PROGRAM MATERIALS
Program #3236
February 17, 2022

Shields Up!
Protecting Private Data in eDiscovery

- Dr. Gina Taranto - ProSearch
- Ryan Costello, Esq. - ProSearch

Celesq® AttorneysEd Center
www.celesq.com

5255 North Federal Highway, Suite 100, Boca Raton, FL 33487
Phone 561-241-1919

PROSEARCH

Shields Up!

Protecting Private Data in eDiscovery

A ProSearch CLE Presentation – February 17th, 2022

Shields Up!

Protecting Private Data and Information in eDiscovery:

An inside look at the latest challenges in protecting private data in discovery, including some best practices and practical steps for supporting your eDiscovery battle plans.

ProSearch Presenters:

- Ryan Costello, Esq., CIPP/E/US: Head of Data Privacy Engagements
- Dr. Gina Taranto, PhD: Director of Applied Sciences

Session Contents

- General Housekeeping and Introduction
- Overview: discovery and data privacy law in 2022
- Greater scrutiny for cross-border discovery?
- Protecting personal data in discovery
- Q & A session
- Conclusions

Gina Taranto, PhD

Director of Applied Sciences

Dr. Taranto leads research and innovation of accelerated learning solutions by directing multidisciplinary teams of technologists, subject matter experts, and data scientists to train the technologies that replicate human decisions. She built the ProSearch Linguistics, Analytics, and Data Science group and oversees the design and implementation of search and automated document review solutions. She is recognized for her expertise in the range of Technology Assisted Review (TAR) technologies, including the application of TAR to emerging issues in protecting private information. She is a published author in the fields of linguistics and information retrieval and frequently presents on topics related to leveraging technology to meet the challenges of discovery.

She received her B.A. from Kresge College at the University of California, Santa Cruz, with honors, and her M.A. and Ph.D. from the University of California, San Diego.

Ryan Costello Esq., CCIP/E/US

Head of Data Privacy

A US-licensed attorney and expatriate based in Europe for more than 10 years, Ryan has cultivated an expertise in data protection and data privacy compliance across a career in eDiscovery and litigation support. With a particular interest in the area where cross-border discovery and data protection intersect, Ryan has worked with a myriad of clients to manage EU-based eDiscovery exercises while navigating data protection compliance challenges on both sides of "The Pond." With the implementation of the General Data Protection Regulation (GDPR) amidst other changes in the regulatory context, Ryan has assisted organizations in remediating cross-border discovery risks at every turn, with an eye toward solutions that utilize best practice technical and organizational measures, data management solutions and innovative technologies. Ryan assists across a range of client engagements, with a focus on assessing protective controls for personal data across the lifecycle of the EDRM. He is also a frequent writer and speaker on the GDPR, as well as data protection compliance topics and challenges in the US and across the globe. Ryan received his BA in English and Communications from Elon University, and his JD from Western New England University.

in

eDiscovery and Privacy Law

Relevant case law and background

The General Data Protection Regulation (GDPR)

– A Primer

Rules for data protection and the free movement of personal data

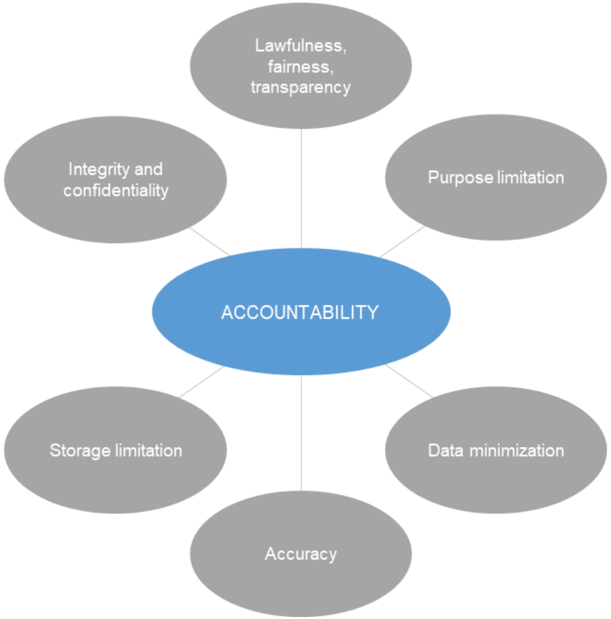
- Consistent across the EU
- Extra-territorial scope
- Significant enforcement potential for DPAs
- Accountability for data controllers AND processors
- Emphasis on rights of data subjects
- Technical and organizational measures for compliance
- Minimal prescriptive guidelines

Personal data: any information related to an identified, or identifiable, natural person

- Opinions, discussion, reviews of a data subject may constitute personal data.
- Special categories of personal data require explicit consent for processing

Processing: any operation or set of operations which is performed on data or data sets

- Retention, preservation, or archiving of data would amount to processing



The GDPR's Accountability Requirement

Processing Principles – Article 5(1)

- Lawfulness, fairness, transparency
- Purpose limitation
  - Personal data collected for specific, explicit and legitimate purposes
- Data minimization
  - Data is adequate, relevant, and limited to what is necessary for the purposes
- Accuracy
- Storage limitation
- Integrity and confidentiality
  - Appropriate technical and organizational measures for security
  - Safeguard data subjects' rights and freedoms

Discovery and Privacy Law: 2022

Post-pandemic surge in legal claims and disputes

- EU Commission/DPA’s – increased sensitivity of risks to the rights and freedoms of data subjects
- Germany – Presidency of Council of the European Union from June 1st 2020
- Greater scrutiny for cross-border discovery?
- Privacy Shield still an open question for transfers

EU – US Cross-border discovery

Aerospatiale – comity analysis for international discovery

1. Importance of the discovery to the litigation
2. The specificity of the request
3. Whether the information sought originated in the US
4. The availability of alternative means to obtain the information
5. Whether the foreign jurisdiction's interest in maintaining confidentiality or data privacy outweighs US interests in discovery

EU – US Cross-border discovery

Aerospatiale's 5-factor test largely favors discovery

- Not a single court has ever excused compliance with a discovery request based on EU data protection objections
- Minimal enforcement risk
- Few major fines since GDPR went into effect
- Data breach (Art. 32) most active area of enforcement
- No DPA cases that deal squarely with discovery in litigation

Potential shift with enforcement

Is there a real risk of prosecution for complying with US discovery order?

- In *re Mercedes-Benz* court: “…whether an EU authority aggressively polices this type of data production in the context of pre-trial discovery in US litigation remains to be seen.”
- *Behrens v. Arconic, Inc.* (E.D. Pa. Mar. 13, 2020): Discovery under Hague Convention where French Blocking Stature applies
- Requirements for producing GDPR-protected data subject to change

Is a Protective Order Enough?

Onus on litigants to ensure adequacy of protections

- *Finjan, Inc. v. Zscaler, Inc. (N.D. Cal. 2019):* Production of UK custodian emails does not violate GDPR, due to protective order, limited search terms, and direct relevancy of data
- *Vancouver Alumni Asset Holdings, Inc. v. Daimler AG (C.D. Cal. 2019):* Defendant to produce unredacted documents, subject to Protective Order
- *In re Mercedes-Benz Emissions Litigation (D.N.J. Jan. 30, 2020):* Protective Order sufficiently balances the EU's interest in protecting its citizen's private data and the US legal system's broad discovery provisions

Accountability in Discovery

GDPR’s impact on US discovery is case-by-case

- *Mercedes-Benz* and *Finjan* likely to be working precedent: Protective Orders applicable where limited, relevant, “benign” personal information at issue
- *Vesuvius USA Corp. v. Phillips (Ohio, June 2020):* Supreme Court denied petition for certiorari
- Incorporating accountability-focused solutions into eDiscovery protocols is critical

Tech solutions in discovery

Battling privacy risk in discovery through innovation and AI

Interim Recap
The Regulatory Landscape

Privacy regulations are simultaneously

- Broad and complex
- Vague and under-defined

Privacy regulations are intended to satisfy a range of purposes

- Protecting individuals’ rights to privacy
- Safeguarding minors
- Protecting consumers and therefore markets

Technology Considerations

- Leverage well-established, best practices for discovery where possible
- One option: treat privacy with the same gravitas given to privilege
- Features to consider when evaluating tools and workflows

Protecting Privacy and Privilege

General, non-technical definitions illustrate some common themes:

- Privacy: a personal choice about whether to disclose information
- Privilege: a legal rule prohibiting the disclosure of private information

In terms of legal discovery, protecting privacy and privilege require:

1. Identifying information in a discoverable data set that requires protection
2. Protecting that information

Practical Considerations

Leverage familiar technology to develop a sound workflow

Recruit tools familiar from protecting privilege in order to protect private information, and focus on putting in place the workflows, tools, and technology that:

1. Identify private information
2. Recognize and extract private information
3. Annotate/label different types of private information

1. Identify private information

Document Review for privilege (as well as for responsiveness) has armed discovery professionals with many tools to assist with identifying private information, and for sampling and measuring techniques to **assess the performance of any solution.**

- Search Terms
- Regular Expressions
- AI/Deep Learning Techniques

Being prepared in the context of *identifying* private information means understanding how the tools you learn work, and ensuring that any workflow builds in the resources required to handle false-positives or provide custom training to off-the-shelf classifiers

1. Identify private information

Traditional Techniques include using search terms or regular expressions. These techniques are known for having high recall – meaning that they are likely to a good job of finding private information – and low precision – meaning that a lot of information that actually isn't private information might be returned by the search. For example, many sixteen digit numbers are in fact credit card numbers, but there are many sixteen digit numbers that are not credit card numbers.

Relying solely on search terms and regular expression will likely require lots of document review to distinguish "actual" private information from "possible" private information.

Search Terms

Broad

- "social", "social security number"
- "date of birth", "birthdate", "birthday"
- "credit card", "credit card number", "cc"

Narrow

- Specific Names
- Known SSNs, Credit Card Numbers

Regular Expressions

- More flexible than search terms, can capture broad ranges of patterns: SSN, CC#, NIN, VIN, etc
- Still potentially very broad
- Performance can be impacted by limitations of search engines

1. Identify private information

More recently developed techniques leverage an approach that is gaining wider acceptance in discovery called *deep learning*. Deep learning algorithms can be used to build models that evaluate complex co-occurrences of multiple features to make a prediction about a piece of information. In the context of identifying private information, deep learning models can be trained to analyze images, text, and even images that contain text. In the context of legal discovery, these techniques can analyze, for example, how likely a sixteen digit number is to be a credit card number, or whether an image is a passport, or a form with handwritten information.

When evaluating deep learning tools, be sure to consider whether and how much training time will be required.

Natural Language Processing (NLP)

NLP models parse identify and leverage linguistic features:

- Parts of speech (noun, verb, adjective)
- Named Entities (real world objects)
  - Persons (Bela, Eloise, Lucia, Mila)
  - Geopolitical Entities (UK, France, Japan)
  - Organizations (IRS, Google, Delta Airlines)
  - Dates (February 17th, 2/17/2022, Thursday)
  - Numbers (22, 6893578078212, four)

Computer Vision (Visual Document Classification)

An “eye” for distinct visual cues of documents

- Size, shape
- Location in a frame
- Structures in an image (points, edges, curves)
- Variations in color/tone

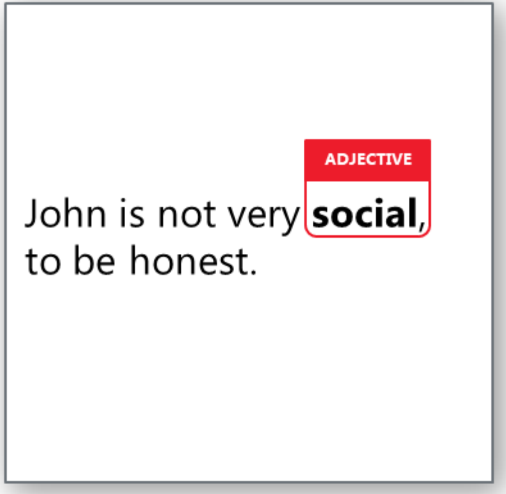
Can find private information that text-based models miss

- Poorly extracted text
- Images that constitute private information (scans of credit cards, IDs, checks)

Natural language processing

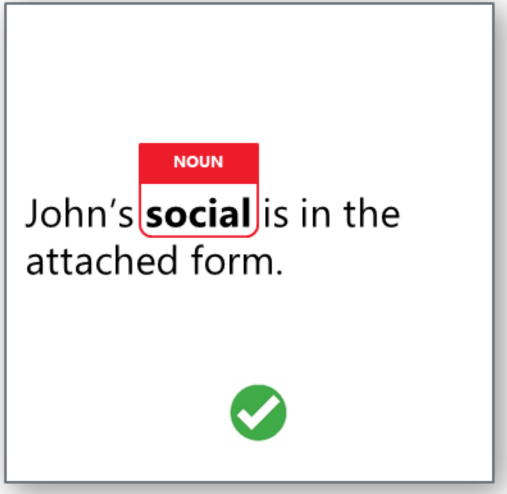
‘Understand’ text and extract various linguistic features

- Parts of speech (noun, verb, adjective, etc.)



John is not very social, to be honest.

Doc 1

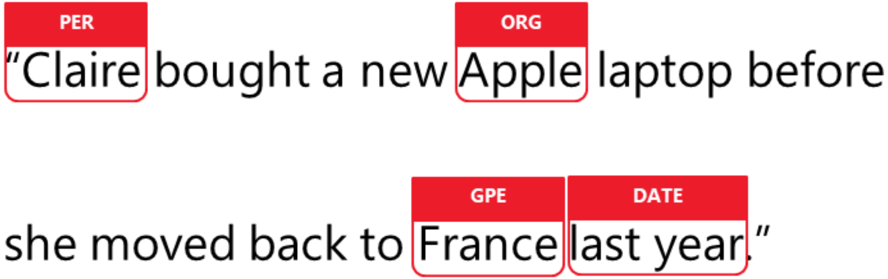


John’s social is in the attached form.

Doc 2

Language processing applied

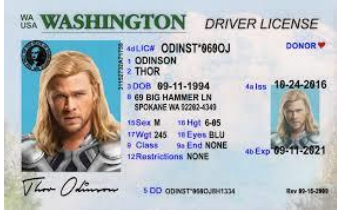
Understanding how language works



"Claire bought a new Apple laptop before she moved back to France last year."

Computer Vision Technology

An “eye” for distinct visual cues of documents



WA USA WASHINGTON DRIVER LICENSE

4d LIC# ODINST*969OJ

DONOR ❤

1 ODINSON

2 THOR

3 DOB 09-11-1994

4a Iss 10-24-2016

8 69 BIG HAMMER LN
SPOKANE WA 92202-4349

15 Sex M 16 Hgt 6-05

17 Wgt 245 18 Eyes BLU

9 Class 9a End NONE

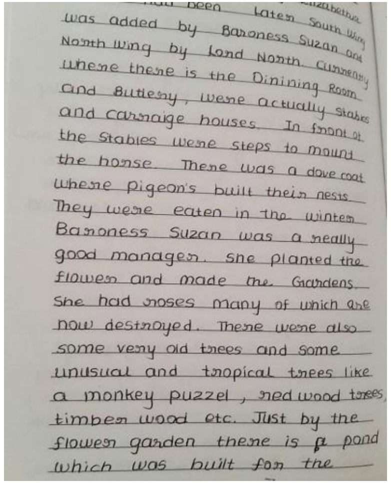
4b Exp 09-11-2021

12 Restrictions NONE

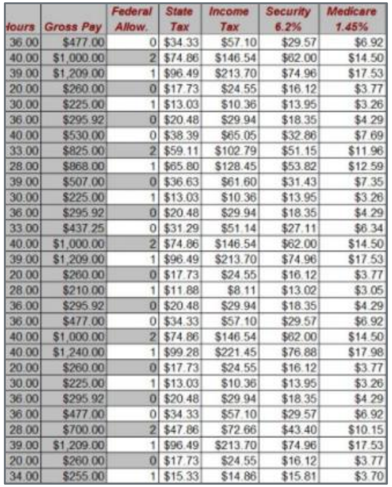
Thor Odinson

5 DD ODINST*969OJBH1334

Rev 09-16-2000



been later izabethan
was added by Baroness Suzan South Wing and
North Wing by Lord North. Currently
where there is the Dinining Room
and Buttery, were actually Stables
and carraige houses. In front of
the Stables were steps to mount
the horse. There was a dove coat
where pigeon's built their nests.
They were eaten in the winter
Baroness Suzan was a really
good manager. she planted the
flower and made the Gardens.
She had roses many of which are
now destroyed. There were also
some very old trees and some
unusual and tropical trees like
a monkey puzzel, red wood trees,
timber wood etc. Just by the
flower garden there is a pond
which was built for the



		Federal	State	Income	Security	Medicare
Hours	Gross Pay	Allow.	Tax	Tax	6.2%	1.45%
36.00	$477.00	0	$34.33	$57.10	$29.57	$6.92
40.00	$1,000.00	2	$74.86	$146.54	$62.00	$14.50
39.00	$1,209.00	1	$96.49	$213.70	$74.96	$17.53
20.00	$260.00	0	$17.73	$24.55	$16.12	$3.77
30.00	$225.00	1	$13.03	$10.36	$13.95	$3.26
36.00	$295.92	0	$20.48	$29.94	$18.35	$4.29
40.00	$530.00	0	$38.39	$65.05	$32.86	$7.69
33.00	$825.00	2	$59.11	$102.79	$51.15	$11.96
28.00	$868.00	1	$65.80	$128.45	$53.82	$12.59
39.00	$507.00	0	$36.63	$61.60	$31.43	$7.35
30.00	$225.00	1	$13.03	$10.36	$13.95	$3.26
36.00	$295.92	0	$20.48	$29.94	$18.35	$4.29
33.00	$437.25	0	$31.29	$51.14	$27.11	$6.34
40.00	$1,000.00	2	$74.86	$146.54	$62.00	$14.50
39.00	$1,209.00	1	$96.49	$213.70	$74.96	$17.53
20.00	$260.00	0	$17.73	$24.55	$16.12	$3.77
28.00	$210.00	1	$11.88	$8.11	$13.02	$3.05
36.00	$295.92	0	$20.48	$29.94	$18.35	$4.29
36.00	$477.00	0	$34.33	$57.10	$29.57	$6.92
40.00	$1,000.00	2	$74.86	$146.54	$62.00	$14.50
40.00	$1,240.00	1	$99.28	$221.45	$76.88	$17.98
20.00	$260.00	0	$17.73	$24.55	$16.12	$3.77
30.00	$225.00	1	$13.03	$10.36	$13.95	$3.26
36.00	$295.92	0	$20.48	$29.94	$18.35	$4.29
36.00	$477.00	0	$34.33	$57.10	$29.57	$6.92
28.00	$700.00	2	$47.86	$72.66	$43.40	$10.15
39.00	$1,209.00	1	$96.49	$213.70	$74.96	$17.53
20.00	$260.00	0	$17.73	$24.55	$16.12	$3.77
34.00	$255.00	1	$15.33	$14.86	$15.81	$3.70

2. Recognize & extract private information

Jurisdictional variation and the evolving interpretation of what counts as private information creates a special challenge when designing workflows for protecting private information. Tools that can identify the type, or category, of private information at a granular level can be called upon in workflows when required.

NLP techniques that can be used to classify bits of information or entities in unstructured text into predefined categories are sometimes called

- Entity identification
- Entity extraction
- Named Entity Recognition (NER)

Tools that include these techniques can help to differentiate between e.g., dates of birth, dates of death, dates of injuries or medical procedures, and dates on an email, and allow each of these to be treated differently in terms of if or how they should be protected.

3. Field or annotate private information

Tools that can identify and recognize private information at a granular level can often annotate that bit of data with it's corresponding label, or field that data into a document review layout. Depending on the discovery need, this can be an invaluable asset. The ability to field information as metadata in a review platform helps reviewers and downstream tools assist in any review, treatment, or QC workflow required as part of a response that protects private information.

Conclusion

Considerations for mitigating data privacy risks in discovery

- Shifting case law and regulations bring uncertainties at-home and abroad
- Accountability: onus on litigants to ensure adequacy of protections
- Search terms and regular expressions may not be enough
- Solutions including natural language processing and computer vision can be game-changers

PROSEARCH

Thank You

For more information visit
www.prosearch.com

or contact a ProSearch expert

PROSEARCH

ProSearch Privacy Suite

A comprehensive approach to identifying PXI.

The ProSearch Privacy Suite uses game-changing deep learning models built on advances in natural language processing and computer vision techniques to identify protected private information.

The right of an individual to data protection and privacy has been addressed by governments and regulatory agencies around the globe with increasing attention and enforcement. ProSearch offers a robust solution for identifying **Personally Identifiable Information** (PII), **Protected Health Information** (PHI), and information subject to **Payment Card Industry Data Security Standards** (PCI-DSS) as well as other private, sensitive, or personal information that is protected by evolving legislation.

How Does it Work?

- Natural language processing techniques move beyond search terms and regular expressions by leveraging statistical models and linguistic features
- Computer vision techniques pick up where NLP stops – identifying scans and facsimiles of documents that contain protected information
- The breadth of expertise at ProSearch brings together data scientists, linguists, and discovery professionals to implement workflows that are optimized for specific PXI identification and response requirements
- Results are integrated into custom Relativity review panels and dashboards to provide easy access to actionable metrics and streamline review workflows
- Ongoing monitoring, maintenance, and training of the deep learning models on your data enable customization and continuous incremental improvement

Benefits Summary

Integrated
Results are accessible in Relativity and seamlessly integrate into review workflows

Specific
Beyond simple document level identification, extraction of PXI values and targeted highlighting support detailed reporting and enhance the reviewer experience

Flexibility
Supported by people, process, and technology, our models and workflows are customizable to a variety of needs related to identifying and protecting PXI

Insight
Detailed reporting about the volumes, types, and sources of PXI in enterprise data supports information governance and data retention planning to minimize future risk

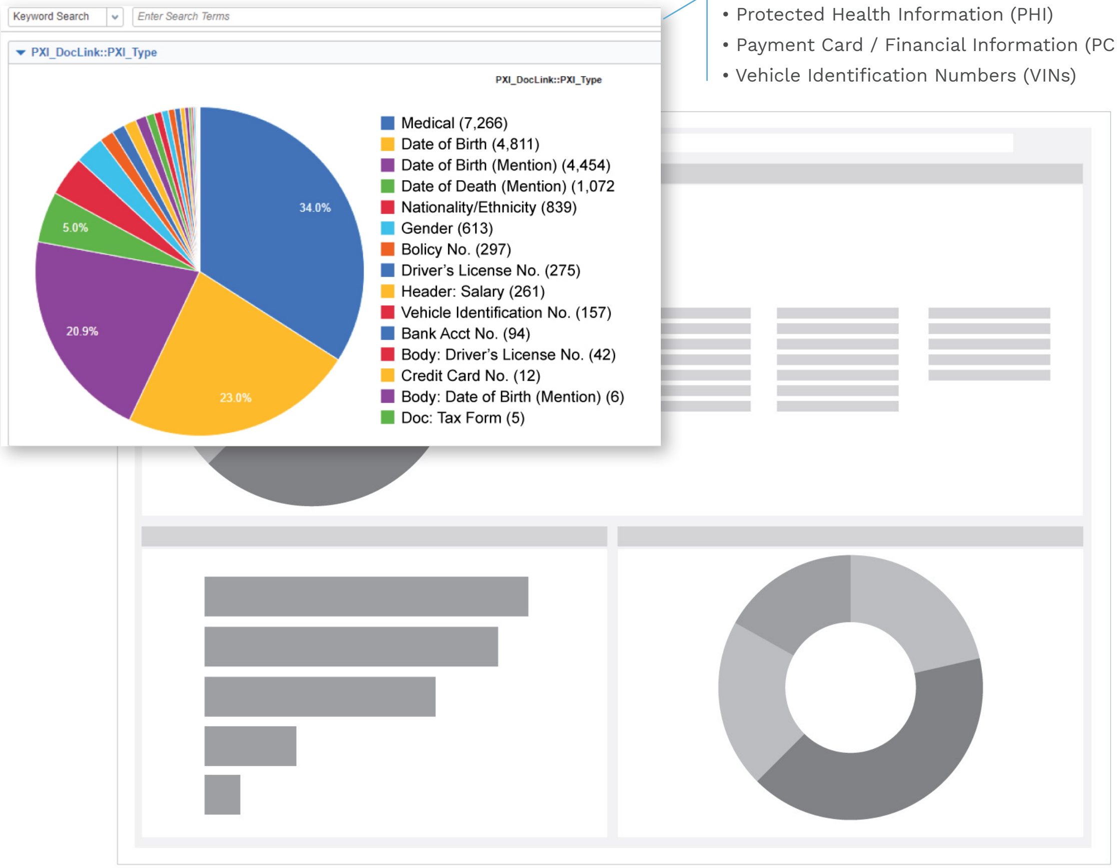
Cost Savings
Superior precision and detailed reporting support maximizing automated next steps and effective workflow planning

Comprehensive PXI Identification

ProSearch Privacy Suite's text and image classifiers identify over thirty types of personal and private information to support compliance with evolving regulation and requirements.

ProSearch Privacy Suite identifies a broad range of PXI, including:

- Personally Identifiable Information (PII)
- Protected Health Information (PHI)
- Payment Card / Financial Information (PCI-DSS)
- Vehicle Identification Numbers (VINs)



Screenshot of ProSearch Privacy Suite's PXI Identification tool

ProSearch Privacy Suite uses game-changing deep learning models built on advances in natural language processing and computer vision techniques to identify protected private information.

LJN
LAW JOURNAL NEWSLETTERS

Cybersecurity
LAW & STRATEGY®

An ALM Publication

SEPTEMBER 2021

How to Address Evolving Privacy Regulations During Discovery

One Recipe for Success: Treat Private Data With the Same Priority Given to Privilege

But for all the coverage that privacy regulations are meant to provide, there is precious little guidance about how to protect private information, and there is very little legal precedent to guide our practices.

By Gina Taranto

In the face of growing data privacy regulations, appetites are growing for solutions that guard protected private information with the same gravitas given to protecting privilege. The challenge? Current privacy regulations are, on the one hand, broad and complex, and on the other, vague and underdefined. This makes implementing systems for preventing the inadvertent disclosure of protected private information an important task.

Jurisdictions around the globe have introduced legislation to serve many similar, but sometimes very different purposes. These range from preserving an individual's most basic right to privacy and the protection of their personally identifiable information (PII), such as with GDPR and CPRA, as well as their private health information (PHI), as with HIPAA. Some legislation is enacted to safeguard minors, while other legislation protects people in their role as consumers. Overlapping with this is the Payment Card Industry Data Security Standard (PCI-DSS).

But for all the coverage these regulations are meant to provide, there is precious little guidance about *how* to protect private information, and there is very little legal precedent to guide our practices. It was only in 2020 that the earliest court decisions began to address these issues and establish precedent.

Further complicating the matter is the sheer number of stakeholders around the general protection of private information in any given organization. A business process to manage the protection of private information could involve any combination of IT, IG, compliance, security, legal and discovery departments. Practically speaking, solving the problem of data protection requires a cross-functional communication plan that is as robust as the actual data protection plan.

But setting the complex communication and planning piece aside, what is an investigator, attorney, or other discovery professional to do when private information requires protection today, during a current litigation or investigation happening now, in the absence of clear guidance?

The situation of being prepared to handle the unique requirements of a specific matter has marked similarities to the challenge culinary professionals face: delivering a range of meals in a variety of styles and flavors to satisfy diverse appetites and specific dietary requirements. The basic technique that kitchens use to manage uncertainty is exactly what we need right now: an emphasis on preparation, or "prep." The uncertainty we face today in responding to our obligations to protect private information is the discovery equivalent of taking charge of our mise-en-place.

Mise-en-place is the French culinary term for having everything in its proper place so that you are ready to go when it's time to start cooking. And in the context of discovery for litigation or investigation, that would be organizing both your data and your discovery toolkit — as your ingredients and utensils — so that when there comes a need to transfer data your response can be swift, complete and accurate. It also ensures the ability to meet the demands of even the pickiest legal procedure or regulating body, localized as necessary to a specific jurisdiction in the matter at hand.

Practically speaking, a strong first pass at a mise-en-place for discovery will include leveraging the technology and workflow familiar from document review for responsiveness and for privilege. When evaluating solutions for managing private information, focus on putting in place the workflows, tools and technology that will allow for:

1. Identifying private information.
2. Recognizing and extracting private information.
3. Annotating/labeling different types (or flavors) of private information.
4. Recruiting the tools familiar from protecting privilege to protect private information.

Identifying Private Information

Document review for responsiveness as well as for privilege has armed discovery professionals with many tools to assist with *identifying* private information. And, perhaps more importantly, through document review we have deep knowledge about the sampling and measurement techniques that can measure the performance of assistive technology. While search terms and regular expressions can be used to ensure broad recall, deep learning models and visual document classification are valuable tools when it comes to minimizing false positives.

Preparing could mean investing in customizing portable data assets — testing off-the-shelf tools designed to identify private information and assessing their performance against samples of your organization's data. You might begin enhancing models in advance of an active matter, or you might just use this information to build time to test models in the context of a live matter and develop a project plan accordingly.

Preparedness in the context of identifying private information means understanding how the tools you use work, so any workflow can build in the resources required to handle false positives or provide additional training to off-the-shelf classifiers to expand coverage on a per-matter basis.

Recognizing Private Information

One of the challenges in defining workflows to protect private information is the jurisdictional variation and the evolving interpretation of regulations. Beyond simply identifying that a document is likely to contain protected private information, real preparedness requires a solution that recognizes the type, or flavor, of private information. This granular level of detail is crucial to handling instances of private information appropriately in the context of a specific matter.

Entity identification, entity extraction and named entity recognition (NER) are some of the names given to a natural language processing technique that classifies bits of information or entities in unstructured text into predefined categories. Being prepared with a privacy solution that can recognize the types of private information your organization gathers can be invaluable.

For example, the ability to distinguish between dates of birth, dates of death, dates of injuries or medical procedures and dates on, say, an email or a financial record allows each be treated differently in terms of how they should be protected.

Annotating/Labeling Different Types (or Flavors) Of Private Information

Once private information has been identified and recognized (extracted), discovery practitioners should be prepared with a solution that annotates documents with the types of potentially protected information existing within them. Technology can do a lot of the work to automate this process, but in most cases will require human support to either train or QC the technology. For example, while assistive technologies can be very effective at identifying dates, distinguishing types of dates — dates of birth, dates of hire, dates of death, and document dates — can sometimes be tricky. The ability to accurately field such information as metadata in a review platform can help practitioners zero in on specific pieces of data or data combinations that require special treatment to protect private information.

Recruiting the Tools Familiar from Protecting Privilege to Protect Private Information

Once the private information is identified, the final step is implementing a solution to protect it. While there isn't much precedent for solutions to protect *private* information, we have plenty of precedent and best practices for what counts as legitimate protection of *privileged* information. Given the absence of more specific guidance, this appears as a good, principled place to start.

Privilege review provides us with mechanisms for redacting as well as withholding privileged information, and an understanding of when to use each. The familiar tools and workflows that help us redact, anonymize or withhold-and-log privileged information can assist us in quickly implementing workflows to protect private information using the same techniques.

In the United States, the number of state laws enacted to address private data sets up a scene of catering to various jurisdictions not unlike a dinner party for guests with diverse dietary requirements, where the host or chef must be prepared to provide various options for the same basic meal. From a few basic ingredients, say beans, salsa, lettuce, and avocado, different plates can be tailored to varying needs. For one guest/jurisdiction, wrap it all in a flour tortilla to make a burrito, while a second guest/jurisdiction is gluten-free, so you'll need put it in corn tortillas (tacos). A third guest/jurisdiction is paleo. Or is it keto? It's gluten and grain-free, so best to throw it on top of a bowl of quinoa. And a fourth guest/jurisdiction follows a low carb diet, so put it on a fresh bed of romaine to make a salad.

Oh, and people either love or hate cilantro, so be prepared for that.

The point is, having all the ingredients prepared and in place at the outset ensures the ability to respond to whatever request or dietary need that comes up. Likewise, ensuring proper attention to privacy regulations when responding to an investigation or litigation matter requires taking the time to organize your data and discovery toolkit first. Do your prep work and use the knowledge and tools you already have for protecting privilege information to treat private data with similar care.

*****

Dr. Gina Taranto*, ProSearch Director, Applied Sciences/Accelerated Learning Solutions, leads research and innovation of accelerated learning solutions by directing multidisciplinary teams of technologists, subject matter experts and data scientists to train the technologies that replicate human decisions. She has been developing teams and solutions in ediscovery for more than 15 years, with experience in the design and implementation of search and automated document review solutions for clients in the financial services, technology and pharmaceutical industries. Taranto received her B.A. from Kresge College at the University of California, Santa Cruz, with honors, and her M.A. and Ph.D. from the University of California, San Diego.*

PROSEARCH

ProSearch.com
info@ProSearch.com
877.447.7291

LJN LAW JOURNAL NEWSLETTERS

Cybersecurity
LAW & STRATEGY®

An ALM Publication

March 2021

A Balancing Act: Mitigating Data Privacy Risks in Cross-Border Discovery

The intersection of foreign laws governing data collection and cross-border discovery operations continues to be a potentially volatile conjunction.

By Ryan Costello

The intersection of foreign laws governing data collection and cross-border discovery operations continues to be a potentially volatile conjunction. Global enterprises have been cautioned to tread carefully when responding to U.S.-driven discovery requests, as expansive discovery exercises, so common in the U.S. under federal and state laws of civil procedure, can be completely foreign and often legally problematic in jurisdictions abroad.

Accordingly, discovery requests implicating custodians and data outside the U.S. can potentially put organizations in a Catch-22: either fall short of their discovery obligations on the one hand or fall afoul of legislation in other nations prohibiting or limiting data collection and transfer to the U.S. on the other. Laws potentially conflicting with discovery obligations include blocking statutes, requirements pertaining to works council agreements and, perhaps most significantly, data privacy regulations.

In particular, it has been EU data privacy regulations, including the General Data Protection Regulation and its predecessor the Data Protection Directive of 1995 that have threatened to pose the most significant potential roadblocks to discovery requests. Given the care with which personal data must be treated under the GDPR (security requirements, data minimization obligations, rights afforded data subjects), accountability for those handling such data and the regulatory and civil fines possible under the regulation, cross-border discovery across the EU seems to warrant an especially heightened level of scrutiny.

View in the Courts

U.S. courts do not view conflict with foreign laws as a de facto bar to discovery and generally will require discovery to proceed, notwithstanding data privacy laws or other foreign legislation that may stand in the way. Relying on a Supreme Court case from 1987, *Societe Nationale Industrielle Aerospatiale v. U.S. Dist. Court for S. Dist. of Iowa*, 482 U.S. 522, courts across the U.S. applying Aerospatiale's five-part balancing test or "comity analysis," which weighs the interests of foreign laws against U.S. discovery, almost always find that the U.S. legal process of pre-trial discovery takes precedent. In fact, as of this writing, there has not been a single case in the U.S. where a party was permitted to fully withhold production of documents based on foreign data privacy regulations.

Discovery can be **limited or curtailed** for factors such as undue burden or expense or even to protect trade secrets and intellectual property. However, the reason that such exemptions generally don't apply for conflict with foreign data protection and privacy laws is largely because the risk of enforcement has been so low.

Enforcement under the GDPR, while always possible, has indeed been limited. Unless courts see a real risk of prosecution for a company under foreign data privacy laws, they are **typically reluctant** to allow limitations to the discovery process or withholding of documents based on GDPR grounds alone. With more significant enforcement action in the future, this may change. For the time being, however, U.S. courts are likely to continue to stipulate that:

1. "[foreign] statutes do not deprive an American court of the power to order a party subject to its jurisdiction to produce evidence even though that act of production may violate that statute"; and

2. "the party resisting the discovery burden bears the burden [of proof] in these cases." *[Rollins Ranches, LLC et al v. Watson*, S.C. May 22, 2020]

Accordingly, while parties will almost assuredly be required to proceed with discovery, they also must consider the requirements of GDPR compliance back in the EU, as well as the specter of enforcement and/or civil proceedings for GDPR violations. So how have parties managed this transcontinental juggling act?

Solutions have comprised varying approaches including:

- Protective orders
- Redactions for anonymizing personal data in datasets
- A "privacy log" that accounts for certain documents withheld from production

Each of these has relative advantages and disadvantages; however, one crucial element underlies all approaches: seeking and identifying the breadth of personal information and personal data scattered across the datasets.

Protective Orders

Protective orders are court instructions included in production sets, such as "attorneys' eyes only," which are intended to protect against the proliferation of personal information or/sensitive data and reduce risk. U.S. courts will typically acknowledge such protections as **sufficient for GDPR compliance purposes** and allow discovery productions to proceed on that basis. However, whether or not such protective orders are sufficient from a European perspective for GDPR compliance remains to be seen.

In any event, a party will want to be clear on what documents contain what information in order to ensure that a protective order stands to reduce GDPR risks for any data subjects implicated. If challenged by an EU regulatory authority or data subject, the producing party must be able to show that it fully complied with GDPR via the protective order and can stipulate precisely what information, and what data belonging to whom, may have been produced.

Redactions

Another approach regularly taken involves redacting personal information in datasets wherever possible (provided the data is not relevant to the discovery request). Courts have, in some cases, allowed **redactions for personal information in discovery**, and it arguably offers even greater protection for EU data subjects, given that data is essentially anonymized.

However, redactions can be costly if applied manually by a review team, error-prone, time-consuming and technically difficult to achieve at scale. Innovative approaches for identifying personal information in datasets prior to, or as a part of, the review workflow can cut down on the expense and difficulty inherent in attempting redaction of personal information. Technical processes for identifying information, prioritizing review and setting aside a workstream that focuses on redaction of personal information and QC checks can be an effective and protective means for meeting discovery obligations and ensuring GDPR compliance.

The Privacy Log

The privacy log can be one of the most comprehensive means for protecting EU personal information. Such a log allows a court to examine a summary of documents that a party wishes to withhold based data privacy grounds. The court is then **better positioned to weigh the interests of discovery against data privacy concerns** on a more concrete basis and can allow for a measured approach that significantly minimizes the GDPR compliance risk.

There is a precedent for using privacy logs with similar processes involving documents withheld for privilege, including attorney-client privilege, bank examiners' privilege and other grounds, such as intellectual property concerns. However, as noted above, courts are reluctant to withhold documents from discovery based on GDPR concerns alone. Parties are therefore encouraged to be extremely precise in noting why documents should be withheld based on data privacy, and a full accounting of the personal data contained in documents and the potential risks to data subjects must be fully understood and indicated in the privacy log.

Further Considerations

Regardless of the approach a party chooses to minimize conflicts of law and impact on data subjects, narrowing the scope of personal data implicated in discovery will be critical at each step of the discovery process. The **Sedona Conference's International Principles on Discovery, Disclosure and Data Protection** (2017) offer great guidance to this end.

However, while this article has focused on the production phase of discovery, a sufficient understanding of personal data implicated in discovery from collection to review and on to production — or at each stage of the EDRM — is essential. With the explosion of chat collaboration tools and medical information related to the COVID-19 pandemic now proliferating datasets across organizations in varied industries, innovation and creative approaches to data privacy considerations and discovery are as valuable as ever.

Conclusion

Absent knowing what personal data may be present in a given dataset, it is difficult for parties to know how best to proceed in a manner that meets the obligations of both U.S. discovery and EU data protection requirements. Solutions and best practices for highlighting personal information implicated in a discovery set, in a manner that's efficient, reliable and cost-effective, have never been more important.

***Ryan Costello**, Esq., CIPP/E/US, is head of data privacy services at ProSearch, a leading provider of comprehensive discovery solutions to corporate legal departments and law firms. A U.S.-licensed attorney and expatriate based in Europe for more than 10 years, Costello has cultivated an expertise in data protection and data privacy compliance. He assists organizations in remediating cross-border discovery risks, utilizing data management solutions and innovative technologies.*

—❖—

APRIL 5, 2021

Legaltech© news

ONLINE

An ALM Publication

Facing New Privacy Regulations Head-On

BY GINA TARANTO, PROSEARCH

As more states enact new data privacy regulations, legal professionals are recognizing the need for new solutions for protecting private information with the same priority given to protecting privilege.

Preventing the inadvertent disclosure of protected private information is more important than ever, but the task is difficult. Current regulations are simultaneously broad, complex, vague and underdefined. Emerging legislation serves many similar but sometimes very different purposes. These range from preserving an individual's most basic right to privacy by protecting their personally identifiable information and private health information to safeguarding children and consumer protection.

For all the intentions behind these regulations, there is minimal guidance about *how* to protect private information, and there is very little legal precedent to guide our practices. It was only in 2020 that we began to see the early court decisions that serve as legal precedent.

Further complicating the matter is the sheer number of stakeholders around the general protection of private information. A business process to manage the protection of private data could involve stakeholders from IT, IG, compliance, security, legal and discovery departments. While cross-functional strategic plans to address the long-term management of private data may take time to come into focus, there are steps that can be taken now to help.

Overcoming the challenge of protecting private data in the context of discovery and compliance starts by taking a step back and getting prepared. Begin by organizing your data and your discovery toolkit, so that when the need to transfer data arises your response can be swift, complete and accurate. Advance preparation ensures the ability to meet the jurisdiction-specific requirements for the matter at hand.

Preparation should include leveraging technology and workflows familiar from document review for responsiveness and for privilege. When evaluating solutions for managing private information, focus on augmenting familiar workflows, tools and technology by incorporating modules that identify, extract and label private data.

1. Identifying private information: Document review for responsiveness and privilege have armed discovery professionals with search and information retrieval tools to assist with identifying private information. Extending this competency to identifying private information means understanding how the tools you use work and where a workflow can accommodate similar techniques designed to identify private information. It also means preparing for the resources required to handle false positives or provide additional training to off-the-shelf classifiers to expand coverage on a per-matter basis.

Preparing might include customizing portable data assets—testing off-the-shelf tools designed to identify private information and assessing their performance against samples of your organization's data. You might begin enhancing models in advance of an active matter, or you might build time into an active matter's project plan for evaluating and refining models.

2. Recognizing and extracting private information: Beyond just identifying documents likely to contain protected private information, real preparedness requires a solution that categorizes the *type* of private information. This granular level of detail is crucial to meeting specific requirements for any matter.

Entity identification, entity extraction or named entity recognition are some of the names given to a natural language processing technique that classifies bits of information or entities in unstructured text into predefined categories. Being prepared with a privacy solution that leverages NER technology can be invaluable, as distinguishing between dates of birth or death and dates on an email or a financial record may allow each to be treated differently in how they are protected.

3. Annotating/labeling different types of private information: Once private information has been identified or recognized, discovery practitioners should look for ways to automate the tracking of this important document metadata. The ability to field information about the type of private information as metadata in a review platform allows reviewers and technologies to take appropriate next steps during review, treatment or QC workflows as part of a data protection response.

4. Recruiting the tools familiar from protecting privilege to protect private information: While there isn't much precedent for solutions to protect *private* information, we have plenty of precedent and best practices for protecting *privileged* information. So this is a good place to start. Privilege review provides mechanisms for redacting, as well as withholding privileged information. Familiar tools and workflows that help us redact, anonymize or withhold-and-log privileged information can help us quickly implement workflows to protect private information using the same techniques.

While the continued introduction of state-level privacy regulations in the U.S. and around the globe may seem a challenge to stay ahead of, discovery professionals are well-poised to handle them. By leveraging the knowledge, workflows and tools already employed for privilege review, legal teams can be confident in their ability to respond to new privacy regulations.

Dr. Gina Taranto is the Director, Applied Sciences/Accelerated Learning Solutions at ProSearch. Gina leads research and innovation of accelerated learning solutions by directing multidisciplinary teams of technologists, subject matter experts and data scientists to train the technologies that replicate human decisions. She has been developing teams and solutions in eDiscovery for more than 15 years, with experience in the design and implementation of search and automated document review solutions for clients in the financial services, technology and pharmaceutical industries.