SoCal Privacy Consultants
Operationalizing Privacy and Security Programs

Data Mapping: Why Important and How to Do It

Revision Date: February 6, 2020 (v.9) includes usage for CCPA and GDPR

A Forrester report[1] of 150 data security professionals published in January 2017:
76% believe their organization has a "mature or very mature" data security strategy,
yet "most companies struggle to encrypt data, audit it for abuse, enforce a strict least privilege model,
classify it, and even understand where it's located" be it on premises, at third parties or in the cloud.

If you don't know where your data is, how can you protect it?

Purpose: The objective is to develop a data flow diagram, resource (data locations) and data inventory, and process inventory ("data map") for each unique data privacy lifecycle (DPLC). For example, Human Resources, mobile apps, websites, IoT/AI/robotics, physical stores and business units/channels/subsidiaries generally have different DPLCs (different data collection, processing, locations, etc.) and require separate data mapping sessions and maps. Data mapping should be a formalized and governed process to ensure sustainability of operational processes require to meet certain privacy legal obligations, such as:

- accurate and current privacy notices
- facilitate the DSAR fulfillment process
- assess lawful processing (see Process Activity Inventory Assessment)

The objectives of data mapping are to:

a) identify what data is collected, used, shared, stored and disposed as well as for what purposes and in what form (personal information ["PI"], encrypted, de-identified or masked (pseudonymized, redacted), etc.;

b) identify how the data is collected (context/method) and from what sources (consumers, data brokers);

c) identify where (data locations or "resources") and how (secure measures) the data is stored;

d) classify the highest level of data sensitivity in each resource to define the required strength of security controls;

e) identify external resources (service providers/data processors, third parties) where data is shared or transferred and for what purposes (provision of service, sale/monetary consideration);

f) inventory DPLC sub-processes and determine if lawfully processed "as is" or whether additional legal obligations are required (i.e. use of professional/employment data outside the context of employment enables full CPA consumer rights, selling PI for valuable consideration requires opt-out);

g) assign a DPLC process owner to each DPLC process for governance purposes;

h) assign a resource owner and custodian to each resource for governance purposes; and

i) identify what data may be transferred across what international borders.

The term "data" is used, as it is important to not only identify personal information ("PI"), but also de-identified and aggregate data to verify that the understanding of the meaning of these terms is consistent with these not being PI as defined by governing laws/regulations.

Resources mean products, services, processes, apps/software, databases, systems, technologies and external resources (service providers/data processors, third parties) housing or containing PI.

Important regulatory notes:

- As of the above revision date, this whitepaper has been updated to assist in complying with the California Consumer Privacy Act of 2018, its amendments and implementing regulations (collectively, "CCPA"). Note at this time, there are many bills under consideration to amend CCPA (if passed and signed into law). As the Office of the California Attorney General has issued one of two required sets of implementing regulations (the first is in draft form) and more amendments are expected, be sure to stay current on changes to CCPA.
- **GDPR**: If data mapping was performed to assist GDPR compliance, it should be updated to account for CCPA requirements which are different. Also see "Usage for GDPR" section at the end of this whitepaper.
- **CCPA**: In addition to identifying known identifiers also include potential data identifiers and data elements that *"identify, relate to, describe, is capable of being associated with, or could reasonably be linked, directly or*

[1] The Data Security Money Pit: Expense In Depth Hinders Maturity
https://info.varonis.com/hubfs/docs/research_reports/Varonis_TLP.pdf

indirectly, with a particular consumer or household." Under CCPA, Personal Information is more broadly defined than under GDPR.

Drawing maps on a whiteboard is not maintainable or available for easy and regular reference and cannot achieve all the following benefits.

Benefits: Most clients have an *inadequate* understanding of their *end-to-end* DPLC processes, and experience several "a-ha!" moments / surprises during the data mapping interview and diagraming process.

- Informs counsel/advisors to advise the business. Enables consultants, privacy/security officials, and internal/external counsel to quickly get their arms around DPLC processes, resources and data inventories, and DPLC sub-process inventories to better identify legal obligations and other risks (threats/vulnerabilities) in order to recommend appropriate improvements in controls.
- Permits proactive risk mitigation during the data mapping process, e.g. refining data collection practices and development of new internal policies, procedures and training.
- Informs controls evaluations and risk assessments, when first reviewing data maps with those participating.
- Informs privacy notices to assure informed, transparent and *accurate* notices to meet required disclosures and avoid material omissions and a deceptive or unfair trade practice and reputation/brand damage. Also, counsel's time to gather DPLC information is more cost-effectively spent reviewing the data map to prepare or update a notice.
- Informs proper design and implementation of consumer opt-ins/outs.
- Informs development and implementation of a centralized data management infrastructure to properly fulfill and respond to consumer rights to: know; request deletion; opt-out/in; and access/portability.
- Informs Privacy/Security-by-Design of new/enhanced resources, when using data maps to guide privacy impact assessments (PIAs, DPIAs in EU) to establish/strengthen privacy *and security* controls.
- Helps establish and maintain governance of DPLC processes, data and resources. During the design of, or planning of changes to a DPLC process, a *future state* data map should also be developed/updated and reviewed for possible changes to the Privacy-by-Design of affected resources, the privacy notice, and any related customer opt-ins/outs and preferences.
- Demonstrates governance and controls when reviewing data maps with regulators and internal/external auditors. Also positions the company to proactively control these discussions from the outset, as opposed to reacting to a barrage of questions. During due diligence discussions, reviewing data maps reassures prospective business partners and investors.
- Shortens new hire learning curve, when reviewing data maps with product/project managers, system administrators, DBAs, engineers and others.
- Facilitates organizational communications and understanding about a DPLC: The act of multiple functions interacting together to perform data mapping helps *breakdown organizational silos* and facilitate communications and improve understanding about the end-to-end DPLC process.
- Aids CCPA and GDPR compliance: Identifies each DPLC process step that should be inventoried to assess for lawful processing purposes, etc.

Process and Methodology: This process and methodology was *developed* well before GDPR (2012)) and has been practically applied with many clients across a wide variety of industries.

Process and functional subject matter experts ("SMEs") should be identified and interviewed using *"follow the data"* questions. The end-to-end DPLC can be documented in a SIPOC, a Six Sigma tool (generally used to get an operational process under control and avoid unintended consequences, such as adverse customer impacts), modified to identify data resources – where data flows from ("inputs") and to ("outputs") during the DPLC process. Only the process steps involving the DPLC - data collected, used, shared, stored and disposed – need to be identified and documented in the SIPOC. A data mapping interview session, with all stakeholders as participants, generally requires about five to six hours to gather sufficient information from scratch to facilitate: development of the data flow diagram(s), the resource and data inventory, and sub-process inventory for each DPLC. However, typically some of the discussion during the work session results from the communication ("breakdown of the silos" as mentioned above) between the process and functional SMEs results in better understanding of the end-to-end DPLC, where data is located, and identifying opportunities to improve the process and strengthen controls. The data mapping documents should be reviewed and signed-off on by the SMEs for accuracy prior to their use, such as, review prior to kickoff risk assessments and create or update privacy notices. **Tip:** Ask resource owners to bring the data schemas for resources they own to the data mapping interviews.

Scope: In addition to data mapping all core business DPLCs, employee data DPLCs and B2B contact information DPLCs should also be data mapped to assess what data is used within what context which determine your organizational legal obligations. See "Usage for CCPA" on page 7.

Raw SIPOC Document: Add as many rows as necessary to capture the DPLC process steps. Using the SIPOC template rename each process step in the language used by the business. (see an example SIPOC on page 6 for a fictitious clinical laboratory processing.

S		I	P	O		C
Data **Suppliers** / Data Sources	**Data Resource From & How** / Data Location	Data **Inputs,** Formats & How Moved / Transferred	**DPLC Process Steps**	Data **Outputs,** Formats & How Moved / Transferred	**Data Resource To & How** / Data Location	Data **Customers** / Endpoints
			Notice			
			Collection, How & Purpose			
			Used / Processed / Accessed & Purpose			
			Used / Processed / Accessed & Purpose			
			Shared / Disclosed & Purpose			
			Cross Border Data Transfers & Purpose			
			Stored / Backed-up			
			Disposed			

Figure 1 - SIPOC Template

Three3. Outputs (collectively along with the SIPOC, these are the "Data Mapping Documents") -

1. **Data Flow Diagram**: The SIPOC information can then be converted to a Visio or like diagram with swim lanes representing organizational or functional control of appropriate parts of the data flow. Diagraming is a little bit of art and a little bit science. Multiple diagrams may be required to fully capture the data flow to the desired level. For instance, at a very high level one may simply be describing how data moves from customers to business units. Other sub-process diagrams may "zoom in" to how data moves within a business unit. For instance, one may have a single diagram showing how personal data from Europe moves into your German datacenter, yet personal data from Asia and the Americas move into an AWS cloud. This diagram would likely not want to try and describe the internals of the datacenter or AWS resources. This is similar to speaking with a CEO or Board: you are able to present only the information they need without noise. If a regulator or someone performing due diligence wants to dig deeper, you can break out the datacenter-specific diagrams and list of resources and vendors.

   Multiple data collection points merging into one data privacy lifecycle can be represented by a single data flow diagram, although any complexity may require multiple layers. Again, mobile apps, websites and physical stores generally have different data privacy lifecycles (different data processing, locations, etc.) and thus would have separate data flow diagrams. The purpose is to define the ***core*** data privacy lifecycle process. One-off processes (exceptions) are normally not included. However, if certain exception processes occur frequently enough, these can also be mapped.

2. **Resource and Data Inventory**: Additionally, a resource and data inventory should be developed for each DPLC process (see table below). Resources include servers, data warehouses/bases, shared files, cloud services, cabinets (paper records), etc. Each resource is specifically named (e.g., ABC server, MNO file share file, etc.), so it is clearly identifiable to all concerned. For each resource, identify: a) the specific data identifiers and highest sensitivity level of data contained within (data sensitivity dictates the required strength of controls); and b) resource owners and custodians establishing governance around these resources. It is surprising how many SMEs do not know resource owner and custodian identities. (While data sensitivity is not the focus of this whitepaper, were highly sensitive data to be compromised it could lead to a reportable breach.)

Resource and Data Inventory		*Identify 11 Specific Data Categories and Data Types within Locations*								
		1	2	3	4	5	6	7	8	9
4 Data Sensitivity Level Classifications	**Resource** (data location):	**Locked file cabinet** in lab - key controlled	**Windows shared file** - encrypted (name)	**Freezer storage** of lab samples	**Database** - encrypted (name)	**San storage** (name)	**Batch report server** (name)	**Report archive server** - encrypted (name)	**Printer in lab** - access & hard drive managed by IT	**Web server** (name)
	Resource Owner:	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD
	Resource Custodian	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD	TBD
		Specific Data Identifiers:	Specific Data Identifiers:	Specific Data Identifiers:	Specific Data Identifiers:	Specific Data Identifiers:	Specific Data Identifiers:	Specific Data Identifiers:	Specific Data Identifiers:	Specific Data Identifiers:
	Highly Sensitive	X	X		X	X	X	X	X	X
	Sensitive			X						
	Slightly Sensitive									
	Non-Sensitive									
		list specific PHI/PII/PD involved	list specific PHI/PII/PD involved	list specific PHI/PII/PD involved	list specific PHI/PII/PD involved	list specific PHI/PII/PD involved	list specific PHI/PII/PD involved	list specific PHI/PII/PD involved	list specific PHI/PII/PD involved	list specific PHI/PII/PD involved

Figure 2 – Resource & Data Inventory

Identify the eleven categories of PI and specific data types in order to fulfill proper disclosure and other consumer rights under CCPA.

11 CATEGORIES OF PI UNDER CCPA

Identifiers	real name/alias, postal/email address, phone#, IP address, account name/#, SSN, driver's license#, passport#, signature, insurance policy#, or other similar identifiers
Commercial information	records of personal property & products/services purchased, obtained or considered, or other purchasing & consuming histories & tendencies
Biometric identifiers	DNA, biological/behavioral data, sleep, health & exercise data (gait)
Characteristics of a protected class under CA or Federal law	race, national origin, ancestry, marital status, sex, gender (sexual orientation), age, religion, physical or mental disability or other medical condition
Internet or other electronic activity information	browsing & search history, & info re: a consumer's interactions with a website, application or advertisement
Unique persistent identifiers that recognize a consumer or family *over time & across* different services	device identifier; IP address; cookies, beacons, pixel tags, mobile ad identifiers & similar technology, customer#, unique pseudonym or user alias; other forms of persistent or probabilistic identifiers used to identify a particular consumer or device
Geolocation data	precise geolocation can be highly sensitive data
Audio, electronic, visual, thermal, olfactory or similar information	voice recordings, transcriptions, keystrokes, etc.
Professional or employment-related information	job applicants, benefits, compensation, performance reviews, education/resume (one-year moratorium)
Inferences drawn	profile of preferences, characteristics, psychological trends, predispositions, behavior, attitudes, intelligence, abilities & attitudes

3. **Process Activity Inventory Assessment:** This serves a similar purpose as the Record of Processing Activity (ROAP) required to comply with GDPR. The assessment of lawful processing ensures PI collected and used by an organization can be justified to meet legal obligations. Common processing rationales may include: purpose consistent with consumer notice and consent; B2B contact information within the context of provision or receipt of goods and services; employee data use consistent within employment context and privacy notice; purpose is not justified under CCPA, etc.

	A	B	C	D
1		**[Company Name]**	Revision: 01/29/2020	Justification Owner
2	**Record of Processing Activities (ROPA) -** legitimate business purpose justification			
3	DPLC Process	Sub-process Steps	Processing Purpose	CCPA Lawful Processing Rationale
4	Activity Tracking/Recording			
5		Record IoT device information	Store recorded data on IoT device	Purpose consistent w/ consumer Notice/Consent
6		Record IoT device information	Transmit recorded IoT device information to company data storage (AWS/DataCenter)	Purpose consistent w/ consumer Notice/Consent
7		QA Review	Transfer data records for quality assurance review	Purpose consistent w/ consumer Notice/Consent
8		Aggregate Data	3rd party data transfer	Purpose consistent w/ consumer Notice/Consent
9		Customer Portal Data Access	Customer data access and reporting	Purpose consistent w/ consumer Notice/Consent
10		Customer Data Transfer	SFTP	Purpose consistent w/ consumer Notice/Consent
11				
12	Productivity & Compliance			
13		Data Analytics	Transfer data to Tableau to perform data analytics and reporting	Purpose consistent w/ consumer Notice/Consent
14		Customer Report Delivery	Transmission of customer reports	Purpose consistent w/ consumer Notice/Consent
15	B2B			
16		Lead Generation	Contact information input to service provider webform	B2B Contact Info w/in context of provision/receipt Goods/Services
17		Salesforce Sync (API)	Service Provider contact information transfer to Salesforce	B2B Contact Info w/in context of provision/receipt Goods/Services
18		Telemarketing	Contact information correction/update	B2B Contact Info w/in context of provision/receipt Goods/Services
19	Human Resources			
20		Candidate Intake	Submission of applicant information to HRIS	Employee Data use consistent w/employment context & Notice
21		Pre-screening Candidates	Review of applicant information in HRIS	Employee Data use consistent w/employment context & Notice
22		In-person Interview process	Interview of candidates by hiring managers	Employee Data use consistent w/employment context & Notice
23		Offer Extended Process	Offer sent to candidate	Employee Data use consistent w/employment context & Notice
24		Post Offer - Backgrd check	Candidate personal information input to BCheck Cloud portal	Employee Data use consistent w/employment context & Notice
25		Onboarding	Identity and employment eligibility verification	Employee Data use consistent w/employment context & Notice
26		Access Badge creation Brivo	Collection of employee photo image for access badge/card	Employee Data use consistent w/employment context & Notice
27		Badge Creation	Creation of an access badge/card	Employee Data use consistent w/employment context & Notice

Figure 3 - Process activity assessment table.

Governance/Ownership: Organizational and operational governance must be established to comply with the legal obligations specified in CCPA and other privacy laws. The governance process and all related roles and responsibilities should be clearly defined in policy. The privacy and security officials are responsible for ensuring all DPLC process owners, resource owners, and resource custodians are identified at all times, including during DLPC process, personnel, and resource changes.

The DLPC process owner is responsible for ensuring:

- the accuracy of the data flow diagram, resource and data inventories, and processing activity inventory at all times;
- proper labeling of the data flow diagram with the DPLC process name, owner and revision date;
- swim lane owners are identified and documented on the data flow diagram;
- proper archival of each version (with a revision date) in a restricted folder in a compliance repository; and
- working with privacy official and counsel to ensure privacy notices are updated and accurate before planned changes to the DPLC process are implemented that affect required disclosures; and
- working with privacy and security officials regarding the Privacy/Security-by-Design of planned future state changes to the DPLCs.

Swim lane owners are responsible for advising the DPLC process owner of any planned changes to ensure their portion of the data map to ensure it is continually maintained as accurate.

Resource owners and custodians establish and maintain RBAC governance, including RBAC rights design, authorization, implementation, and periodic review. In some cases, they may have overall responsibility for the Privacy/Security-by-Design of assigned resources. When resource owners and custodians are identified prior to performing a controls evaluation and risk assessment, the responsible individuals can properly participate and be better engaged in these activities.

Partially filled in SIPOC: This is a fictitious clinical laboratory processing example. Using business process mapping architecture, this is an example of a “level 0” process. This could be further broken into “level 1”, “level 2”, “level 3”, etc., sub-level processes.

Privacy Data Mapping SIPOC

SoCal Privacy Consultants
Lean. Sustainable. Legally Defensible.

Revision Date 13 Mar 2016

Process Name: PHI/PIIPII Processing Sample

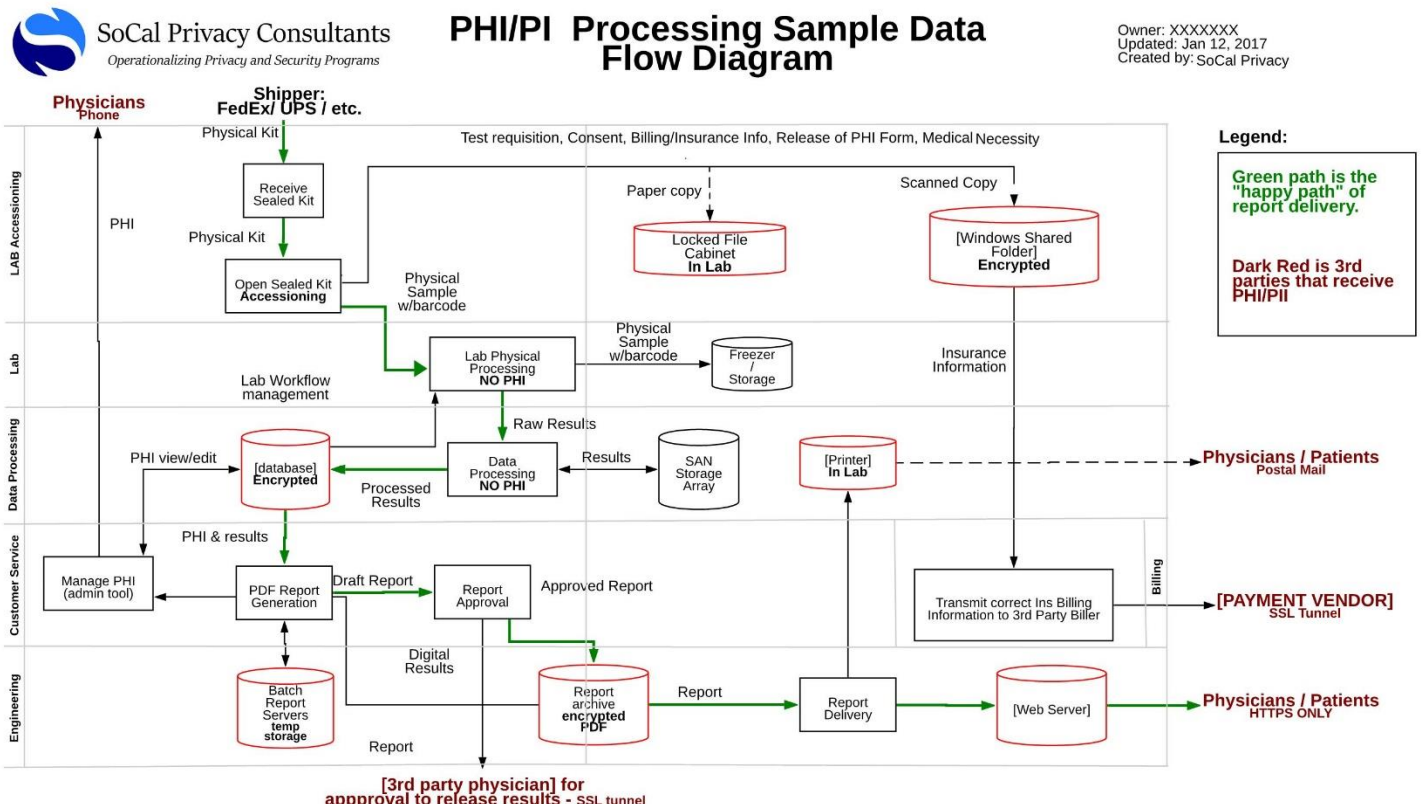
Process Owner: [Privacy Official or delegate]

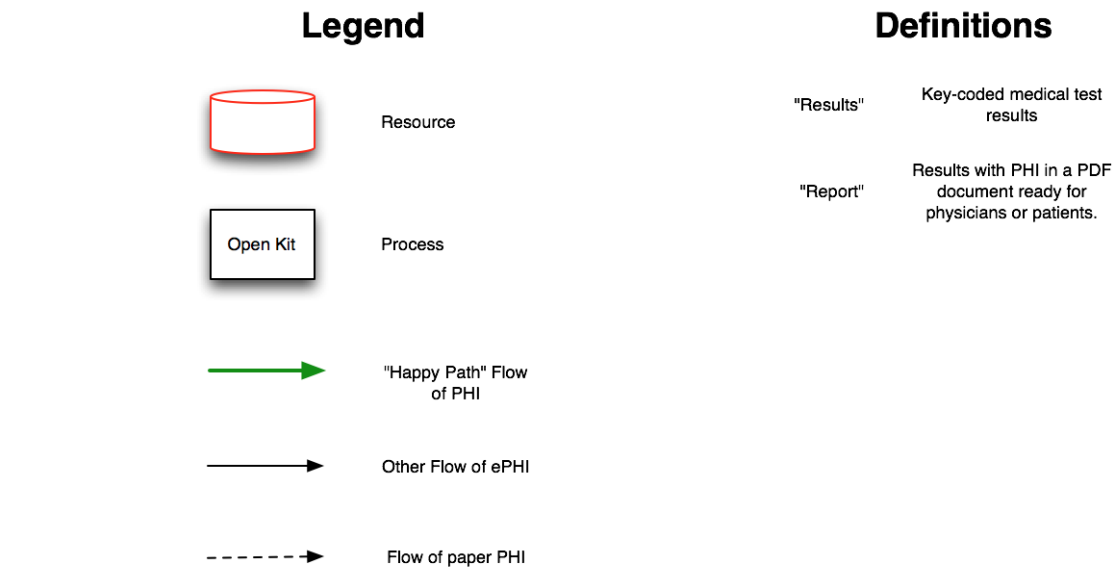
Data States: At rest; in motion; at endpoints; at disposal/desruction

Process Steps

S		I	P	O		C
Data Suppliers / Sources	**Location From / Data State**	**Data Inputs**	**Data Flow Steps**	**Data Outputs**	**Location To / Data State**	**Data Customers / Endpoints**
Clinic / Practice	US / INTL	Signed Consent	**Notice**	Signed Consent	Locked File cabinet, Scanned to:[Shared Folder:/consents/month-Year]	Legal
Clinic / Practice	FedEx / UPS	sample kit, activation code, Test Requisition Form ("TRF") (name gender dob, demographics, physician info), clinical history, consent / auth	**Receive Sealed Sample Kit**	sample kit, activation code, TRF (name gender dob, demographics, physician info), clinical history, consent / auth	Receiving Area	Accessioning
Accessioning	Receiving Area	sample kit w/Patient name, activation code	**Open Kit: Assign barcode to sample**	sample kit w/Patient name, activation code, barcode (accessionNo)	Lab Work Bench	Lab
Accessioning	Lab Work Bench	TRF (name gender dob, demographics, physician info), clinical history, consent / auth, activation code	**Open Kit: Scanning paperwork with sample kit**	TRF (name gender dob, demographics, physician info), clinical history, consent / auth, activation code	[Backend Admin Tool], Shared Folder: "Sample Tracking", [Shared Folder:/consents/month-Year]	Customer Service, Billing
Lab	Lab Work Bench	sample w/Patient name, barcode, activation code	**Lab Processing**	sample w/Patient name, barcode, activation code; raw output	Sample into freezer; SAN storage Array	Data Processing
			•			
			•			
			•			

Sample data flow diagram created from SIPOC: The fictitious clinical laboratory processing data flow diagram follows.





Limitations: The quality and accuracy of data maps developed through this process are dependent upon having the right people identified and participating in the data mapping interviews, properly vetting the diagrams for sign-off, and ultimately owning these going forward.

- Automated "data mining" tools: We have observed clients' use of automated data mining reports which produced false positives and did not capture all types of personal data. However, these tools may be useful to check and validate the data elements captured during the data mapping interview process.

- Automated "data mapping" tools: It remains to be seen whether future automated data mapping tools can replicate all the benefits from using an interactive, highly participatory data mapping interview process, as espoused at the beginning of this whitepaper.

However, automated tools scanning and identifying (inventorying) specific data elements may make sense for large organizations with vast amounts of unstructured information residing in numerous data repositories.

Business models: Organizations can act as a data controller in one set of circumstances and in another set of circumstances as a service provider/data processor or third party. When data mapping, it is important to identify these different roles and data map each DPLC process separately as each has different legal obligations. It may also be appropriate to assign different DPLC owners. Similarly, organizations acting in a DTC manner in one set of circumstances and in a B2B manner in another should separately map these DPLCs.

Usage for CCPA: Data map all the core business DPLCs, employee data DPLCs, and the B2B contact information DPLC(s).

- Employee Data: During 2020, there is a limited one-year moratorium on full CCPA applicability to employee data that is used in the context of employment, including benefit information and emergency contact information. However, privacy notices must be provided to job applicants and employees. There are generally about six (6) different employee DPLCs that should be separately data mapped. These include pre-employment, onboarding, employment and post-employment processes requiring about six (6) hours combined to data map. During data mapping, identify when employee data may be used outside the context of employment, e.g. perks, employee discount programs, etc., in which case CCPA is fully applicable, e.g. consumer rights. Lastly, keep in mind that all employees retain the right to a private right of action for breaches of employee data resulting from inadequate "reasonable security" of such data. During data mapping, you may find opportunities to better protect employee data and avoid the use of unstructured data to mitigate risk to the organization as well as its job applicants and employees.
- B2B Contact Information: During 2020, there is also a limited one-year moratorium on full CCPA applicability to information solely collected and used in B2B communications and transactions with other organizations or government agencies obtained in the context of due diligence, or provision or receipt of goods or services. This data is often housed in a CRM system, such as Salesforce. It is important to identify and separately data map any B2B contact information used outside this context with creates full CCPA applicability.
- Marketing: Don't forget to data map any PI used for marketing purposes as a core business DPLC process.

Usage for GDPR: Due to its extra-territorial reach, many businesses not headquartered in the EU must comply with the EU's General Data Protection Regulation ("GDPR"). While this data mapping process was developed eight years ago, well prior to GDPR, it can help identify data collected and all the DPLC process steps that should be inventoried as part of a GDPR gap assessment. An additional SIPOC process step should be added to identify cross border data transfers. A separate Excel table can be used for purposes of determining and documenting the lawful processing purposes for collecting, using, sharing, cross-border data transfers, and storing of personal data. The UK and French DPAs (data protection authorities), ICO and CNIL respectively, have published example templates that can be used for this purpose.

Evidence of Good Faith Compliance: Data mapping helps uncover risks that when properly addressed helps establish a defensible privacy and security posture.

Feel free to reach out to us if you have any questions.
SoCal Privacy Consultants perform gap and risk assessments and help organizations establish practical, sustainable, defensible and trustworthy privacy and security programs.

Michael Cox, CIPP/US | CEO and Founder
SoCal Privacy Consultants | www.socalprivacy.com
mcox@socalprivacy.com | m: 619.318.1263