**Jane Bambauer, *Is Data Speech?*, 66 STAN. L. REV. ____ (forthcoming) (excerpts)**

## I. WHAT IS "DATA," AND WHAT IS "SPEECH"?

This section addresses two important definitional questions. Since the terms "data" and "speech" invite multiple meanings, defining them with precision here will focus and constrain the rest of the discussion.

### A. Data

When this article uses the term "data," it means a fixed record of a fact.

***

This article will use "information" to mean any objective representation of something that has occurred—representation of a fact.[24] It need not be fixed. A man yelling "timber" has momentarily created information that a tree is falling, though it will soon dissolve into memory and the air. It also need not be manmade, as a fossil can represent facts as clearly as any drawing or description of the original specimen.

This is an expansive definition, but it is constrained, importantly, by factual representation. Utterances sometimes convey facts ("timber!"), but not always. Thus, when a person yells "Republicans suck" or "Beauty is truth, truth beauty," they are not representing a fact and, thus, are not uttering information about Republicans or beauty. These are opinions, ideas, hypotheses, and artistic expressions.[25] Information reports only something that has occurred.

The scope of this article is narrower still. We will investigate the constitutional treatment of data—*information that somebody has deliberately caused to be captured and recorded into a fixed, man-made format*. Though we will spend much of our energy thinking about electronic data that has recorded some type of event—the time and details of a file transfer, or the time and location of a cell tower ping—data does not have to be electronic. Handcollected tallies, cash register tapes, and photographic portraits are data. Even parts of Keats' Ode to a Grecian Urn are data to the extent it describes an actual urn.

The article confines its investigation to data for two reasons. First, manmade data has more similarities to traditional speaker-listener arrangements than other types of information (though there are some limits to the similarities), so it might have a more promising claim to the First Amendment's protections. Second, unlike other forms of information, the quantity of data has exploded in the digital age.[26] The sudden change in the status quo makes data a likely candidate for policy debates and regulation. In other words, this is the information that governments worry about.

The article will also refer to records (as in "business records" or "medical records.") These are data created with the specific purpose of preserving information. Thus, they can be distinguished from "functional data", such as server data, that is produced in order to carry out a function or service.[27]

Functional data is data too; it came to be because browser programs or other software were deliberately coded to create the data. However, they are conceptually distinct from records because they were not necessarily created for the purpose of preserving information. These machine-to-machine communications were originally produced to carry out the steps necessary for web browsing. Though the data turn out to be very valuable for their information content, they were not always originally created with that purpose in mind. So, records and functional data are distinct subsets of data. Many scholars, myself included, have homed in on the distinction between functional data and other types of records as a useful boundary for law.[28] But as this Article explains, First Amendment doctrine should not capitalize on this distinction.

***

### B. Speech

Equally important is the definition of "speech." The goal of this Article is to understand whether

regulations on the collection or transfer of data implicate the First Amendment, requiring the government to justify and narrowly tailor the regulation. It is not particularly important whether First Amendment interests are raised under the banner of "speaker rights", "listener rights", or under some other concept, such as the semi-developed "right to receive information" (discussed at length later in the article). Ultimately the article seeks to understand whether legal restraints on the creation or dissemination of data draw First Amendment scrutiny on any basis.

Because some forms of data are already very familiar forms of expression—maps, almanacs, photographs and the like—it is quite natural to refer to these data as "speech." So this article will use the term "speech" as convenient shorthand for First Amendment protection (under the speech clause) on any basis. For our purposes, if data is constitutionally protected at all, it is "speech."
***
Granted, the Supreme Court's choice to distinguish not only between speech and conduct (which, while difficult, is necessary) but also between protected and unprotected speech breeds some uncertainty about whether classifying something as "speech" automatically means that it receives judicial protection.33 Since courts are inconsistent in defining the boundary between protected and unprotected speech, Schauer and Richards are justified in casting doubt on whether First Amendment scrutiny applies by default.
***
No account of the First Amendment treatment of data, whether normative or descriptive, can be complete without some reflection on the qualities of data—qualities that make it at once similar and very different from other more familiar forms of expression.
***
Data communicates. It tells a narrative just as effectively as prose, imagery, and music, to those with the training to interpret it. Its style is dry, but this does not interfere with its ability to light up the mind. A database can be interpreted directly by a person with the help of a codebook, and it can also be translated into other, more familiar forms of expression like maps, charts, graphs, and descriptive sentences. Lest there be any doubt about data's intimate connection to other forms of expression, one may recall that the very first form of writing were data: the accounting records of traders in ancient Mesopotamia.46 Data were the building blocks of the rest of written language.

Modern data collection practices raise uniquely difficult constitutional linedrawing problems. When data is created to inform a broad public audience, or even to inform a specific audience, its conceptual similarities to pamphlets and other types of traditional speech are obvious. But when it is created without the intent to inform anybody on anything in particular, the parallels to traditional speech begin to break down.

***

## II. DATA IN FIRST AMENDMENT PRECEDENT

The descriptive analysis of the First Amendment case law as it relates to data runs into contradictions. When data already exists, and is transmitted from one person or entity to another, the case law strongly supports the conclusion that raw facts are protected by the First Amendment.
***
One might assume that the creation of data tends to receive protection as well, since the collection of data is a necessary precursor to having and sharing it. Indeed, some cases suggest this is so, but at present the authorities are in conflict. A long strand of privacy case law confidently concludes that data creation is conduct insufficiently expressive to garner First Amendment protection.
***

### A. Existing Data

Last year, privacy advocates and data aggregators watched anxiously as the Supreme Court decided Sorrell v. IMS Health.47 Justice Kennedy's opinion very nearly resolved the question whether data is

protected speech. Big Data, held by Big Pharma, was at the center of the case. The state of Vermont passed legislation prohibiting pharmaceutical companies from receiving and using prescription data to customize their advertising to doctors. IMS Health, one of the data aggregators that preferred to sell their prescription data to the pharmaceutical companies, brought a First Amendment challenge. Six of the U.S. Supreme Court Justices agreed the law was unconstitutional. Part of the opinion suggested that the restriction on transfers of data between willing givers and receivers was automatically a restriction of speech. Data are facts, after all, and facts are "the beginning point of much constitutionally protected expression." According to Kennedy, strict scrutiny "arguably" applied on that basis alone. But in the end, Kennedy opted to resolve the case on narrower grounds.[48]

Why couldn't Justice Kennedy pull the trigger and declare that data gets speech protection?[49] Kennedy no doubt had misgivings about the broad and unanticipated consequences that such a declaration might bring about.

Justice Kennedy's reservations notwithstanding, courts have shown they are quite prepared to apply the First Amendment to already existing raw information that one individual or entity wishes to share with another. Prior to IMS, the cases addressing the issue have insisted that raw information is speech without hesitation. Occasionally they have done much more, recommending that free speech rights protect uninhibited analysis and access to information, along with the freedom to disseminate it.

This subpart catalogs the cases in three areas of law where the constitutional protection of data-sharing has naturally had reason to surface: privacy, advertising, and copyright. The results across each of the disciplines consistently protective the free communication of data.

**1. Gleanings From Privacy Law**

The Court wrote on a fairly clean slate when it decided Sorrell v. IMS.[50] Very few cases have raised First Amendment challenges to data privacy statutes. But the challenges that have been brought have concluded unequivocally that the communication of raw data is speech.

In Trans-Union Corporation v. Federal Trade Commission, a credit reporting agency challenged the constitutionality of the Fair Credit Reporting Act, which forbids companies from sharing consumer credit reports except for a specified list of purposes.[51] The court never doubted that the consumer reports were speech. Relying on earlier Supreme Court precedent, the D.C. Circuit found that consumer reports received protection, but warranted more lenient intermediate scrutiny as commercial speech.[52] Existing business records receive First Amendment protection of some form.

The Tenth Circuit case U.S. West v. FCC goes further still.[53] U.S. West also addressed a First Amendment challenge to a privacy regulation—section 222 of the Telecommunications Act of 1996, which restricts telecommunications providers from disclosing or using customer data outside of a narrow set of purposes. The telecommunications data at issue in U.S. West included any "information that relates to the quantity, technical configuration, type, destination, and amount of use of a telecommunications service subscribed to by any customer[]."[54] This definition is expansive. It includes not only the sorts of information that the telecommunications provider records in its bills and other business records (though billing information is certainly covered[55]), but also data describing when, where, and to whom a customer places a call. This is information generated in the process, and for the purpose, of providing telecommunications service. It is, in other words, functional data.[56]

The Court found that Section 222 infringed on the telecommunications service providers First Amendment rights. But it did not do so by concluding that the customer data itself was speech. Instead, the court located the First Amendment interest in the communications that U.S. West intended to make to its customers with the aid of the data. The First Amendment attached to the advertising U.S. West wished to tailor to their customers' usage habits. So technically, the court never identified the data as a source of First Amendment rights. Instead the court focused its analysis on the indirect

burden imposed on U.S. West's traditional commercial speech.[57]

But if the ultimate question is whether data receives First Amendment protection, the distinction is inconsequential. If Section 122 infringed on U.S. West's right to advertise to its customers, it did so only indirectly through its effect on data. The fact that the court treated the indirect burden as a First Amendment problem reveals that data has a sufficiently close relationship to free expression to merit derivative protection.

***

**B. Data Creation as Nonexpressive Conduct**

***

First Amendment rules on the creation of data can be explored through the court's treatment of documentary photographs, since they are a familiar subset of data.

***

(S)ome courts have justified treating photography as nonexpressive conduct because it can be done without an audience in mind. Indeed, security cameras create images without the involvement of anybody resembling an author. Arguably, the First Amendment should require, at the very least, somebody we can identify as a speaker, and somebody else who could be an intended audience. As one court puts it:

> To achieve First Amendment protection, a plaintiff must show that he possessed: (1) a message to be communicated, and (2) an audience to receive
> that message, regardless of the medium in which the message is to be expressed.[97]

Courts have used tests like this one to find that photographers have no First Amendment right to take photographs or videos in public places. In Porat v. Lincoln Towers Community Association [98] , a man who was ticketed for photographing a public courtyard lost his Section 1983 claim, based on his First Amendment rights. He lost because he had described himself as a "photo hobbyist," a phrase that unexpectedly doomed his case. According to the district court, "he effectively disclaims any communicative property of his photography as well as any intended audience by describing himself as a 'photo hobbyist,' and alleges that the photographs were only intended for 'aesthetic and recreational purposes."[99]

In contrast, in Pomykacz v. Borough of West Wildwood, a woman who followed around the town's mayor photographing him, and whose conduct satisfied the elements of criminal harassment, successfully defended herself on First Amendment grounds because she claimed her photographs were part of her political activism, and would be used to corroborate her theories that the mayor engaged in nepotism (if, indeed, these theories were borne out.)[100]

These cases certainly craft a clear enough rule, and they just as certainly train future photographers about what to say during a deposition (claim to be an activist, never a hobbyist), but should the First Amendment shield only speakers who have audiences from government interference?

***

In any event, even if protected speech requires a speaker and an audience, photography bans ought to attract scrutiny by their nature. The very purpose of a photography ban is to prevent a wider audience from seeing the scene. With the exception of bans on flash photography (which are usually designed to prevent the deleterious effects of light on sensitive objects), photo bans are designed to cut down on the communicative potential.[103]

In time, the rule that mechanically capturing information is nonexpressive conduct will prove to be unworkable. All-party wiretap statutes have already begun to fall apart under the weight of increased judicial scrutiny. These cases and others like it are discussed next.

**C. Data Creation as Expressive Conduct**

Over the last decade several state prosecutors, exercising exceedingly poor judgment, have chosen to bring criminal charges against citizens who recorded the conduct of on-duty police.[104] Some (though not all) of the courts dismissed the charges based on a First Amendment right to record. However, with one exception, the right was crafted narrowly, as a right to record public officials performing their public duties. [105] The Department of Justice has written opinion letters recognizing the right on narrower terms still—protecting a person's right to record "police activity."[106]

But last year, in ACLU v. Alvarez[107], the Seventh Circuit enjoined the enforcement of Illinois' all-party wiretap statute out of recognition to a broader right to record. To record anything.[108] The majority (over a thoughtful dissent by Richard Posner) arrived at their position because a well-working First Amendment must unite speech and the creation of speech:

> The act of making an audio or audiovisual recording is necessarily included within the First Amendment's guarantee of speech and press rights as a corollary of the right to disseminate the resulting recording. [] There is no fixed First Amendment line between the act of creating speech and the speech itself. [] This observation holds true when the expressive medium is mechanical rather than manual. [] Audio and audiovisual recording are communication technologies, and as such, they enable speech. Criminalizing all nonconsensual audio recording necessarily limits the information that might later be published or broadcast — whether to the general public or to a single family member or friend — and thus burdens First Amendment rights.

Other courts have inched toward the same position by advancing a right to access information. In S.H.A.R.K. v. Metro Parks Serving Summit County[109], an animal rights group brought a civil rights claim against a county for (accidentally) disposing their video footage of deer culling in a state park. The activists had set up video cameras in trees to record during hours that the park was closed to visitors. The Sixth Circuit found that the animal rights activists had a First Amendment interest in recording the deer culling, and more generally, in "accessing information". However, the right was differentiated from, and given significantly less scrutiny than, the right to expression.[110]

***

Though the Supreme Court has not directly addressed the question whether mechanical recordings are protected speech, it has recognized that information-gathering is a necessity for speech. In Branzburg v. Hayes, although the Court found that neutrally applicable laws apply to the press, it recognized that "without some protection for seeking out the news, freedom of the press could be eviscerated."[111]

***

This is not to say that the state is powerless to create laws that intentionally obstruct access to information. (T)hey can with proper justification. The outcome of Dietemann may survive scrutiny if the state can show it has a strong interest in protecting its constituents' seclusion by outlawing surreptitious recordings in one's own home. Mr. Dietemann is a sympathetic plaintiff because the humiliating photographs were taken without his knowledge, and on his own turf.[114] Mr. Dietemann has a liberty interest in keeping surreptitious recording equipment out of his home, and such an interest could compete in court with the reporters' interests in recording information to which they lawfully observed. Given the long-held expectations that the home provides sanctum[115], Mr. Dietemann may have won his case despite the First Amendment arguments.

Outside our homes, on the other hand, enforcement of an expectation that people will not record what we say or do is in conflict with their liberty interest in knowledge creation.

To this day the right to access information is underdeveloped.[116] Its relationship to full speech rights is awkward. Courts recognize that the right to free speech is hollow without access to information, but the constitutional protection of information has yet to achieve coherence.