SoCal Privacy Consultants
Operationalizing Privacy and Security Programs

Effective and Practical Data Mapping and Governance of Data Processing Lifecycles, Resources and Data

Enabling Accurate Privacy Notices, Privacy/Security-by-Design,
Data Minimization, Data Retention/Disposal, Service Provider Management,
Consumer Privacy Rights Fulfillment and More

Rev. Date: 02JUNE2022 (first performed in 2013)
Includes usage for HIPAA, CCPA/CPRA (CA), VCDPA (VA), CPA (CO), UPA (UT) and CTDPA (CT) and GDPR

Defined Terms

- Data = personal information ("PI":CA), protected health information ("PHI"), personal data ("PD":other state laws/GDPR)
- DPLC (data processing lifecycle) "follows the data" through collection (sources/suppliers), usage/access, sharing/transfer, storage (resources) and retention/disposal ("processing") as well as for what purposes
- Resource means assets (apps/software, databases, systems, technologies), internal resources (file cabinets) and external resources (service providers/data processors, third parties) housing/containing data
- Data mapping is about developing a visual DPLC diagram and inventories of associated resources/data and service providers (data processors/HIPAA business associates)

Goals and Benefits

Key Goals (Primary Benefits)

1. Understand what data is collected/processed by company and for what purpose, what data is processed by its service providers, and in what countries.
2. Determine applicable laws and regulations specifying company obligations and individual rights.
3. Establish data processing governance roles and responsibilities.
4. Operationalize privacy requirements, e.g., accurate privacy notices, Privacy/Security-by-Design, data minimization, data retention/disposal, service provider management, consumer privacy rights fulfillment and more.
5. Improve security posture by understanding where data is located and mitigating breach risk.

Important Secondary 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 process.

- Informs counsel/advisors to better advise the business. Enables consultants, internal/external counsel and privacy/security officials to quickly get their arms around DPLC processes, and inventories of resources/data and service providers to better identify legal obligations and other risks (threats/vulnerabilities) in order to recommend appropriate improvements in controls.
- Informs privacy notices to assure informed, transparent and *accurate* notices to meet required disclosures and avoid material omissions (a deceptive or unfair trade practice with resulting 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.
- Becomes single source of truth and formalizes authorized decision makers (vs. an ad-hoc person).
- Improves operational process (design), staff efficiency and effectiveness, and customer experiences.
- Informs proper design and implementation of consumer opt-ins/outs.
- Informs controls evaluations and risk assessments, when reviewing data maps with participants.
- Permits proactive risk mitigation, e.g., refining data collection practices and development of new internal policies, procedures and training.
- Informs Privacy/Security-by-Design as part of the SDLC. During design of planned changes to a DPLC, a *future state* data map should be developed and reviewed to guide development of a privacy impact assessment (PIA/DPIA in EU) to establish and strengthen privacy *and security* controls.
- Helps establish and maintain governance of DPLC processes, resources, data, and service providers. Formally designated (and communicated) decision makers with clear roles and responsibilities generally make better risk and control decisions than ad-hoc decision-makers.
- Informs establishment of a centralized data management framework to properly fulfill and respond to consumer rights: access/portability; deletion; opt-out/in; etc.

- Demonstrates governance and controls when reviewing data maps with regulators, auditors and other internal/external stakeholders. 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.
- Reduces data silos and shadow IT and associated cost. Improves data quality.
- Facilitates data minimization and data retention/disposal (new legal obligations)
- Shortens new hire learning curve, when reviewing data maps with product/project managers, process/operations analysts, system administrators, DBAs, engineers and others.
- Facilitates organizational understanding about a DPLC: Helps *breakdown organizational silos* and facilitates communications to improve understanding about the end-to-end DPLC process.
- Aids regulatory compliance: Facilitates assessment of processing purposes and organizational ownership of regulatory compliance at a DPLC process level.
- Informs breach investigations and response: If for example a server is compromised, it is critical to know quickly what data, and its sensitivity, is located there.

Ensuring Customer and Brand Trust

Data mapping brings visibility to the end-to-end data processing lifecycle and is foundational to establishing governance around DPLCs, resources and data. For many SMBs, this is a move away from managing in an ad-hoc and risky manner and towards developing a systematic discipline that: a) drives operational efficiency and effectiveness (operational excellence); b) improves the customer experience to reinforce and maintain trust; and c) facilitates risk management to assure expected outcomes (no surprises). This means avoiding adverse customer impacts, compliance issues and other risks that destroy customer/brand trust and reputation. It establishes a baseline from which to evaluate alternative ways of doing things to continually improve. It also facilitates identifying who is accountable and responsible for changes, risk management and compliance. Lastly, along the way the actual DPLC, resource and data subject matter experts (SMEs) are identified to others who may have looked to the wrong people for answers which can be highly problematic and the root cause of unintended impacts.

Common Issues with Ineffective Data Mapping

- Drawing maps on a whiteboard is not maintainable, available for easy/regular reference, and cannot achieve the benefits described above.
- Using questionnaires to collect data mapping information from SMEs often leads to incomplete and conflicting information vs. getting to the source of truth by interviewing all appropriate SMEs at the same time.
- Data mapping at the enterprise level, rather than DPLC-level, is ineffective in terms of establishing effective data processing governance.
- Automated data mining/discovery tools cannot identify all data types and produce countless false positives making manual data mapping necessary. Automating an incomplete data mapping scheme, e.g., "data lineage", does not make it better and does not facilitate data processing governance.

Key Objectives

1. Define and name discreet data processing lifecycles (DPLCs) within an organization.
2. Develop a thorough understanding of what data is collected, how and from whom (source) it is collected, processed/used, shared/transferred and retained/disposed, including backups.
3. Create a visual data flow diagram for each DPLC. This helps you tell your data story to key internal and external stakeholders, including business partners, investigators and auditors.
4. Develop comprehensive inventories of resources, data, and service providers and classify the data's sensitivity level.
5. Establish a practical data governance strategy with clear roles and responsibilities for DPLC process owners and resource owners/custodians to facilitate accurate privacy notices, Privacy/Security-by-Design, proper privacy rights fulfilment, etc.

Effective data mapping and data governance are foundational to a sustainable and defensible privacy and security program, complying with privacy and security laws, regulations, standards and policies, and protecting the company, including its brand/reputation, consumers, investors, board/executives and other stakeholders.

Organizational Governance of DPLCs, Resources and Data

Formalized data mapping identifies, defines and maintains discreet DPLCs. DPLCs are vehicles for driving sustainable governance into business operations to facilitate meeting certain privacy legal obligations, such as:

- assessing what law or regulation applies to each discreet DPLC
- transparent and always current/accurate privacy notices
- DSAR (consumer rights request) and HIPAA access rights fulfillment processes
- data retention and disposal program (required by CPRA) based on a DPLC's purpose
- Privacy/Security-by-Design of planned changes using privacy impact assessments ("PIAs")
- other legal obligations, e.g., data minimization, consumer opt-out to limit use of sensitive PI
- "reasonable (defensible) security" of highly sensitive data (know where it is located to protect it)
- proper oversight and management of service providers

Regarding the first point, companies that serve in multiple roles, e.g., data controller and processor, and/or are subject to multiple laws/regulations, e.g., HIPAA in the B2B context and various state privacy laws in the D2C context, require a construct to facilitate effective governance. This cannot be done at the enterprise-level. However, it can be effectively managed at the DPLC-level. Where multiple DPLCs are involved, a "DPLC Dashboard" is helpful. More on this later.

Defining Discreet DPLCs and Scope

Defining a DPLC

Understanding the business model – meaning how it obtains consumers/clients and makes money - is an important first step to defining whether there is one or more discreet DPLCs.

- If an organization collects data in multiple ways, e.g., social media, mobile app, website, IoT/AI/robotics and physical stores, that feed into the same platform/infrastructure using common resources, this may be a single DPLC depending upon the complexity involved. If any of the collection processes are complex, these will likely be separate DPLCs from the infrastructure/platform DPLC.
- If an organization acts as both a data controller and a service provider (data processor) and/or acts both in a D2C manner and a B2B manner, these different roles should be considered separate DPLCs as each likely has different legal obligations.
- Different business units/channels/subsidiaries are generally separate DPLCs because each collects different data, uses data for different purposes, processes data differently, may share/transfer data differently and/or may store data in different locations. For example, marketing uses data very differently than business operations.
- Any massively complex processing should be broken into separate DPLCs.

Additional Scope for California: If CPRA applicability thresholds are triggered, also data map:

- **HR Data DPLC** (job applicants, employees, and consultants/contractors residing in CA): The moratorium on full CCPA applicability to HR data (used in the context of employment) sunsets January 1, 2023 when CPRA becomes effective triggering full applicability to such data. There are generally about six HR Data mini-DPLCs, during pre-employment, onboarding, employment and post-employment phases, requiring about six hours to data map.
- **B2B Contact Information DPLC** (CA residents): The moratorium on full CCPA applicability to B2B contact data also sunsets January 1, 2023 when CPRA becomes effective triggering full applicability to such data. B2B contact data is often obtained in the context of due diligence or provision or receipt of goods or services and often housed in a CRM system, such as Salesforce.

Consider data mapping HR Data anyway. All affected "individuals" have a private right of action for breaches of HR Data resulting from inadequate "reasonable security" of such data. Data mapping the HR DPLC has facilitated identification of opportunities to better protect HR Data and mitigate breach risk to the organization and its job applicants, employees and consultants/contractors. Employee data is often not securely managed and is generally invisible to those responsible for privacy and security.

Marketing: Don't forget to data map any PI used for marketing purposes as a business DPLC process.

Data Mapping Interview Process and Procedures

Interview Process

Process and functional 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, processed/used, shared, transferred and stored/disposed – need to be identified and documented in the SIPOC. A data mapping interview session, with all stakeholders as participants, generally requires about three to five hours depending upon the DPLC's complexity to gather sufficient information from scratch to facilitate development of the data flow diagram(s), the resource/data inventory, and service provider 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 during a validation session and final documents signed-off on by the SMEs for accuracy prior to their use, such as, review prior to kickoff of risk assessments and to create or update privacy notices.
Tip: Ask resource custodians to bring the data schemas for their assigned resources to the data mapping interviews.

Data Mapping Procedures - Each DPLC requires separate data mapping sessions typically with different participants who collectively understand the data collected/processed and its purposes.

1. Identify what data is collected, used, shared, x-border transferred, stored-retained/disposed as well as for what purposes and in what form (encrypted, de-identified, pseudonymized, redacted).
2. Identify how the data is collected or obtained (context/method) and from what sources/suppliers (consumers, data brokers, social media).
3. Identify where ("resources" or data locations) and how (secure measures) the data is stored.
4. Classify the levels of data sensitivity in each resource based on <u>context (data classification).</u>
    a. **Highly sensitive data** (defined by 50 state data breach laws) is data that were it to be compromised could lead to a reportable breach and thus requires "reasonable security" controls for defensibility.
    b. **Sensitive data** (defined by U.S. state omnibus privacy laws) requires: a) additional organizational obligations, such as notice, opt-in/out consent, etc.; and/or b) consumer rights, such as ability to limit profiling and unnecessary/secondary use.
        i. For HIPAA, consider **Designated Record Sets** to be sensitive data due to patients having rights of access and correction. Since 2019, OCR has focused on enforcement of these rights.
    c. **PI Types and Categories**, as defined by CCPA/CPRA, which must be disclosed in privacy notices

    Note: Traditionally, data classifications were *exclusive*, meaning data would be classified as being in a single classification. However, we now need to shift this paradigm to recognize that *data can be in multiple classifications* based on *context*, that affect consumer rights and/or company obligations.
5. Identify external resources (service providers/data processors/business associates-HIPAA, third parties) where data is housed and for what purposes (provision of service, sale/monetary consideration).
6. Designate a DPLC process owner to each DPLC process for governance purposes.
7. Designate a resource owner and custodian to each resource for governance purposes.
8. Identify what data may be transferred across what international borders.

<u>Process Activity Analysis</u>

For the **U.S.**, in many cases, particularly for SMBs, discreetly defining DPLCs generally permits analysis of what jurisdictional or sectoral law/regulation applies (see "DPLC Dashboard" below). However, certain DPLC sub-processes should be identified and addressed that may:

- trigger different obligations within the same law or regulation, such as making the entity a data controller when the entity is primarily acting as a processor in a particular DPLC
- trigger a different law or regulation when there is a change in data usage that is outside the scope of the law or regulation that primarily has jurisdiction over a particular DPLC, for example if the data use

is not consistent with GLBA it may trigger a state privacy law that grants data-level (vs. entity-level) GLBA exemptions

- trigger a different law or regulations when there is a change in context, for example when HR data is used outside the context of employment and a state privacy law only grants an exemption if used within the context of employment

The **EU**'s GDPR requires a Record of Processing Activities ("**ROPA**"), essentially requiring an inventory of processes for lawful processing evaluation. This data mapping process can be used to identify and inventory all the DPLC sub-process steps to facilitate such an assessment. The UK and French DPAs (data protection authorities), ICO and CNIL respectively, have published example ROPA templates that can be used for this purpose. Similarly, an additional SIPOC process step should be added to identify cross border data transfers to facilitate development of required transfer impact assessments.

A privacy consultant or lawyer can assist with such an analysis.

Data Mapping Tools, Methodology and Outputs

This process and methodology were *developed in 2013* (well before GDPR required a ROPA) and has been practically applied with many clients across a wide variety of industries.

Prework: Use a DPLC Data Inventory and Classification Tool to Identify and Classify Data
Develop and use a data inventory and classification tool to identify data housed in resources within a DPLC and properly classify the data's sensitivity. *SoCal Privacy's tool is only made available to clients.*

State Omnibus Privacy Laws: Identify and classify:

- **Specific data types and categories** (based on the eleven categories of PI defined by CCPA/CPRA) that must be disclosed in a privacy notice and to which consumers have access rights, include any deidentified data Deidentified data and aggregate data should also be identified to verify that the understanding of these terms is consistent with how governing laws/regulations define these exemptions.
- **Sensitive PI** which triggers additional consumer rights and company obligations (based on a compilation of various state omnibus privacy laws that should be updated as new laws emerge)
- **Highly sensitive data** that were it to be compromised could lead to a reportable breach as well as any deidentified data (based on a compilation of the 50 state breach notification laws)

HIPAA: Identify and classify:

- **18 PHI Identifiers** and deidentified data
- **Designated Records Sets** to which patients have access and correction rights
- **Highly sensitive data** that were it to be compromised could lead to a reportable breach

Interview Tool: Use a Modified SIPOC to Document Responses to Follow-the-Data Questions
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 next page 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

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.

SIPOC Data Flow Mapping Worksheet — **Example BioPharm - Proprietary and Confidential**

Client Name: **Example BioPharm** Rev Date: **7/22/2021**

Process Name: **Billing** Process Owner: **Bill Williams**

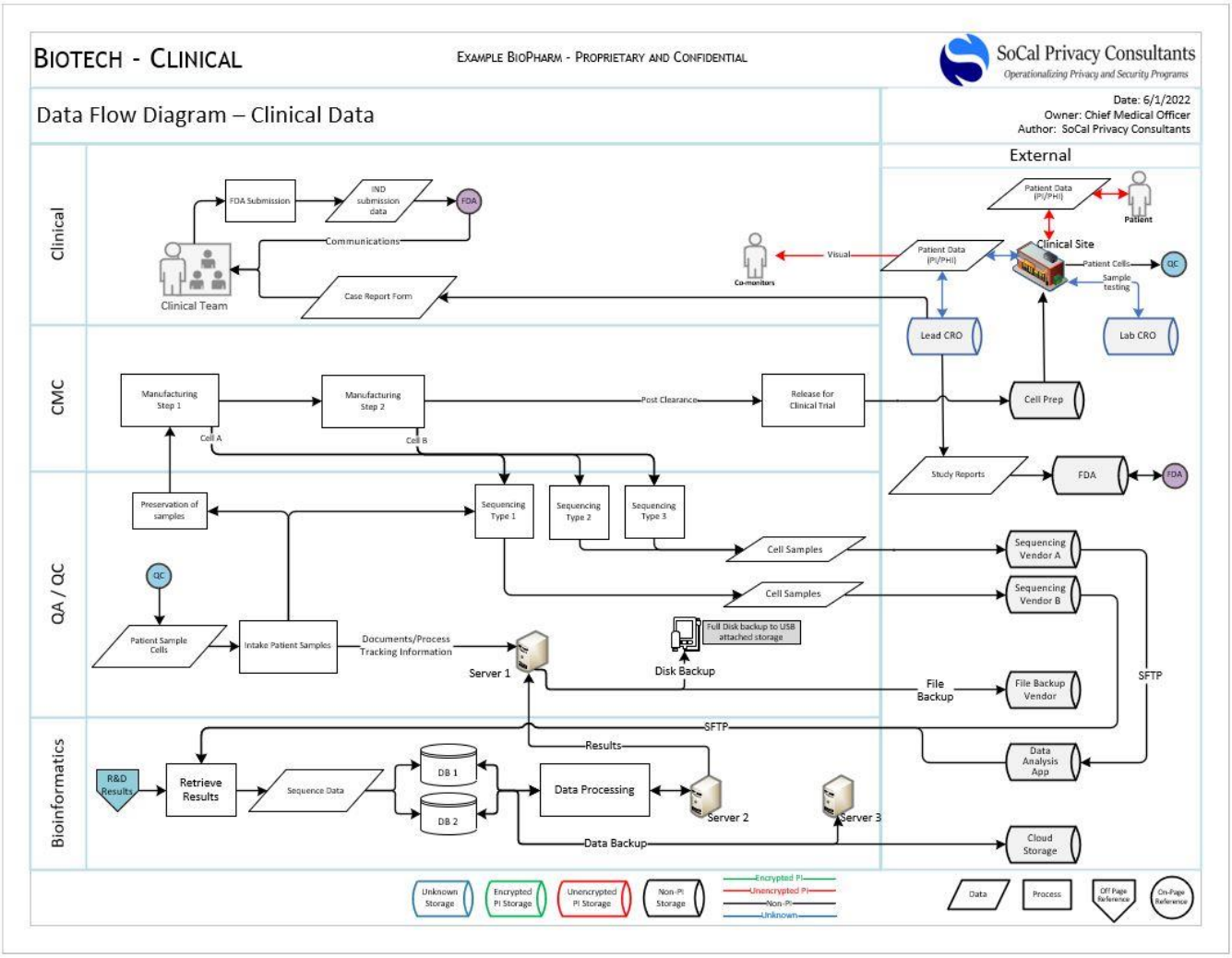
SoCal Privacy Consultants Lean. Sustainable. Legally Defensible.

S			I	P	O		C	
Data **Suppliers**/Sources	Service Provider (if applicable)	Data Resource **From** (data location)	Data **Inputs** (Data fields sent)	**Process** Step Name	Data **Outputs** (Data fields sent)	Data Resource **To** (data location / endpoint)	Data Customers	Data format/ transmission type (HTTPS/SFTP/HTTPS API/MS-SQL etc)
patient		sales	name, dob, address, lic #, gender, email, crimiinal background, ssn, chemical dependency, signatures,	**New Business Data Entry**	name, dob, address, lic #, gender, email, crimiinal background, ssn, chemical dependency, signatures,	Oasis	All of company	http
patient		sales	name, dob, address, lic #, gender, email, crimiinal background, ssn, chemical dependency, signatures,	**New Business Data Entry**	name, dob, address, lic #, gender, email, crimiinal background, ssn, chemical dependency, signatures,	Perceptive	All of company	http
patient		Data Warehouse	gender, email, crimiinal background, ssn, chemical dependency, signatures,	**Acct validation**	gender, email, crimiinal background, ssn, chemical dependency, signatures,	AR Accountants	All Accountants	http
AR Accountants		Data Warehouse	service charges/fees, current balance, amt due, name, addr, doc names, acct number,	**Invoicing**	service charges/fees, current balance, amt due, name, addr, doc names, acct number,	Insured (admin, finance dept, CPA etc) and brokers	Insured (admin, finance dept, CPA etc) and brokers	email, pdf file
AR Accountants		Data Warehouse	service charges/fees, current balance, amt due, name, addr, doc names, acct number,	**Invoicing**	service charges/fees, current balance, amt due, name, addr, doc names, acct number,	Insured (admin, finance dept, CPA etc) and brokers	Insured (admin, finance dept, CPA etc) and brokers	USPS
IT		X File Share	service charges/fees, current balance, amt due, name, addr, doc names, acct number,	**Bank eBill upload**	service charges/fees, current balance, amt due, name, addr, doc names, acct number,	Primary Bank	Primary Bank	pdf file, sftp
Primary Bank	Primary Bank	Bank eBill	payment confirm num, payment date, insured name, payment amt, acct number	**eBill Payment File download**	payment confirm num, payment date, insured name, payment amt, acct number	Finance shared drive	All Accountants	auto script, csv file, sftp
Secondary Bank	Secondary Bank	Bank Lockbox	sometimes a copy invoice pymt stub) and text of the data from the checks	**Bank Lockbox file download**	sometimes a copy invoice pymt stub) and text of the data from the checks	Finance shared drive	All Accountants	https browser, image files (pdf) and csv file
Finance Shared drive		Bank confirmataion files	payment confirm num, payment date, insured name, payment amt, acct number	**Finance Tools upload**	payment confirm num, payment date, insured name, payment amt, acct number	Finance Tool	All Accountants	smb
Finance Tool		text file generated by Finance Tool	payment confirm num, payment date, insured name, payment amt, acct number	**Lockbox Process**	payment confirm num, payment date, insured name, payment amt, acct number	Data Warehouse	All Accountants	Y shared drive
Insured		Insured's bank	acct info, payment info	**Wire transfer payment process**	acct info, payment info	Secondary Bank	All Accountants	Wire transfer

Outputs/Deliverables (collectively along with the SIPOC, these are the "Data Mapping Documents")

1. **Data Flow Diagram**: Convert the SIPOC information 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. When multiple data collection points merge into a single infrastructure using common resources, e.g., an ERP or Salesforce platform, this may be a single DPLC represented by a single data flow diagram, although any complexity may require multiple layers. Complex collection points might be considered mini-DPLCs. When processes use different infrastructures and/or resources, these should be separate DPLCs requiring separate data flow diagrams, e.g., B2B and D2C business channels and acting as data controller vs. service provider. This is about data mapping the *core* data processing lifecycle process. One-off processes (exceptions) are normally not included. However, if certain exception processes occur frequently enough, these can also be mapped.

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



2. **Resource and Data Inventory**: Additionally, a master resource and data inventory should be developed permitting filtering for each DPLC process (see table below). Resources include servers, data warehouses/bases, shared files, cloud services, cabinets (paper records), etc. Each resource should be 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 types and classify the sensitivity levels of data contained within; and b) assigned resource owners and custodians establishing governance around these resources. It is surprising how many SMEs cannot identify resource owners and custodians as these have not been designated.

Resource and Data Inventory **Example BioPharm - Proprietary and Confidential**

SoCal Privacy Consultants
Lean. Sustainable. Legally Defensible.

Resource Inventory Owner: _____

Document Owner: _____

DPLC Name	Resource Name	Host/Provider	Internal or External?	Resource Owner	Resource Custodian	Data Privacy Classification Level(s) (refer to DPLC Data Checklist)	Data Level Identifiers
Customer Service	SQL Database		Internal	Joyce Johnson	John Williamson	PI	name, title, address, email, phone, org
Customer Service	Data Warehouse		Internal	Joyce Johnson	John Williamson	PI	name, title, address, email, phone, org
Customer Service	CRM	CRM	External	Bill Williams	Kevin Turner	PI	Provider and Broker name, speciality, address, email, phone, POC name and gender, DOB
Customer Service	Bulk eMail	beMailIT	External	Mari Madrigal	Kevin Turner	PI	name, email, speciality, address, age, dob, title, gender, clinic name
Customer Service	US Mail	USPS	External		Kevin Turner	PI	Name, mailing address
Billing	RDBMS		Internal	Bill Williams	Joyce Johnson	Highly Sensitive, Sensitive, and PI	name, dob, address, lic #, gender, email, employment, signatures
Billing	Data Warehouse		Internal	Bill Williams	Joyce Johnson	Highly Sensitive, Sensitive, and PI	name, dob, address, lic #, gender, email, employment history, criminal background, signatures
Billing	Bank e-Billing	Bank	External	Bill Williams	John Collier	PI	insured name, addr, acct number, acct balance, payments
Billing	Finance Tool application		Internal	Bill Williams	John Collier	PI	insured name, addr, acct number, payments
Billing	Finance Shared Drive		Internal	Joyce Johnson		PI	insured name, payment amt, acct number, balance, check images
Billing	US Mail	USPS	External	Bill Williams		PI	insured name, addr, doctor names, acct number, acct balance
Billing	Email	Microsoft	External	Bill Williams	Kevin Turner	PI	insured name, addr, doctor names, acct number, acct balance
Risk Management	websites database		Internal	Joyce Johnson	Kevin Turner	Sensitive and PI	Name, email, account id, password
Risk Management	Video Conferencing	Cisco (Webex)	External	Joyce Johnson	Tina Turner	No PI	fn, ln, prof designation, org
Risk Management	Video Conferencing	MS (Teams)	External	Joyce Johnson	Tina Turner	No PI	fn, ln, prof designation, org
Risk Management	Video Conferencing	Goto Meeting	External	Joyce Johnson	Tina Turner	No PI	fn, ln, prof designation, org
Risk Management	RDBMS		Internal	Bill Williams	Kevin Turner	PI	member's fn, ln, professional designation, org association, phone, email
Risk Management	Data Warehouse		Internal	Bill Williams	Kevin Turner	PI	member's fn, ln, professional designation, org association, phone, email

Figure 2 – Resource & Data Inventory

3. **Service Provider Inventory**: If no contract management system is in place, this should be created to track agreements/amendments and initial and any periodic assessments to assure and document compliance. For each service provider, the inventory should also identify related data types, data sensitivity classification, and assigned resource owner.

DPLC Dashboard: For organizations with multiple DPLCs, creating and maintaining a DPLC side-by-side dashboard (Excel works) facilitates visibility and governance. It should identify the DPLC's name, the DPLC Process Owner, business model (B2B, D2C), role (controller, processor), countries/states, applicable laws/regulations, data classification types, etc.

Manual Data Mapping 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 and Mapping Limitations

- Automated data discovery tools: We have observed clients' use of automated data discovery reports which produced many 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. The effectiveness of these tools should improve over time using AI.
- Automated data mapping tools: Many of these automated tools today simply create "data lineage" maps showing how data moves from one resource to another. While helpful, the data lineage maps we have observed do not represent complete data flow diagrams of data processing lifecycles. Thus, establishing governance around DPLCs as described within this whitepaper cannot be fully achieved. 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.

Automated scanning, identifying and inventorying of specific data elements makes sense for large organizations with vast amounts of data residing in numerous data repositories. However, complete reliance on such automation is not advisable at this time.

Data Processing Governance Roles and Responsibilities

There are many dimensions to organizational and operational governance that should be established to defensively comply with legal and regulatory obligations. These governance processes and all related roles and responsibilities should be clearly defined in policy and communicated in roles-based training. This section strictly focuses on the data governance dimension of an effective privacy and security program.

As background, IT and Security are generally focused on resource and data governance. However, privacy is concerned about the compliance and risk issues associated with data notice, purpose, and processing, e.g., use, access, sharing, x-border transfer, and retention/disposal (individually and collectively, "data processing"). This document describes our recommended key roles and responsibilities for *operationalizing* data processing governance.

Privacy and Security Officials cannot manage privacy and security in a vacuum. Privacy and security should be embedded within business operations through *distributed* roles and responsibilities. To facilitate this, we recommend using the Enterprise Risk Management (ERM) model where management is responsible for the risks and compliance issues inherent in their organization's process, people and technology. The ERM governance model has been used over the last twenty+ years by banks and insurance companies and in more recent years its adoption has been increasing in many other industries. To implement this model, key role-holders require some risk management and compliance training as well as the support of privacy, security and compliance officials. In 2020, the NIST Privacy Framework validated our approach with its data mapping and inventory, data processing ecosystem risk management, and governance control categories.

Privacy and Security Officials should be responsible for ensuring that DPLC Process Owners are designated for every DPLC (data processing lifecycle). As DPLC Process Owners leave the organization or assume other roles, the Privacy and Security Officials should ensure new DPLC Process Owners are designated. Preferably, these officials should either be certified professionals, be encouraged and supported to become certified, or have access to certified professional consultants and/or lawyers. Example certifications include: CIPM and CIPP/US for privacy; and CISM and either CISSP, CISA or GIAC for security. (As CHSPE is HIPAA-specific, we recommend the broader privacy and security certifications.)

DPLC Process Owners, similar to the ERM model, are responsible for the risk and compliance issues inherent in their assigned DPLC processes. DPLC Process Owners should understand the purposes and uses of data for a given DPLC and thus generally represent the business side (vs. IT/engineering). DPLC Process Owners should:

- Ensure Resource Owners and Custodians are designated for each Resource utilized by their DPLC.
- Ensure the data flow diagrams, and resource/data and service provider inventories are continually accurate, even though the responsibility to maintain these may be delegated to others. Process owners should:
    - Ensure proper labeling of the data flow diagram with DPLC process name, owner and revision date.
    - Swim lane owners are identified who can ensure their portion of the DPLC is always accurate and their names are documented with their swim lane on the data flow diagram.
    - Ensure proper archival of each version (with a revision date) of the diagram and inventories in a restricted folder in the compliance repository.
- Ensure all data processing – notice, collection, use, access, sharing, any x-border transfer, and retention/disposal – related to their DPLC meets policy, legal and regulatory requirements *working with the Privacy Official and counsel as appropriate*, including but not limited to:
    - Ensuring privacy notices are updated and accurate before planned changes to the DPLC process are implemented that affect required disclosures.
    - Authorizing the use of on-premises and off-site or cloud-based Resources (including Resources used by Business Associates for HIPAA) that may contain Personal Data/Information ("PD/PI") or Protected Health Information ("PHI"), *also* with guidance from the Security Official.
    - Ensuring deidentification standards and requirements are sustainably met if data is to be deidentified.
- Ensure the Privacy/Security-by-Design of planned changes to their assigned DPLC that has privacy and/or security implications working with the Privacy and/or Security Officials (and in some cases appropriate Resource Owners and Custodians). To facilitate this, ensure that a future state data map of planned changes is developed and reviewed with these Officials, as appropriate, during the design phase of the S-SDLC (secure software development lifecycle).

Resource Owners should:

- Ensure the appropriate RBAC rights design (usually assigned to workforce members in certain functional roles) for their assigned Resource, based on the minimum necessary rule (part of Privacy-by-Design).

- Authorize what workforce members (usually in certain functional roles) should have access to PD/PI or PHI (and Designated Record Sets for HIPAA) located in these Resources, and their level of access (read only, downloadability/printability, editability, etc.).
- Periodically review the assigned RBAC rights to ensure continued viability. The latter includes the removal of those whose roles have changed and no longer require access or the same level of access and those who have been terminated as well as evaluating whether additional RBAC roles should be added as the company scales over time to ensure the minimum necessary access to PD/PI or PHI (and Designated Record Sets).

Resource Owners usually represent the business side in order to understand how to properly design and sustainably maintain RBAC roles and determine who should have access and for what purposes consistent with the minimum necessary rule.

Resource Custodians (aka, data stewards) should facilitate:

- Periodic review of RBAC rights assigned to workforce members working with the Resource Owners.
- Security-by-Design of their assigned Resource(s) working with the Security Official.
- Fulfillment of consumer or patient rights requests with a DSAR (Data Subjects Access Rights) Coordinator or HIPAA Rights Coordinator respectively.

In order to have a separation of duties and avoid a conflict of interest, Resource Custodians should be in IT/engineering, preferably under the supervision of the same manager to develop common protocols/standards and not report within the organizational control of Resource Owners.

DPLC Process Owners and key Resource Owners and Custodians should participate as appropriate in periodic controls evaluations and risk assessments.

DSAR Coordinator/HIPAA Rights Coordinator should receive and coordinate consumer/patient rights requests (for HIPAA whether received from HCPs or patients) and coordinate with Resource Custodians to fulfill these requests. Rights Coordinators could be the Privacy Officials in small organizations or report to Legal/Privacy Counsel. If HR Data triggers CPRA's (California) applicability thresholds (HIPAA does not apply to HR Data), HR should have its own DSAR Coordinator.

While there is more than one way to accomplish these required activities, always be sure to define and designate all of the above responsibilities in policy and reinforce them through roles-based training.

Let us know if you have any questions about data mapping and/or data governance. We'd also be happy to provide consulting about how to operationalize the larger privacy and security governance framework, of which this is a part of, as well as provide roles-based training.

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, Chief Privacy Consultant
SoCal Privacy Consultants | www.socalprivacy.com
mcox@socalprivacy.com | m: 619.318.1263