Design Assurance Standard

PUBLIC DRAFT

Date: 10/1/2024

Standards Committee

Institute of Operational Privacy Design

NOTE TO REVIEWERS

Please focus primarily on substance. The formatting will change. We will fix any minor grammatical or spelling errors.

- Look for logical inconsistencies.
- Look for gaps in logic.
- Look for confusing language or statements.

Table of Contents - Note: Numbering will be updated in final version

Introduction

The Institute of Operational Privacy Design (IOPD) is dedicated to the creation and adoption of privacy design standards to protect privacy. The IOPD's mission is threefold: to evangelize privacy by design through education and standards, to provide accountability through certification mechanisms, and to publicly recognize good privacy practices by companies around the globe.

The IOPD is a non-profit, membership-based professional organization primarily run by a volunteer Board of Directors. The Board seeks input from advisors who sit across industries and have varied roles within organizations. As advocates for privacy design standards, the IOPD and its members vow to respect the privacy of individuals in all their practices. We hold ourselves to the highest standard, which we expect from all businesses recognized publicly for good privacy practices.

This Design Assurance Standard (the "Assurance Standard") follows two years of effort by the IOPD's Standards Committee after its adoption of the Design Process Standard (Process Standard) in January 2023. While the Process Standard details the design process components needed to incorporate privacy considerations and reduce privacy risks, this Assurance Standard uses an assurance case to confirm an organization's claim that a specific product, service, or business process has been designed, developed, or deployed with privacy aforethought. In other words, the Assurance Standard doesn't apply to an organization but to a specific object of evaluation. The intent of this certifiable standard is for organizations to demonstrate that they have achieved reasonable assurance around "privacy by design and default" claims.

In theory, a product, service, or business process that has been designed, developed and deployed using the Process Standard should meet the Assurance Standard. In practice, this may not be the case because, for the Process Standard, organizations may select their own risk model, whereas this Assurance Standard uses a defined risk model. The singular risk model, and use of an assurance case more generally, provides a common measure enabling relative comparison of privacy respecting qualities between disparate products, services and business processes.

In addition to providing a measure for comparison, this Assurance Standard provides methodologies to determine whether the evidence supporting a purported claim has been satisfied. Organizations whose products, services or business processes satisfy these evidentiary burdens can apply for the IOPD's Privacy by Design and Default Trust Mark.

This Standard uses a "claims, arguments, and evidence" (CAE) structured notation for the assurance case.[1] Although the IOPD's Standards Committee has done most of the hard work constructing the assurance case (i.e., the set of claims, arguments, and necessary evidence), there will be a customization aspect necessary from organizations utilizing this Assurance

[1] See Section 'Assurance Cases' for an explanation of CAE structured notation.

standard. Organizations will have to select in-scope risks and identify specific controls to address those risks, as well as identify their customers and the configurations supplied to customers. Organizations will need to go further still to support two claims: They will need to provide their Arguments (and Evidence) that benefits outweigh residual risks (Claim 6) and that the configuration delivered to customers better balances risks and benefits than other possible configurations (Claim 7). The variety of potential arguments for these claims precludes a generic prescription in this standard. Thus, organizations are required to make their own. While arguments must be sound, logically valid, and supported by evidence, the primary motivation is to ensure that organizations are making arguments balancing the benefits and risks and not just charging forward with features and configurations unexamined.

The IOPD hopes you find the Assurance Standard approach both refreshing and innovative. The IOPD has crafted the Assurance Standard for organizations to utilize cases to clarify the connections between the evidence and claims. This Assurance Standard approach is novel in comparison to the current landscape where many standards and certifications rehash the Fair Information Practice Principles (FIPPs) or specify requirements or controls without consideration of whether those controls address risks that are created by the product, service, or business process in question.

Privacy professionals who work in "data protection" rather than privacy may find the language used in the Standard particularly jarring, though some terms (e.g. "proportional") will seem familiar. First and foremost, this Standard is meant to address the broader domain of privacy, not specifically data protection (or even "information privacy"). Second, many of the terms are drawn from systems engineering, threat modeling, and other approaches to risk management that are not well represented in a vernacular coming from compliance and rights.

That being said, looking under the hood of this Standard, many of the concepts align very closely with those in data protection. For instance, one can read the defined term Interactions[2] as equivalent to "data processing." Threat Actors include controllers, processors, and third parties, such as cybercriminals. Many Harms in the risk model, such as exclusion[3], mirror data protection rights. After publication of this Standard, the IOPD will begin working on guidance for use in complying with Article 25 of the European Union's Regulation 2016/679/EU (i.e. General Data Protection Regulation or GDPR).[4]

[2] Defined as "An action by and between a Threat Actor and At-Risk Parties or their proxies, such as data related to those At-Risk Parties."

[3] Not defined in this Standard, but Exclusion in the Solove Taxonomy is failure to let an individual know about data processing or participate in its use.

[4] Article 25 of GDPR only applies to controllers and not manufacturers, deployers and processors. However, to the extent that those parties make decisions about the purpose and means of processing, they can be considered controllers. The IPOD will have a full discussion of this in a future guidance on applying this standard in the European Union.

Purpose

The intended audiences of this Standard are privacy professionals, organizations designing, developing, configuring or deploying products, services and business processes, implementers and Assessors as well as privacy and data protection regulators. This Standard serves several purposes for these four distinct audiences:

- **For the privacy professional,** the Standard serves to illustrate an aspirational and achievable objective with regard to the design, development, or deployment of products, services and business processes. It can help with discussions of how to improve privacy in organizations.

- **For organizations**, the Standard represents an achievable objective with which organizations can measure designs and qualify whether a design is privacy- respecting. This can be for the purpose of internal improvement, brand differentiation, compliance with obligations, and/or satisfying ethical imperatives. For those organizations wanting to assert that they have accomplished “Privacy by Design and Default,” the Standard represents a rigorous set of externally validated criteria to back up that claim.

- **Implementers and Assessors** can utilize the Standard in support of their client engagements to assist those clients in achieving “Privacy by Design and Default.”

- **Regulators** can use the Standard as a benchmark by which to review claims and public statements of “Privacy by Design and Default” in the designs of products, services, or business processes by organizations.

Assurance Cases

Assurance cases are a formal approach to establishing confidence in a belief or assertion. Assurance cases are used to explain why a target (system, product, service, or process) is believed to have certain qualities. Historically, the principal property of concern in assurance cases has been safety. Indeed, assurance cases as a generic term emerged out of safety cases. Over time, the approach has been extended to other properties, notably security, and more recently privacy. One of the benefits of using an assurance case is flexibility where a rigid prescriptive requirements-based standard may not be contextually relevant. Given the vagaries of privacy concerns and context, assurance cases seem well suited to privacy.

Assurance cases are based on structured argumentation, a technique that dates back over half a century and in its original form is attributed to British philosopher Stephen Toulmin. Structured argumentation consists of decomposing the different elements of an argument and mapping them and their relations to one another. Thus, claims are specified, evidence supporting those claims described, and the reasoning connecting evidence and claims is articulated. Qualifiers, counterclaims, and counter-evidence may also be included. Structured argumentation aims to explicitly document all aspects of an argument in a way that supports its systematic evaluation[5].

Assurance cases typically employ defined graphical languages for the purpose of documenting their arguments. The two most widely used are Goal Structuring Notation (GSN) and Claim, Argument, Evidence (CAE). A variety of tools are available to support the construction of assurance cases using such languages. This standard utilizes CAE for specifying its privacy by design and default assurance case.

In CAE, a claim is an assertion about an attribute of the system. Arguments can be deductive (i.e. asserting the truth of the claim based on the truth of other claims), inductive (i.e. flowing from supporting evidence), or simply a rephrasing of the claim to support better analysis (i.e. tautological). Evidence flows from the argument and supports the validity of the claim.

The reliance on assurance cases in this standard reflects recognition of both the variety of targets to which it might be applied and long-standing deficiencies with the way privacy risk is typically approached. Efforts to move privacy from a compliance to a risk-based approach notwithstanding, prescriptive approaches (which specify controls regardless of context) continue to predominate, as do privacy failures. This trend has only intensified as the complexity of the socio-technical environment has increased.

Using an assurance case helps to demonstrate that solid grounds exist that support privacy has been appropriately addressed. Its use accommodates the variety of systems, products,

[5] Colloquially, an assurance case can be likened to a legal case, where an advocate presents a claim (“the defendant is guilty”). The advocate then presents an argument as to the claims validity (“the defendant is guilty because they had opportunity, means and motive’). Finally, the argument is supported by the evidence (“the defendant was in the ballroom, with the candlestick and hated the decedent”). The weight of the evidence supports the argument and ultimately the claim.

services, and processes to which this Standard might be applied. Assurance cases are agnostic with respect to the nature of the target of concern and both enable and compel completely customized explication in a standard format of how privacy risks are addressed, independent of any prescriptive list of measures. Ultimately, the assurance case for a given target aims to convey why confidence in the completeness and sufficiency of those measures is warranted based on the privacy risk model that is also part of this Standard. In other words, the goal is to make sure we have confidence in the measures taken to protect privacy, based on a clear understanding of the risks involved.

Typographical Conventions

The first reference to any term in a block of text will include a link to its definition. The text containing that link will be underlined to make clear that it contains a link. The following constructs are defined here:

'A or B' means 'either A or B or both';
'A and B' means 'both A and B';
'A xor B' means 'A or B but not both'

R = the set of in-scope Risks from the Risk Model
R' = the set of Residual Risks after Controls are applied
r_x = specific Risk
r'_x = specific Residual Risk after Controls are applied

C = the set of Controls
c_y = a specific Control

I = the set of interactions between Threat Actors and At-Risk Parties by virtue of the Target System
i_z = a specific Interaction

J = the set of justifications
j_z = a specific Justification

Italics are for non-normative text, typically used for examples or supplementary information, such as analogies or references to concepts in common understanding.

Definitions

Where definitions come from external sources, those sources are referenced in footnotes. Definitions contain cross-reference links where appropriate. In each definition, only the first reference to another definition is linked if the same cross-reference appears more than once in that definition.

Applicant

The Role that applies the standard to the Target System. The Applicant designs, develops, or deploys the Target System.

Argument

Reasoning that provides the bridge between what is known or is assumed (Subclaims, Evidence) and the *Claim* being asserted. The argument used depends on the type, trustworthiness, and extent of available Evidence and the nature of the Claim. Note that "Argument" is an overloaded word. It is used with a specific meaning here.[6]

- **Reasoning step** (supported by subclaims) - an Argument is a reasoning step if a Claim can be deduced from a set of Subclaims.[7] *Example: the animal can reach tree- tops if the animal is a giraffe or the animal is a flying bird.*
- **Evidentiary step** (supported by evidence) - an Argument is a evidentiary step if Evidence makes the Claim more likely than not, based on inductive reasoning. Evidentiary steps may provide a method to measure the confidence of the claim. *Example: the animal is a flying bird if (Evidence it is a bird) and (Evidence of it flying).*
- **Tautological step** (restatement of a Claim for clarity or specificity) - an Argument is a tautological step if a Subclaim simply defines or restates a Claim. *Example: the zoo has an aviary if the zoo has a place to keep birds.*

Assessor

The party that evaluates an Applicant's conformance to the standard. Assessors may be internal (a department or individual employed by the Applicant) or external (a party contracted by the Applicant to review their conformance).

[6] Adelard (NCC Group), CAE Framework. Available at https://claimsargumentsevidence.org/notations/claims-arguments-evidence-cae/

[7] Rushby, J.M. (2015) The Interpretation and Evaluation of Assurance Case. Available at https://www.csl.sri.com/~rushby/papers/sri-csl-15-1-assurance-cases.pdf

At-Risk Party

A Role impacted by a Harm because of their Role in the Target System. While generally an individual (*i.e. natural person*) is at risk, the term here is not limited and may be used, in context, by the Applicant to refer to a non-natural person, such as a business, that can be impacted by a Harm.

Benefit

A desired consequence of an Interaction.

Bystander

A Role whose existence is immaterial to the operation of the Target System. *An example would be a person in the background of a photograph.*

Claim

An assertion about a property of the Target System. A Claim is **Mandatory** if it is required by this Standard. A Claim is **Selective** if required by this Standard, but which allows the Applicant to select the specifics of the Claim. A **Subclaim** is a Claim that is made as part of an Argument supporting another Claim.

Configurability

The ability to change settings in the Target System. In the illustration below, a lowered switch cover prevents changes to the configuration of the switches. As used in this standard, an Applicant chooses the configurability of the system (which switch covers are up or down) to enable the Customer to configure the system (turn switches on and off). The Applicant also chooses the configuration of the system as delivered to the Customer.

Figure 1: A physical set of switches which can be enabled or disabled. The settings of those switches represent a potential configuration of the system. The switch covers represent the configurability of the system. Lowering a switch cover is analogous to removing the configurability of a particular setting (the switch is set to whatever setting was made before the switch cover was closed).

Configuration

System settings choices made regardless of Configurability. In the illustration under Configurability, the switch settings (whether they are on or off) represent the Configuration, regardless of whether it is further configurable (which is dependent on the position of the switch cover). Configuration is often plural to denote that there may be multiple sets of settings delivered to different types of Customers.

Consequence

A desired (Benefit) or undesired (Harm) result of an Interaction.

Consumer

A Role that receives Benefit from the Target System (i.e. they consume the output of the Target System).

Contracted Party

A Role in contract (directly or indirectly) with the Applicant. *Contracted Parties include vendors, clients, partners, employees, contractors, and their vendors', clients, employees, and contractors, and others*.

Control

An action taken by the Applicant to reduce Risk. Controls are organized into two types, System Controls and Environmental Controls, though some actions may satisfy both types. For the purposes of this Standard, Controls are limited to System Controls, and any reference to Controls means System Controls. Environmental controls are often implemented by the Applicant, Customer, or Other Parties and affect the environment in which the Target System operates. *For instance, a Control that restricts the sale of the Target System in repressive regimes would be an Environmental Control. Similarly, another Environmental Control would be a process to conduct a risk assessment of the Target System or an access management policy*. System Controls are implemented within a Target System. As systems may be sociotechnical, System Controls can, but need not be, technical. *Examples include probabilistic, risk based, access controls, or a contract dictating terms with a vendor within the Target System.*

Environmental Controls	*System Controls*
Access Management Policy	*Access Controls*
Conducting a Risk Assessment on the Target System	*Probabilistic, risk based, access controls*
Vendor Management Process	*Contract terms for a vendor who performs some system functions.*

Table 1 Comparison Of Environmental and System Controls

Customer

A Role that receives, from the Applicant, a Configuration of the Target System. The term 'receives' includes license, lease, purchase, and other forms of procurement and does not require payment. An Applicant may have multiple types of Customers, depending on sales channels, markets, industries, segments, and verticals. Customers need not pay for the Target System. A Customer may have other roles in the Target System (e.g. Threat Actor or At-Risk Party). Customers need not operate the Target System. They may be a distributor, resellers, installer, or otherwise repurpose the Target System for further delivery.

Evidence

An artifact that establishes facts that can be trusted and lend confidence to the truth of a *Claim*. In projects, there can be many sources of information, but what makes this evidence is

the support or rebuttal it gives to a Claim.[8]

Functional Requirement

A defined constraint on a system that affects the system's environment outside the system boundary.

Harm

An undesired Consequence of an Interaction.

Interaction

An action between a Threat Actor and At-Risk Parties or their proxies (e.g. data related to those At-Risk Parties).

Justification

A statement supplied by the Applicant as to why an Interaction should be allowed in light of the potential Risks. *In GDPR parlance, Justifications are a combination of purposes of processing activities and legal bases*

Necessary

A characteristic of a proposed Interaction based on the need to meet Functional Requirements or Non-Functional Requirements of the Target System.

Non-Functional Requirement

A defined constraint on a system that affects its operations within the system boundary.

Non-contracted Party

A Role not in contract with the Applicant but contemplated as part of the Target System. *An internet service provider (ISP) for an internet connected device would be a Non-contracted party, contemplated as needed by the Target System, but not in contract with the Applicant*

[8] Ibid

designer.

Other Party

Any Role not performing a function in the Target System but that could interact with an At-Risk Party or their proxy, such as data related to the party, by virtue of the Target System's operation.

Proportionate

The notion that a measure of the Applicant's supplied Justification exceeds a measure of the Risk against which it is compared. Note that measures need not be a simple risk calculation, but may include factors of equity, fairness, or other ethical concerns.

Operator

A Role that operates the Target System to produce Benefit for others. Conventionally, this is a worker whose labor is used to produce the output of the Target System.

Resource

A party whose existence is material to the Target System. *An example would be a data subject of a data brokerage service*.

Risk

A measure of likelihood and severity of Harm using the Risk Factors under the Risk Model.[9] For the purposes of this standard, Risk refers only to privacy related risks.

Residual Risk

A measure of Risk remaining after a change in the context, such as applying Controls.

Risk Factor

A characteristic used in a Risk Model as an input to determining the level of risk in a risk

[9] Institute of Operational Privacy Design, Inc., Design Process Standard, v 1.0 (2021). Available at https://instituteofprivacydesign.org/certification-standard/

assessment.[10]

Risk Model

A representation that elaborates key terms and abstract factors that contribute to or negate Harms (see NIST definition).[7] This standard uses a specified Risk Model (see section 6).

Role

A party's relationship to the Target System. A party may play multiple roles within one Target System.

Target System

The system designed, developed, or deployed and scoped for evaluation under this Standard.

Threat Actor

A Role whose action could result in a Harm to an At-Risk Party. This standard defines four categories of Threat Actors: Applicant, Contracted party, Non-Contracted Party, Other Party. Threat Actors may be further classified by the Interactions they engage in (*e.g. Contracted party call centers*).

Threat

A potential action by a Threat Actor that, if realized, could result in Harm(s) to At-Risk Parties. For the purpose of the Risk Model used in this Standard, Threats are implied from the category of potential Harm: processing of data, dissemination of data, attempted collection of data, and invasions into personal space or autonomy.

Vulnerability

A condition or state that puts a party at risk of experiencing Harm. For the purposes of the Risk Model used in this standard, there are two vulnerabilities that arise from Interactions: (1) Threat Actor interacts with a party or their data, and (2) Threat Actor has control, though not necessarily possession, of a party's data.

[10] NIST Cybersecurity Resource Center, Glossary, entry for risk factor. https://csrc.nist.gov/glossary/term/risk_factor

Figure 2 is provided to help readers visualize the relationships between various defined entities and attributes. The primary relationship is between the Applicant and the Target System (shown in bold outline), of which there is only one of each, for the purposes of this Standard.

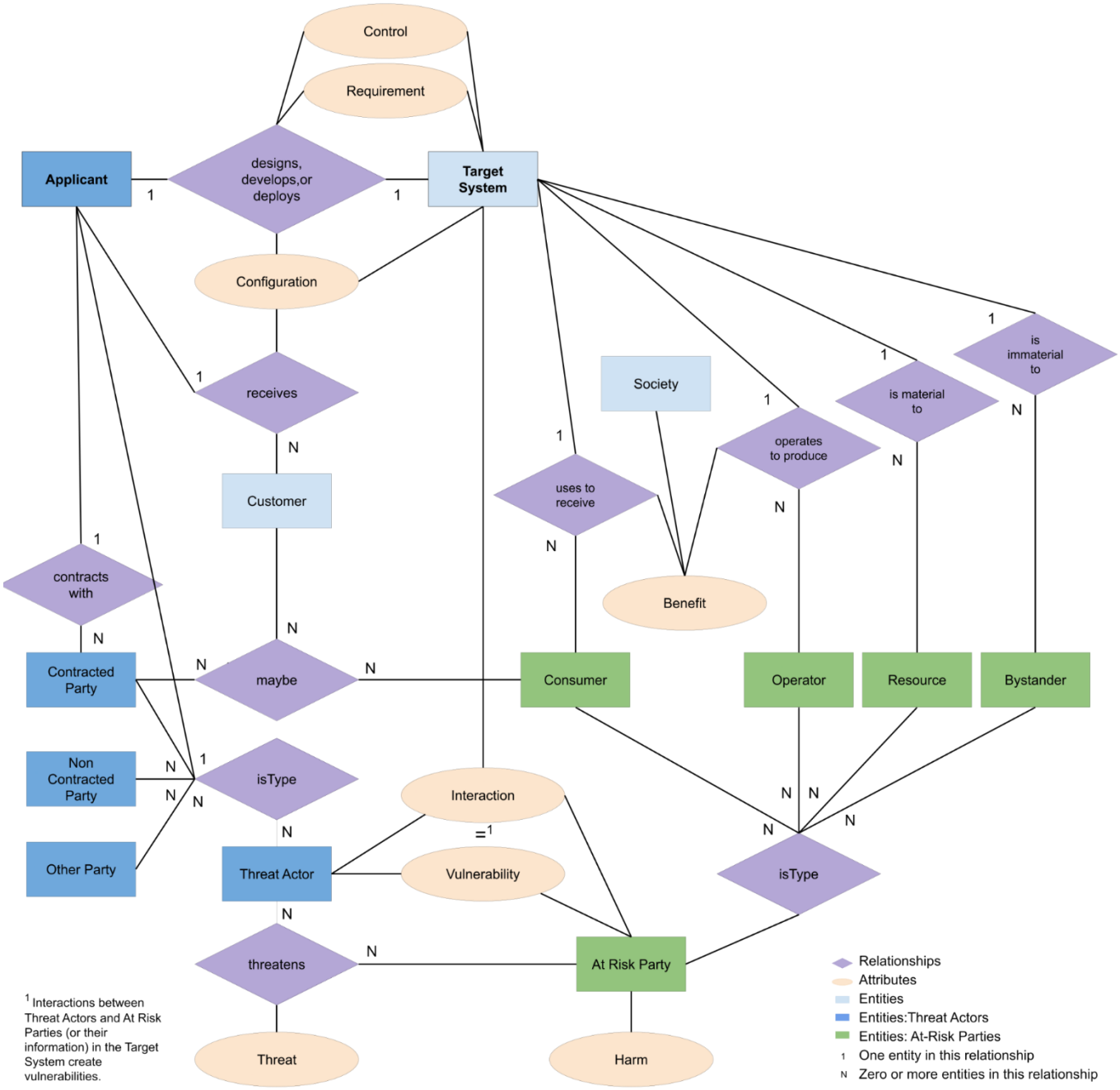

Figure 2 Entity Relationship diagram

Risk Model

A risk model is a construct that provides the basis for risk assessment of a product, service, or business process. A complete risk model consists of component models for threats, vulnerabilities, and adverse consequences, along with ways of representing likelihood and impact severity. The component models reflect the chain of elements that result in risks, in which threats exploit vulnerabilities resulting in adverse consequences. Risk models are essential for risk assessment as they specify risks of concern in a given domain and define how those risks can manifest. Not every threat will be capable of exploiting every vulnerability, nor will every exploitation lead to every possible adverse consequence. The purpose of the risk assessment is to identify those alignments of threats, vulnerabilities, and consequences that are viable combinations for the target.

There are a variety of pre-existing privacy risk models, though most are missing one or more components and are therefore incomplete. Applicants may leverage a pre-existing model or develop one or more that are tailored to their domains of operation.[11] The assurance case at the heart of this standard, however, must be constructed with reference to a specific privacy risk model to standardize the analysis of the Claims, Arguments, and Evidence constituting the case. Customized risk models would require significantly more review and analysis of the case to determine if the risk model sufficiently and completely addresses the risks. Such flexibility would also increase subjectivity and opportunities for gaming the standard, causing confusion in the marketplace and regulators looking for standardization. The standard would be rendered largely meaningless if each Applicant could utilize a different privacy risk model in their assurance case, and consistently evaluating those cases would become unmanageable. Therefore, this Standard requires Applicants to employ the common privacy risk model defined here. This model is general enough to accommodate all Applicants and Targets. The model leverages Solove's Taxonomy of Privacy[12] problems, a widely held and used model of privacy, and defines different types of:

- Vulnerabilities
- Threats
    - Actions
    - Actors
- Consequences
    - Harms

[11] The IOPD Design Process Standard v 1.0 allows organizations complete flexibility in defining their risk model and selecting their risks.

[12] Solove, Daniel, A Taxonomy of Privacy (2006), University of Pennsylvania Law School. Available at https://scholarship.law.upenn.edu/penn_law_review/vol154/iss3/1/

- At-Risk Parties

Explication of likelihood and severity is left to the Applicant. The included table (Figure 3) details this model. The model defines two Vulnerabilities, each of which can be exploited by two actions (resulting in four threat actions). The two Vulnerabilities and corresponding threat actions are:

- **Vulnerability 1:** Threat Actor interacts with At-Risk Parties or data related to At-Risk Parties
  - **Threat action I:** Threat Actor invades personal space or disrespects autonomy
  - **Threat action II:** Threat Actor attempts to collect or solicit information
- **Vulnerability 2:** Threat Actor has control of data related to At-Risk Parties
  - **Threat action III:** Threat Actor processes information
  - **Threat action IV:** Threat Actor disseminates information

These four threat actions can be taken by any of four types of Threat Actors: the Applicant, Contracted Parties, Non-contracted Parties, and Other Parties. This results in 16 potential Threats (e.g. a potential action by a Threat Actor).

Each of the different threat actions results in exactly one of a set of related Harms. Those Harms follow the categorization of privacy harms under Solove's taxonomy.

Threat action I: Threat Actor invades personal space or disrespects autonomy (which can lead to)

- physical or psychological intrusion or interference with decision making (i.e. invasion harms)

Threat action II: Threat Actor attempts to collect or solicit information (which can lead to)

- surveillance, interrogation (i.e. collection harms)

Threat action III: Threat Actor processes information (which can lead to)

- aggregation, secondary use, identification, insecurity, and exclusion (i.e. information processing harms)

Threat action IV: Threat Actor disseminating information (which can lead to)

- disclosure, exposure, increased accessibility, distortion, breach of confidentiality/trust (i.e. information dissemination harms)

The model contains four different types of At-Risk Parties, which are: Consumers, Operators, Resources and Bystanders. Each of these four groups of At-Risk Parties can be impacted by any of the four threat actions (and corresponding harms) perpetrated by any of the four Threat Actors leading to a total of 64 (4*4*4) potential risks. Not all risks may be pertinent to all Target

Systems. Furthermore, the Applicant will scope the Risks they wish to address when applying the Standard, descoping risks due to negligible likelihood or impacts or due to other considerations, such as analysis costs or market demands.

Example risk: the Applicant (a Threat Actor) processes information (a threat action) of a Consumer (an At-Risk Party) considered a secondary use of data (a Harm).

Example risk: a potential client (a Non-Contracted Party / Threat Actor) of the Applicant asks (a threat action) an employee (an Operator / At-Risk Party) during a marketing call (an Interaction which creates a Vulnerability) being made by the employee a personal question (a Harm, specifically a collection harm).

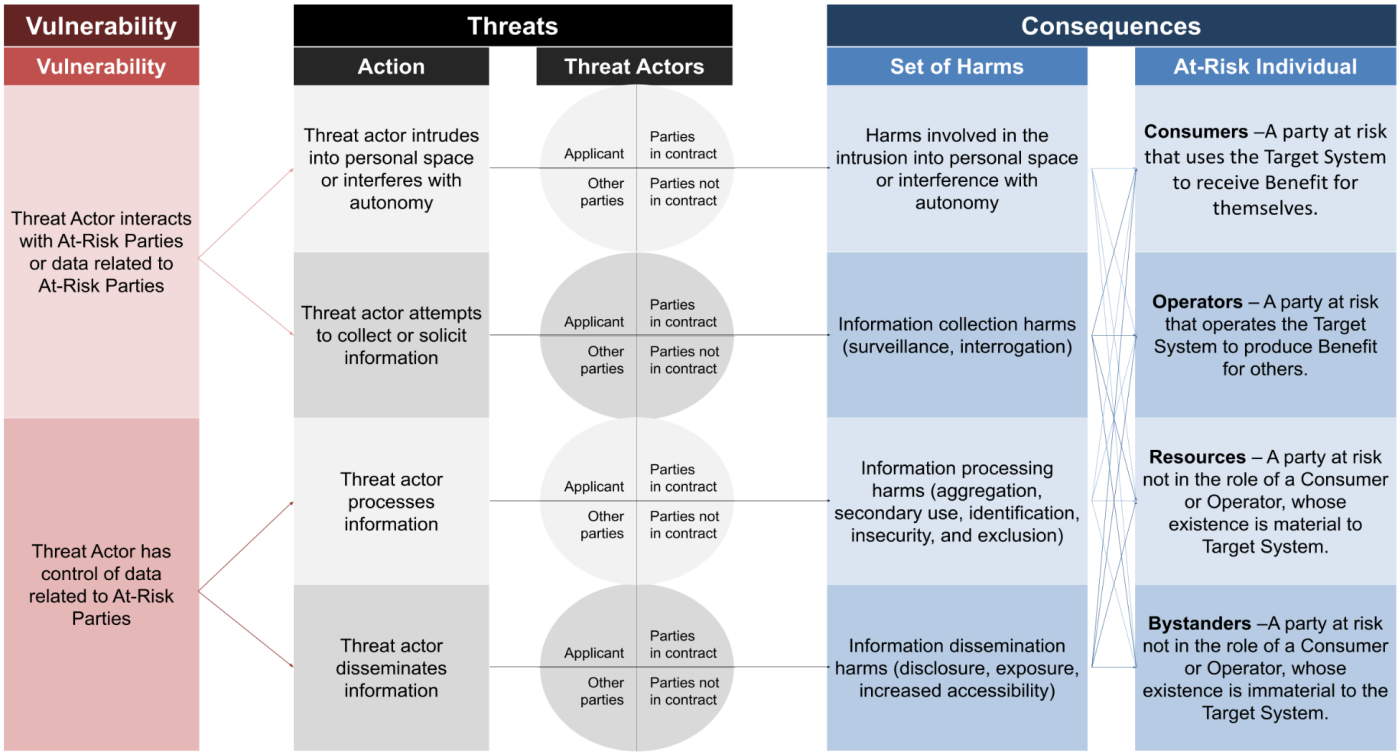

Figure 3 Risk Model illustrating relationships between Vulnerabilities, Threats and Consequences in the model.

Note that Roles (both Threat Actors and At-Risk Parties) may shift depending on the perspective taken from the context of the Applicant, the Target System, and that system's requirements. *If the perspective is that of a platform owner as Applicant ("Applicant A1"), for example, then the Target System is a service (e.g., the payroll platform). The recipient company is the Customer. HR staff would constitute Consumers (i.e.they receive the benefit of the platform as it makes their job easier), company employees whose information has been*

uploaded by the HR staff would be considered Resources (i.e. they are material to the service). See Figure 4.

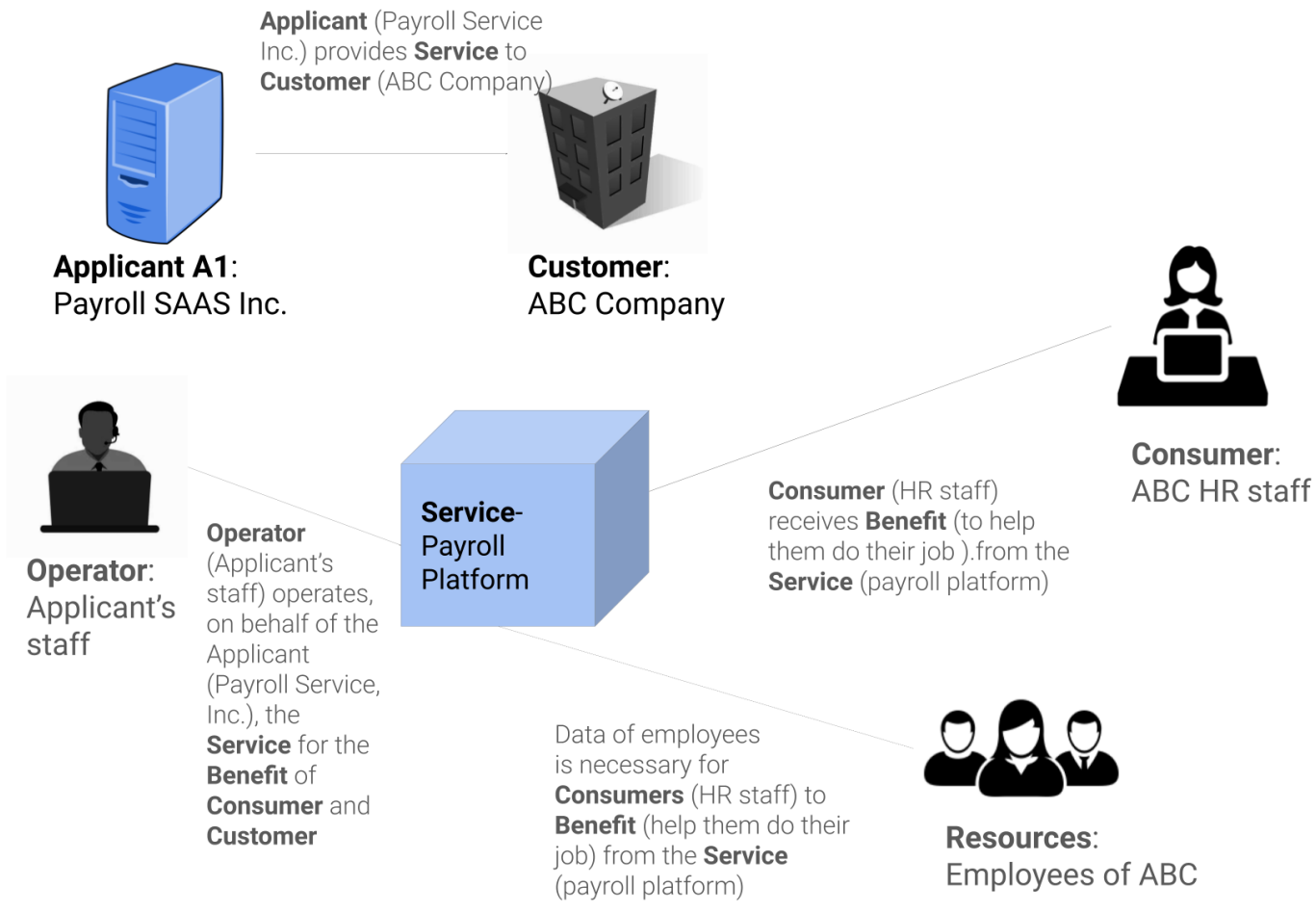

Figure 4 Example of a Payroll Platform service showing the relationships of various parties.

If, however, the Applicant is the company ("Applicant A2") using the platform to provide payroll to their employees, the analysis shifts. The platform owner is a Contracted Party and the company's employees might now be considered Consumers.Why? Because the context of the Target System has changed. The service has shifted. Whereas the platform owner was providing a payroll system service, Applicant A2 is providing a business process (e.g., payroll process). HR staff are Operators, operating the payroll process to pay employees. Employees are now the Consumers, receiving the benefit of the business process. The platform is merely a tool in the business process of paying employees. This is a nuanced understanding, but required to construct the appropriate chain of actors and actions. See Figure 5.

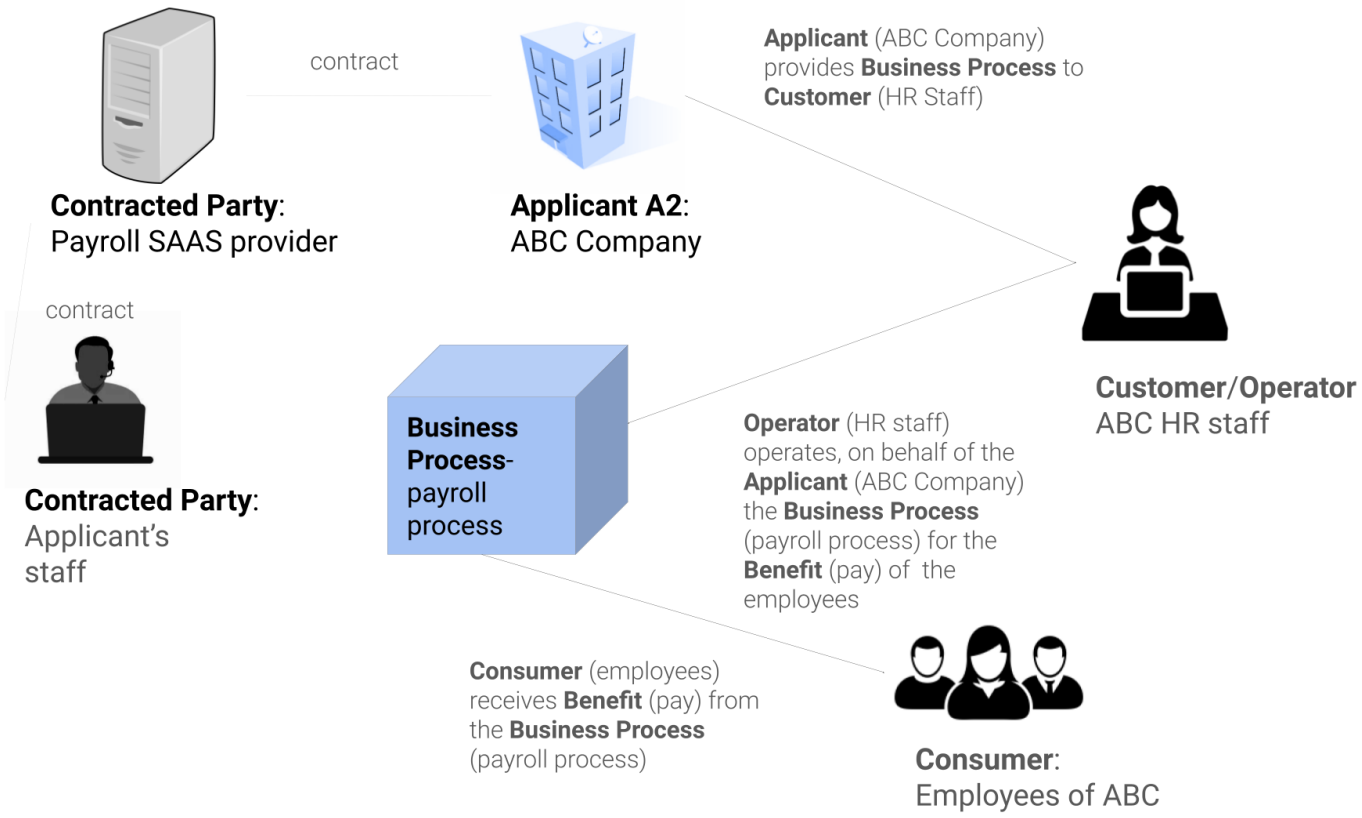

Figure 5 Example of a payroll process using the payroll platform service in Figure 4 showing the relationships of various parties.

Consider another short example. In a rideshare app, where the company providing the app is the Applicant, both drivers and passengers are Consumers of the app. The Target System is the app whose requirement is to match drivers and passengers. But if a driver were to apply the Standard to their service, a privacy-respecting ride, the perspective changes. The driver is the Operator of the ride service, the passenger is the Consumer. The app isn't providing the ride, they are just a Contracted Party and, consequently, a Threat Actor.

Context is crucial in the analysis of the Risk Model and definition of the Target System.

Scope Selection

Scoping is a critical first step, before applying the Standard to a Target System. Scoping includes

- Defining the Target System
- Identifying the intended Customer(s)
- Identifying the classes of Risk treated (including a delineation of the Threat Actors and groups of At-Risk Parties.)

Defining the Target System

The Applicant must first identify the product, service, or business processing being provided by the Applicant to the intended Customer. This can be done using a plain English description of the product, service, or business process, including intended purposes, uses, and basic functionality. Where the boundaries are unclear, such as systems that interface with other components, related or consuming systems, the Applicant's responsibilities should be made clear. The Applicant should also make clear whether their role is as a designer (i.e. they make basic design decisions and document the result of that decision), a developer (i.e. they take a design and construct a live functioning instantiation of that design), or the deployer (i.e. they take an instantiation and deploy it in an operational environment) of the Target System. Applicants may take on multiple roles, but the description of the Target System should make clear which roles Applicants play and for which components of the system.

Applicants must further delineate the Target System with Functional and Non-Functional Requirements. These specific constraints define what the system is supposed to do and how it is supposed to do it, which helps narrow the scope (*e.g. defining a Target System as a computer program developed by the Applicant is not as helpful as a computer program which adds numbers with up to 10^{100} and with a precision to 99 decimal places*). Functional and Non-Functional Requirements that have no bearing on the system's Risk need not be included, though the Applicant should conduct an analysis to ascertain that. Note that providing Functional and Non-Functional Requirements is also part of Evidence 2.1.1 in Claim 2.

The Target System should be scoped from the perspective of the Applicant. *For example, the WordPress Foundation designs and develops the software product, WordPress.Customers of this organization, such as WordPress.com, deploy the software in various environments. Some of them have customers that may design websites deployed on WordPress instances.*

The Applicant could be WordPress Foundation, in which case the Target System is the software (i.e. product), Wordpress.com, in which case the Target System is the hosting platform (i.e. service), or the website using wordpress.com, in which case the Target System is the website (i.e. service).

Identifying the Customers

Applicants may have different sales channels, markets, industries, verticals, or other segments. Applicants must identify the distinct types of Customers that they serve, at least to the degree that those Customers receive distinct Configurations. Customers may warrant differing Configurations because of their varying risk profiles (e.g. government customer channels will have different risk profiles than private sector customer channels), Identifying Customers can be complex for large, multi-channel products and services. Applicants need NOT scope every Customer channel to apply this Standard. For instance, an Applicant may decide to only analyze or certify their consumer market.

Customers should not be confused with Consumers, though they may sometimes be the same. Customers receive the product or service from the Applicant. Consumers use the product or service for some function. For example, Customers for a business process would be the recipient of the output of the process (e.g., the customer of the budgeting process would be the department receiving the budget; the customer of a shipping process would be the fulfillment department; the customer of a marketing campaign development process would be the department whose product is being marketed). A Consumer of a business process is the one who uses the business process to perform its function (e.g., the marketing department uses the marketing campaign development process to develop a marketing campaign). The Operators of a business process are the parties providing the process to the Consumer. In the marketing example, the Operators include the company (who provides the people to complete the process), marketing department management (who provides the steps of the process), and the information technology department (who provides the technology).

Identifying the Risks

The last part of scoping involves determining in-scope Risk. The Risk Model provides for 64 distinct risks (four Threat Actors x four sets of Harms x four types of At-Risk Parties). It is not expected that Applicants address all 64 risks. Instead, Applicants are encouraged to narrow the scope. There are two standard approaches to selecting risks. Scoping Risk cannot simply reflect whether Controls are in place, ignoring Risk(s) which have not been treated.

Approach One: Risk Subset Selection based on Intended Audience

The Applicant selects Risks relevant to the audience for whom the Standard is being applied. For instance, if the Applicant wishes to showcase to their target market the Applicant's reduction of Risk from particular types of Threat Actors then the Applicant is free to choose those limited risks. Perhaps also the Applicant is only concerned about data sharing related risks and thus selects harms related to information dissemination. Whatever the scoping decision, the Applicant should provide a reasonable Justification for the scope.

Approach Two: Risk Subset Selection based on relevance of Risks

The Applicant selects all 64 risks in the Risk Model. Then the Applicant systematically reviews

each risk and eliminates risks that are 1) not relevant to the Target System or 2) where the risks are negligible.

Some risks may not be relevant to a particular Target System. For instance, maybe the Target System does not involve any Bystanders or Bystander data. Making note of this could eliminate whole classes of risk. In the example where Bystanders are not relevant, 16 risks are eliminated from consideration.

Some risks may be negligible, in either likelihood of occurrence or impact to At-Risk Parties. This would not include Residual Risks, the measure of risk after Controls have been applied, but Risks at the outset. For example, the Applicant might assert that risks related to information sharing harms are negligible because information is only shared internally[13]. In general, scoping of risks is at the discretion of the Applicant, who must describe the basis for their decisions.

[13] This is not to say this is a valid Justification for dismissing information sharing related harms.

Organization of the Assurance Case

The assurance case in this standard follows a Claims, Argument, Evidence (CAE) structure. Claims are subject to Arguments, which are supported by Subclaims or Evidence. Usually, the Applicant need not create a privacy assurance case from scratch - the preliminary work has been done by the IOPD Standards Committee in drafting this standard. Because the case has been laid out, Applicants are also not free to alter the verbiage of the Claims, Arguments, or Evidence statement, except where latitude is granted to populate the statement with selected options. Applicants are tasked with selecting certain arguments relevant to their Target System. Applicants are also tasked with selectively applying Claims, Arguments, and Evidence to in-scope Risks. Ultimately, the Applicant must provide the Evidence based on their selections. There are two deviations from this that Applicants should be aware of.

For the purpose of Claim 6 “Benefits Outweigh Residual Risk”, the Applicant must, for each in scope Residual Risk, construct Argument 6.1 for their particular Target System and context in which that Target System is designed, developed, or deployed. Depending on the Argument, the Applicant will then need to construct Subclaims, additional Arguments, and Evidence to support the Claim that Benefits outweigh Residual Risk.

Similarly, in the Privacy by Default prong of the case, Applicants must construct an argument for Claim 9 “Changes to the Configuration(s) as delivered to the Customer(s) would create an undesirable balance of Benefits and Risks”. Applicants may use the same argument structure for each Configuration delivered, though supported by differing Evidence, or Applicants may provide distinct arguments for different Configuration(s). Regardless, each Configuration delivered must be supported by the Claim that alterations would be undesirable.

Mandatory Claims, Arguments, and Evidence are indicated in the following case descriptions. They are illustrated as solid borders in the diagram in the following section. Selective Claims, Arguments, and Evidence (i.e. those where the Applicant must make a selection from in scope risks or identified controls) are indicated as such in the case descriptions and illustrated with dashed borders in the diagram.

Case: Privacy by Design and Default

The Privacy by Design and Default Case includes one root Claim, three supporting Claims, and 10 Subclaims. Each Claim is supported by arguments which are further supported by Subclaims or Evidence. Figure 6 contains the complete case flow from Evidence to the root Claim (indicated in green). Shapes in blue indicate where the Applicant must complete the argument with their own construction. Diagrams 2 and 3 detail the Privacy by Design and Privacy by Default prongs, respectively, including the Claim, Argument, and Evidence statements.

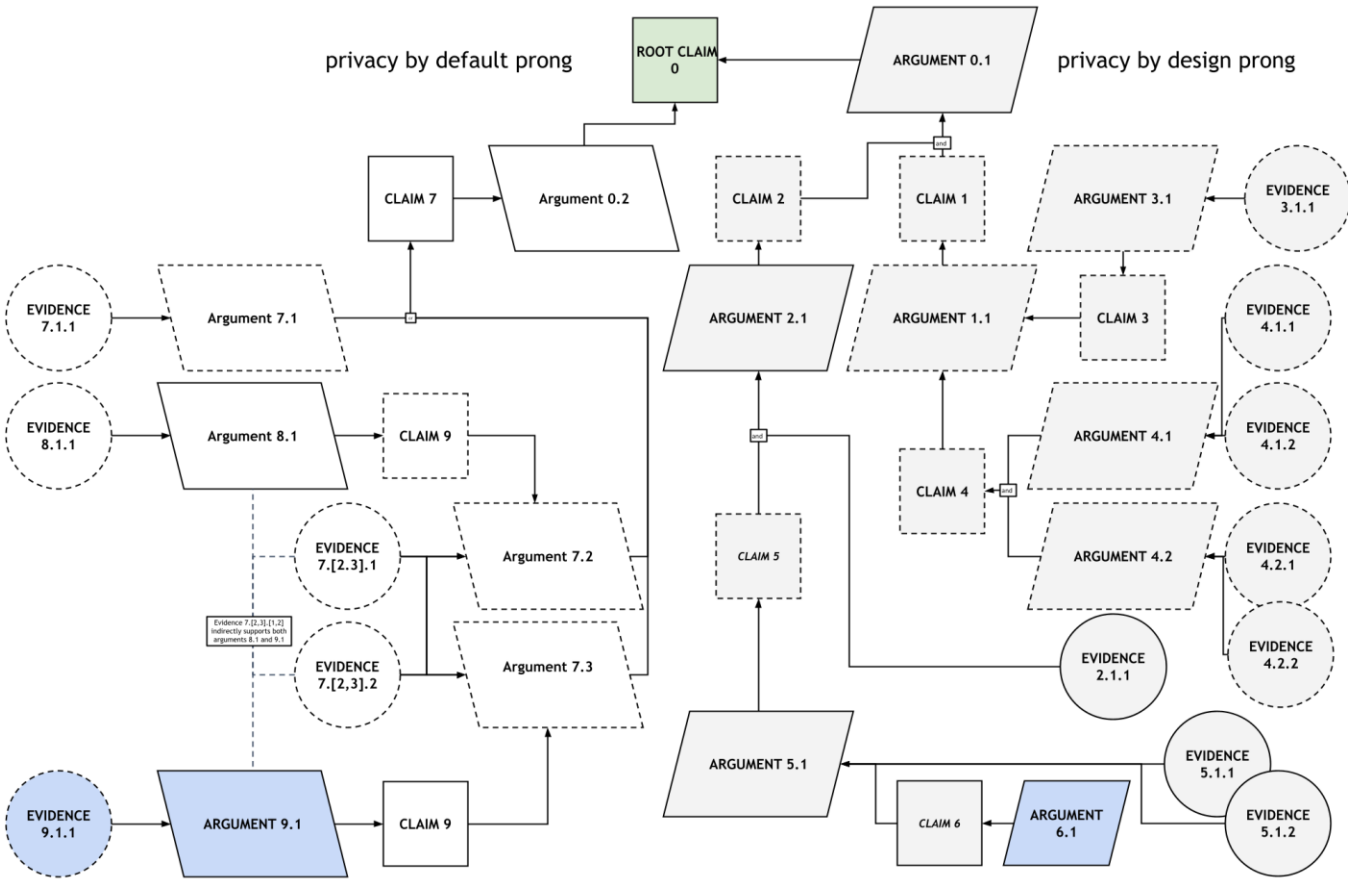

Figure 6 Privacy by Design and Default Case Structure

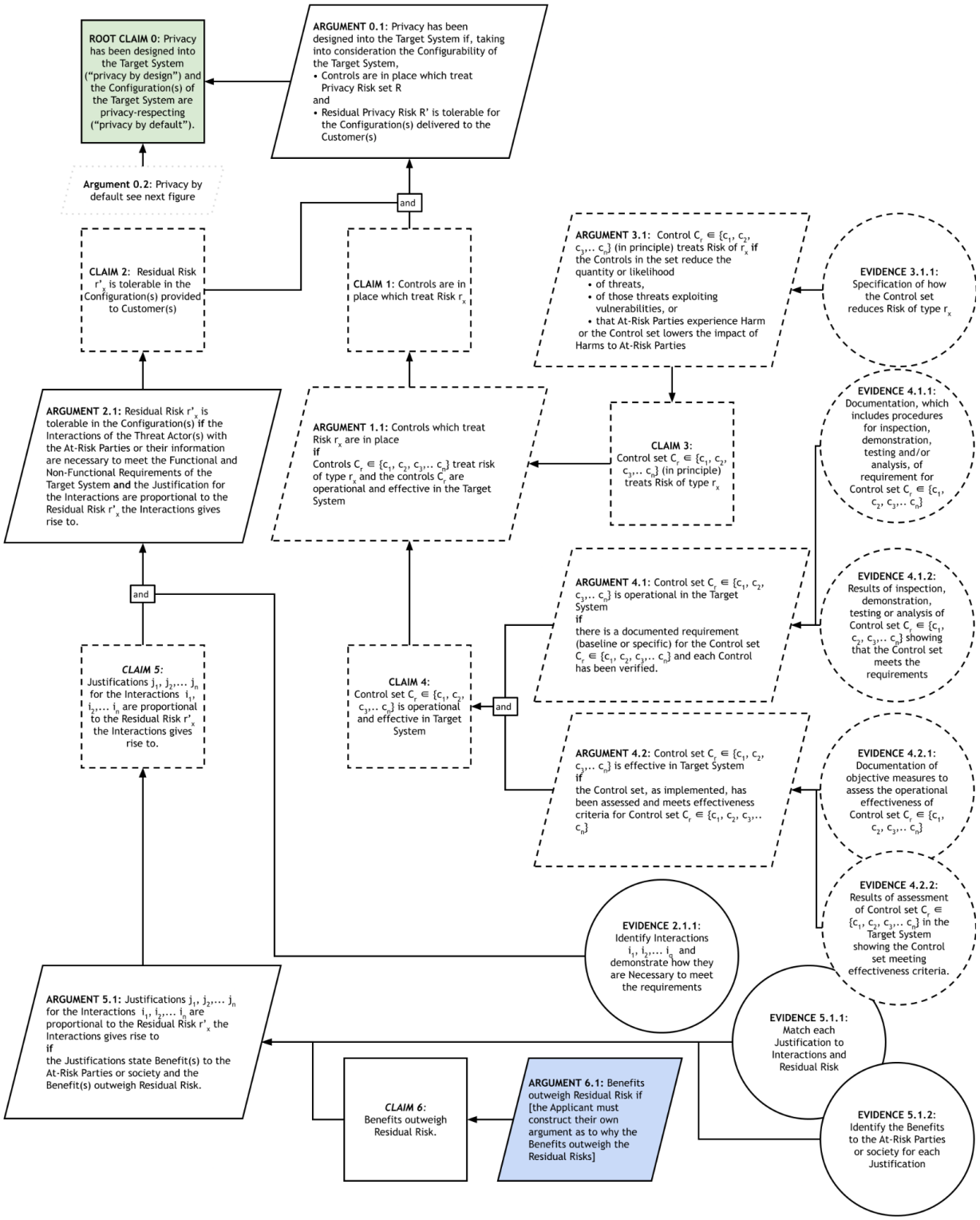

Figure 7 Privacy by Design Case prong

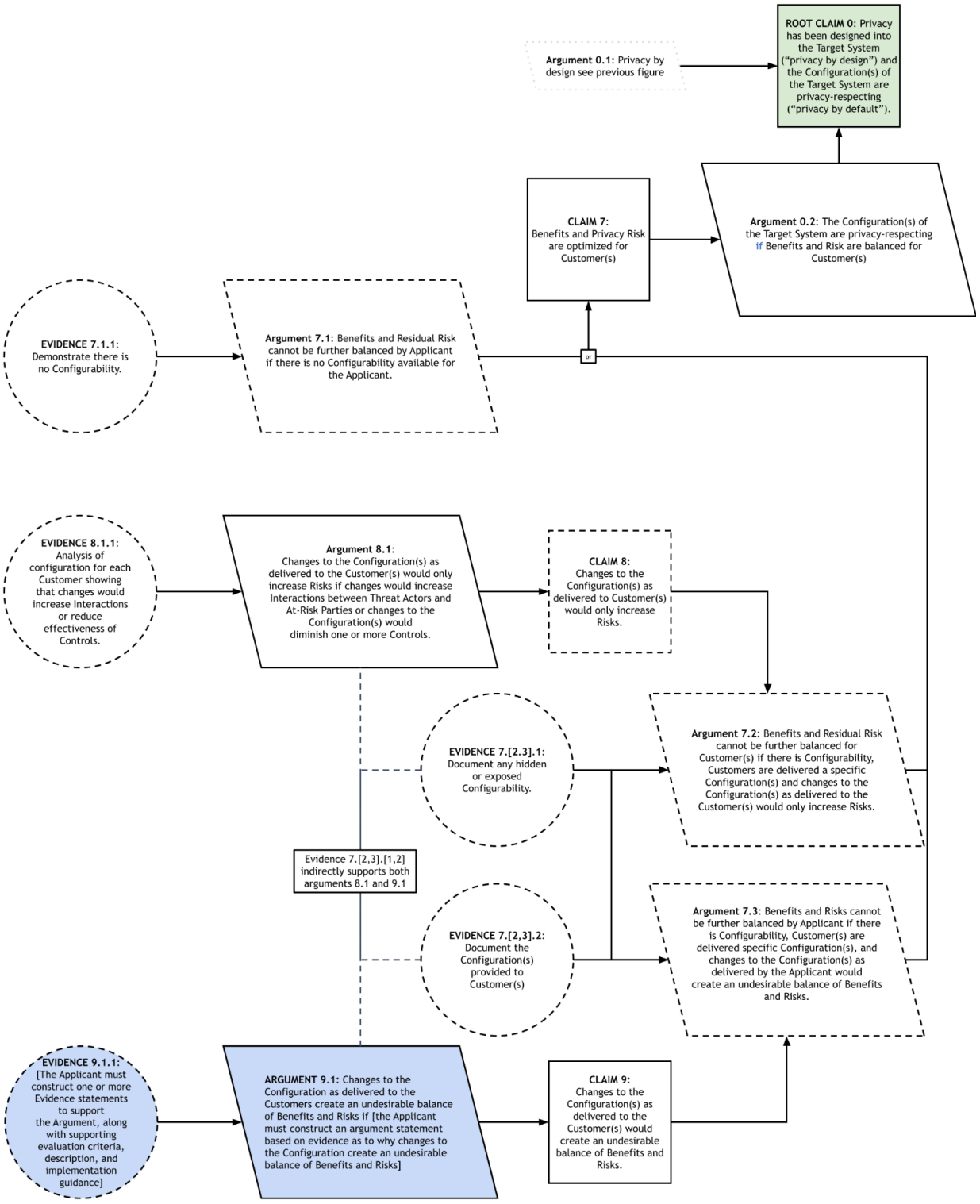

Figure 8 Privacy by Default Case prong

600 # Root Claim

Claim 0	Privacy has been designed into the Target System ("Privacy by Design") and the Configuration(s) of the Target System are privacy-respecting ("Privacy by Default").
Mandatory	**Description:** The root claim is essentially a restatement of the concept of privacy by design and default. There are two parts to the claim. First, that privacy has been designed into the Target System (i.e. the object of evaluation). When applied to the design, "privacy" can be thought of as a quality of the system. In other words, there is a thoughtfulness in the design that addresses privacy concerns. Applicants may note that, though the claim is about the design of the system, the subsequent arguments, subclaims, and evidence are not about the process of the design but rather the end results. The IOPD's Design Process Standard covers the design, development, and deployment of systems with privacy taken into consideration at the beginning instead of being bolted on after. The second part of this claim restates what is meant for a Target System to exhibit "Privacy by Default." Privacy by Default is more complex than Privacy by Design. It essentially means that, as delivered to Customers of the Applicant, the Target System's settings strike a balance between Benefits and Risks.
Argument 0.1 Reasoning Step	Privacy has been designed into the Target System if, taking into consideration its Configurability, Controls are in place which treat Privacy Risk set R and Residual Risk set R' is tolerable for the Configuration(s) delivered to Customer(s).
	Description: What does it mean to design privacy into a system to address privacy concerns? First, there is an understanding by the Applicant designing, developing, or deploying the Target System that systems create a set of Risks for parties, denoted R in the Argument statement. In other words, there is a chance that some Interaction(s) within the system will occur that will negatively impact a party and broadly that these Interactions fall under the umbrella of what's considered a Privacy Harm. Once Risks are understood, the Applicant seeks to address those risks through Controls. Controls rarely eliminate Risks but are designed to reduce them. What's left is the set of Residual Risks, denoted R' in the Argument statement. Where Risks cannot be eliminated, the Residual Risks must be tolerable (see Claim 2).

	R denotes the set of risks selected in scoping by the Applicant. R' denotes the set of risks in R after controls have been applied. Individual risks within these sets are denoted r_x and r'_x, respectively. The references to Configurability and Configuration in the Argument statement take into consideration that some controls or functionality of the system may be enabled or disabled by default and enabled or disabled by the Customer or others. Controls need not be enabled in the Configuration to treat risks by default. They could be disabled because, for instance, the control limits the functionality of the system. **Subclaims:** Claim 1 Controls are in place which treat Risk r_x Claim 2 Residual Risk r'_x is tolerable
Argument 0.2 Tautological Step	The Configuration(s) of the Target System are privacy-respecting if Benefits and Risks are balanced for Customer(s).
	Description: Privacy-respecting is not an absolute. There is a balance between Benefits (to the Applicant, to Customers, to At-Risk Parties, to society, and other stakeholders) and Risk to parties. To be privacy respecting, the Applicant must balance these Benefits and Risks when delivering the Target System for use by the Customer(s). This is done within the confines of the Configurability of the Target System, giving Customers a Configuration to meet their needs without creating undue Risks for At-Risk Parties. **Subclaims:** Claim 7 Benefits and Risks are balanced for Customers

601

602 # Claim 1: Controls are in place which treat Risk r_x

Claim 1	Controls are in place which treat Risk r_x
Selective	**Description:** Controls are actions that reduce risk. The crux of this claim is that the controls are in place in the Target System.
Argument 1.1 Reasoning Step	Controls that treat Risk r_x are in place if Controls $C_r \in \{c_1, c_2, c_3,..c_n\}$ treat Risks of type r_x and the Controls C_r are operational and effective in the Target System.
	Description: To be in place, Controls must be designed and operating effectively to appropriately reduce or eliminate risks to parties. Not every Control treats every risk type. Appendix II provides a noncomprehensive mapping of common privacy controls and the manner in which they address risk in this standard's Risk Model. Assuming a Control treats a type of risk, that Control must be implemented, functional, and functioning effectively in the Target System. If all of these are true, then the claim can be justified. Note: having a Control in place does not mean that a Control must be actively preventing risk. A Configuration may enable or disable a Control. The claim here is that the controls, if enabled, will treat risk. The determination of whether a Control needs to be enabled by default is made as part of Claim 2 Residual risk is tolerable in the Configuration(s) provided to Customers. The Control may address risk not present in the default, may only be triggered if a risk materializes, or may address risk for particularly risk averse parties. "In place" merely means it is functionally available should it be needed. **Subclaims:** Claim 3 Control $C_r \in \{c_1, c_2, c_3,.. c_n\}$ (in principle) treats Risks of type r_x Claim 4 Control c_y is operational and effective in Target System

603

Claim 2: Residual Risk r'_x is tolerable in the Configuration(s) provided to Customers(s)

Claim 2	Residual Risk r'_x is tolerable in the Configuration(s) provided to Customer(s)
Selective	**Description:** Each Residual Risk remaining after application of Controls must be tolerable. The inclusion of Configuration(s) recognizes that not all risks may be present in the default due to some functionality being absent, and that not all Controls need be active if the risks they treat are not present. Customers are free to change the Configuration, subject to the Configurability of the system, to enable functionality or disable Controls, to meet their own needs and risk tolerance. But, the Configuration(s) as delivered to Customer(s) must be tolerable "out of the box".
Argument 2.1 Reasoning Step	Residual Risk r'_x is tolerable in the Configuration(s) provided to Customer(s) if the Interactions of the Threat Actor(s) with the At-Risk Parties or their information are Necessary to meet the Functional and Non-Functional Requirements of the Target System and the Justification for the Interactions are proportional to the Residual Risk r'_x the Interactions give rise to.
	Description: Interactions resulting from the Configuration(s) must include only those Necessary to meet the requirements of the Target System. While additional Interactions, which introduce additional risk, may be enabled, the concern here is the Interactions contemplated while the Target System is in the specific Configuration. Necessity is the key to this part of the argument. If it is not Necessary, it should be left to subsequent configuration rather than enabled. For each of those Interactions, there must be a Justification, beyond its necessity to system requirements. Without a Justification, an unfounded system requirement could be established necessitating an Interaction (e.g. requirement: collect email addresses). Justification provides the reasoning behind the requirement, and ultimately, the Interaction (e.g. justification: to communicate with the user about their account). Further, Justification for those Interactions, both individually and collectively, must be Proportional to the Residual Risk(s) resulting from those Interactions. The proportionality of each Residual Risk is measured in Claim 5.

Argument 2.1 *Continued*	**Subclaims:** Claim 5 Justifications j_1, j_2,... j_p for the Interactions i_1, i_2,... i_q are Proportional to the Residual Risk r'_x the Interactions give rise to. **Evidence**: Evidence 2.1.1 Identify Interactions and demonstrate how they are Necessary to meet the requirements
Evidence 2.1.1	Identify Interactions i_1, i_2,... i_q and demonstrate how they are Necessary to meet the requirements
	Description: For each Interaction by each Threat Actor, in all Configuration(s) delivered to Customer(s), the Applicant must: 1. document how a Functional or Non-Functional Requirement is directly dependent on that interaction; and 2. demonstrate that the same Functional or Non-Functional Requirement cannot be achieved without that interaction. **Evaluation Criteria:** The Assessor must review, on a pass or fail basis, whether the Applicant has 1. explicitly and comprehensively defined and documented each Interaction, 2. directly linked every Interaction in every Configuration delivered to Customer(s) to one or more Functional or Non-Functional Requirement, and 3. demonstrated how the removal of each Interaction results in at least one Functional or Non-functional Requirement being unachievable or severely impaired. **Implementing Guidance:** During design, development, and deployment, the Applicant should review any interactions to ensure they support one or more Functional or Non-Functional Requirements. Additionally, Applicants should complete the following steps: 1. Document each Interaction. 2. Document each unique type of Threat Actor and each of their potential Interactions with the At-Risk Parties or their data. Note that types of Threat Actors are more granular than the four classes of Threat Actors in the Risk Model used. A type of Threat Actor is a grouping of unique Threat Actors that shares a common profile of Interactions (e.g. customer

[redacted]	service agents or call centers, where the Applicant employs more than one call center). 3. Directly link each Interaction with the At-Risk Parties or their data by each group type of Threat Actor to at least one requirement. For each Interaction, document how removing that Interaction results in at least one Requirement being unachievable.

606

607 # Claim 3: Control set C_r treats Risk of type R_x

Claim 3	Control set $C_r \in \{c_1, c_2, c_3, .. c_n\}$ (in principle) treats Risk of type r_x
Selective	**Description:** The intent of implementing controls is to reduce risk. This claim concerns a particular Control set treating a particular type of risk. This claim must be made for each Control set claimed to be in place by the Applicant and must address each Risk from the Risk Model that the Applicant has selected as in scope. While treatment of risk generally offers some broader means (e.g. transferring, accepting), the treatment in this standard is limited to the application of Controls which reduces risk per the factors in the argument below. This Claim is selective in that Applicants must itemize the Controls they have in place in the Target System, map them to each of the Risks they have determined to be in scope, and select this Claim for each set of those Controls.
Argument 3.1 Tautological Step	Control set $C_r \in \{c_1, c_2, c_3, .. c_n\}$ treats risk of type r_x if the Controls in the set reduces the quantity or likelihood • of threats, • of those threats exploiting vulnerabilities, or • that At-Risk Parties experience Harm or the Control set lowers the impact of Harm to At-Risk Parties.
	Description: Risk treatment (i.e. reduction) can occur through one of four means: Reducing Threats: The Control set could reduce the opportunity of the Threat Actors (e.g. the Applicant, Contracted Party, Non-contracted Party, or Other Party) to act. Without opportunity, there is no threat. *If the associated risk is an action on data, deleting the data removes the Threat Actor's ability to act on that data (i.e. no data, no opportunity).* Reducing Exploitation: The Control set could reduce the motivation of the Threat Actor (i.e. likelihood that the threat exploits vulnerabilities). *A contract clause to terminate the contract in case of breach of terms will disincentivize a Contracted Party to take advantage of data they may have on a party.* Reducing Harms: The Control set could reduce the likelihood an At-Risk Party (e.g. Consumers, Operators, Resources, and Bystanders)

	experience Harms. This factor concerns threat materialization (i.e. threat materializing into a Harm). This occurs where there is some impediment (or difficulty) of the potential threat to materialize into something that impacts the party. *A Threat Actor may have data about a party and want to do something with it (i.e. the threat and the desire to exploit their possession), but if the data is encrypted, the Threat Actor will have a harder time thus the party is less likely to experience any Harm from the Threat Actor's actions.* Reducing Impacts: The Control set could lower the impact, not just the likelihood of an impact. *Reducing the specificity of a medical record from "the patient visited an HIV specialist" to "the patient visited a doctor" may reduce the tangible impact should that information be shared with someone.* **Evidence:** Evidence 3.1.1 Specification of how the Control set reduces risk of type r_x
Evidence 3.1.1	Specification of how the Control set reduces Risk of type r_x
	Description: This evidence is about providing a defensible statement that each Control in the set reduces risk through one of the four means listed in Argument 3.1. The Applicant need not provide proof nor real-world evidence of risk reduction but must provide a reasonable argument that the Control addresses the Risk in the way specified. See the description of Argument 3.1 for examples. **Evaluation Criteria:** The Assessor will review each statement as to how each Control reduces corresponding in scope Risks. Each statement must: • identify the Control, • identify the Risk, including the Threat Actor, At-Risk Party, and Consequence, • identify the means of risk reduction (threats, exploitations, harms, or impact), state how that Control achieves the means. Example: *"Deleting data about Consumers held by the Applicant reduces the quantity of Information Processing harms because future information processing cannot occur without data" includes the*

	Control, the Risk (Applicant, Consumer, Information Processing harms), means (reduction of threats) and achievement of means ("future information processing cannot occur without data"). The Assessor will further review the soundness of the achievement of means and reject those that are logically flawed or not based on available evidence. **Implementing Guidance:** Applicants should think about how Controls reduce risk during the selection (i.e. requirements phase) and design of Controls, but ultimately, the statement construction may occur solely for the benefit of an Assessor. The Applicant should review the available literature (e.g. requirements documentation, external sources about control effectiveness) to construct the statement for the benefit of post-hoc review.

608

Claim 4: Control set C_r is operational and effective in Target System

Claim 4	Control set $C_r \in \{c_1, c_2, c_3, .. c_n\}$ is operational and effective in Target System
Selective	**Description:** Claim 3 covers whether there are Controls that treat Risk, but to have those risks addressed in the operation of a particular Target System, those Controls must be operational, meaning they work (*e.g. no good having a lock on a door that's broken*), and they are effective, meaning they actually reduce the risk they are meant to reduce (*e.g. the working lock isn't easily bypassed by strong push*). One important note is that a Control need not be enabled (*e.g. the door can presently be unlocked*). This is covered by the default state (*i.e. Configuration*) in which the Target System is delivered to Customers. *Continuing with the analogy, an organization could deliver a building with the door locked or unlocked, per the Customer's needs, this claim is about whether that lock works and works effectively at preventing people without keys from entering.* This Claim is selective in that Applicants must itemize the Controls they have in place in the Target System and select this Claim for each Control set. This Claim has two parts, operationality and effectiveness. Each part is supported by a separate Argument. Both Arguments must be made to establish the Claim.
Argument 4.1 Evidentiary Step	Control set $C_r \in \{c_1, c_2, c_3, .. c_n\}$ is operational in the Target System if there is a documented requirement (baseline or specific) for the Control set $C_r \in \{c_1, c_2, c_3, .. c_n\}$ and each Control has been verified.
	Description: To claim operationality, there must be both a requirement for a Control set and verification that the Control set, as implemented in the Target System, meets the requirement. A requirement can come in the form of a baseline requirement that supports all systems of the Applicant or a specific requirement in the context of the Target System. To support the existence of the requirement, the requirement must be documented, and evidence (e.g. Evidence 4.1.1) of that documentation (e.g. entries in a supporting tool, written requirements) must be presented. Further, to be considered operational, each Control must be verified, through inspection, demonstration, testing, or by whatever means the

	requirement document or test case documentation specifies to evidence that the Control works. Verification should show that the Control is operational as designed. Effectiveness is addressed in Argument 4.2. **Evidence:** Evidence 4.1.1 Documentation, which includes procedures for inspection, demonstration, testing, and/ or analysis of the requirement for Control set $C_r \in \{c_1, c_2, c_3,.. c_n\}$ Evidence 4.1.2 Results of inspection, demonstration, testing, or analysis of Control set $C_r \in \{c_1, c_2, c_3,.. c_n\}$ showing that the Control set meets the requirements
Evidence 4.1.1	Documentation, which includes procedures for inspection, demonstration, testing, and/or analysis of requirements for Control set $C_r \in \{c_1, c_2, c_3,.. c_n\}$
	Description: To claim a requirement exists for a particular Control in the Target System, there must be documentation that records this requirement. The Applicant must have a copy of or reference to available documentation. Further, merely having a requirement doesn't mean that the requirement was implemented, hence the need to verify that the requirement has been met and the Control set has been implemented. Verification of Control sets usually involves gathering evidence, validating evidence, analyzing evidence, and concluding whether it's operational. Evidence can be gathered through inspection, demonstration, testing, and/ or analysis. The process of verification must also be documented. **Evaluation Criteria:** To sufficiently evidence the existence of a requirement, the presented documentation must contain the following: • a description, with enough specificity to facilitate implementation, of the Control set; and • one or more methodologies for verifying the implementation of the Control set in a system. The methodologies for inspection, demonstration, testing, and/or analysis must be written with enough specificity to provide an objective conclusion as to whether the Control set has been properly implemented. An Assessor should review the documentation for each Control set presented to assess the sufficiency of the documentation in meeting the above criteria. Results are rendered as sufficient or insufficient.

	Implementing Guidance: Many organizations have undocumented requirements, especially when it comes to Non-Functional Requirements. While it is important to document requirements, it's imperative when designing for privacy. Being able to clearly explain why system components affecting Risk are in place is central to the claim that Residual Risk is tolerable (see Claim 2). Documentation can come in the form of some baseline system requirements policy or standard (*e.g. "all systems must be resilient and have 99.99% uptime"*). System specific Functional Requirements are typically found in system design documents or product/sprint backlogs in Agile development. Requirements may not be formal but must be recorded in a form accessible to the designers, developers, and deployers. Quality attributes (i.e. Non-functional Requirements) are commonly found in baseline requirements documentation, such as a corporate system standards document, including any external standards the Applicant applies. For applicable external standards, there should be some documentation or evidence, such as a policy document, that demonstrates the Applicant actually uses the standard.
Evidence 4.1.2	Results of inspection, demonstration, testing, or analysis of Control set $C_r \in \{c_1, c_2, c_3, .. c_n\}$ showing that the Control set meets the requirements
	Description: Control set requirements or test documentation must include a method of assessing the operation of the Control set in a particular system. This evidence is about showing that such an assessment took place and that the results of the assessment show that the Control set is operational. The method of assessment is left to the Applicant, the Control designer, or an independent body. **Evaluation Criteria:** The methodology of the Control set assessment must identify the required evidence to support evaluating Control set operationality. To determine if this evidence is sufficient, • the Control set assessment must match the assessment methodology, and • a conclusion must be rendered (and supported by evidence) that the Control set is operational In the event of a large number of Control sets, it is sufficient for the Assessor to randomly select a sufficient number of sample Control set assessments to give a 95% confidence level that all the Control sets meet the evidence criteria. Any assessment that uses statistical

	sampling must include the methodology and resulting confidence level. **Implementing Guidance:** Because of the vagaries of assessment of Control set operationality, it is important to have a central repository and a process for assessment performance. This process should include timely re-reviews and version control, in the event of potential material changes to the Target System. The repository should note the date the assessment was performed and any relevant Target System version, and maintain copies of the assessment evidence and a conclusory statement. In the event of a conclusion of non-operationality of a Control set, the Control set should be reassessed after completion of any remedial actions undertaken.
Argument 4.2 Evidentiary Step	Control set $C_r \in \{c_1, c_2, c_3,.. c_n\}$ is effective in Target System if the Control set, as implemented, has been assessed and meets effectiveness criteria for Control set $C_r \in \{c_1, c_2, c_3,.. c_n\}$
	Description: Operational effectiveness of a Control set describes how well an implemented Control set is functioning in order to mitigate specific Risks the Control set intends to treat. Effectiveness must be measured against an objective standard (i.e. Evidence 4.2.2). **Evidence:** Evidence 4.2.1 Documentation of objective measures to assess the effectiveness of Control set $C_r \in \{c_1, c_2, c_3,.. c_n\}$ Evidence 4.2.2 Results of assessment of Control set $C_r \in \{c_1, c_2, c_3,.. c_n\}$ in the Target System showing the Control meeting effectiveness criteria.
Evidence 4.2.1	Documentation of objective measures to assess the operational effectiveness of Control set $C_r \in \{c_1, c_2, c_3,.. c_n\}$
	Description: The measures of operational effectiveness of a Control set describe whether the Control set operates consistently to a specified degree. Ideally, these measures should be objective, independently available, such as through a recognized standard, and independently verifiable. The Applicant may provide their own objective measures, though this may result in enhanced scrutiny by an Assessor. This evidence is about providing either external support for the objective measures used or providing internal documentation as to how the measures work, how they are

objectively measured, and how they assure operation effectiveness.

Evaluation Criteria: For each Control set, a methodology must be presented to assess the operational effectiveness of the Control set. This methodology may be internal or external (*e.g. NIST 800-53A Rev.5 Assessing Security and Privacy Controls*). More scrutiny should be given to internal assessment methodologies that have not undergone peer review.

To determine if the evidence is sufficient, the presented methodology must:

- be based on objective criteria,
- be deemed applicable by an internal or external Assessor,
- be documented in a way that is understandable and self contained,
- be performable in a finite amount of time,
- state the following:
  - the type of the evidence to be gathered
  - the methods for gathering the evidence
  - the criteria for evaluating the reliability and sufficiency of the evidence
  - the process for assessing the Control set based on the evidence
  - the dependencies on which the assessment may rely on
  - the measure for which the Control set can be deemed effective or not effective

In the event the provided methodology presents a way to measure operational effectiveness on a spectrum, but without criteria to state whether the Control set is effective or not (e.g. “the Control set is x% effective”), the documentation must demonstrate the Applicant’s determination of effectiveness with sufficient justification, in context, of why that determination was made.

Implementing Guidance: Control effectiveness assessment methodologies should ideally be externally provided, either through a recognized standards body or from an independent entity with subject matter expertise in the Control’s functions. Similarly for Control Set effectiveness assessment methodologies, though this may be less likely, because of the combinatorial explosion possible for sets of controls. Having said that, a standard approach (e.g. boolean algebra, ‘but for’ analysis, etc.) may be applicable in general, obviating the need for custom predefined assessment methodologies for specific sets of controls. Independent parties need not be the ones conducting the assessment, though this provides stronger

	evidence that the results (evidence 4.2.2) are unbiased.
Evidence 4.2.2	Results of assessment of Control set $C_r \in \{c_1, c_2, c_3, .. c_n\}$ in the Target System showing the Control set meeting effectiveness criteria.
	Description: This evidence is about showing that an assessment of operational effectiveness has been conducted for each Control, and the set as a whole, and that the results of the assessment confirm that the Control set is holistically effective. The method of assessment must match the documentation per Evidence 4.2.1. **Evaluation Criteria:** The methodology used to assess the effectiveness of the Control set must identify the required evidence to evaluate whether the Control set is effective or not. To determine if this evidence is sufficient, • the Control set assessment must match the assessment methodology, and • a conclusion must be rendered (and supported by evidence) that the Control set is effective. While the Control set assessment need not be conducted by an independent party, the relationship between the parties conducting the individual Control and/or Control set assessment and the operation of the Control set should be taken into account. Results should be scrutinized for any potential bias. In the event of a large number of Control sets, it is sufficient for the Assessor to randomly select a sufficient number of sample Control sets assessments to give a 95% confidence level that all the Control sets meet the evidence criteria. **Implementing Guidance:** Because of the vagaries of assessment of Control set effectiveness, it is important the Applicant maintain a central repository and a process for assessment performance. The repository should note the date the assessment was performed, maintain copies of the assessment evidence, and include a conclusory statement. In the event of a conclusion of non-effectiveness of a Control set, the Control set should be reassessed after completion of any remedial actions undertaken.

611

Claim 5: Justifications for the Interactions are proportional to the Residual Risk r'_x the Interactions give rise to

Claim 5	Justifications $j_1, j_2, \ldots j_p$ for the Interactions $i_1, i_2, \ldots i_q$ are proportional to the Residual Risk r'_x the Interactions give rise to.
Selective	**Description:** There is a complex relationship between Interactions and Risks. One Interaction can lead to multiple related Risks. Similarly, multiple Interactions may contribute to one Risk. For each in-scope Risk, denoted r_x, the set of Interactions (between a potential Threat Actor and At-Risk Party) that contribute to that Risk, must be justified by the Applicant. A Justification is a statement indicating the reason the Interaction takes place in the Target System. Furthermore, each Justification must be Proportionate to the Residual Risk (i.e. the Risk remaining after Controls have been applied). In other words, the greater the Residual Risk, the stronger the Justification required.
Argument 5.1 Evidentiary Step	Justifications $j_1, j_2, \ldots j_p$ for the Interactions $i_1, i_2, \ldots i_q$ are proportional to the Residual Risk r'_x the Interactions gives rise to if the Justifications state Benefit(s) to the At-Risk Parties or society and the Benefit(s) outweigh(s) the Residual Risk.
	Description: The key element of this argument is that proportionality hinges on benefits to the At-Risk Party or society and that those benefits outweigh the Residual Risk. The Justification statements must have an explicit or implied Benefit. This Benefit is either directly to the At-Risk Party or society. Benefits to the Applicant cannot be considered here. Any Benefits to the Applicant come through meeting requirements, as shown in Evidence 2.1.1. Outweighing the Benefits need not be strictly based on utilitarian comparisons but may include ethical considerations, as described in Claim 10. Further, Benefits need not be siloed and consideration of multiple Justifications may be pooled to judge the proportionality of Interaction(s) and Residual Risk(s). **Evidence:** Evidence 5.1.1 Match each Justification to Interactions and Residual Risk Evidence 5.1.2 Identify the Benefits to the At-Risk Parties or society for each Justification **Subclaims:** Claim 6: Benefits outweigh Residual Risk.

Evidence 5.1.1	Match each Justification to Interaction(s) and Residual Risk(s).
	Description: The Applicant must perform a matching exercise to ensure that every Interaction in the Target System and every in-scope Risk has at least one Justification attached. **Evaluation Criteria:** For each in-scope Risk put forth by the Applicant, the Applicant must identify the Interaction(s) in the Target System that give rise to that Risk. Further, the Applicant must link the sets of Interaction(s) and Risk(s) to one or more Justification statements. **Implementing Guidance:** Ideally, Risks are identified, mitigated and justified as part of a risk management process during the design, development, and deployment of systems. Justifications for Interactions can generally be identified early on. The Applicant should have a good sense of the Justification prior to creating the Target System, but should employ, as part of its risk management practices, a procedure to ensure that Justification continues to align with and outweigh Residual Risks.
Evidence 5.1.2	Identify the Benefits to the At-Risk Parties or society for each Justification
	Description: Each Justification statement must explicitly or implicitly include a Benefit to At-Risk Parties or society. **Evaluation Criteria:** The Assessor will review Justification statements to ensure they: • include one or more Benefits (If they do not include an explicit Benefit, the Benefit should be reasonably obvious to the Assessor.), • clearly identify the beneficiary or have a reasonably obvious beneficiary that is apparent to the Assessor, and • social benefits must provide external validation (e.g. law, policy paper, advocacy group) **Implementing Guidance:** The Applicant should compose a list of Justification statements. For each Justification statement the list should explicitly identify each Benefit and beneficiary. An example list is provided below.

Justification Statement	*Benefit*	*Beneficiary*
The interactions allowed for a personalized shopping experience.	*Personalization, reduction in time spent finding items of interest (explicit).*	*Consumers ("Shoppers")*
The interactions disincentivize shoplifting.	*Reduction in law enforcement expenditures to investigate crime.* *Enforcement of social contract to pay for goods and services (implicit).*	*Society*

Table 2 Example of justifications, benefits and beneficiaries

614

Claim 6: Benefits Outweigh Residual Risks

Claim 6	Benefits outweigh Residual Risks
Mandatory	**Description:** The Benefits of the product, service, or business outweigh the Residual Risks to the At-Risk Party. The measures of benefits and risks need not be done on a purely utilitarian scale, but may include ethical, societal, and/or other considerations.
Argument 6.1 Evidentiary Step	Benefits outweigh Residual Risk if [the Applicant must construct their own argument as to why the Benefits outweigh the Residual Risks]
	Description: The Applicant must construct an argument as to why the benefits outweigh any risks remaining (i.e. Residual Risks) after all controls have been applied. Only benefits to At-Risk Parties or society may be considered. Benefits to the organization are tied to the necessity of Functional and Non-Functional Requirements found in **Evidence 2.1.1**. Such an argument may include a balance of interests, policy concerns, ethical factors, and opinions of the stakeholders. There need not be one argument for all benefits and risks, but these can be classified and grouped, and the Applicant may provide multiple arguments to cover the entire range of benefits and risks in the Target System. Arguments should be written in the abstract and not include specific activities and risks. Evidence as to whether specific activities and risks have benefits that outweigh those risks should be documented in 6.1.[]. The argument may also be supported by subclaims. In this case, the sub- claims should be intuitively obvious such that they need no additional supporting arguments or evidence. **Evaluation Criteria:** While normally reserved for evidence, evaluation criteria are provided here because Assessors will be tasked with evaluating the Argument statements provided by Applicants. Arguments must: • use inductive or deductive reasoning as to why benefits outweigh risks. Such reasoning must be logically consistent and based on available evidence. • include objective evidence to support the conclusions **Implementing Guidance:** Applicants should begin with a well-founded rationale of why they feel the benefits outweigh the risks. This

	rationale should be formalized into a logical argument that can be objectively validated. For instance, the argument might be that beneficiaries judged the benefits to them as worth the risk. This argument would be supported by evidence of the individual choice and their informed adjudication of this choice. Note, this may be a high bar. While obvious in many voluntary activities (*e.g. a party chooses to rock climb knowing the risks of death or injury*), many risks may not be so obvious, intuitive, or reasonably explained to affected parties. Particular consideration should be given to power imbalance and vulnerable groups (e.g. children, ethnic minorities, LGBTQ+, those with disabilities, or those with no choices, or who cannot stop or remove such a service, such as when dealing with the public sector, or who have to take on the service based on their Postal Code). **Evidence:** Evidence 6.1.[ ] [Evidence statement to be provided by the Applicant consistent with their Argument]
Evidence 6.1.[ ]	[The Applicant must provide evidence statements to support their Arguments]
	Description: The details of the evidence will depend on the arguments provided by the Applicant in Argument 7.1. The Applicant may provide multiple evidence statements in support of their Argument. **Evaluation Criteria:** The Applicant must provide Evaluation Criteria upon which the Assessor must review the evidence. The Assessor will review two items. First, they must review whether the evidence statement logically supports the Arguments. Second, they must evaluate whether the evaluation criteria assess the sufficiency of the evidence enough to support the Arguments. Additionally, once the evidence statement and evaluation criteria are assessed, the evidence needs to be assessed against the evaluation criteria provided by the Applicant. **Implementing Guidance:** The Evidence statement should be a direct restatement of the elements supporting the Applicant's constructed argument. Evaluation criteria should be written in plain language such that an external Assessor can evaluate the sufficiency of the evidence to support the Argument and, ultimately, the claim. Evaluation criteria can consider the existence of discrete elements or an analysis of elements resulting in some level of confidence as to the truth of the elements. Applicants need not supply Implementation Guidance to themselves, though such guidance may be helpful to standardize

[redacted]	processes related to meeting the Standard in the future, especially where such evaluation must be applied to multiple Target Systems and for multiple risks that may evolve over time. An output similar to the ICO's Legitimate Interest Assessment[14] balancing test should be considered;

616

[14] https://ico.org.uk/media/about-the-ico/disclosure-log/4017958/ic-109330- z1w4-attachment-2.pdf

Claim 7: Benefits and Residual Risks cannot be further balanced by Applicant

Claim 7	Benefits and Residual Risks cannot be further balanced by Applicant
Selective	**Description:** In the context of this prong of the assurance case ("privacy by default"), balancing Risks and Benefits for Customers takes into consideration the Configuration handed to Customers and the further Configurability of the Target System. Claim 7 in the "privacy design" chain covers Benefits outweighing Residual Risk, so the assumption here is that has occurred. Therefore, this claim covers whether any system Configurability provided to Customers requires them to affirmatively make changes in the Configuration to increase risk, but presumably with commensurate Benefits. In other words, the design as delivered balances risk and benefits, but allows for changes to that balance. This claim is supported by one or more of these Arguments. Applicants may choose any or all but must make at least one Argument. The Applicant could, for instance, say there is no configurability in this feature of the Target System; there is Configurability in this feature, but changes would only increase risk; or there is Configurability in the remainder of the Target System but as delivered risks and benefits are balanced. Applicants should note the use of the term Customers. Customers are the recipients of the Target System and may or may not be Consumers. For instance, if the Applicant's Customers are business entities who then provide a service to Consumers, the Customers are not the Consumers of the Applicant's Target System. See definitions to determine overlap.
Argument 7.1 Evidentiary Step	Benefits and Residual Risk cannot be further balanced by Applicant if there is no Configurability available for the Applicant.
	Description: In situations where there is no Configurability (i.e. no changes which could affect Risk), any Configuration delivered to Customer(s) may be considered balanced within the scope of Target System's design, developed version or deployed version. The reason this is considered balanced is because this is a comparative analysis and if there are no other options to compare against (because there is no Configurability), changes to balance are moot. Design, development, or deployment changes **cannot** be considered at this stage, in the privacy by default analysis. The Target System version is set in the privacy by design analysis and

	any questions of reduced risks in the version (rather than the Configuration) are addressed in Claim 2: Residual Risk r'_x is tolerable in the Configuration(s) provided to Customer(s). **Evidence:** Evidence 7.1.1 Demonstrate there is no configurability.
Evidence 7.1.1	Demonstrate there is no Configurability.
	Description: Changes to balance are moot where there is no Configurability (and the design addresses Risks, see Claim 2), but the Applicant must demonstrate that there is no Configurability. This may be a high burden since many systems have internal settings that may be adjusted (variables and such). The question becomes which of those settings have been exposed to the Applicant. No ability to (reasonably) change settings equates to no Configurability. This does not include the ever present ability of an Applicant to reengineer a system to alter its behavior. It is sufficient if settings are not exposed in a way that the Applicant would normally engage. Designers generally have much more leeway, developers a little less so, and deployers of systems generally have the least ability to configure. Due to the vagaries of systems, demonstration may come in many forms (e.g. screenshots of non adjustable settings, operations manuals). The Applicant should pick a method of demonstration that reasonably conveys a lack of Configurability. **Evaluation Criteria:** The Assessor should review the provided evidence and make an inference that it reasonably conveys that the Applicant has no available settings at their disposal. The Applicant need not provide incontrovertible proof nor need they demonstrate that they cannot, through extraordinary means, alter the behavior of the system. *For example, a non- technical operator of a website need not consider the ability to alter the code running the website (or inject unexpected commands through a webform). The same would not be the case for the developer building a web application.* **Implementing Guidance:** Proper demonstration of no configurability will depend on the context of the Target System and the relationship of Applicant. Since designers have the broadest leeway, it will be difficult for them to argue that the design constraints limit configuration of the design. Developers may be able to argue that design requirements limit their developers. Deployers will have the easiest time demonstrating that the developed system they are provided by the developers provides no Configurability.

	A designer of a striking device (e.g. a hammer, mallet, etc) is limited by the materials and material sciences. They may further be limited by more specific goals (be able to strike a steel spike into concrete without breaking). But otherwise, the design is highly configurable. The designer may pass on some configurability to the engineer (i.e. the development task will bring the design to life). The engineer will still have some development options to consider. The end-user deploying the striking device at a worksite may be left with little configuration options, the striking device is a static tool. As a counter example, the design could allow for different headpieces depending on what's being struck, allowing for configuration by the end-user.
Argument 7.2 Evidentiary Step	Benefits and Residual Risk cannot be further balanced for Customer(s) if there is Configurability, Customers are delivered a specific Configuration(s) and changes to the Configuration(s) as delivered to the Customer(s) would only increase Risks.
	Description: If the system allows for the settings to be changed (i.e. the system has Configurability), changes will only increase the Risks, with no commensurate Benefit. For instance, a setting may turn off a control, with no upside. A setting could also create a Threat where one didn't exist (such as collecting data). If there is a change in Benefits, Applicants must look to Argument 7.3 to argue that any changes would upset the balance in an undesirable way. **Evidence:** Evidence 7.2.1 Document any hidden or exposed Configurability Evidence 7.2.2 Document the Configuration(s) provided to Customer(s) **Subclaims:** Claim 8 Changes to Configuration(s) as delivered to the Customer(s) would only increase Risks.
Argument 7.3 Evidentiary Step	Benefits and Risks cannot be further balanced by Applicant if there is Configurability, Customer(s) are delivered specific Configuration(s), and changes to the Configuration(s) as delivered by the Applicant would create an undesirable balance of Benefits and Risks.
	Description: This argument considers the balance between Benefits and Risks with the specific Configuration provided to specific Customers. Context may vary depending on the market (e.g. business to business, business to consumer, business to government), industry, vertical, or other factors that may adjust the

	types of Threats, Vulnerabilities, Consequences, Threat Actors, or At-Risk Parties. Benefits need not necessarily outweigh Risk as other factors, such as ethics, fairness, or social policy, may contribute to the analysis. **Evidence:** Evidence 7.3.1 Document any Configurability hidden or exposed to Customers Evidence 7.3.2 Document the Configuration(s) provided to Customer(s) **Subclaims:** Claim 9 Changes to the configuration as delivered to the Customers would create an undesirable balance of Benefits and Risks
Evidence 7.[2,3].1	Document any hidden or exposed Configurability.
	Description: To assess the balance of a Configuration, Configurability (i.e. available settings) must be identified. This includes settings visible to Customers, Consumers, or others, whether or not they can easily access or modify those settings. It also covers any hidden options requiring advanced configuration or developer tools, that may be available, and utilized by the Applicant to change the default Configuration(s) for Customer(s). Each configurable element must be explained along with its options, function(s), and effect(s) within the Target System. The Applicant may also provide Justification why some available settings are included in the documentation, such as they are not intended to be accessible settings or are beyond their normal skills and activities (*e.g. configuration files or changing values in code where the Applicant is not intended to be making such alterations*). **Evaluation Criteria:** The Assessor will review the documentation to determine that each setting is described in sufficient detail to ascertain: • how that setting is set • what options are available • to which Functional or Non-functional Requirement the setting relates • what effect(s) the setting has on the Target System (e.g. turn on or off functionality, security or privacy controls). The Applicant need not address Configurability that has a negligible

	effect on Risks (*e.g. changing the color mode, unless changing the color mode has an effect on Threat Actor making it more difficult for them to collect data*). The Assessor may make an independent review of the Target System, including any environments into which the system is deployed, to ensure completeness of the documentation. This is especially important if the documentation provided by the Applicant contains noticeable gaps, lacks specificity, or has contradictory information. Design and development decisions and settings that are not visible or readily accessible may also need to be addressed, therefore Assessors should be familiar with the design and development process to determine if decisions have an impact on the Configurability (*e.g. the decision to develop a feature for iPhone and not Android distinguishes Configurations between two market segments*). **Implementing Guidance:** The nature of this documentation will depend on whether the Applicant is a designer, developer, or deployer of the Target System. Deployers should look for settings provided in the Target System by the developer, either those clearly available (*e.g. an Administrators dashboard*) or described in documentation (*e.g. a configuration file*). The deployer should also consider the environment into which the system is deployed and whether settings in the broader environment might also be considered part of the Configuration of the Target System (*e.g. deploying on various databases where configuration of the database will also impact the risks of the data stored here*). While not part of this evidence, but rather Evidence 8.[2,3],2, deployers must document not only the Configurability but what options were chosen. Developers, having much more leeway, need to consider not only their decisions to include Configurability into the components they develop but how that decision may also be a Configuration option. Of course, if the decision is made solely by the designers and the developers have no authority to make decisions, there is no need to document that as part of the Configurability. It becomes extremely important that developers document decisions because some of those decisions may have significant impact.
Evidence 7.[2,3].2	Document the Configuration(s) provided to Customer(s)

	Description: This evidence requires specifying the specific Configuration(s) provided to specific types of Customer(s), detailing setting options chosen. This list should mirror that provided in 8.[2,3].1 and extends that list to include the specific options chosen and match those options to the type of Customer(s) that option was selected for. **Evaluation Criteria:** The Assessor will review the documentation for completeness with the following considerations: • has the Applicant identified all of the Customer segments to which different Configurations are applied, and • has the Applicant identified all of the selected settings consistent with the Configurability of the Target System as described in 7.[2,3].1. **Implementing Guidance:** While documentation after the fact may be done (for instance in a retroactive analysis of the Target System for conformance with this Standard), it is recommended that first, the Applicant keeps a running record of the Configurability of the Target System, including effects, for Evidence 7.[2,3].1, and second, has a configuration repository to retain Configuration by Customer segment. This can be automated, in part, for large, diverse deployments. An even more robust documentation system could include Justifications for those Configuration selections, to avoid post- hoc Justification being provided to satisfy Claim 2.

619

Claim 8: Changes to Configuration(s) as delivered would only increase Risks

Claim 8	Changes to Configuration(s) as delivered to Customer(s) would only increase Risks
Mandatory	**Description:** The Target System, in the Configuration(s) that it is delivered to Customers, is set to minimize Risks and, therefore, cannot be changed to reduce said risks further. Any change(s) made to the Configuration will raise risk, which could either increase the likelihood of occurrence or impact should risk materialize.
Argument 8.1 Evidentiary Step	Changes to the Configuration(s) as delivered to the Customer(s) would only increase Risks if changes would increase Interactions between Threat Actors and At-Risk Parties or changes to the Configuration(s) would diminish one or more Controls.
	Description: Any alteration(s) to the Configuration(s) as delivered to Customer(s) increases likelihood of Harm if said alteration(s) enables one or more Threat Actors to more easily or effectively engage with an At-Risk Party or their proxy, such as data, or such alteration(s) disables or reduces the effectiveness of one or more Controls. **Evidence:** Evidence 8.1.1 Analysis of configuration for each Customer showing that changes would increase Interactions or reduce effectiveness of Controls (Evidence 7.2.1) and (Evidence 7.2.2). As part of the argument preceding Claim 8, the Applicant must provide Evidence 7.2.1 and Evidence 7.2.2, which are incorporated to support this Argument.
Evidence 8.1.1	Analysis of Configuration for each Customer showing that changes would increase Interactions or reduce effectiveness of Controls
	Description: Applicant provides evidence, in the form of analysis, screenshots, source code with visual output, or other attestation that reasonably supports the Configuration provides effective Controls. Additionally, they must demonstrate that any alteration to said Configuration will weaken Controls or raise the likelihood that one or more Threat Actors will effectively engage with At-Risk Parties. **Evaluation Criteria:** The Assessor shall adjudicate whether the evidence supplied by Applicant reasonably supports the conclusion that changes to the Configuration for each Customer would increase

[redacted]	Interaction or weaken Controls. This is with respect to proving that Configuration for each Customer is set to minimize Threat Actor engagement with At-Risk Parties and any alteration to said Configuration weakens Controls. **Implementing Guidance:** For each possible setting change, the Applicants need to review how that change would affect Interactions between Threat Actors and At-Risk Parties. Applicants are not required to consider every possible Configuration but should reasonably anticipate where individual changes to settings do not increase Interactions, but where multiple setting changes could have such an effect. The Applicant needs to also consider the effects of setting changes on Controls. Similarly, the primary focus is on individual settings disabling or directly weakening Controls. However, it is important to reasonably consider that the cumulative effect of multiple settings changes could weaken a Control where individual settings may not.

622

Claim 9: Changes to Config(s) would create undesirable balance of Benefits and Risks

Claim 9	Changes to the Configuration(s) as delivered to the Customer(s) would create an undesirable balance of Benefits and Risks
Mandatory	**Description:** Settings are features of the Target System that can be enabled or disabled. The state of these settings is a Configuration. Different Configurations may be delivered to different Customers. Whether the position of a setting enhances or diminishes Benefits or increases or decreases Risk depends on the specific setting's effect in the context of the design. This claim statement makes the assertion that changes to settings would be undesirable when considering the Benefits and Risks involved. Note that this is not a one size fits all for every Configuration delivered to every type of Customer. Different Customers operate in different markets, industries, and verticals, and thus engender different risks, with varying likelihoods and impacts to At-Risk Parties. Similarly, Benefits may be heavily dependent on these contextual factors as well, thus each Configuration should be viewed in light of the particular context in which it is deployed.
Argument 9.1 Evidentiary Step	Changes to the Configuration as delivered to the Customers create an undesirable balance of Benefits and Risks if [the Applicant must construct an argument statement based on evidence as to why changes to the Configuration create an undesirable balance of Benefits and Risks]
	Description: The Applicant must construct an argument as to why the balance between Benefits and Risk would be undesirable if the Configuration delivered to the Customer were altered. Only Benefits to At-Risk Parties or society may be considered. Benefits to the organization are tied to the necessity of Functional and Non-Functional Requirements found in Claim 2. Such an argument may include a balance of interests, policy concerns, or ethical factors, and opinions of the stakeholders. There need not be one Argument for all Configurations, Benefits, and Risks, but these can be classified and grouped, and the Applicant may provide multiple Arguments to cover the entire range of Benefits and Risks in the delivered Configurations. The Argument may also be supported by subclaims. In this case, the sub- claims should be intuitively obvious such that they need no additional supporting Arguments or evidence.

Evaluation Criteria: While normally reserved for evidence, evaluation criteria are provided here because Assessors will be tasked with evaluating the Argument statements provided by Applicants. Arguments must:

- use inductive or deductive reasoning as to why the balance between Benefits and Risks would be undesirable. Such reasoning must be logically consistent and based on available evidence.
- include objective evidence statements to support the conclusions

The question of desirability need not be made from any party's perspective. In other words, the argument does not need to consider the subjective desires of any one party. The argument should appeal to normative ethical principles, legal or moral obligations, considerations of fairness, equity, and justice, and utilitarian weighing of Benefits and Risks. There is currently no agreed upon construction of an objective argument for the undesirability of a resulting Configuration change, thus it is up to the Applicant to demonstrate that they have thought about it and the Configuration was not the result of an accident, ignorance, or unsavory motivations, but rather careful deliberation.

Implement Guidance: Applicants should begin with a well-founded rationale of why they believe changes to the Configuration would be undesirable from a Benefits and Risk perspective. This rationale should be formalized into a logical argument that can be objectively validated. For instance, one argument might be that the affected parties judged the benefits to them as worth the risk. This Argument would be supported by evidence of the parties' choice and their informed adjudication of this choice. Note, this may be a high bar. While obvious in many voluntary activities (e.g. an individual chooses to rock climb knowing the risks of death or injury), many risks may not be so obvious, intuitive, or reasonably explained to individuals.

Evidence:

Evidence 9.1.[ ] [Applicant must provide Evidence Statement, Description, Evaluation Criteria and Implementation Guidance]

Evidence 7.3.1 and Evidence 7.3.2 As part of the Argument preceding Claim 9, the Applicant must provide Evidence 7.3.1 and Evidence 7.3.2, which are incorporated to support this Argument.

Evidence 9.1.[ ]	[The Applicant must construct one or more Evidence Statements to support the Argument, along with supporting evaluation criteria, description, and implementation guidance]
	Description: The details of the evidence will depend on the arguments provided by the Applicant in Argument 10.1. The Applicant may provide multiple evidence statements in support of their Argument. **Evaluation Criteria:** The Applicant must provide evaluation criteria upon which the Assessor must review the evidence. The evaluation criteria supporting this Argument should be substantially the same as the evaluation criteria provided in Evidence 6.1[ ]. The Assessor will review two items. First, they must review whether the evidence statement logically supports the Arguments. Second, they must evaluate whether the evaluation criteria assess the sufficiency of the evidence enough to support the Arguments. Additionally, once the evidence statement and evaluation criteria are assessed, the actual evidence needs to be assessed against the evaluation criteria provided by the Applicant. **Implementing Guidance:** The Evidence statement should be a direct restatement of the elements supporting the Applicant's constructed argument. Evaluation criteria should be written in plain English such that an external Assessor can evaluate the sufficiency of the evidence to support the Argument and, ultimately, the claim. Evaluation criteria can consider the existence of discrete elements or an analysis of elements resulting in some level of confidence as to the truth of the elements. Applicants need not supply implementation guidance to themselves, though such guidance may be helpful to standardize processes related to meeting the Standard in the future, especially where such evaluation must be applied to multiple Target Systems and for multiple risks that may evolve over time.

625