some interesting license arrangements.

• Consider the below from Cloudera:

5. Ownership. As between Cloudera and Customer and subject to the grants under this Agreement, Cloudera owns all right, title and interest in and to: (a) the Product (including, but not limited to, any modifications thereto or derivative works thereof); (b) all ideas, inventions, discoveries, improvements, information, creative works and any other works discovered, prepared or developed by Cloudera in the course of or resulting from the provision of any services under this Agreement; and (c) any and all Intellectual Property Rights embodied in the foregoing. For the purpose of this Agreement, “Intellectual Property Rights” means any and all patents, copyrights, moral rights, trademarks, trade secrets and any other form of intellectual property rights recognized in any jurisdiction, including applications and registrations for any of the foregoing. As between the parties and subject to the terms and conditions of this Agreement, Customer owns all right, title and interest in and to the data generated by the use of the Products by Customer. There are no implied licenses in this Agreement, and Cloudera reserves all rights not expressly granted under this Agreement. No licenses are granted by Cloudera to Customer under this Agreement, whether by implication, estoppels or otherwise, except as expressly set forth in this Agreement.

Hold On To What You Got

• Note that Cloudera owns:
  – Modifications and derivative works of its Product;
  – all ideas, inventions, discoveries, improvements, information, creative works and any other works discovered, prepared or developed by Cloudera in the course of or resulting from the provision of any services under this Agreement
  – any and all Intellectual Property Rights [including] patents, copyrights, moral rights, trademarks, trade secrets

• You own:
  – data generated by the use of the Products

• Does “data” include IP or exclude it?
Data Just Wants to be Free

• Hague Declaration on Knowledge and Discovery in the Digital Age

“Intellectual property was not designed to regulate the free flow of facts and ideas, but has as a key objective the promotion of research activity; ... Licenses and contract terms should not restrict individuals from using facts, data and ideas.”
What Is the IP?

• “Raw Data” to the extent purely factual should not be subject to any copyright protection
  – "Raw facts may be copied at will."
  – “[O]riginality, not “sweat of the brow,” is the touchstone of copyright protection in directories and other fact-based works.

• Original selection, coordination and arrangement of the facts may be protected.
What Is The IP?

- Databases generally have some copyright protection in the US
  - Per the Compendium sec. 1002.6—"[A] database generally consists of two fundamental elements: (i) a data set or multiple data sets, and (ii) an information retrieval program or system that serves as the sole entry point into the underlying data, information, or files. Typically, the party that created the information retrieval program or system is not the same party that created the copyrightable content contained within the database. An application to register a database typically covers the selection, coordination and/or arrangement of data, information or files, but does not cover the data, information or files unless they are specifically claimed in the application."
What Is the IP?

- Databases generally have copyright protection
  - Per case law: “[Although] factual information generally accessible to the public is not protected by copyright law, the compilation of those facts may be copyrightable.” *Gemel Precision Tool, Co. v. Pharma Tool Corp.*, 1995 U.S. Dist. LEXIS 2093 (E.D. Pa. Feb. 13, 1995) (citing *Feist*).
What Is the IP?

The EU Approach

- Right subsists in a database if there has been “substantial investment in obtaining, verifying or presenting the contents of the database”

EU Directive 96/9/EC on the Legal Protection of Databases
Joint Authorship Issues

• Ordinarily one must make an independently copyrightable contribution to be a joint author.

• For example, merely supplying the idea for a software program does not make one a joint author. *Woods v. Resnick*, 725 F. Supp. 2d 809 (W.D. Wis. 2010)

• However, in rare circumstances courts find that collaborative processes can lead to joint authorship through mere idea submission. *Gaiman v. McFarlane*, 360 F.3d 644, 658-59 (7th Cir. 2004)
New Works or Derivations?

- Derived data - Copyright owner has exclusive right to create a derivative work.
- Transformative work – Altering the original with new expression, meaning or message.
- Are insights from data transformative?
Data as Trade Secret

• Customer lists, software architecture, database of client preferences, etc.

*Computer Associates Int’l, Inc. v. American Fundware, Inc.*, 831 F. Supp. 1516 (D. Colo. 1993) (common law hired to invent doctrine applied to ownership of accounting software developed for hiring company by independent contractor and treated as trade secret)
Other U.S. approaches to protect data


Patentability?

• Patentability of algorithms, methods of analysis being called into question with recent Supreme Court decisions.

• Is identification of patterns in data patentable? What about the insights derived from the patterns?
Still Much To Be Done

• 80% of enterprise data is “unstructured”
• Sometimes called “dark data”, it refers to the data that is unanalyzed, unformatted, or sometimes not collected.
• This does not mean that it is “uncopyrightable”, just that it does not conform to specific, pre-defined data models
• Can include:
  – Raw survey inputs
  – Customer data
  – Previous employee data
  – Financial statements
  – Email conversations
  – Chat transcripts
  – Call center transcripts
  – Account data
Which Means….

- Enterprise clients will need help in deciding what data should be analyzed and in the analysis of that data.
Many Players, Many Hands

• There are a vast array of consultants, analysts, software providers, and open source solutions that manage, format, repackage, analyze, and provide advice on such data.

• For example Gartner predicts that by 2017, more than 30% of enterprise access to broadly based big data will be via intermediary data broker services

Use Case

• Assume
  – hospital A contracts with EMR company B to format, de-identify and store raw medical claim data;
  – Hospital A sends formatted data to big data SAS platform company C for further formatting to use Hadoop protocol;
  – Hospital A also contracts with big data analytics company D to create analysis tools for MapReduce optimization;
  – Hospital A asks population health analytics firm E to spot trends in de-identified data with resulting data set
Use Case

• Hospital A– no underlying copyright protection for raw data supplied with no selection or arrangement

• Trade secret protection?
Use Case

- EMR Company B
  - Copyright protection for formatting of data?
  - What degree of direction and/or expression from Hospital A in choosing data fields?
Use Case

• Big Data SAS Platform Company C
  – Does reformatting data for use in Hadoop create new expression?
  – Patent ownership issues?
Use Case

• Big Data Analytics Company D
  – Does reformatting data again to meet MapReduce process create new expression?
  – Patent ownership issues?
Use Case

• Population Health Analytics Firm E
  – Aggregation of data with data sets from other clients create new protectable data base?
  – Derivative work?
A Useful Framework?

• Access vs. Ownership
• Licensing as the solution
A Useful Framework?

- IP
- Regulatory
- Contract
Regulatory and other issues

- Contractual
  - Ownership
  - Use
  - Confidentiality
- Privacy
  - Protected data types
  - Applicable privacy policies
  - Regulatory and self-regulatory
- De-identification
- Security and breach
Panel Discussion

• Peter Lam, Intellectual Property Counsel, General Electric Company (GE Power)
• Brian Endter, Senior Counsel, Concur Technologies, an SAP company
Authorship and Artificial Intelligence

• In 2014 the Copyright Office released the Third Edition of the Compendium, which is “the governing administrative manual for registrations and recordations issued by the Office on or after that date.”

• It addressed “fundamental principles of copyright law,” including who can be an author and what kinds of content are copyrightable.
Authorship and Artificial Intelligence

• One issue addressed in the Compendium was the “Human Authorship Requirement”

• This section 306 identified several examples of content created by non-humans that would not be subject to copyright

• Examples include:
Is AI Just a Dumb Ape?

Examples:

- A photograph taken by a monkey.
Is AI Just a Dumb Ape?

- A mural painted by an elephant.
Is AI Just a Dumb Ape?

Similarly, the Office will not register works produced by a machine or mere mechanical process that operates randomly or automatically without any creative input or intervention from a human author.
How Did We Get Here?

• Commission on New Technological Uses of Copyrighted Works (CONTU)
  – Commissioned in 1974 to study, among other things, copyrightability of software for 1976 Copyright Act
  – Not completed until 1978
  – Definition of “computer program” added in 1980
How Did We Get Here?

• “The development of this capacity for ‘artificial intelligence’ has not yet come to pass, and, indeed, it has been suggested to this Commission that such a development is too speculative to consider at this time.”

CONTU Final Report, at p. 45
The Future Is Now-ish

GIZMODO

Artificial Neural Networks Can Day Dream—Here's What They See

http://gizmodo.com/these-are-the-incredible-day-dreams-of-artificial-neura-1712226908
The Future Is Now-ish

- Advanced Natural Language Generator (NLG)
- Narrative Science Story Builder

Dominion worked with Narrative Science to create a vehicle description product called StoryBuilder. StoryBuilder takes individual vehicle data, such as make, model, year, mileage, color and features, and pulls together a short, consumer-focused story describing the vehicle. Based on the intelligent strength of Quill and Dominion's ability to expand existing data resources, new content is continually added to StoryBuilder. Dominion captures new requirements based on direct customer feedback and uses it to improve the content, so it becomes more personalized and relevant for readers. Given Quill's flexibility, it can be easily customized to meet these specific needs and generate large amounts of output, resulting in individualized stories about each vehicle.

http://resources.narrativescience.com/h/i/78223077-case-study
The Future Is Now-ish

“A Series Of Unfortunate Tech Predictions - Artificial Intelligence and IOT are inseparable

- Big data analytics for IOT software revenues will experience strong growth, reaching $81 billion by 2022 says Strategy Analytics
- Smart Cities will use 1.6 billion connected things in 2016 says Gartner
- By 2025 IOT will be a $1.6 trillion opportunity in Healthcare alone says McKinsey
- 50 billion+ connected devices will exist by 2020 says Cisco
- Data captured by IOT connected devices will top 1.6 zettabytes in 2020 says ABI Research
- There are 10 major factions fighting to become the interoperating standard for IOT

Timeline of Studies and Amendments

- 1973 AI Winter
- 1974 CONTU Commissioned
- 1976 Copyright Act
- 1978 CONTU Report
- 1980 Computer Software Copyright Act
- 1986 OTA IP in an Age of Electronics and Information
- 1998 Digital Millennium Copyright Act
- 2013-2015 Hearings on “Next Great Copyright Act”
Not Just a U.S. Problem

- An Australian case illustrates the reality of the problem and consequences
  - Employees used program that auto-generated Material Safety Data Sheets for hazardous materials
  - The MSDS was generated by pulling from materials and regulatory information loaded in a database and rendering the MSDS in HTML
Not Just a U.S. Problem

• Court found no copyrightable expression in HTML code generated by program—
  – It was not as though the authors … having in mind the source code they desired to write, used the computer to that end. They were not computer programmers, and there is no suggestion that they either understood source code or ever had a perception of the body of source code which was relevant to the MSDSs on which they worked
But Who Really Is the Author?

- There are at least five AI output ownership allocation possibilities:
  - computer,
  - the user,
  - the author of the generator program,
  - both jointly,
  - or no one.

Who Is an Author?

• “There has to be some original expression contributed by anyone who claims to be a co-author, and the rule … is that his contribution must be independently copyrightable”

  Gaiman v. McFarlane, 360 F.3d 644, 658 (7th Cir. Wis. 2004)

• "Originality merely requires independent creation by the author and just a scintilla of creativity."

Programmer?

- **Ownership in programmer?**
  - Programmer sets parameters in which AI functions
  - Similar to the video game cases:
    - The player of a video game does not have control over the sequence of images that appears on the video game screen. He cannot create any sequence he wants out of the images stored on the game's circuit boards. The most he can do is choose one of the limited number of sequences the game allows him to choose. He is unlike a writer or a painter because the video game in effect writes the sentences and paints the painting for him; he merely chooses one of the sentences stored in its memory, one of the paintings stored in its collection.

Midway Mfg. Co. v. Artic Int'l, Inc., 704 F.2d 1009, 1012 (7th Cir. Ill. 1983)
Programmer?

- Ownership in Programmer
  - Are the AI’s outputs a derivative work of the code?
  - But where is programmer’s creative expression in the output?
  - Has the programmer just created a tool?
User?

- Ownership in User also compelling:
  - Deploys AI similar to use of other devices
    - Videographer of a live performance
    - AI simply a tool that creates new, unique expression
  - Could the work-for-hire doctrine extend to the end-user’s use of AI?
  - Could the user and AI be joint authors similar to the “psychographic cases”? *(See, e.g., Cummins v. Bond, 1 Ch. 167 (1927); Penguin Books U.S.A., Inc. v. New Christian Church of Full Endeavor, Ltd., No. 96 Civ. 4126, 2000 U.S. Dist. LEXIS 10394 (S.D.N.Y. July 21, 2000); Urantia Found. v. Maaherra, 114 F.3d 955, 958 (9th Cir. 1997)*
Programmer and User?

• Ownership in Both Programmer and User?
  – Is AI like the creation of a comic book character, warranting special rules?
  – “The decisions that say… that each contributor to a joint work must make a contribution that if it stood alone would be copyrightable weren't thinking of the case in which it couldn't stand alone because of the nature of the particular creative process that had produced it.”

  Gaiman v. McFarlane, 360 F.3d 644, 658 (7th Cir. Wis. 2004)

• Avoid “peeling the onion until it disappeared.”
A Sixth Ownership Possibility?

- 9(3) of the UK Copyright, Designs and Patents Act 1988 Act provides that the author of a computer-generated work ‘shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken.’
A Sixth Ownership Possibility?

- **Nova Productions Ltd v Mazooma Games Ltd**, [2006] RPC 379
  - Video game designer was deemed author of various screens.
  - Player’s input was “not artistic in nature and he has contributed no skill or labour of an artistic kind... All he has done is to play the game
Is the US Behind?

- Provisions similar to the UK Act are in statutes in Hong Kong, New Zealand, South Africa, and India
- On May 9th Japan approved the Intellectual Property Promotion Plan 2016
  - “This plan will revise the copyright system to make it possible to use copyrighted works under set conditions in order to promote the collection and usage of big data toward a fourth industrial revolution. Moving forward, we will continue to consider the extent of intellectual property rights and for whom they will be granted, in relation to creative works such as music and novels produced by artificial intelligence.”

http://japan.kantei.go.jp/97_abe/actions/201605/09article3.html
Opportunity to Amend?

- In 2013 the Register of Copyrights called for Congress to consider enacting the “Next Great Copyright Act”
- Hearings held throughout 2013-2015
- Stalled but not abandoned
Thank You

Gina Culbert, rculbert@bakerlaw.com
Chad Rutkowski, crutkowski@bakerlaw.com
Alan Friel, afriel@bakerlaw.com