Advancing Robot Lawyers

Towards Interoperable Protected Research Data

Micah Altman

Joint work with Stephen Chong and Alexandra Wood

Supported by the NSF and the Sloan Foundation

The Dataverse Project

Use and share data

Data Tags

Data Use Agreements

Social Scientist

?

Privacy Legislation

Analysis in legal memos

Best practices

Statutory text

Data use agreements (DUAs)

- Contracts that govern transfer of data containing personal information that is subject to some restriction on its use.
- Limits on use, obligations to safeguard, liability for harm arising from use or misuse, publication requirements and restrictions, privacy rights

Wood, O'Brien & Gasser, “Contractual Approaches to Research Data Sharing: A Comparative Analysis of Institutional Data Sharing Agreements and Policies”, Working Paper (forthcoming)

DUA Drivers

Statutory Law

1

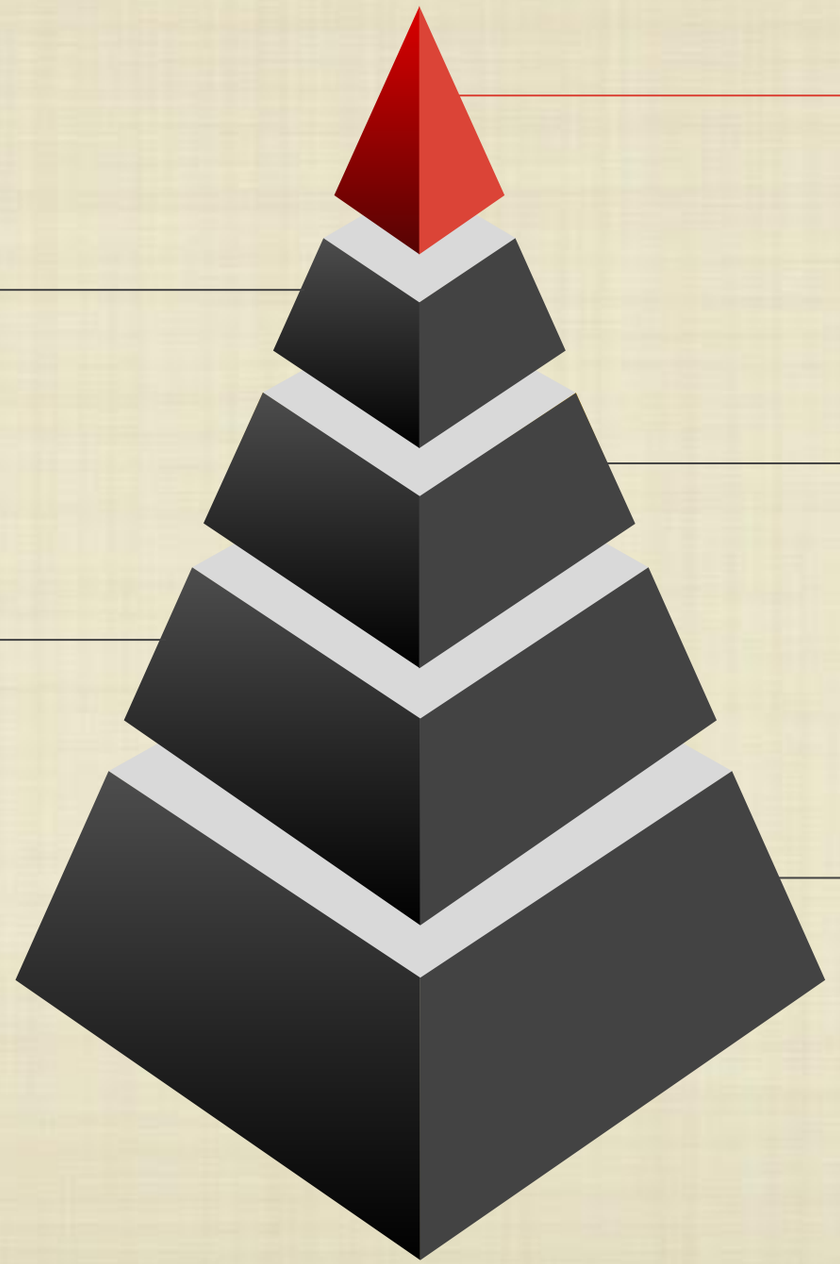
Laws may require that agreements are established before certain actions, such as data transfer occur.

Regulations may provide express requirements and guidelines

Case Law; Regulatory Guidance

May establish further requirements, establish terminology, establish sufficiency of technologies or processes.

2



Standards

3

External, published, community-recognized standards or specifications may establish, terminology, technology, processes.

Industry Practices

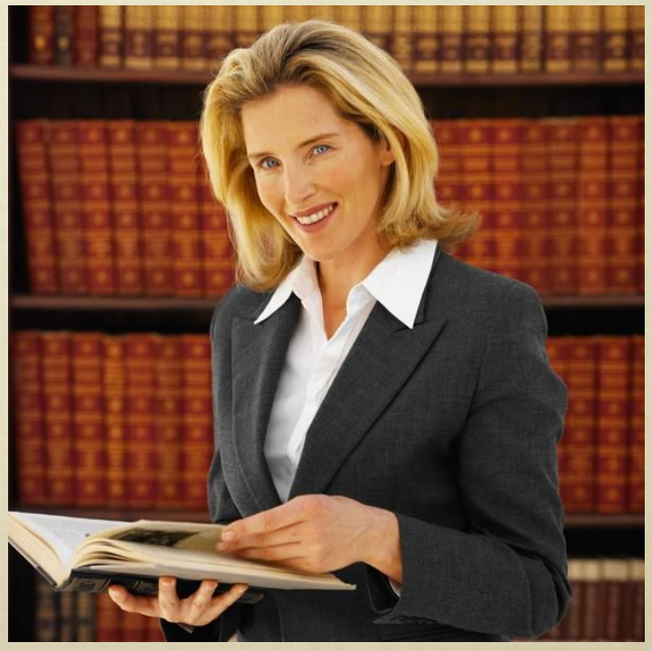
4

Local Practice and Motivation

5

DUA Standardization

- Standardized data use agreements are common
  - Boilerplate and clickthrough agreement(s)
  - Typically used by large data repositories
  - Coarse-grained, don't accurately capture permissions/obligations for specific dataset
  - Often present barrier to data integration and downstream reuse
- Customized data use agreements are
  - Customized for the specific data transfer
  - Typically requires human (lawyer) effort
  - Requires expertise in privacy legislation (i.e., can still be inaccurate)
- Ideally: Automated generation of modular DUAs

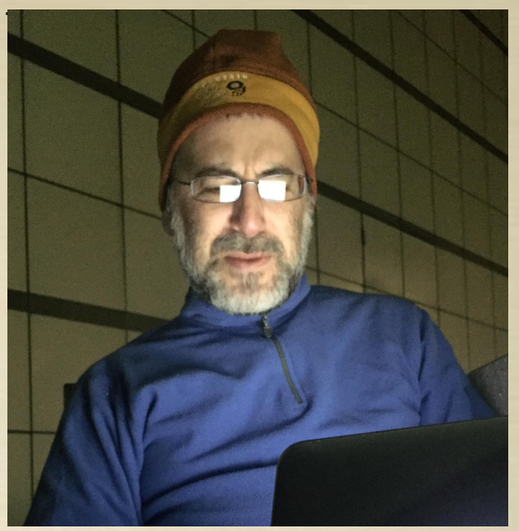


The Dataverse Project

Use and share data

Data Tags

Data Use Agreements



Social Scientist

This work

Tag characterization

License generation

Formal model
of regulatory requirements and best practice

Privacy Legislation

Analysis in
legal memos

Best practices

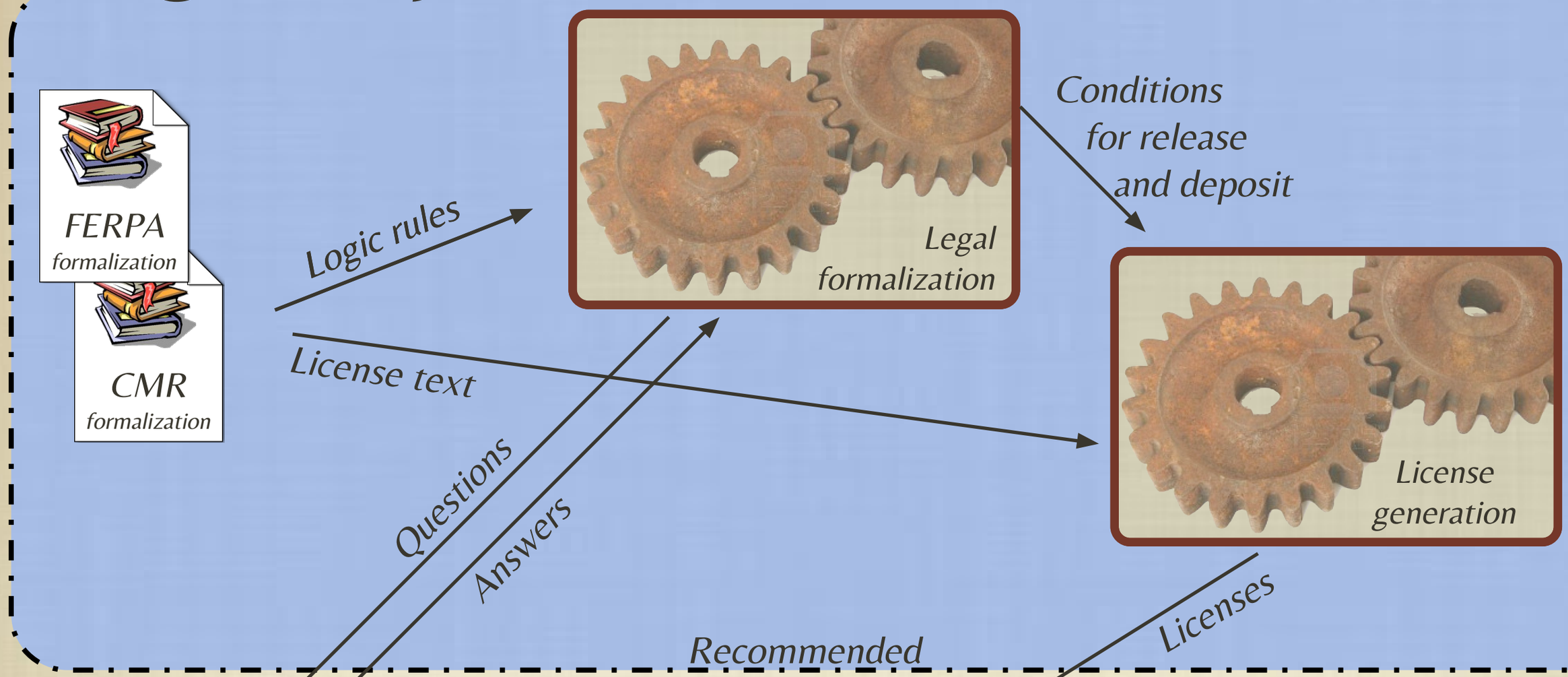
Statutory text

Computer Science

A Practical Approach

- Limited domain
  - Data repositories
    - Aspects of privacy legislation related to storing and transferring data
- Focus on subset of conditions that can be automatically handled
  - Not intended to handle 100% of situations
  - May need to be escalated to a human...
- Iterative development...
  - Prototype implementations of components, formal logic
  - Deployment through dataverse, interview tool

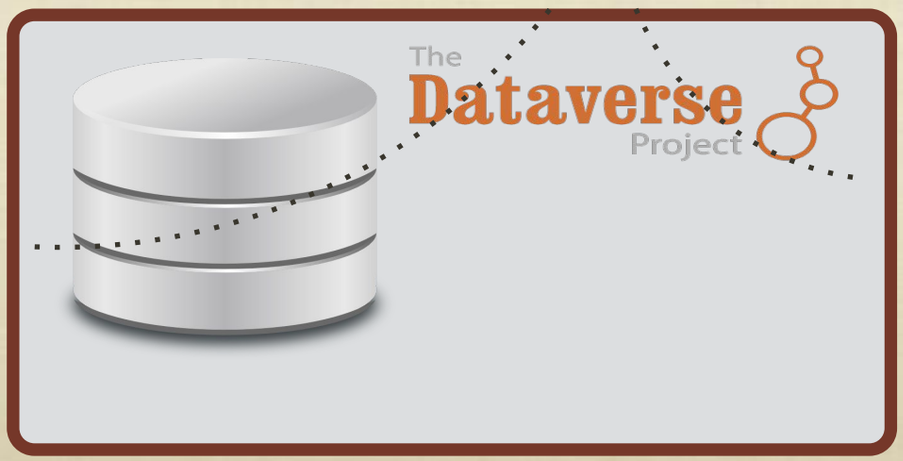
Logic System overview



Data Owner

Data deposit

Deposit license

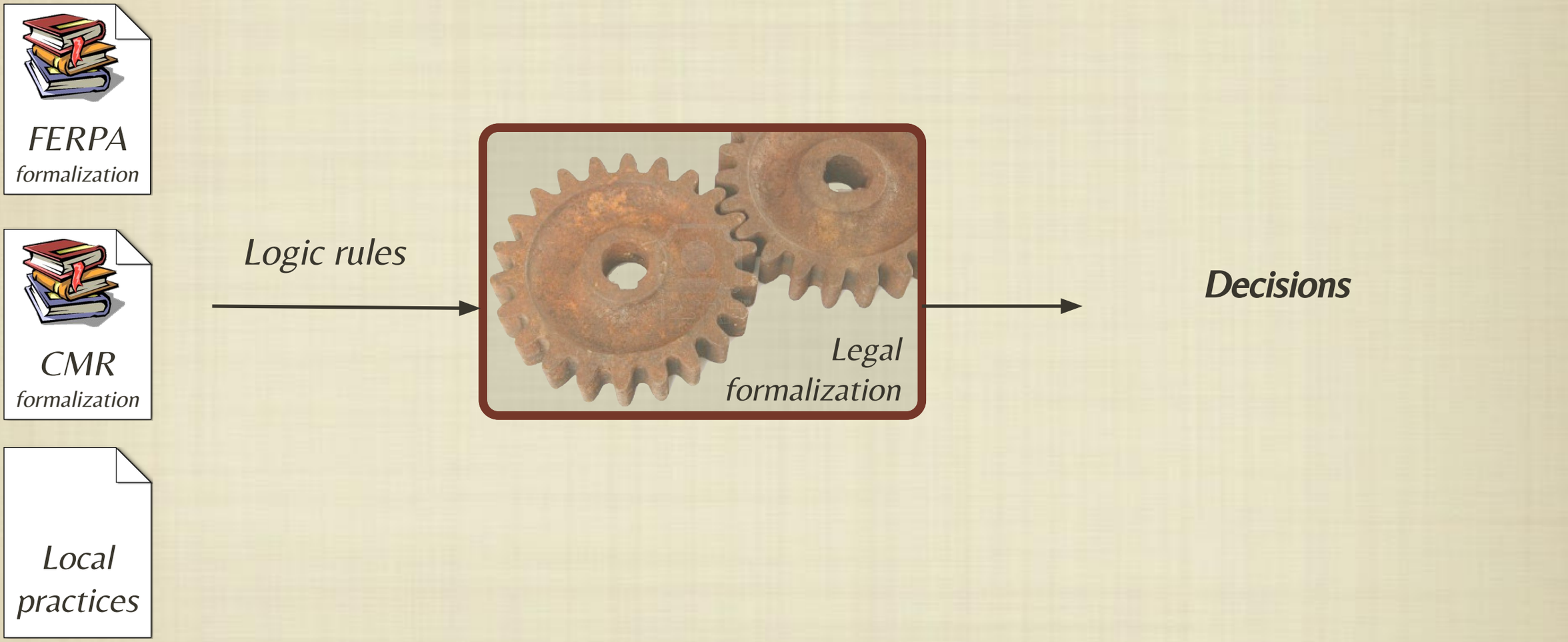


Data use
agreement

Data

Data User

Logic System overview



Legal Formalization

- Formalize aspects of privacy legislation
  - Using a logic programming language
- Answer whether legislation/best practice permits or denies specific actions on data sets
  - Expert-system-like ability
- Combines
  - computer science (formal modeling),
  - law (legal research & analysis),
  - social science (survey design),
  - information science (taxonomies, workflows)

Formalization Process

Identify Critical Use Cases	Framing: Data Repository (Controller) Decisions	Context: Data lifecycle	Key Use Cases: Deposit, Retention, Transformation, Release
Characterize Key Elements of Cases	Actors: depositor, repository, data subject, data consumer	Actions: offer; accept; store; destroy ...	Entities: dataset; records; assertions ...
Summarize Targeted Area of Law	Identify relevant documents: code, regulation, guidance, caselaw ...	Highlight: how law restricts key actors, actions, entities, use cases	Summarize: Legal brief describing and commenting on restrictions
Encode Summary using Formal Logic	Permissions: Encode any action that is controlled by law	Conditions: List the types of conditions under which action is permitted or denied	Standardize: Develop standard license terms; affirmations; taxonomic classification that can satisfy conditions

Formal model: Actions

`dd` : Data depositor

`r` : Repository

```
Deposit(dd, ds, r, cs)
Accept(r, ds, dd, cs)
Release(r, ds, du, dd, cs)
...
```

`ds` : Dataset

`du` : Data user

`cs` : Condition set
(provides further details about action)

Permitted or Denied

- Actions can be **permitted** or **denied** by legislation

```
Permitted(leg, a)
Denied(leg, a)
```

`leg` : Legislation

`a` : Action

  - Or neither permitted or denied
- E.g., `Denied(ferpa,`

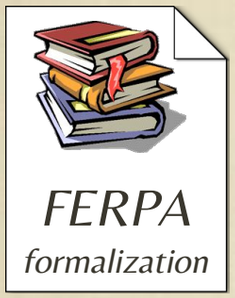
```
        Release(harvardDataverse,
                cs152grades-2015sp,
                jon@doe.com,
                chong@seas.harvard.edu,
                [dataverseClickthrough]))
```

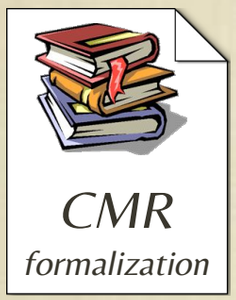
Example formalization

```
Let dd be the data depositor
Let du be the data user
Let ds be the data set
Let r be the repository
Let cs be a set of conditions

IF  CMR:depositorInScope(dd, ds)
AND CMR:identifiable(ds)
AND NOT (CMR:secure(r)
         AND CMR:isAcceptableConditionsForRelease(cs))
THEN DENIED(cmr, Release(r, ds, du, dd, cs))


Let l be a license
Let cs be a set of conditions
IF  License(l) ∈ cs
AND licenseImplies(l, CMR:TransmissionEncrypted)
THEN CMR:isAcceptableConditionsForRelease(cs)
```

License system overview

Conditions
for release
and deposit





Standardized
Terms

Licenses

Deposit Agreement

Repository TOS

Access License

License generation

- **Framework:**
  - Rules base - for each law:
    - Rules restricting actions
    - Conditions definitions
    - Terms for license conditions
  - License template
    - Outline/structure of license
  - License terms
    - Text
    - Conditions under which to include
  - Repository specifics
    - Which laws are in scope
    - Repository conditions satisfied
- **Input**:
  - Legal interview
    - Questions
    - Affirmations
  - Transaction specific
    - Name of repository, name of recipient, dates of study, ...
- Output:
  - Human-readable license supporting core actions
    - Deposit Agreement
    - Repository TOS (Retention & Destruction)
    - Release (Access & Use)

Data Release Example License

DATA LICENSE AGREEMENT

This Agreement is made and entered into by and between [repository:supplied:name] (hereafter the “Data Provider”) and [dataUser:supplied:name] (hereafter the “Data Recipient”) and establishes procedures relating to an exchange of data (hereafter the “Data”) between the Parties in a manner consistent with [TERMS:RELEVANT LAWS].

[TERMS:PREAMBLE]

The Parties agree to the following terms and conditions:

[TERMS:DEFINITIONS]

Period of agreement. This Agreement shall begin on [repository:supplied:startDate], or date of execution, whichever is later, and end on [repository:supplied:endDate], unless terminated in writing by either Party.

License. Subject to the terms of this agreement, the Data Provider grants to the Data Recipient, and the Data Recipient accepts, a non sub-licensable, non-assignable, non-transferable, non-exclusive license to use the Data. No ownership interest in the Data is bestowed to the Data Recipient pursuant to this Agreement. The License shall be automatically voided and terminated without notice to Data Recipient if Data Recipient violates the terms of this Agreement.

The Data Provider may terminate the License at any time and for any reason.

[TERMS:USE]

Nothing in this Agreement may be construed to allow either Party to maintain, use, disclose, or share personal information in a manner not permitted under federal or state law.

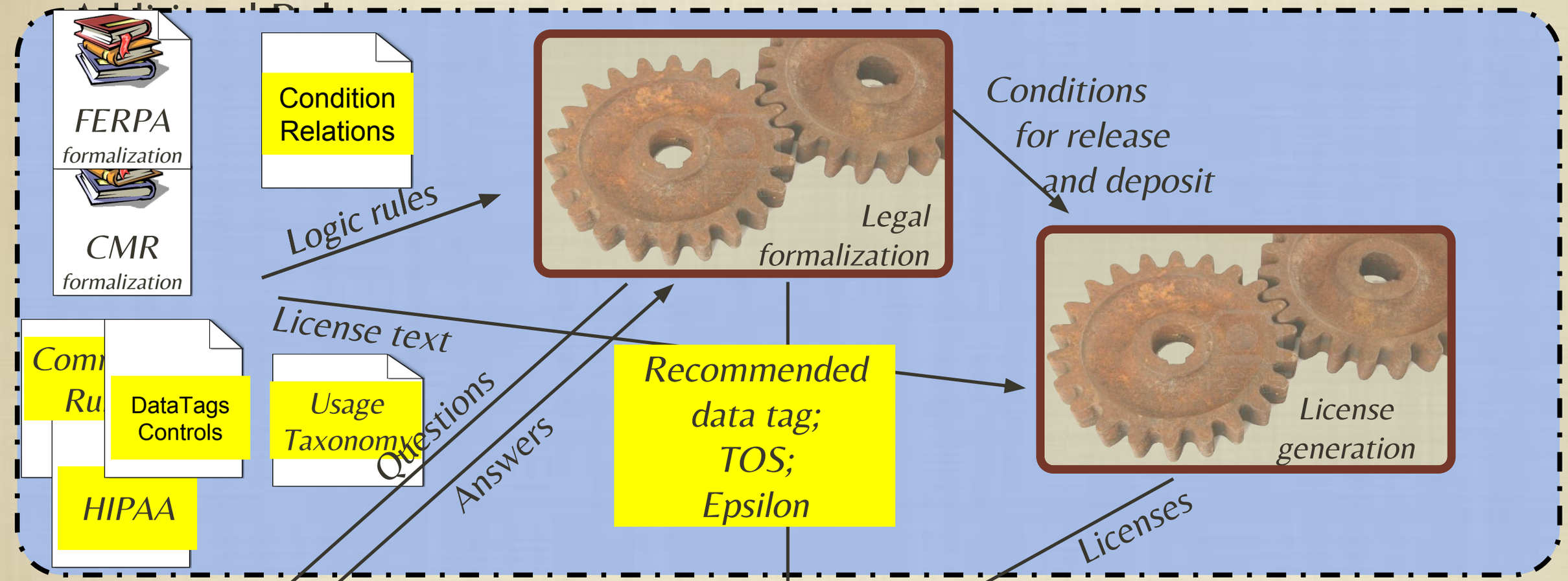
Permitted uses. Data Recipient is granted a nonexclusive, revocable license to have and use the Data provided the Data Recipient shall comply with all of the terms and conditions of this License.

[TERMS:PERMITTED USES]

...

Future work

- Additional rulesets
- Modular interoperable licence design
- Enabling differentially private releases
- From prototype to toolkit



Additional Rulesets

- Laws & Regs
  HIPAA; Common Rule; GDPR
- DataTags Controls
  - Incorporating assigned security controls in licenses
  - Inferring required security controls from legal requirements
- Local Practices
  Repository specific practice; Industry practice

Action	Actor	Rule
Use & Transfer Rules		
1. Deposit	Data Depositor → Repository	IF data:HIPAA:inScope AND data:HIPAA:DeIDsafeharbor THEN PERMITTED (Deposit(dd,ds,r,c)) IF data:HIPAA:inScope AND data:HIPAA:ExpertDeID THEN PERMITTED (Deposit(dd,ds,r,c)) IF data:HIPAA:inScope AND data:HIPAA:limitedDataSet AND data:HIPAA:DUA THEN PERMITTED (Deposit(dd,ds,r,c)) IF data:HIPAA:inScope AND data:HIPAA:decedentsexception THEN PERMITTED (Deposit(dd,ds,r,c)) IF data:HIPAA:inScope AND data:HIPAA:preparatory THEN PERMITTED (Deposit(dd,ds,r,c)). IF data:HIPAA:inScope AND data:HIPAA:waiver THEN PERMITTED (Deposit(dd,ds,r,c)). IF data:HIPAA:inScope AND data:HIPAA:businessassociatecontract THEN PERMITTED (Deposit(dd,ds,r,c)). If data:HIPAA:inScope AND NOT data:HIPAA:waiver AND NOT data:HIPAA:businessassociatecontract AND NOT data:HIPAA:preparatory AND NOT data:HIPAA:decedentshistorical AND NOT data:HIPAA:decedentsrecent AND NOT data:HIPAA:limiteddataset AND NOT data:HIPAA:ExpertDeID AND NOT data:HIPAA:DeIDsafeharbor THEN DENIED (Deposit(dd,ds,r,c)).

Enabling Differential Private

- Constructs for reasoning over derived data
- Baseline epsilon (practices rulesets)
- Epsilon in law-specific ruleset

RepositoryPractice:SufficientEps=.0001

IF **derivedFrom**(*ds*, _, *tool*) AND **dpComputation**(*RepositoryPractice:SufficientEps*, *tool*) THEN **diffentiallyPrivate**(*RepositoryPractice:SufficientEps*, *tool*))

IF **data:RepositoryPractice:inScope**(*ds*) AND **differentiallyPrivate**(*RepositoryPractice:SuffientEps*, *ds*) THEN **permits**(*RepositoryPractice*, **Release**(*ds*))

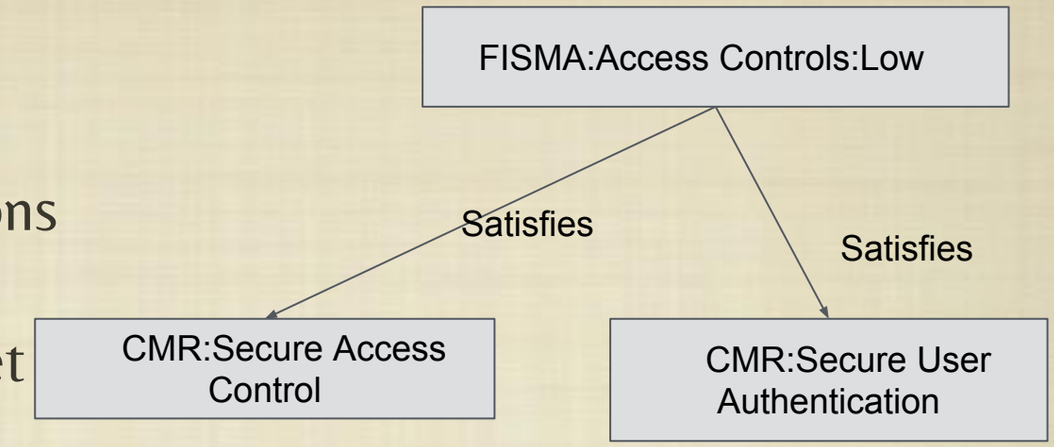
K. Nissim, A Bembenek, A Wood, M Bun, M Gaboardi, U. Gasser, D O'Brien, T Steinke, and S. Vadhan. 2016. "Bridging the Gap between Computer Science and Legal Approaches to Privacy." In Privacy Law Scholars Conference. Privacy Law Scholars Conference, Washington D.C., 2016

Kobbi Nissim, Thomas Steinke, Alexandra Wood, Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, David O'Brien, and Salil Vadhan. 3/2017. Differential Privacy: A Primer for a Non-technical Audience.

Interoperable license design

Now: Layered rulesets

- Each ruleset has scoping condition
- Each ruleset encodes restrictions on common actions
- Conditions are ruleset-specific
- Actions are denied if denied by **any** in-scope ruleset



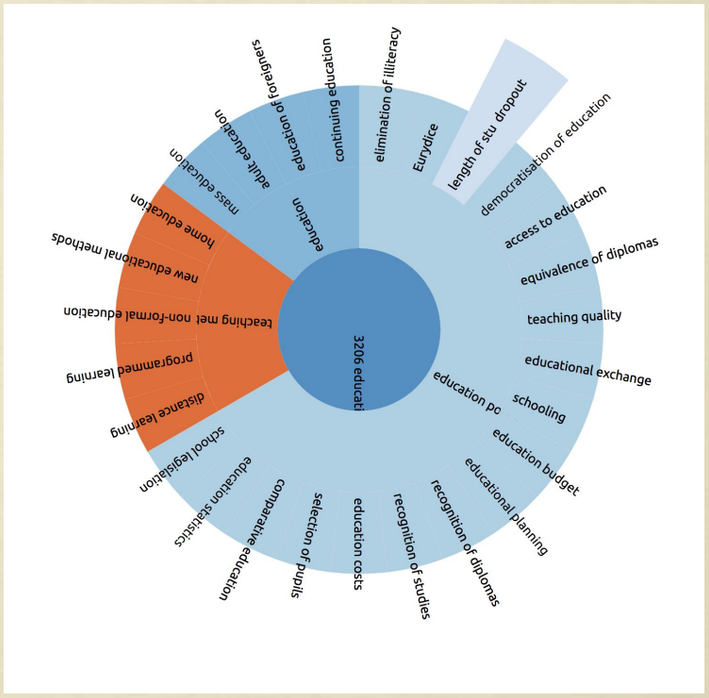
Future interoperability goals

- Third parties may use the data without seeking case-by-case authorization from rights holders
- Reasoning over derivatives and combination works
- Construct machine-actionable DUA's, consent, etc.

Interoperability Approaches

- Common conditions on derived data
- Individual Condition substitution
- Taxonomies
- Iterative case implementation & validation

Condition Substitution



Use Domain Taxonomy

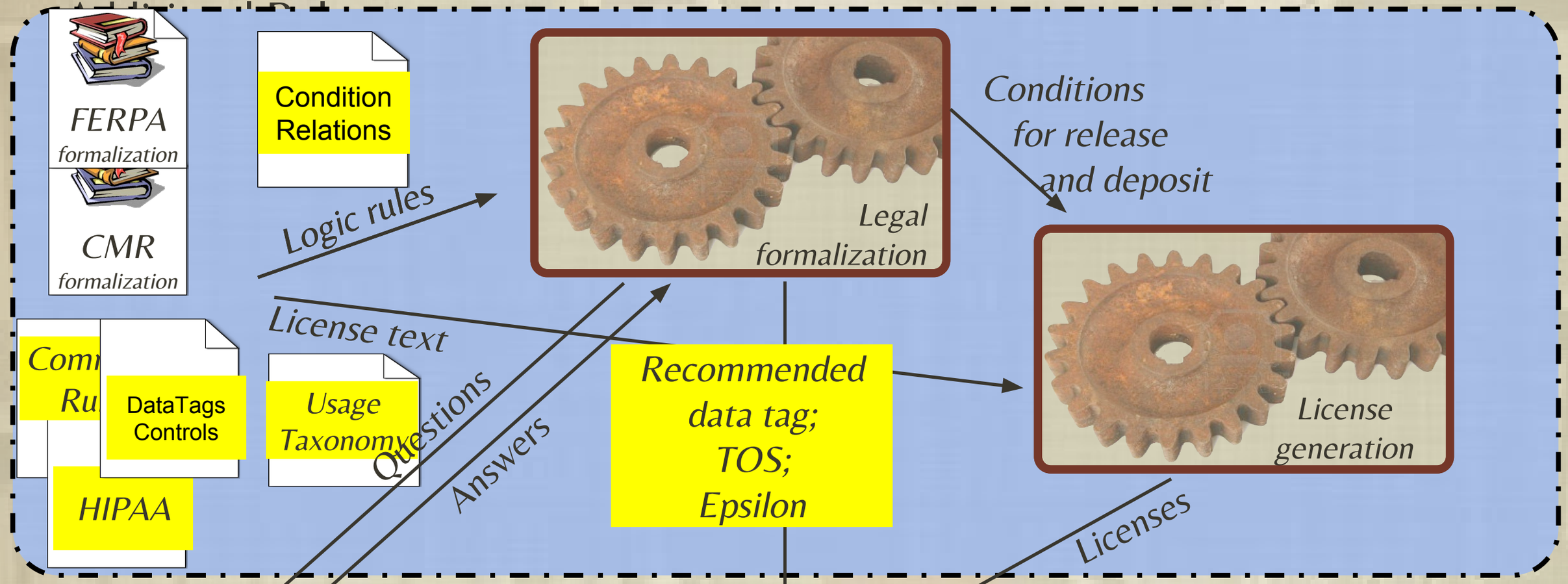
Related Work

- K. Nissim, A Bembenek, A Wood, M Bun, M Gaboardi, U. Gasser, D O'Brien, T Steinke, and S. Vadhan. 2016. “Bridging the Gap between Computer Science and Legal Approaches to Privacy.” In Privacy Law Scholars Conference. Privacy Law Scholars Conference, Washington D.C., 2016
- Kobbi Nissim, Thomas Steinke, Alexandra Wood, Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, David O'Brien, and Salil Vadhan. 3/2017. Differential Privacy: A Primer for a Non-technical Audience.
- Wood, O'Brien & Gasser, “Contractual Approaches to Research Data Sharing: A Comparative Analysis of Institutional Data Sharing Agreements and Policies”, Working Paper (forthcoming)

Work by Others

- “The British Nationality Act as a Logic Program” Sergot et al. (1986)
- “Privacy and Contextual Integrity: Framework and Applications” Barth et al. (2006)
- “Privacy APIs: Access control techniques to analyze and verify legal privacy policies” May at al. (2006)
- “Analyzing regulatory rules for privacy and security requirements” Breaux and Anton (2008)
- “Reasoning about conditions and exceptions to laws in regulatory conformance checking” Dinesh et al. (2008)
- “A formalization of HIPAA for a medical messaging system” Lam et al. (2009)
- “Experiences in the Logical Specification of the HIPAA and GLBA Privacy Laws” DeYoung et al. (2010)
- “PriCL: Creating a Precedent. A Framework for Reasoning about Privacy Case Law” Backes et al. (2015)
- “A Semi-Automated Methodology for Extracting access control rules from the European Data Protection Directive” Faterna et al. (2016)

Advancing Robot Lawyers:
Towards Interoperable Protected Research Data



Demo of License Generation