**NIST Big Data Public Working Group (NBD-PWG)**

**NBD-PWD-2015/M0449**

**Source: Security & Privacy Subgroup**

**Status: Draft**

**Title: NBDWG Big Data Privacy Vocabulary – A to M (to be continued)**

**Author: Mark Underwood and S&P Team**

NBDWG Big Data Privacy Vocabulary

Issues

# Mark’s Approach

* Assume that “privacy fabric” can be developed into a meaningful Big Data concept that can be mapped to the NBDRA.
* Orchestration, PII traceability, remains an integral part of the NBDRA
* Integrate concerns expressed by past less technical members re: privacy, with particular attention to disadvantaged, disenfranchised groups, sociological context
* Rely on risk management frameworks to introduce “user” role, including individual citizen responsibility and training.
* Develop a privacy-only taxonomy drawing from existing standards or drafts. May require crosswalks to other accepted terms – e.g., “confidentiality,” and distaste that some have for *Status: Slow going. Still reviewing CNSSI-4009 and NIST.IR.7298r2*
* Proceed on two “tiers (levels, depths) as being considered below. One tier is a for business analysts (non-developers), and the other is designed for a modestly deeper adoption by architects.
  + (1) a checklist of [role+NBDRA component] or [data+NBDRA component]. Where possible only develop checklists where a clear use case narrative exists (in the NBDWG or elsewhere.)
  + (2) BPMN or SysML schemas (but shallow)
* After developing a draft, revisit the Security Fabric and harmonize the approaches
* Send interim drafts to this sub-sub group.

# BD system timeline

Design-time (functional requirements) Possible subtopics: resilience, risk, compliance

Run time

Management (can be hybrid, or multiple time slices)

Audit (live, forensic – implying after-the-fact)

Remediation / Restoration

In the ISO document table, see “Implementer,” “User”

These concerns are handled in the [federal enterprise architecture](https://en.wikipedia.org/wiki/Federal_enterprise_architecture), though the system development timeline seems a bit submerged for a standards effort.

# Relationship to Security

Privacy may have 1:1 analogs to security measures, or it may be somewhat different. E.g., consider “assured information sharing” (CNSSI-4009).

# Design Patterns for Workflow

Using BPEL or other schemata, identify typical orchestration patterns, roles, data elements, security measures, risks, etc. May need both snapshot (cursory) and detailed versions to be useful. (See also SysML).

# Governance

Possible headway (not “solutions”):

# Privacy Fabric

**Discuss**There should be an opt-out of the privacy fabric conformance; e.g., for astronomy (but what about PI’s and organization-owned roles for data?).

**Discuss** Does the presence of intellectual property (for data, code) imply a nearly omnipresent privacy fabric.

# Privacy Use Cases

**Discuss** Can we use the existing ones from V1?

# Privacy Taxonomy

| Term | Source | BDRA | Comments |
| --- | --- | --- | --- |
| Privacy disassociability | NISTIR 8062 (draft) Appendix A | “Privacy fabric” for purposes of this analysis | Needs refinement. “Enabling the processing of personal information or events without association to individuals or devices beyond the operational requirements of the system” |
| Privacy subsystem predictability | 8062 Appendix A |  | Needs refinement |
| Privacy subsystem manageability | 8062 Appendix A | TBD | Needs refinement |
| Role: privacy subsystem oversight |  |  |  |
| Role: privacy subsystem operations |  |  |  |
| Role: privacy subsystem design |  | Architect responsibilities call-out | NIST 8062 groups ops & design; we separate |
| personal information |  |  | “For the purpose of risk assessment, personal information is considered broadly as any information that can uniquely identify an individual as well as any other information, events or behavior that can be associated with an individual. Where agencies are conducting activities subject to specific laws, regulation or policy, more precise definitions may apply.” |
| Privacy risk |  |  | Roughly, adverse impact X likelihood of occurrence, scoped |
| Privacy controls: administrative |  |  |  |
| Privacy controls: technical |  |  |  |
| Privacy controls: physical |  |  |  |
| Adverse privacy event |  |  |  |
| privacy context: system |  |  |  |
| Privacy engineering |  | Use for narrative. May not have normative value beyond describing collection of system features, workflow elements | “Privacy engineering is an emerging field, but currently there is no widely-accepted definition of the discipline. For the purposes of this [publication], privacy engineering is a collection of methods to support the mitigation of risks to individuals arising from the processing of their personal information within information systems.” |
| NIST privacy risk model | 8062 Appendix C |  |  |
| Privacy metasystem issues |  |  | 8062 used “Summary Issues.” “Initial contextual analyses about data actions that may heighten or decrease the assessment of privacy risk.” |
| Privacy attack vector |  |  | Attack against Personal Information, a privacy subsystem, role, etc. |
| Owner/originator |  |  | System component, role or individual originating a data element. |
| Access\* | NIST.IR.7298r2, SP 800-32 | Includes access to workflow, orchestration |  |
| Role: Access authority\* | CNSSI-4009 |  | Person or software |
| Access Control | FIPS 201 |  |  |
| ACL\* | FIPS 201, CNSSI-4009 | Consider local vs. global big data ACLs |  |
| Access control mechanism\* | CNSSI-4009 |  |  |
| Access type\* |  |  |  |
| Accountability | 7298 |  | Grouped subprocesses: traceability, non-repudiation, deterrence, fault isolation, intrusion detection, intrusion prevention, after-action recovery, legal action. |
| Active content | 7298r2 |  | “Electronic documents that can carry out or trigger actions automatically on a computer platform without the intervention of a user.“ |
| Active/passive security testing |  | Big data exchanges will often entail passively tested, or passive assurance for exchanges between components[[1]](#endnote-1) |  |
| Administrative Safeguards | 7 |  |  |
| Advisory |  | Big Data may require a “new” grouping of advisories | “Notification of significant new trends or developments regarding the threat to the information systems of an organization. This notification may include analytical insights into trends, intentions, technologies, or tactics of an adversary targeting information systems.” |
| Privacy agent |  | Program acting on behalf of person or organization to automate a privacy-related process[[2]](#endnote-2) |  |
| Allocation | SP 800-37 | Useful for workflow in determining privacy responsibilities: design-time, governance-time | The process an organization employs to determine whether security controls are defined as system-specific, hybrid, or common.  The process an organization employs to assign security controls to specific information system components responsible for providing a particular security capability (e.g., router, server, remote sensor). |
| Application | SP 800-37 | How would a BDRA app be different |  |
| Assessment | SP 800-53A | Apply to BDRA privacy (also sec?). How different from audit? | Grouping of terms: findings, method, object, objective, procedure, Security Control Assessor |
| Assurance | SP 800-27, SP 800-53A, CNSSI-4009 | Can we map to Privacy Assurance (i.e., map to analogous goals?) | “Grounds for confidence that the other four security goals (integrity, availability, confidentiality, and accountability) have been adequately met by a specific implementation. “Adequately met” includes (1) functionality that performs correctly, (2) sufficient protection against unintentional errors (by users or software), and (3) sufficient resistance to intentional penetration or by-pass.” |
| Assurance Case (for privacy) |  | Can we map to Privacy Assurance (i.e., map to analogous goals?). Also see below. | “A structured set of arguments and a body of evidence showing that an information system satisfies specific claims with respect to a given quality attribute. “ |
| Assured Information sharing |  | Analogy for privacy sharing | “The ability to confidently share information with those who need it, when and where they need it, as determined by operational need and an acceptable level of security risk.” |
| Attack, sensing, warning; attack signature (for privacy)[[3]](#endnote-3) |  | Attack signature for privacy is not the same as a general attack | “Detection, correlation, identification, and characterization of intentional unauthorized activity with notification to decision makers so that an appropriate response can be developed. “ |
| Audit, audit data, audit log, reduction tools, audit review, audit trail |  | Subset created for privacy. Could be a smaller problem to solve, or a larger one, depending.[[4]](#endnote-4) |  |
| Authentication (various terms) |  | Could be needed to allow “owner” of privacy data to see or correct their own data. |  |
| Authority |  |  |  |
| Authenticity |  |  | Provenance |
| Authorization |  |  | Time-limited authorization to access, or use privacy data |
| Authorization to operate |  |  | Interop issues for BD concerning privacy data |
| Automated privacy monitoring |  | To Do | Use of automated procedures to ensure security controls are not circumvented or the use of these tools to track actions taken by subjects suspected of misusing the information system. |
| Back door (privacy) |  | Use of BD variety to circumvent privacy safeguards |  |
| Baseline security (for privacy controls) |  |  | The minimum security controls required for safeguarding an IT system based on its identified needs for confidentiality, integrity, and/or availability protection. |
| Behavioral outcome (for privacy fabric training) |  | Useful for cross-org privacy |  |
| Biometric information |  | Special concern for privacy in any system? |  |
| Body of Evidence (for privacy controls adherence) |  |  | “The set of data that documents the information system’s adherence to the security controls applied. The BoE will include a Requirements Verification Traceability Matrix (RVTM) delineating where the selected security controls are met and evidence to that fact can be found. The BoE content required by an Authorizing Official will be adjusted according to the impact levels selected.  “ |
| Boundary; boundary protection |  | Needed for BDRA?[[5]](#endnote-5) |  |
| Browsing (for identity info) |  |  |  |
| Business impact assessment (for privacy fabric) |  |  | “An analysis of an information system’s requirements, functions, and interdependencies used to characterize system contingency requirements and priorities in the event of a significant disruption.” |
| Certificate (esp. identity certificate) | CNSSI-4009 | No different meaning vs. security, but perhaps more urgent context? | Certificate management may be different in privacy fabric when individual citizens (including children) are involved |
| Certification (see also baseline), certifier |  | Identify a baseline point at which privacy fabric controls were applied & certified as operational | “A comprehensive assessment of the management, operational, and technical security controls in an information system, made in support of security accreditation, to determine the extent to which the controls are implemented correctly, operating as intended, and producing the desired outcome with respect to meeting the security requirements for the system. |
| Chain of Custody |  | IoT plus Big Data for privacy | “A process that tracks the movement of evidence through its collection, safeguarding, and analysis lifecycle by documenting each person who handled the evidence, the date/time it was collected or transferred, and the purpose for the transfer. |
| Chain of Evidence |  | IoT plus Big Data for privacy. Same, but applied to privacy data subset | “A process and record that shows who obtained the evidence; where and when the evidence was obtained; who secured the evidence; and who had control or possession of the evidence. The “sequencing” of the chain of evidence follows this order: collection and identification; analysis; storage; preservation; presentation in court; return to owner. |
| Chief Privacy Officer |  | TBD |  |
| Classified information (\*privacy subset) | SP 800-60, EO 13292, CNSSI-4009 | Adapt meaning from US mil to apply to privacy subset |  |
| Classified (privacy) data spillage |  |  |  |
| Clearance for access to privacy data or tools (both?) |  | Useful to identify fabric roles permitted to access privacy data, or to use re-identifying tools. Obvious: Data access, tools access aren’t the same. See access, authorization. | “Formal certification of authorization to have access to classified information other than that protected in a special access program (including SCI). Clearances are of three types: confidential, secret, and top secret. A top secret clearance permits access to top secret, secret, and confidential material; a secret clearance, to secret and confidential material; and a confidential clearance, to confidential material. |
| Common Control / Security Control Inheritance / Common criteria |  | Across app and data providers possibly spanning organizations. “Common criteria” is a document for privacy fabric requirements | “A security control that is inherited by one or more organizational information systems. |
| Common Control Provider (role for privacy) |  | Role responsible for inherited privacy controls | “An organizational official responsible for the development, implementation, assessment, and monitoring of common controls (i.e., security controls inherited by information systems). |
| Common Misuse Scoring System for Privacy |  | A rough metric for potential privacy fabric weaknesses | “A set of measures of the severity of software feature misuse vulnerabilities. A software feature is a functional capability provided by software. A software feature misuse vulnerability is a vulnerability in which the feature also provides an avenue to compromise the security of a system. |
| Community of Interest for privacy data |  | A CoI may be a class of users in the privacy fabric. E.g. tribal, disabled, genetic abnormalities, high medical cost | “A collaborative group of users who exchange information in pursuit of their shared goals, interests, missions, or business processes, and who therefore must have a shared vocabulary for the information they exchange. The group exchanges information within and between systems to include security domains. |
| Community risk for privacy |  | Add – privacy fabric | “Probability that a particular vulnerability will be exploited within an interacting population and adversely impact some members of that population. |
| Compartmentalization (see DHHS meaning) |  |  | “A nonhierarchical grouping of sensitive information used to control access to data more finely than with hierarchical security classification alone. |
| Compromise – As applied to privacy |  | Especially re-identificaition | “Disclosure of information to unauthorized persons, or a violation of the security policy of a system in which unauthorized intentional or unintentional disclosure, modification, destruction, or loss of an object may have occurred. |
| Compromising Emanations (for privacy data) |  |  | “Unintentional signals that, if intercepted and analyzed, would disclose the information transmitted, received, handled, or otherwise processed by information systems equipment.  “ |
| CND |  | Different for privacy fabric? |  |
| Confidentiality | SP 800-53; SP 800-53A; SP 800-18; SP 800-27; SP 800-60; SP 800-37; FIPS 200; FIPS 199; 44 U.S.C., Sec. 3542 | Traditional meaning for privacy embodied in numerous standards, despite its problems. | “Preserving authorized restrictions on information access and disclosure, including means for protecting personal privacy and proprietary information. |
| Contamination |  | Scenario: a de-identified DB is placed into a “system” containing potentially re-identifying resources | “Type of incident involving the introduction of data of one security classification or security category into data of a lower security classification or different security category. |
| Continuous monitoring (of privacy fabric) |  |  | “The process implemented to maintain a current security status for one or more information systems or for the entire suite of information systems on which the operational mission of the enterprise depends. The process includes: 1) The development of a strategy to regularly evaluate selected IA controls/metrics, 2) Recording and evaluating IA relevant events and the effectiveness of the enterprise in dealing with those events, 3) Recording changes to IA controls, or changes that affect IA risks, and 4) Publishing the current security status to enable information-sharing decisions involving the enterprise. “ |
| Controlled interface |  | Control at the BDRA interface for privacy fabric (different?) | “A boundary with a set of mechanisms that enforces the security policies and controls the flow of information between interconnected information systems. |
| Covert testing (of privacy fabric) |  |  |  |
| Credential, credential service provider |  |  | “A trusted entity that issues or registers Subscriber tokens and issues electronic credentials to Subscribers. The CSP may encompass Registration Authorities (RAs) and Verifiers that it operates. A CSP may be an independent third party, or may issue credentials for its own use. |
| Criticality, criticality level |  | Not all privacy data elements or tools may be equal |  |
| Cryptographic binding |  |  | “Associating two or more related elements of information using cryptographic techniques. “ |
| Conformance to privacy fabric XXX |  | Refer to Frank’s initiative |  |
| Data integrity (privacy corruption) |  | Mis-identification (e.g., TSA list) |  |
| Default classification (for privacy data, or privacy tooling) |  |  |  |
| Digital forensics |  | As applied to privacy fabric: still emerging; check academic lit |  |
| End-to-end privacy XXX |  | TBD |  |
| Event (privacy) | CNSSI-4009 | Subset of events appropriate to privacy | “Any observable occurrence in a system and/or network. Events sometimes provide indication that an incident is occurring. |
| External provider, external network | SP 800-37; SP 800-53 | Critical for privacy data/controls preservation in big data across clouds, across organizations | “A provider of external information system services to an organization through a variety of consumer-producer relationships, including but not limited to: joint ventures; business partnerships; outsourcing arrangements (i.e., through contracts, interagency agreements, lines of business arrangements); licensing agreements; and/or supply chain exchanges. |
| False Acceptance |  | Mis-identification (?) | Biometric domain in 800-76 |
| Hacker – Identity hacker |  |  |  |
| Health Information Exchange | NISTIR-7497 | Important as a de facto BD Variety source for re-identification due to U.S. ubiquity. See also UnitedHealthCare Optum | “A health information organization that brings together healthcare stakeholders within a defined geographic area and governs health information exchange among them for the purpose of improving health and care in that community. “ |
| Identification | SP 800-47 | TBD – Needs refinement | “The process of verifying the identity of a user, process, or device, usually as a prerequisite for granting access to resources in an IT system. “ |
| Identifier | FIPS 201, CNSSI-4009 | Identifiers can be automated, e.g., biometric theft, or photo recognition | “A data object - often, a printable, non-blank character string - that definitively represents a specific identity of a system entity, distinguishing that identity from all others. “ |
| Identity |  | Our V1 documents may not follow this standard usage. | “The set of attribute values (i.e., characteristics) by which an entity is recognizable and that, within the scope of an identity manager’s responsibility, is sufficient to distinguish that entity from any other entity. “ |
| Identity-based Security Policy |  |  |  |
| Identity Binding |  |  |  |
| Identity-based access control |  |  |  |
| Identity proofing |  |  |  |
| Identity token |  |  |  |
| Identity validation |  |  |  |
| Identity verification |  |  |  |
| Impact, impact level, impact value | 800-60, CNSSI-4009, SP 800-34, SP 800-30 | Same concepts but mapped to privacy fabric |  |
| Incident |  | Same meaning, covered under “confidentiality” | “An occurrence that actually or potentially jeopardizes the confidentiality, integrity, or availability of an information system or the information the system processes, stores, or transmits or that constitutes a violation or imminent threat of violation of security policies, security procedures, or acceptable use policies. “ |
| Incident handling for privacy incidents |  | Subset, but could be different from superset |  |
| Indicator |  | Recognized signal that an adversary might be attempting to compromise privacy fabric |  |
| Information assurance for privacy |  |  | “Measures that protect and defend information and information systems by ensuring their availability, integrity, authentication, confidentiality, and non-repudiation. These measures include providing for restoration of information systems by incorporating protection, detection, and reaction capabilities. “ |
| Information Domain |  | Needs to be enlarged for BD privacy fabric | “A three-part concept for information sharing, independent of, and across information systems and security domains that 1) identifies information sharing participants as individual members, 2) contains shared information objects, and 3) provides a security policy that identifies the roles and privileges of the members and the protections required for the information objects. “ |
| Information Operations (as applied to identity disruption) | CNSSI-4009 |  | “The integrated employment of the core capabilities of electronic warfare, computer network operations, psychological operations, military deception, and operations security, in concert with specified supporting and related capabilities, to influence, disrupt, corrupt, or usurp adversarial human and automated decision-making process, information, and information systems while protecting our own. “ |
| Information owner |  |  |  |
| Information sharing environment |  | Highlight as a potential area for variety-enabled identification | “ISE in its broader application enables those in a trusted partnership to share, discover, and access controlled information. “ |
| Information Security Architect (sub: privacy) | SP 800-39 | Identifies design-time role. Architecture refers to the design. |  |
| Information Steward (for confidential data, tools) |  |  | “An agency official with statutory or operational authority for specified information and responsibility for establishing the controls for its generation, collection, processing, dissemination, and disposal.” |
| IS Resilience |  | Does this notion apply to identity attacks specifically? |  |
| IS Security Risks (privacy subset) |  |  | “Information system-related security risks are those risks that arise  through the loss of confidentiality, integrity, or availability of  information or information systems and consider impacts to the  organization (including assets, mission, functions, image, or  reputation), individuals, other organizations, and the Nation.” |
| Information Value |  |  | “A qualitative measure of the importance of the information based upon factors such as: level of robustness of the Information Assurance controls allocated to the protection of information based upon: mission criticality, the sensitivity (e.g., classification and compartmentalization) of the information, releasability to other countries, perishability/longevity of the information (e.g., short life data versus long life intelligence source data), and potential impact of loss of confidentiality and integrity and/or availability of the information.” |
| Insider threat for confidentiality breaches |  | E.g., access to personnel records, authentication systems, ACLs |  |
| Intellectual property |  | Especially IP connected to or owned by a person, but also IP treated the same way as “privacy” data. Further study.[[6]](#endnote-6) |  |
| Interconnection Security Agreement | SP 800-47, CNSSI-4009 |  |  |
| Interface Control Document |  | Different for privacy? |  |
| Internal network privacy controls |  | Use cases are different |  |
| IT privacy awareness and training program |  |  |  |
| IT privacy policy (three + types) | SP 800-12 | Program policy; issue (context specific) policies; system- or device- or app-specific policies | 1) Program Policy—high-level policy used to create an Program policy - organization’s IT security program, define its scope within the organization, assign implementation responsibilities, establish strategic direction, and assign resources for implementation.  2) Issue-Specific Policies—address specific issues of concern to the organization, such as contingency planning, the use of a particular methodology for systems risk management, and implementation of new regulations or law. These policies are likely to require more frequent revision as changes in technology and related factors take place.  3) System-Specific Policies—address individual systems, such as establishing an access control list or in training users as to what system actions are permitted. These policies may vary from system to system within the same organization. In addition, policy may refer to entirely different matters, such as the specific managerial decisions setting an organization’s electronic mail (email) policy or fax security policy. “ |
| Key terminology: list, loader, management, logger, exchange, escrow, etc. |  | TBD – Map to confidentiality-specific ? |  |
| Least trust |  | Metric needed for trust components & disclosed to originator/owner | “The principal that a security architecture should be designed in a way that minimizes 1) the number of components that require trust, and 2) the extent to which each component is trusted. “ |
| Line-of-business privacy guidelines | OMB, SP 800-60, OB Business Reference Model FEA V2.3 | Domain- or discipline-specific privacy best practices[[7]](#endnote-7) | Lengthy |
| List-oriented object privacy protection | CNSSI-4009 |  |  |
| Major / Minor application (for privacy) | OMB Circular A-130 Appendix III; SP 800-18 | What makes it major / minor in the BDRA? |  |
| Masquerading privacy data (see identity) | SP 800-19 |  |  |
| <<< Not Finish Yet:  To Be Continued >>> |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

1. For example, identifying where there is no active testing available (e.g., encryption assurance). [↑](#endnote-ref-1)
2. We have not fully considered this in the NBDRA directly. Use cases would help. A privacy “process” covers a diverse set of processes. [↑](#endnote-ref-2)
3. Useful: Notion of a privacy attack vector is a useful big data discriminator, and may be highly system-specific. [↑](#endnote-ref-3)
4. Audit for privacy could entail audit for a small subset of a larger database, or audit intended to verify that security or privacy controls are being enforced. [↑](#endnote-ref-4)
5. Possibly needed for BDRA [↑](#endnote-ref-5)
6. IP protections, defenses, risks are similar but also different from individual human privacy. [↑](#endnote-ref-6)
7. LOB or Domain-specific privacy. See also incidents, events, etc. Needs improved definition and examples. [↑](#endnote-ref-7)