# NBDIF Transparency Requirements

Transparency is succinctly defined in ISO 16759:2013 as “open, comprehensive, *accessible,* clear and understandable presentation of information.” This definition reflects a general purpose, lay understanding of transparency. The definition is one among several important dimensions of a transparency framework.

More detail is specified in this NBDIF framework. Additional context is usually required as data may be aggregated or disaggregated across and between big data systems. Application of algorithmic processing on data creates additional responsibilities for data owners. Changes in ownership, governance and system configurations over time are an integral part of big data security and privacy fabric.

In addition to the System Communicator, NBDIF support for transparency is buttressed by optional *System Learner Models* and *Interaction Profiles*.

The NBDIF considers these additional big data transparency considerations:

* Data may consist of information that is fully or partially anonymized.
* Data sources can include current but also legacy data sources (e.g., earlier versions of IoT devices) or systems
* Promises regarding transparency and privacy must be retained across enterprises and original system architects.
* When data is shared, transparency and privacy data must travel with it at equivalent or better provenance and granularity.
* ISO 16759:2013 does not address risk. Stakeholders, users and data providers must be provided with risk as part of transparency. Risk is high domain-specific, thus additional metadata and modeling data will likely accompany mechanisms that support transparency.
* Transparency is further discussed in NIST 800-37 Rev 2, which asserts one goal for the NIST Risk Management Framework as “To support consistent, informed, and ongoing authorization decisions (through *continuous monitoring*), reciprocity, and the *transparency* and *traceability* of security and privacy information” (Public draft, Chapter 1, p. 3, October 2018, italics added). Added challenges associated with the information supply chain were also highlighted: “Effective risk decisions by authorizing officials depend on the *transparency* of controls selected and implemented by external providers and the quality and efficacy of the assessment evidence produced by those providers. *Transparency* is essential to achieve the assurance necessary to ensure adequate protection for organizational asset” (Ibid, Appendix G, italics added). The NBDIF specifies guidelines to support transparency, traceability and monitoring of data, algorithms, ownership and relevant system attributes.
* Audit records (e.g., when transparency was disclosed, to whom, what was disclosed) shall be retained beyond individual system life cycle design patterns. System life cycle status can be a critical component of transparency disclosures.
* Provide transparency for withdrawal of consent. For instance, compliance with GDPR specifies a right to be forgotten, but there will be practical or system limitations. Full transparency would include what data has been quarantined (“forgotten”), but as noted elsewhere in the NBDIF, big data will often persist beyond its originating system(s), and this process creates transparency requirements.
* In some cases where extensive granularity is required, support for dual privacy and transparency can itself require or add significantly to big data systems.
* Timelines must be maintained for significant transparency events, such as changes to algorithms, data ownership, increased or decreased data governance, configuration management.
* Changes to algorithms, such as joins with geospatial or other data sources, typically mandate additional transparency.
* Big data systems over time may experience increases or decreases in risk. For example, a small data set could be merged with a much larger data set, or when data is moved from a high security data center to a lower security data center. Shifts in risk profile shall be disclosed as part of transparency conformance.

Transparency may have necessary vs. sufficient considerations. For instance, regulators may mandate that lenders explain why credit is denied, even though credit decisions may be fully or partially supported by algorithms (e.g., Fair Credit Reporting Act, 15 U.S.C. § 1681).

**Big Data Transparency** Explanations will often focus on data providers and data provider processes. For example, in a clinical setting, an explanation for why a particular Rx was prescribed could be different for the patient, patient family, a clinical decision support system, the primary care physician, a radiologist, a pharmacist, public health official, or malpractice attorney. For a big data system, explanations of a real time big data stream to a data consumer may be needed for future system implementers to understand how that data source should be ingested. In addition, some big data systems will need an explanation of the processes that include how data is being collated with other sources.

To support transparency, a big data system provider output should include,at a minimum, a natural language explanation that is understandable to the identified target user population(s). When the explanation is challenging to offer (e.g., “what a system does”) the best alternative may be to explain what it is (how the process works, came about, etc.) or to provide representative scenarios.

For machine-to-machine implementations, transparency may be best achieved by implementing domain-specific languages which can be dynamically linked to scenarios, images or natural language text. Ad hoc solutions will likely fail to scale in systems with Big Data variety or specialized domains which frequent software releases or changes in the science, technology or regulatory landscape.

For Big Data systems, a layered approach is required to provide a safe, scalable, composable security and privacy transparency fabric. The NBDIF specifies three levels of voluntary conformance to big data system transparency:

1. **Transparency Level 1 Conformance** Level 1 utilizes the System Communicator to provide online explanations to users and stakeholders. These explanations, subject to other security and privacy guidelines and constraints, include explanation of the output of system processes to include, most commonly, a natural language explanation understandable by identified target user populations. “User populations” roughly follow the definition of roles in ISO 27nnnn (see NIST list). Transparency contracts and explanations shall be retained with the system, along with a record of what has been disclosed, accepted or rejected. Granularity shall be sufficient to meet the needs of the identified user populations. This shall be achieved through NBDIF Interaction Profiles at individual user granularity. Accompanying disclosure records may, for instance, include information requested but not provided due to system constraints or regulation, but Interaction Profiles are recommended at Level 1. Interaction Profiles will likely include elements derived from baselines and profiles specified in the NIST Cybersecurity Framework (SP 800-53B Revision 5 control baselines).
2. **Transparency Level 2 Conformance** Level 2 specifies a domain-specific system model, along with System Communicator protocols included in Transparency Level 1. Each system domain has potentially unique roles, attributes, phases, elements and dependencies which must be modeled. In addition, Big Data Interaction Profiles are mandatory at Level 2 and shall include a full, privacy-preserving record of all transparency-related transactions with a Big Data system. (Interaction Profile integrity may be ensured using Big Data techniques discussed in this document, such as blockchain). Level 2 conformance shall also include a System Learner Model for individual users (Knijnenburg, B.P., Cherry, D., Wilkinson, D., 2017). This model “teaches” what a Big Data system does, what risks may be involved, what impacts on privacy or security should be considered, how data may be shared and “learns” more. A continuously evolved “System Learner model” is preserved and tightly linked to the domain model of the application. While privacy is a key part of the model, the security and privacy fabric must include other facets of Big Data systems as they evolve over time and touch other aspects of their interactions with users or systems.
3. **Transparency Level 3 Conformance** Level 3 incorporates Level 2 practices plus digital ontologies for the associated domains and learner models. Automated reasoning systems at Level 3 allow for fully traceable explanations that are system-, learner-, feature-, time- and domain-dependent. Level 3 conformance may require linkage to a natural language processing subsystem. The System Learner Models and Interaction Profiles shall permit automated reasoning such as that specified in ISO 18629, and automation of processes outlined in NIST SP 800-162 for attribute-based security. The additional capabilities enable automated escalation of system processes based upon elevated risk, safety, adjustment of user interfaces for impaired users or children, automation of notification and alerting, and ease of interoperability with legacy systems such as metadata management or compliance engines.

In the NBDPWG Big Data framework, “information” has a broader meaning that is normally associated with systems design. Hence, transparency has a broader implication as well. For instance, transparency may include anthropological elements (Bellestero, 2018). Empirical methods may be needed to provide for measurement of transparency effectiveness, so that tuning and improvements can evolve with big data systems deployments in DevOps. They may incorporate empirically based “effective information design” (Seizov, Wulf & Luzak, 2018). These capabilities in turn demand measurement data which contributes both to a Big Data system’s purview, but also enlarges the scope of the security and privacy fabric.

## Related Standards

### Under development

IEEE P7001, IEEE P7003, IEEE P7007, NIST 800-30 V2, IEEE P1915.1, IEEE P2675

### Extant

NIST SP 800-53 Security and Privacy Controls for US Federal Agencies

NIST SP 800-30 Guide for conducting risk assessments

NIST SP 800-37 Risk Management Framework

NIST SP 800-160 Security Engineering

NIST SP 800-162 Guide to Attribute Based Access Control

ISO 15288 Systems engineering life cycle management

ISO 15408-1, -2, -3 Security requirements and controls

ISO 29148 Requirements Engineering

ISO 27001 Security Management Systems Requirements

[ISO 10303-11](https://docs.google.com/document/d/1vi2kB3humyKJfwJ81_RaF6UPYAfcaB3K3fkqD6iih6o/edit#bookmark=id.9nhlb210q1pr) Product data management, including models

[ISO 11179](https://docs.google.com/document/d/1vi2kB3humyKJfwJ81_RaF6UPYAfcaB3K3fkqD6iih6o/edit#bookmark=id.cltat2offvid) Metadata registries

[ISO 12207](https://docs.google.com/document/d/1vi2kB3humyKJfwJ81_RaF6UPYAfcaB3K3fkqD6iih6o/edit#bookmark=id.8yj5fywmhc8p) Software life cycle processes

IEEE P16085 Risk management in systems life cycle

ISO 14001 Monitoring measurement for environmental processes

ISO 18629 Automated reasoning about business processes

ISO/PAS 19450 ISO TC184/SC5 Object Process Methodology (conceptual modeling standard)

ISO/IEC 19788-5:2012(en) Information technology — Learning, education and training — Metadata for learning resources — Part 5: Educational elements | [ISO page](https://www.iso.org/obp/ui#iso:std:iso-iec:19788:-5:ed-1:v1:en)

ISO 26000 Social aspects of systems processes[ISO 27018](https://docs.google.com/document/d/1vi2kB3humyKJfwJ81_RaF6UPYAfcaB3K3fkqD6iih6o/edit#bookmark=id.1fzzzev2jr4d) Protection of PII

[ISO 33001](https://en.wikipedia.org/wiki/ISO/IEC_33001) Systems life cycle engineering,replaced 15504

IEC 61508 Electronic systems safety reliability

ISO 21932 Sustainability

ISO 16759 Carbon footprint specification for print

ISO 15836-1 Dublin Core Metadata Initiative

-- Include list of ISO standard usage for “transparency.”

Bibliography

1. J. Powles, "New York City's bold, flawed attempt to make algorithms accountable," *New Yorker*, Dec. 2017. [Online]. Available: <https://www.newyorker.com/tech/annals-of-technology/new-york-citys-bold-flawed-attempt-to-make-algorithms-accountable>
2. A. Ballestero, Transparency, Sep. 2018. [Online]. Available: <http://dx.doi.org/10.1002/9781118924396.wbiea1505>
3. O. Seizov, A. J. Wulf, and J. A. Luzak, The Transparent Trap: A Multidisciplinary Perspective on the Design of Transparent Online Disclosures in the EU, Oct. 2018. [Online]. Available: <http://dx.doi.org/10.1007/s10603-018-9393-0>
4. Wilkinson, D., Sivakumar, S., Cherry, D., Knijnenburg, B.P., Raybourn, E., Wisniewski, P., Sloan, H. (2017): User-Tailored Privacy by Design. Work-in-progress paper at the Usable Security Mini Conference (USEC), San Diego, CA, DOI: [10.14722/usec.2017.23007](http://dx.doi.org/10.14722/usec.2017.23007) .
5. Knijnenburg, B.P., Cherry, D., Wilkinson, D.,Sivakumar, S., Ghaiumy Anaraky, R., Namara, M., Ash, E., Davis, R.M., and Raybourn, E.M. (2017) “Privacy Support for the Total Learning Architecture: Operational Characteristics”. PS4TLA Specification Document, v0.1, Clemson University, Clemson, SC. Available at: http://www.usabart.nl/PS4TLA/spec0.1.pdf.