Comparing policies for open data from publicly accessible international sources

Line C. Pouchard[1], Megan Sapp Nelson, and Yung-Hsiang Lu

Abstract

The Continuous Analysis of Many Cameras (CAM2)[2] project is a research project at Purdue University for Big Data and visual analytics. CAM2 has the ability to collect over 60,000 publicly accessible video feeds from many regions around the world. These video feeds were originally collected for improving the scalability of image processing algorithms and are now becoming of interest to ecologists, city planners, and environmentalists. With CAM2's ability to acquire millions of images or many hours of videos per day, collecting these large quantities of data raises questions about data management. The data sources have heterogeneous policies for data use, and some sources have no policies. Separate agreements had to be negotiated between each video stream source and the data collector. In this paper, we propose to compare data use policies that are attached to the video streams and study their implications for open access. The need for common points of legal guidance for webcam stream users and publishers is demonstrated through this analysis of usage agreements.

Keywords

Video, CCTV, public access, open data, re-use, privacy, webcams

Introduction

Thousands of network cameras have been deployed by many organizations for different purposes. For example, many departments of transportation (DoT) deploy cameras along highways or congested streets (City of New York, 2015). National parks deploy cameras showing the views from visitor centers. Some zoos deploy cameras showing animals' activities. Graham et al. (Graham, Riordan, Yuen, Estrin, & Rundel, 2010) used geo-located cameras for a plant phenology monitoring system. The National Park Service of the United States deploys cameras observing air quality (National Park Service). Nearly 20,000 cameras from a single site (Weather Underground) allow users to see weather worldwide. Another site has more than 40,000 cameras (OPAG) watching tourist attractions. The data are available to the public through the Internet: anyone connected to the Internet can see the data (image or video) from these cameras. The data may provide insightful information about our world, such as traffic congestion, air quality, weather, etc. An important value proposition for this data may lie in the ability to extract and compare information from multiple sources. Even though the data are publicly available, it is not easy using the data for scientific research due to accessibility challenges: there is no single site where the data from disparate sources are available. In the United States, the department of transportation of each state has its own website showing the traffic cameras. Different websites have different data formats and require

different ways to retrieve the data. More importantly, different organizations have different policies and restrictions about how data may be used. This paper describes the challenges of using the globally available camera data for scientific research.

In order to facilitate research using the data from global camera networks, we have been building a system that allows large-scale analysis of image and video (Kaseb et al, 2015a; Kaseb et al, 2015b; Hacker et al, 2014). CAM2 (Continuous Analysis of Many CAMeras,) is a research tool for solving some of problems mentioned above: it is a single site through which the data from many heterogeneous cameras can be retrieved and analyzed. The cloud-based computing engine can handle large quantities of data. An event-driven programming interface offers the flexibility to execute diverse programs analyzing the data. Users execute the analysis, using either their own scripts or with scripts provided by the system on the computing engine at the back-end of CAM2, and download the analysis results.

In most cases the data are intended to be viewed by human eyes through web browsers. CAM2 uses only publicly available data -- no password is required to access the individual feeds, only to login into CAM2. A big challenge in constructing CAM2 is to obtain the permissions granting usage of the data for scientific research. It is possible to write computer programs to retrieve the data from the sources' web sites but for courtesy, we requested explicit permissions from the data owners.

In remote sensing and video stream applications, data arrive at very high frequency. They exemplify the volume and velocity characteristics of Big Data, characteristics that have implications for infrastructure and privacy policies. Volume refers to the amount of storage needed to accommodate the data, and velocity to the speed at which the data is produced and transferred through networks (Jagadish et al., 2014). Issues of data retention, sharing and re-use arise for Big Data that are scantily addressed in the legislation on privacy. Individual policies sometimes address these issues but each data owner has a different policy and set of restrictions. Thus we research the terms that data owners use to articulate access and re-use of their data in those policies and how these policies in turn affect re-use of the data for scientific research.

This paper reports our results from the analysis of the policies we obtained from disparate data sources. We frame the discussion with examples of data privacy law in the US and EU where a comprehensive, unified framework is given to Member States by the European Council directive. We perform qualitative analysis on the 15 policies we obtained and present the results in this paper. The discussion shows a comparison of the policies with a focus on terms of use. The CAM2 project was designed to test and improve image analysis algorithms in real time, but other researchers, such as environmental scientists, city planners, and ecologists are becoming interested in the data as a resource for additional scientific analyses. We are illustrating the implications of those policies for other researchers interested in re-using the data. One of our findings is that these policies tend to be sparse in terms of re-use, so we are also trying to understand the implications of these gaps for researchers interested in re-using the data.

Literature Review

We analyzed the UK's data protection policy as applied to CCTV for insight into areas that may be addressed by a CCTV usage policy. This policy was selected due to the widespread adoption in the

UK and the multiple iterations that this policy has gone through in response to appeals to the European Court of Human Rights.

The framework for the UK data protection policy is provided by the European Union Data Protection Directive of 1995 (European Council, 1995) and enforceable through the EU Member States since 2004. The 1995 EU Data Protection Directive and its amendments provide three important items with regard to the video streams we are interested in:

1. images or voice are considered personal data (article 29, Working Party);
2. an all-encompassing definition of processing is provided, including "collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction" (article 2(b));
3. when data is transferred to another country, member states must ensure that this country affords equal protection to individual privacy.

The EU Data Protection Framework Decision applies to cross-border exchanges of personal data processed in the framework of police and judicial cooperation in criminal matters but does not apply to domestic data (European Digital Rights, 2009).

The UK CCTV Code of Practice describes the many factors that a CCTV operator must consider and the rules that must be followed before the implementation of a CCTV system (Information Commissioner of the UK, 2014, 2014a). Operators should conduct a privacy impact assessment and ensure that effective administration with decision-making power about the storage, possible encryption, retention and use of the images is put into place. Images should be retained in a secure location for no longer than strictly necessary as prescribed by the purpose of recording them, accessible only by authorized personnel in a controlled location, and erased once the defined period of retention is reached. Any operator putting images on the internet must consider the possible disclosure of individuals' personal data and proceed accordingly. It is also recommended that operators remain in control of the information and conduct periodic audits to review the many provisions of the Code of Practice. Another important aspect of these Practices is the emphasis on making the use fit the purpose: operators are to ensure that the images are used according to the purpose for which the system was put in place (often surveillance).

But the surveillance and video systems in place in public places evolve rapidly with new technology and a greater acceptance of the public. Examples of emerging technology which impacts the application of regulations include ubiquitous computing and wearable imaging devices, automatic licence (or number) plate recognition, unmanned aerial vehicles, and cameras equipped with direct access to the internet. These new technologies and the increased ability to link information challenges the protections afforded by the EU Data Directive and the framework put in place by member states.

As a consequence of a more interconnected society, re-use of information for purposes completly different than the ones for which the information was captured is becoming common (Coudert, 2009). The principle of collecting and retaining data specifically for the purpose described at inception of the system is being eroded. Safeguards put into place are not respected or impractical

in the context of interconnected video networks. The UK government itself recognized that the lack of definitions of data sharing hampered communication between recovery efforts in the aftermath of the 2005 London bombings (Coudert, 2009). Some scholars argue that the protections are toothless, the regulatory bodies lack the resources to enforce them and rely upon the goodwill and cooperation of those they are regulating (Gras, 2004). Nonetheless, in the UK, a legal and regulatory framework exists that is relatively homogeneous, and affords some amount of protection.

There is no equivalent data and privacy protection framework in the US at the federal level but instead what many call a ”patchwork” of regulators and regulations, common law, federal legislation, the US Constitution, state law and certain state constitutions. Privacy is regulated primarily by industry on a sector-by-sector basis and US regulation of the private sector is minimal (Levin and Nicholson, 2005). The major components to privacy law as it applies to data include the Fair Information Practice Principles, the bulk of which were adopted in the early 70s and 80s. The Federal Trade Commission issued a series of streamlined principles at the beginning of the 21st century, focused on online privacy, including the notions of notice and consent for personal information collection. Notice and consent cover the requirements to inform consumers that private information is collected in the course of providing a service and to give them the choice of accepting that their information may be used for other purposes. One observes a trend that these principles have been weakened to focus on the procedures of giving notice and consent rather than on substance (Strandburg, 2014).

The US is more concerned with protecting the privacy of its citizens from government than regulating industry practices. At the same time, industries are pushing back from any regulation and privacy protection is more the result of market constraints and unpredictable common law (Levin and Nicholson, 2005). As an example of the piecemeal approach, many regulatory acts issued by various regulatory bodies currently govern privacy and could be applied to data (See Appendix 2). Privacy protection in public places, which is where video surveillance cameras operate, does not exit, following a Supreme Court and numerous federal court rulings that there are no privacy expectations in a public place (Slobogin, 2002). Data use is largely unregulated except in a few exceptions (health care, credit reporting) and re-use of Big Data renders meaningless the notions of notice and consent due to the complexity and numerous possibilities of aggregating the data (Ohm, 2014).

Analysis

We analyzed the contents of fifteen usage agreements with a variety of different entities, both international and United States, government and business entities. As several of these agreements specify that the terms shall not be made public, no entity will be identified here, other than as a representative of a class (government entity, business entity, US or international). The agreements were analyzed with Nvivo using a coding structure that was developed based upon the terms present in the sample of usage agreements.

The coding schema developed in this project is included in this article as Appendix 1. The nodes were identified based upon the terms that were present in the collected agreements and the context in which those terms were presented (term creator; creator classification; technical

guidelines and specific information that we sought as researchers including data sharing and recording or duplication).

As a basic classification scheme we divided formal from ad hoc policies. By formal policies, we mean that contractual agreements were signed between the researcher at Purdue University or a university representative and a legal representative of a data providing entity. These policies tend to be longer and contain many disclaimers protecting the provider entity from any liability (costs, disputes, responsibility for the actions of a third party) from the use of the feeds. They also tend to assert that the policy itself is protected by copyright and in one case not to be made public without express written permission from the provider. Ad hoc policies are those created by the manager of a data stream. These tend to be focused upon the technical aspects of accessing the data stream, are not formally ratified, and provide little guidance about how the stream may be used. We had 10 ad hoc policies and 5 formal ones.

Figure 1 shows the terms present in the formal policies and the total number of codes recorded for those terms. For instance, the code "Use restriction" is assigned 12 times over 3 separate policies. Branding is assigned 7 times over 5 policies.

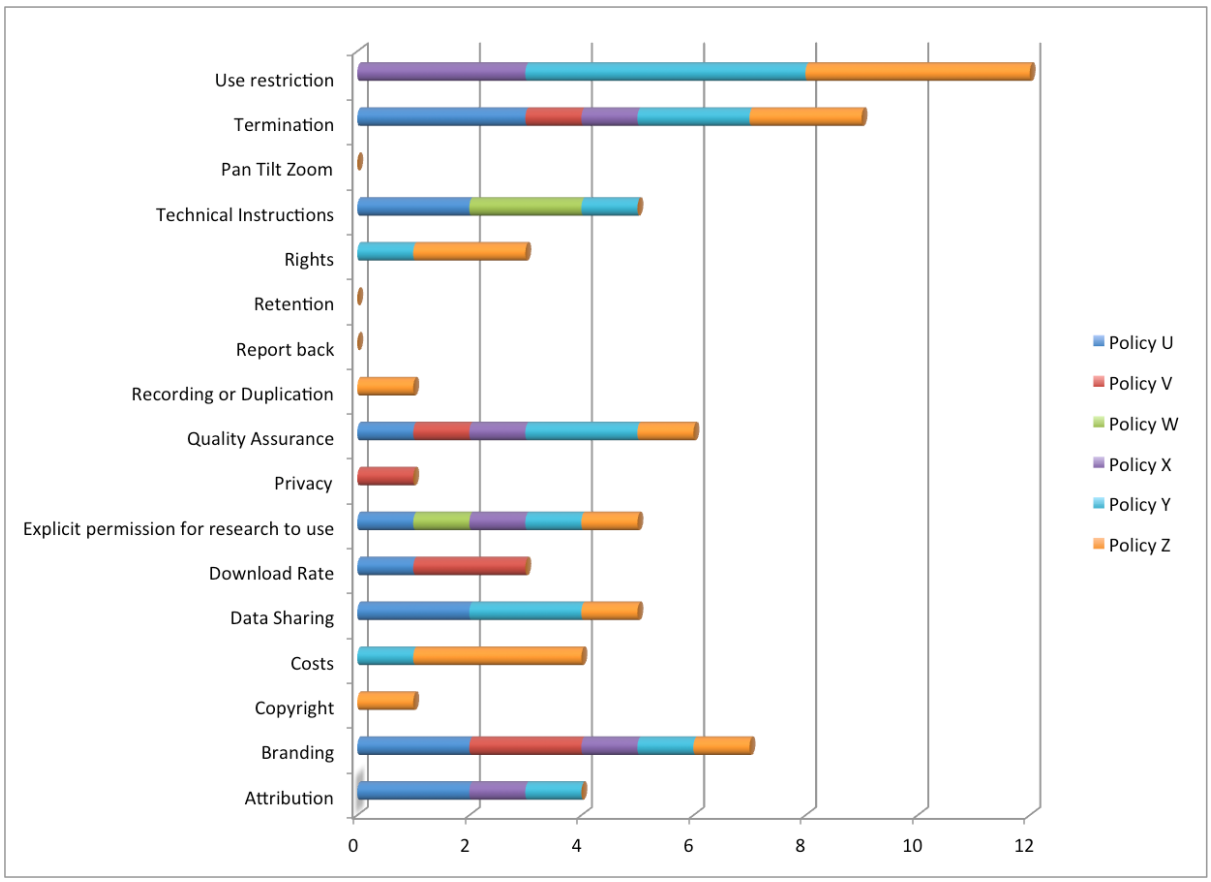

Figure 1: Formal policies and coding distribution

The ad hoc policies are much less formal in content and form. Often they are in the form of an email to the researcher granting the researcher the right to use the video feeds. They are often technical

in content, explaining how to access the feeds, providing APIs, and sometimes expressing concerns or providing limitations about the download rate so as not to slow down their systems. Because of the technical content of these emails, it often appears that they might have been written by the developers or maintainers of the system.

Figure 2 shows the distribution of coding across the ad hoc policies documents and the total number of times a particular code is encountered.

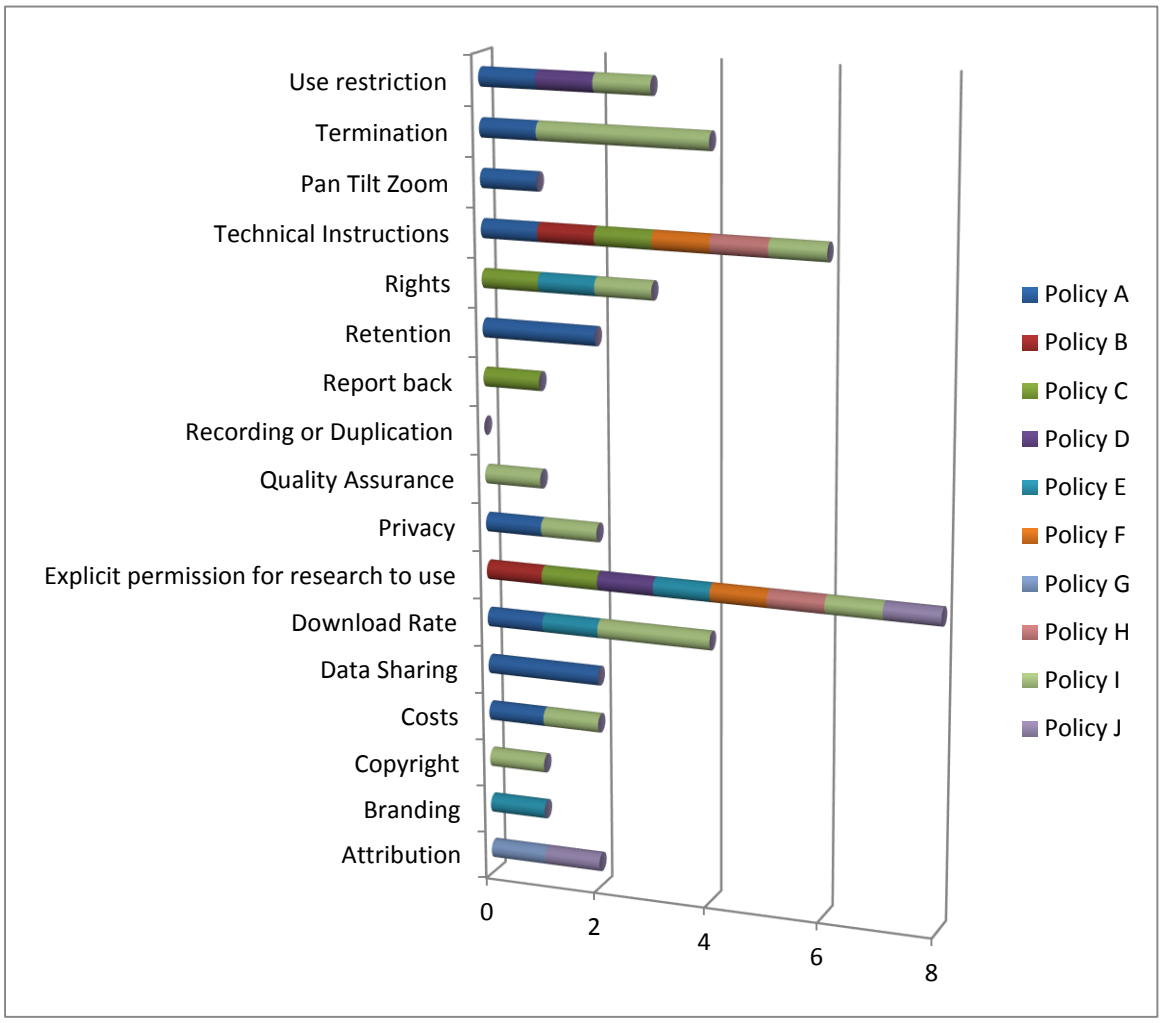

Figure 2: Ad hoc policies and coding distribution

Discussion

The policies were divided into two primary classifications, formal and ad hoc. Formal policies were developed by law entities as contracts specifying usage guidelines. These required signatories for both the user and the data stream provider to sign a legally binding agreement prior to accessing the data stream. These legal agreements included many usage restrictions and guidelines in common,

including branding or marketing, attribution expectations, re-use guidelines, data retention restrictions, and access termination.

Ad hoc policies were created on the fly by the data stream provider. They did not require formal signatures. Generally they included minimal guidelines for use and focused on technical guidelines and explicit permission for the end user to access the data stream.

In the case of the ad hoc policies, the permission was given as a person-to-person communication in the form of an email or other written communication. The ad hoc policies were focused upon granting permission to use the web streams to Dr. Lu.

Use restrictions were imposed in seven different agreements, the five formal policies, and two ad hoc policies. There were several different types of restrictions. The most common restriction was a technical restriction. To protect servers from being overwhelmed, the rate of download was restricted, generally either in time between downloads and file size that should be downloaded and occasionally both.

The download rate restrictions included are in the table below.

Table 1

Examples of time limit or file size limit
A picture will not be captured more than once every two minutes
Allowed one picture per hour per camera
Allowed one 320 x 240 jpeg per second
No camera will be accessed more than once every five minutes
No more than a cumulative 24 hours of images that are no more than one week old.

For the researcher, each of these separate restrictions has to be built into the automated retrieval system. As shown above, a “universal” rate of download could be one picture per hour, as all other exceptions fall within that rate. However, the researcher would be missing a large quantity of data, as demonstrated by the allowable download rate of one image per second. Therefore the CAM2 system needs multiple use categories built in so that different cameras can be accessed at different rates. This leads to a significantly more complicated download script.

Additional restrictions are focused on attribution and branding. For both government and business entities, attribution of the original data stream was requested. However, there was no general agreement in how that attribution should be carried out. In some cases a hyperlink on the CAM2 website was requested. In others, a logo should be placed on the CAM2 website where it is visible

during the search of camera data streams. Others generally requested that the camera owner be identified but did not specify where or how. In other cases an academic citation to the source was requested on publications or websites. Again, this broad range of requests for attribution makes it difficult for the researcher to efficiently meet these terms. In general, the "universal" policy may be to represent the data owner with logo and hyperlink on the screen for all cameras, and cite all data providers in academic articles where that data was used. However, unless a script is run that identifies the data owner for all cameras used in a given simulation, the sheer number of cameras (60,000 and growing) that may be included in the analysis make this unwieldy.

Data sharing was the primary provision of interest to the researchers. Perhaps not surprisingly, this was not on the radar of most of the entities. It did appear in four formally developed policies. In those cases, one entity indicated that data sharing was not required. Two entities expressly prohibited data sharing and reuse. The fourth entity allowed reuse only in the case of broadcast journalism. This is obviously discouraging to researchers who may want to continue to develop a video analysis system to create truly reproducible research results on specified sets of camera and image data. In the case of Big Data, reproducibility is an ongoing area of concern and research. Further developments in this aspect of Big Data research may help to deal with some of the issues presented by these data sharing provisions.

There were also restrictions on the ability to retain, copy or duplicate footage. One policy explicitly prevented retaining more than 5 consecutive minutes of images, and no more than 24 hours of footage that is a week old or more. But this policy does not prohibit data sharing explicitly. Another restricts all right to copy or duplicate in any way the video feeds unless the user is a media outlet. These restrictions on retention and copying place an additional burden on the possiblity of re-use of the data by other researchers.

The policies provided to the researcher were for the most part unable to provide the researcher with the rights, permission, or guidance needed to extend use of the data streams to other research projects. Data sharing and re-use was built into only one ad hoc policy and two formal policies. In those cases, data sharing and re-use was explicitly allowed to broadcasters only in one formal policy, explicitly disallowed in another formal policy, and is specifically left open to the discretion of the end user in the ad hoc policy. There is no similar language between the three policies. While these video streams present many interesting possibilities for large scale analysis having to do with climate, economics, and engineering, the policies that mention data sharing for the most part restrict it. Many more policies do not mention sharing one way or another. Even if the researcher took the most liberal interpretation of that hole in the policy, it leaves the researcher with the problem of having to build restrictions within the CAM2 system for which camera feeds can be re-used by other researchers. Another solution is to not use CAM2 data feeds that explicitly prohibit re-use.

Additionally recording and duplication is prohibited by two formal usage policies. This means that specific considerations for feeds that cannot be re-used beyond the CAM2 purpose would have to be built in as well.. In those cases, the relative value of video streams that cannot be reused may need to be reconsidered and these feeds may need to be dropped from the project, because the real value of the CAM2 system goes beyond the simultaneous analysis of data feeds for improving image

processing algorithms. If those streams cannot be included for future scientific use, are these video streams actually useful in this context?

The policies add value to the video streams. Those policies that do have specific information about the re-use of data streams and duplication of images are giving the researcher the chance to do extensive new science. The lack of policy indicates a video stream that may end as a liability to the research project due to lack of clear guidance on appropriate duplication, re-use, and preservation actions that can be taken.

Suggested Actions

We propose that an agency such as NISO, the National Information Standards Organization, should create a template usage agreement that highlights terms that encourage scientific reuse of video data. The template may include language giving permission for the creation of derivative works and information regarding appropriate storage and retention as well as preservation (including protection of privacy and metadata to preserve quality and identify date and time of creation.) This template could then be offered by the researcher to the video data producers when no formal policy already exists for that entity. The template could also be adopted by local and state government agencies as the starting point for their formally adopted policies, thereby ensuring that the relevant terms needed for scientific reuse are included in these legally binding agreements.

We propose the following as key components of this template:

Data provider identification

> One problem that we identified was that in many adhoc policies the data owner was not identified clearly. This would help with both proper attribution and accountability.

Download rate

> Many different download rates may be technically feasible due to the differences in camera and server specifications but for ease of use in planning and building data analysis systems, we propose that a specified rate of download be identified by NISO for a generic usage agreement, with specific higher or lower rates of download negotiated on a project by project basis between the data owner and data user.

File size

> As in download rates, many different file sizes may be technically feasible due to differences in the specifications of the equipment used by both the data owners and the data users. We suggest that the template include language that guides the clear negotiation of file size.

Statement of re-use that allows for general scientific investigation

> A key finding of our investigation was that scientific or academic re-use was sometimes prohibited partially or in full. A statement that the data stream can be re-used for scientific or academic research generally would enable multidisciplinary research investigations on the same data set. Additionally, a statement that the data

streams may be used to create derivative data sets would help further scientific research, as subsets of collated data sets could then be created to accompany publications.

Privacy

A statement governing appropriate use of the data set regarding individual's privacy should be included in all terms of video data re-use. We believe that the default should be that individuals' privacy will not be infringed upon by the re-use of the streaming video data. For projects that seek to use video to develop face identification algorithms and similar technologies, these terms should be negotiated on a project by project basis.

Quality Control

If data producers are worried about data streams being re-distributed in a misleading way, a date and time stamp, metadata regarding the source of the original data stream, as well as a branding icon may be required on still images or video clips. Those requirements should be spelled out in the terms of use.

Attribution

A suggested attribution including an academic citation that is generated as part of the usage agreement would go a long way towards insuring that data users have an easy, consistent way to refer to the data stream that they are accessing. As part of this, data providers may consider instituting a persistent URL for the website hosting their streaming data. Additionally the date or time stamp may be referenced in the attribution as well.

Retention and preservation

Data streams may only be valuable if analysis can be performed over extended collections of data representing days, months or years worth of data. Suggested language may include negotiable terms for the storage and preservation of streaming data for use in longitudinal data analysis systems.

Accountability

The template should include the responsibilities of the data user to the data provider whether that be proper attribution, reports back to the data provider on how the data is used, or assurance that quality control measures have been put in place. These terms should also be negotiated and included in the agreement.

Conclusion

This paper reported our experience requesting and obtaining permissions using publicly available video data for scientific research. Even though the data are already publicly available on the Internet, the heterogeneous sources of data and the terms of use specified or missing create many administrative difficulties. There is no consistent policy among different states or cities of the same country.

Several changes could be made at the regional and national level to facilitate the reuse of data. To start with, we recommend that the Government Accountability Office or other similar agency develop a single video data reuse policy that is applicable to all agencies. On the international level, central governments should establish a common set of rules governing video data. Among different countries, an international standard could be established for the appropriate terms that enhance scientific reuse to be included in usage policies.

Ideally, a legal framework should be created that will protect both video data producers and end users. This requires a culture change that encourages additional legal guidelines governing the privacy and data management practices of public and private companies and government agencies. Legal scholars are currently calling for this sort of legal framework but some are more concerned with citizens' privacy rights to be protected from government action (Slobogin, 2002; Greer, 2012) than with sharing scientific data. With the advent of Big Data, other legal and privacy scholars are lately calling for a comprehensive framework and national discussion about the scope and foundational concepts of privacy in the context of Big Data re-use (Lane, Stodden, et al, 2014).

If templates and policies similar to these were adopted, many more scientific uses of video data streaming on the internet could be embarked upon, ranging from environmental and ecological studies to economic assessments based upon the movement patterns of individuals showing up on cameras in a variety of public places. Until these policies are specified, the legal liability of the researcher that presumes too much is too great to enable these more advanced, longitudinal studies of streaming video data.

References

City of New York (2015). NYCDOT - Real Time Traffic. Retrieved March 6, 2015, 2015, from http://nyctmc.org

European Council (1995). Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Official Journal of the European Union, No L 281/31.

European Digital Rights (2009). Data protection framework decision adopted. Retrieved March 23, 2015 from http://history.edri.org/edri-gram/number7.3/data-protection-framework-decision.

Graham, E. A., Riordan, E. C., Yuen, E. M., Estrin, D., & Rundel, P. W. (2010). Public Internet-connected cameras used as a cross-continental ground-based plant phenology monitoring system. *Global Change Biology, 16*(11), 3014-3023.

Greer, O. J. (2012). No Cause of Action: Video Surveillance in New York City *Michigan Telecommunications and Technology Law Review, 589*.

Hacker, T. J., & Lu, Y.-H. (2014). An Instructional Cloud-Based Testbed for Image and Video Analytics. Proceeding of the 2014 IEEE 6th International Conference on Cloud Computing Technology and Science (CloudCom), Singapore, December 2015.

Information Commissioner of the UK (2014). CCTV Code of Practice - Draft for consultation. Retrieved March 23, 2015, from https://ico.org.uk/media/about-the-ico/consultations/2044/draft-cctv-cop.pdf

Information Commissioner of the UK (2014a). In the picture: a data protection code of practice for surveillance cameras and personnel information. Retrieved March 23, 2015, from https://ico.org.uk/media/for-organisations/documents/1542/cctv-code-of-practice.pdf

Jagadish, H., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., & Shahabi, C. (2014). Big data and its technical challenges. *Communications of the Acm, 57*(7), 86-94.

Kaseb, A. S., Berry, E., Rozolis, E., McNulty, K., Bontrager, S., Koh, Y., Delp, E. J. (2015a). An interactive web-based system for large-scale analysis of distributed cameras. Proceedings of Imaging and Multimedia Analytics in a Web and Mobile World conference, San Francisco, February 2015.

Kaseb, A. S., Chen, W., Gingade, G., & Lu, Y.-H. (2015b). Worldview and route planning using live public cameras. Proceedings of the Imaging and Multimedia Analytics in a Web and Mobile World, San Francisco, February 2015.

Lane, J., Stodden, V., Bender, S., & Nissenbaum, H. (2014). *Privacy, Big Data, and the Public Good: Frameworks for Engagement*: Cambridge University Press.

Levin, A., & Nicholson, M. J. (2005). Privacy law in the United States, the EU and Canada: the allure of the middle ground. *U. OTTAWA L. & TECH. J., 2*, 357.

National Park Service (2015). NPS explore nature. Retrieved March 6, 2015, from http://www.nature.nps.gov/air/webcams/ .

Ohm, P. (2014). Changing the rules general principles for data use and analysis. In J. Lane, V. Stodden, S. Bender & H. Nissenbaum (Eds.), *Privacy, Big Data, and the Public Good: Frameworks for Engagement*: Cambridge University Press.

OPAG (2015) Webcams.travel. Retrieved March 6, 2015, from http://www.webcams.travel/ .

Scherr, C. (2007). Government Surveillance in Context, for E-mails, Location, and Video: You Better Watch Out, You Better Not Frown, New Video Surveillance Techniques are Already in Town (and Other Public Spaces). *ISJLP: A Journal of Law and Policy for the Information Society, 499*(3).

Slobogin, C. (2002). Public Privacy: Camera Surveillance Of Public Places And The Right To Anonymity. *Mississippi Law Journal, 72*.

Strandburg, K. (2014). Monitoring, Datafication and Consent: Legal Approaches to Privacy in a Big Data Context. In J. Lane, V. Stodden, S. Bender & H. Nissenbaum (Eds.), *Privacy, Big Data, and the Public Good: Frameworks for Engagement*: Cambridge University Press.

Weather Underground (2015) Weather Webcams. Retrieved March 6, 2015, from http://www.wunderground.com/webcams/ .

Wright, D., Friedewald, M., Gutwirth, S., Langheinrich, M., Mordini, E., Bellanova, R., Bigo, D. (2010). Sorting out smart surveillance. *Computer Law & Security Review, 26*(4), 343-354

Appendix 1

Table 2: List of nodes and instances of coding references

Node	Sources	References
AdHoc Policy	10	13
Article example	2	6
Formal Policy	8	8
Still Photos	5	7
Videos	4	5
Attribution	9	10
Branding	9	10
Copyright	1	1
Costs	4	7
Data Sharing	5	11
Download Rate	5	7
Explicit permission for research to use	14	14
Privacy	2	3
Public Domain	2	2
Quality Assurance	7	8
Recording or Duplication	2	2
Report back	1	2
Retention	1	5
Rights	6	8
Technical Instructions	10	14

Termination	8	15
Use Restrictions	7	16

Appendix 2

Table 3: Examples of the US regulatory bodies who have issued Acts that affect privacy.

The Federal Trade Commission enforces	Fair Credit Reporting Act	Consumer Reporting Agencies must maintain accurate records and can forward records to anyone with a legitimate interest.	1970
Department of Justice	Privacy Act	Regulates the use of data by government agencies.	1974
The Federal Trade Commission enforces	Financial Modernization Act	Financial institutions must have and share a privacy policy by which customers can decline sharing their personal information with third parties.	1999
The Federal Communication Commission enforces	Cable Communications Policy Act	Cable companies are not allowed to collect or share personal information without individuals' consent.	1984
The Federal Communication Commission enforces	Video Privacy Protection Act	Video stores cannot disclose their customers' rental history.	1988
The Department of Health and Human Services	Health Insurance Portability and Accountability Act (HIPAA)	Protects patients' health information from being released to potential employers.	1996
State of California	Online Privacy Protection Act	One of the most comprehensive laws. Websites' privacy policies must be highly visible and customers must be informed of third party use of their data.	2003

End-notes

[1] Line C. Pouchard (pouchard@purdue.edu) is Assistant Professor, and Computational Science Information Specialist at Purdue University Libraries. She specializes in Big Data. Megan Sapp Nelson (mrsapp@purdue.edu) is Associate Professor of Library Sciences and Engineering Librarian at Purdue University Libraries. Yung-Hsian Lu (yunglu@purdue.edu) is Associate Professor of Computer and Electrical Engineering at Purdue University and ACM Distinguished Scientist.

[2] https://cam2.ecn.purdue.edu/