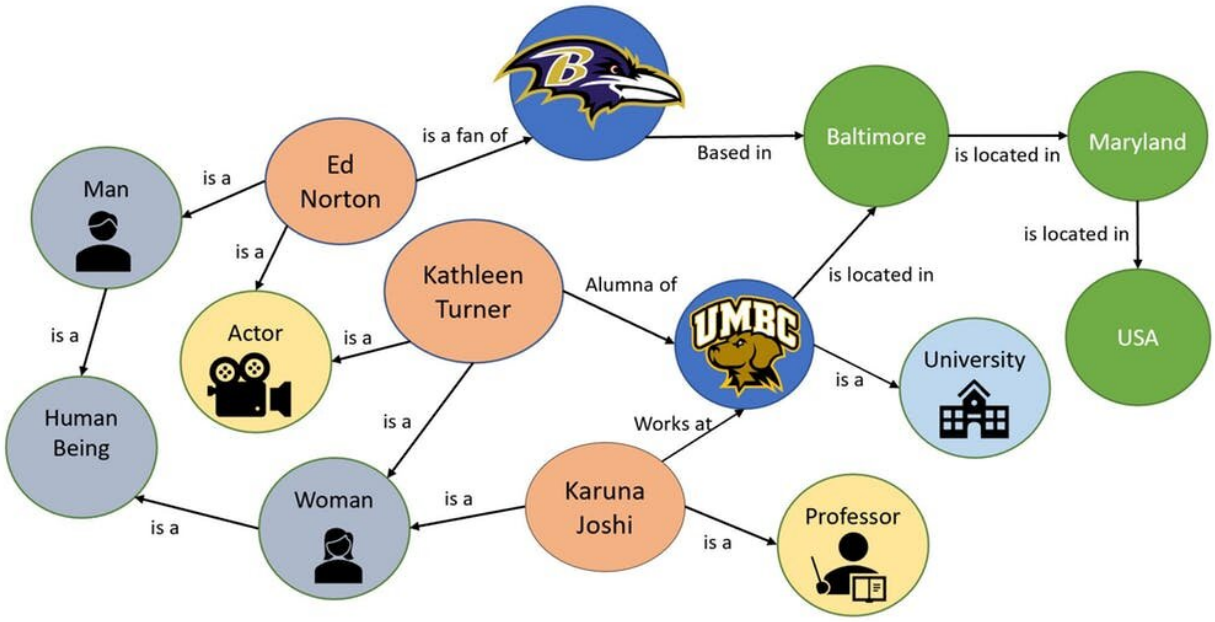
Data compliance could be enforced by AI scan of internet for privacy violations

February 10 2020, by Karuna Pande Joshi



An example of a simple knowledge graph. Credit: Karuna Pande Joshi, CC BY-ND

You're trailing bits of personal data—such as credit card numbers, shopping preferences and which news articles you read—as you travel around the internet. Large internet companies make money off this kind of personal information by sharing it with their subsidiaries and third parties. Public concern over online privacy has led to laws designed to control who gets that data and how they can use it.

The battle is ongoing. Democrats in the U.S. Senate recently introduced a bill that includes penalties for tech companies that mishandle users' personal data. That law would join a long list of rules and regulations worldwide, including the Payment Card Industry Data Security Standard that regulates online credit card transactions, the European Union's General Data Protection Regulation, the California Consumer Privacy Act that went into effect in January, and the U.S. Children's Online Privacy Protection Act.

Internet companies must adhere to these regulations or risk expensive lawsuits or government sanctions, such as the Federal Trade Commission's recent US$5 billion fine imposed on Facebook.

But it is technically challenging to determine in real time whether a privacy violation has occurred, an issue that is becoming even more problematic as internet data moves to extreme scale. To make sure their systems comply, companies rely on human experts to interpret the laws—a complex and time-consuming task for organizations that constantly launch and update services.

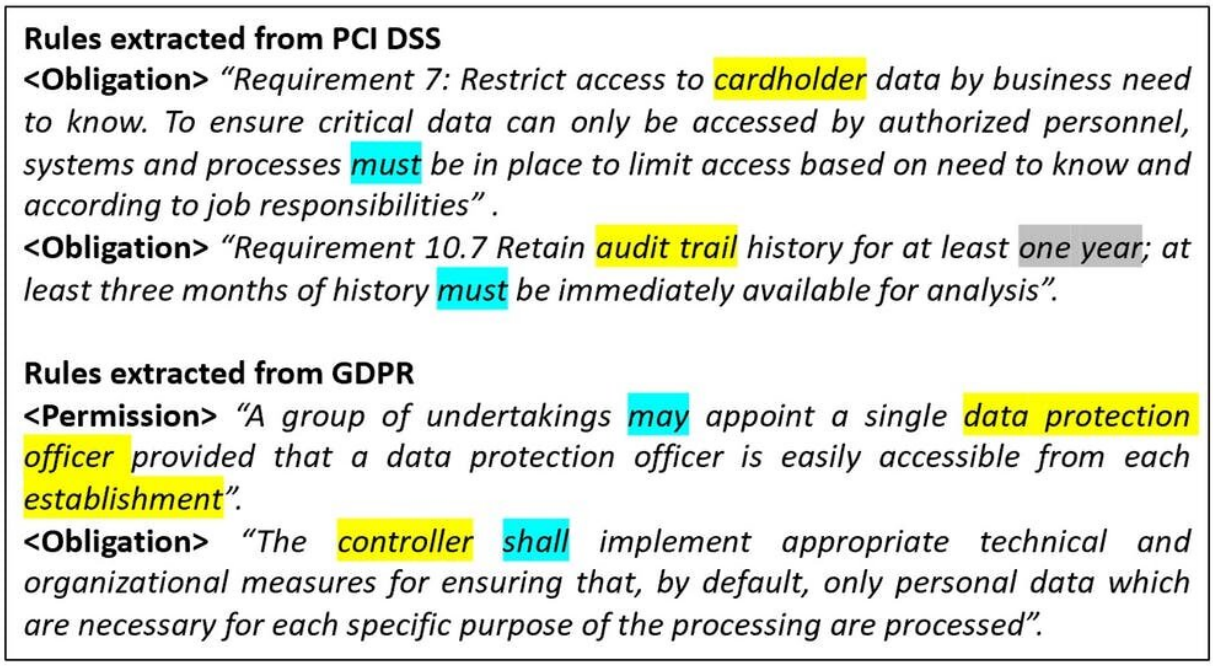
My research group at the University of Maryland, Baltimore County, has developed novel technologies for machines to understand data privacy laws and enforce compliance with them using artificial intelligence. These technologies will enable companies to make sure their services comply with privacy laws and also help governments identify in real time those companies that violate consumers' privacy rights.

Helping machines understand regulations

Governments generate online privacy regulations as plain text documents that are easy for humans to read but difficult for machines to interpret. As a result, the regulations need to be manually examined to ensure that no rules are being broken when a citizen's private data is analyzed or

shared. This affects companies that now have to comply with a forest of regulations.

Rules and regulations often are ambiguous by design because societies want flexibility in implementing them. Subjective concepts such as good and bad vary among cultures and over time, so laws are drafted in general or vague terms to allow scope for future modifications. Machines can't process this vagueness—they operate in 1's and 0's—so they cannot "understand" privacy the way humans do. Machines need specific instructions to understand the knowledge on which a regulation is based.



The researchers' application automatically extracted Deontic rules, such as permissions and obligations, from two privacy regulations. Entities involved in the rules are highlighted in yellow. Modal words that help identify whether a rule is a permission, prohibition or obligation are highlighted in blue. Gray indicates the temporal or time-based aspect of the rule. Credit: Karuna Pande Joshi, CC

One way to help machines understand an abstract concept is by building an ontology, or a graph representing the knowledge of that concept. Borrowing the concepts of ontology from philosphy, new computer languages, such as OWL, have been developed in AI. These languages can define concepts and categories in a subject area or domain, show their properties and show the relations among them. Ontologies are sometimes called "knowledge graphs," because they are stored in graphlike structures.

When my colleagues and I began looking at the challenge of making privacy regulations understandable by machines, we determined that the first step would be to capture all the key knowledge in these laws and create knowledge graphs to store it.

Extracting the terms and rules

The key knowledge in the regulations consists of three parts.

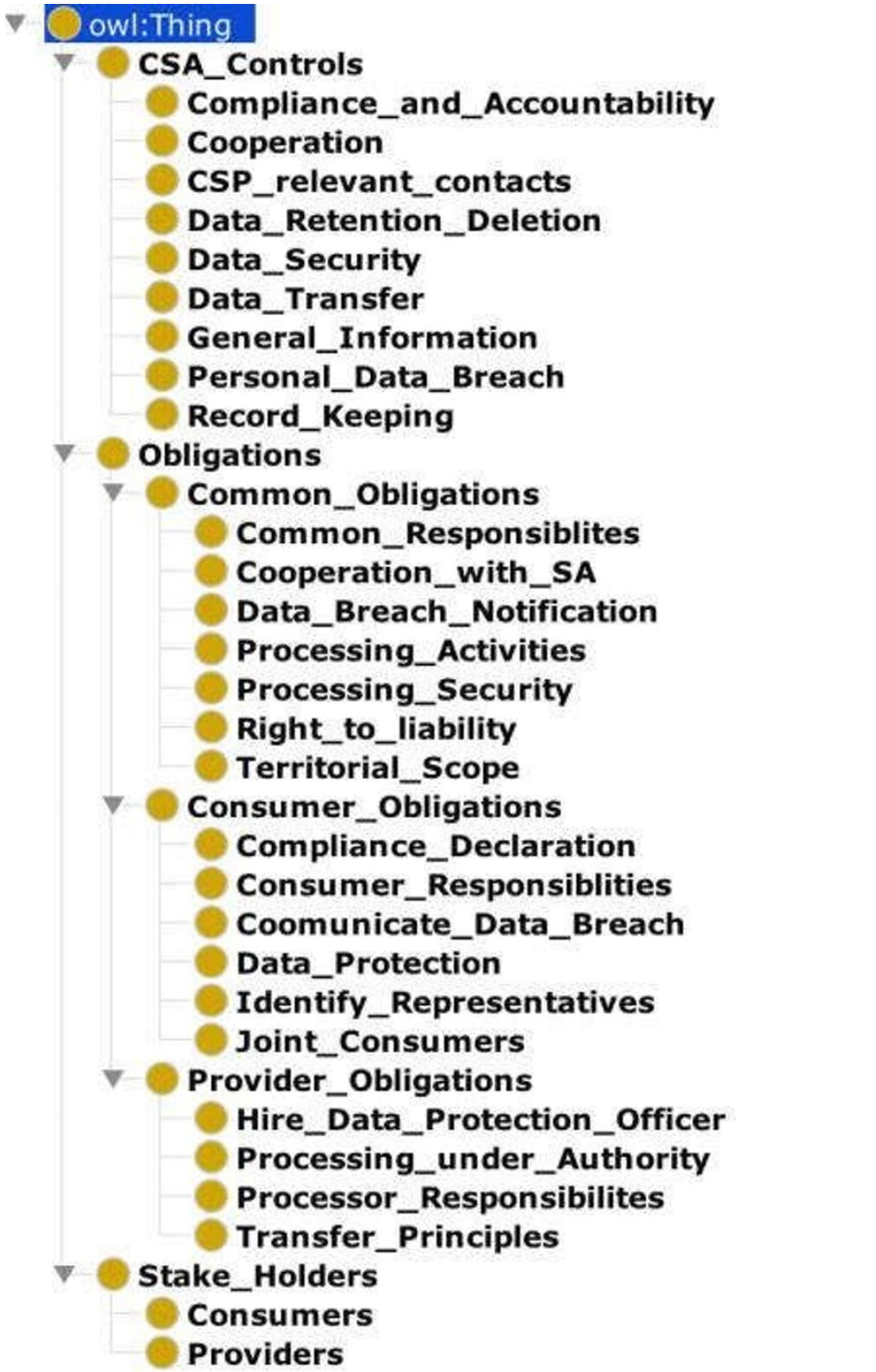
First, there are "terms of art": words or phrases that have precise definitions within a law. They help to identify the entity that the regulation describes and allow us to describe its roles and responsibilities in a language that computers can understand. For example, from the EU's General Data Protection Regulation, we extracted terms of art such as "Consumers and Providers" and "Fines and Enforcement."

Next, we identified Deontic rules: sentences or phrases that provide us with philosophical modal logic, which deals with deductive behavior. Deontic (or moral) rules include sentences describing duties or obligations and mainly fall into four categories. "Permissions" define the

rights of an entity/actor. "Obligations" define the responsibilities of an entity/actor. "Prohibitions" are conditions or actions that are not allowed. "Dispensations" are optional or nonmandatory statements.

To explain this with a simple example, consider the following:

- You have permission to drive.
- But to drive, you are obligated to get a driver's license.
- You are prohibited from speeding (and will be punished if you do so).
- You can park in areas where you have the dispensation to do so (such as paid parking, metered parking or open areas not near a fire hydrant).


owl:Thing
CSA_Controls
Compliance_and_Accountability
Cooperation
CSP_relevant_contacts
Data_Retention_Deletion
Data_Security
Data_Transfer
General_Information
Personal_Data_Breach
Record_Keeping
Obligations
Common_Obligations
Common_Responsiblites
Cooperation_with_SA
Data_Breach_Notification
Processing_Activities
Processing_Security
Right_to_liability
Territorial_Scope
Consumer_Obligations
Compliance_Declaration
Consumer_Responsiblities
Coomunicate_Data_Breach
Data_Protection
Identify_Representatives
Joint_Consumers
Provider_Obligations
Hire_Data_Protection_Officer
Processing_under_Authority
Processor_Responsibilites
Transfer_Principles
Stake_Holders
Consumers
Providers

knowledge graph for GDPR regulations. Credit: Karuna Pande Joshi, CC BY-ND

Some of these rules apply to everyone uniformly in all conditions; while others may apply partially, to only one entity or based on conditions agreed to by everyone.

Similar rules that describe do's and don'ts apply to online personal data. There are permissions and prohibitions to prevent data breaches. There are obligations on the companies storing the data to ensure its safety. And there are dispensations made for vulnerable demographics such as minors.

My group developed techniques to automatically extract these rules from the regulations and save them in a knowledge graph.

Thirdly, we also had to figure out how to include the cross references that are often used in legal regulations to reference text in another section of the regulation or in a separate document. These are important knowledge elements that should also be stored in the knowledge graph.

Rules in place, scanning for compliance

After defining all the key entities, properties, relations, rules and policies of a data privacy law in a knowledge graph, my colleagues and I can create applications that can reason about the data privacy rules using these knowledge graphs.

These applications can significantly reduce the time it will take

companies to determine whether they are complying with the data protection regulations. They can also help regulators monitor data audit trails to determine whether companies they oversee are complying with the rules.

This technology can also help individuals get a quick snapshot of their rights and responsibilities with respect to the private data they share with companies. Once machines can quickly interpret long, complex privacy policies, people will be able to automate many mundane compliance activities that are done manually today. They may also be able to make those policies more understandable to consumers.

Provided by The Conversation