

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

X.1148

(09/2020)

SERIES X: DATA NETWORKS, OPEN SYSTEM
COMMUNICATIONS AND SECURITY

Secure applications and services (1) – Web security

Framework of de-identification process for telecommunication service providers

Recommendation ITU-T X.1148

ITU-T X-SERIES RECOMMENDATIONS
DATA NETWORKS, OPEN SYSTEM COMMUNICATIONS AND SECURITY

PUBLIC DATA NETWORKS	X.1–X.199
OPEN SYSTEMS INTERCONNECTION	X.200–X.299
INTERWORKING BETWEEN NETWORKS	X.300–X.399
MESSAGE HANDLING SYSTEMS	X.400–X.499
DIRECTORY	X.500–X.599
OSI NETWORKING AND SYSTEM ASPECTS	X.600–X.699
OSI MANAGEMENT	X.700–X.799
SECURITY	X.800–X.849
OSI APPLICATIONS	X.850–X.899
OPEN DISTRIBUTED PROCESSING	X.900–X.999
INFORMATION AND NETWORK SECURITY	
General security aspects	X.1000–X.1029
Network security	X.1030–X.1049
Security management	X.1050–X.1069
Telebiometrics	X.1080–X.1099
SECURE APPLICATIONS AND SERVICES (1)	
Multicast security	X.1100–X.1109
Home network security	X.1110–X.1119
Mobile security	X.1120–X.1139
Web security	X.1140–X.1149
Security protocols (1)	X.1150–X.1159
Peer-to-peer security	X.1160–X.1169
Networked ID security	X.1170–X.1179
IPTV security	X.1180–X.1199
CYBERSPACE SECURITY	
Cybersecurity	X.1200–X.1229
Countering spam	X.1230–X.1249
Identity management	X.1250–X.1279
SECURE APPLICATIONS AND SERVICES (2)	
Emergency communications	X.1300–X.1309
Ubiquitous sensor network security	X.1310–X.1319
Smart grid security	X.1330–X.1339
Certified mail	X.1340–X.1349
Internet of things (IoT) security	X.1360–X.1369
Intelligent transportation system (ITS) security	X.1370–X.1389
Distributed ledger technology security	X.1400–X.1429
Distributed ledger technology security	X.1430–X.1449
Security protocols (2)	X.1450–X.1459
CYBERSECURITY INFORMATION EXCHANGE	
Overview of cybersecurity	X.1500–X.1519
Vulnerability/state exchange	X.1520–X.1539
Event/incident/heuristics exchange	X.1540–X.1549
Exchange of policies	X.1550–X.1559
Heuristics and information request	X.1560–X.1569
Identification and discovery	X.1570–X.1579
Assured exchange	X.1580–X.1589
CLOUD COMPUTING SECURITY	
Overview of cloud computing security	X.1600–X.1601
Cloud computing security design	X.1602–X.1639
Cloud computing security best practices and guidelines	X.1640–X.1659
Cloud computing security implementation	X.1660–X.1679
Other cloud computing security	X.1680–X.1699
QUANTUM COMMUNICATION	
Terminologies	X.1700–X.1701
Quantum random number generator	X.1702–X.1709
Framework of QKDN security	X.1710–X.1711
Security design for QKDN	X.1712–X.1719
Security techniques for QKDN	X.1720–X.1729
DATA SECURITY	
Big Data Security	X.1750–X.1759
5G SECURITY	X.1800–X.1819

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T X.1148

Framework of de-identification process for telecommunication service providers

Summary

Telecommunication organizations collect, manage, use, and share data about individuals, including personally identifiable information. As a result, they utilize data de-identification techniques to protect individuals' data. Recommendation ITU-T X.1148 describes a framework of de-identification process with operational steps and specifies data release models and data stages in a de-identification process for telecommunication service providers based on data lifecycle model and the roles of stakeholders.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T X.1148	2020-09-03	17	11.1002/1000/14249

Keywords

Data principal, de-identification, de-identification process, k-anonymity, l-diversity, PII protection, release models, t-closeness.

* To access the Recommendation, type the URL <http://handle.itu.int/> in the address field of your web browser, followed by the Recommendation's unique ID. For example, <http://handle.itu.int/11.1002/1000/11830-en>.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at <http://www.itu.int/ITU-T/ipr/>.

© ITU 2020

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

Table of Contents

		Page
1	Scope	1
2	References.....	1
3	Terms and definitions	1
	3.1 Terms defined elsewhere	1
	3.2 Terms defined in this Recommendation.....	3
4	Abbreviations and acronyms	3
5	Conventions	3
6	Overview of de-identification process.....	3
	6.1 Data lifecycle model and de-identification phase	3
	6.2 De-identification considerations.....	5
7	De-identification process framework	6
	7.1 Step 1 – Preliminary review	7
	7.2 Step 2 – Applying de-identification	8
	7.3 Step 3 – Adequacy assessment for de-identification process.....	9
	7.4 Step 4 – Follow-up management.....	9
8	De-identified data utility.....	10
	8.1 De-identified data stages	10
	8.2 Data release models.....	11
	8.3 Relation between data release model and data stage.....	13
Annex A – Adequacy assessment procedures.....		14
	A.1 Preparing basic documents	15
	A.2 Organizing an assessment panel.....	15
	A.3 Conducting assessment.....	15
	A.4 Additional de-identification measures.....	16
	A.5 Data utilization	16
Annex B – Unstructured de-identification approaches		17
Appendix I – Examples of typical de-identification techniques		19
	I.1 Statistical tools for de-identification techniques	19
	I.2 Cryptographic tools for de-identification techniques	19
	I.3 Suppression techniques	19
	I.4 Pseudonymization techniques	19
	I.5 Generalization techniques	19
	I.6 Randomization techniques.....	20
	I.7 Synthetic data	20

	Page
Appendix II – De-identification process approaches	21
II.1 Data-centric de-identification approach	21
II.2 Role-centric de-identification approach	22
Bibliography.....	24

Introduction

With the rapid development of Internet based information and communication technologies and services, large amounts of data are generated, transmitted, and stored with explosive growth. Data are generated by many sources: not only sensors, cameras or network devices, but also web pages, email systems or social networks, and many others. Datasets are becoming so large, complex and are arriving so fast that traditional data process methods and tools are no longer adequate. Efficient data analytics within tolerable delay becomes very challenging. A paradigm called big data analytics are being developed to resolve the above issues.

Telecommunication organizations collect, manage, use, and share data about individuals, including personally identifiable information. As a result, they utilize data de-identification techniques to protect individuals' data. The relationships between parties participating in the data flow for data exchange affect whether data de-identification needs to be performed before its collection, after its collection but before its storage, or only before it is shared with the next party in data exchange. Accordingly, telecommunication service providers need to provide data de-identification as a service in a timely, efficient, and safe way to data customers.



Recommendation ITU-T X.1148

Framework of de-identification process for telecommunication service providers

1 Scope

This Recommendation provides an overview of de-identification process based on data lifecycle model, specifies a de-identification process framework with operational steps and roles of stakeholders in the de-identification process. It further discusses data release models and data stages in a de-identification process and includes various de-identification approaches and examples in its annexes and appendices.

This Recommendation does not address issues related to regulation.

2 References

None.

3 Terms and definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 aggregated data [b-ISO/IEC 20889]: Data representing a group of data principals, such as a collection of statistical properties of that group.

3.1.2 anonymization [b-ISO/IEC 29100]: Process by which personally identifiable information (PII) is irreversibly altered in such a way that a PII principal can no longer be identified directly or indirectly, either by the PII controller alone or in collaboration with any other party.

3.1.3 attribute [b-ISO/IEC 20889]: Inherent characteristic.

3.1.4 dataset [b-ISO/IEC 20889]: Collection of data.

3.1.5 de-identification [b-ISO 25237]: General term for any process of reducing the association between a set of identifying data and the data subject (see clause 3.2.4).

3.1.6 de-identification process [b-ISO/IEC 20889]: Process of removing the association between a set of identifying attributes and the data principal.

3.1.7 de-identification technique [b-ISO/IEC 20889]: Method for transforming a dataset with the objective of reducing the extent to which information is able to be associated with individual data principals.

3.1.8 de-identified dataset [b-ISO/IEC 20889]: Dataset resulting from the application of a de-identification process.

3.1.9 de-identified information [b-NISTIR 8053]: Record that have had enough PII removed or obscured such that the remaining information does not identify an individual and there is no reasonable basis to believe that the information can be used to identify an individual.

3.1.10 differential privacy [b-ISO/IEC 20889]: Formal privacy measurement model that ensures that the probability distribution of the output from a statistical analysis differs by at most a specified value, whether or not any particular data principal is represented in the input dataset.

NOTE – More specifically, differential privacy provides:

- a) a mathematical definition of privacy which posits that, for the outcome of any statistical analysis to be considered privacy-preserving, the analysis results from the original dataset are indistinguishable from those obtained if any data principal is added to or removed from the dataset; and
- b) a measure of privacy that enables monitoring of cumulative privacy loss and setting of an upper bound (or "budget") for loss limit. A formal definition is as follows. Let ϵ be a positive real number, and M be a randomized algorithm that takes a dataset as input. The algorithm M is said to be ϵ -differentially private if for all datasets $D1$ and $D2$ that differ in a single element (i.e., the data for one data principal), and all subsets S of the range of M , mml_m1 , where the probability is taken over the randomness used by the algorithm.

3.1.11 identifier [b-ISO/IEC 20889]: Set of attributes in a dataset that enables unique identification of a data principal within a specific operational context.

NOTE – See Annex B for a discussion of how this definition relates to those given in other standards.

3.1.12 identifying attribute [b-ISO/IEC 20889]: Attribute in a dataset that can contribute to uniquely identifying a data principal within a specific operational context.

3.1.13 privacy stakeholder [b-ISO/IEC 29100]: Natural or legal person, public authority, agency or any other body that can affect, be affected by, or perceive themselves to be affected by a decision or activity related to personally identifiable information (PII) processing.

3.1.14 pseudonymization [b-ISO/IEC 20889]: de-identification technique that replaces an identifier (or identifiers) for a data principal with a pseudonym in order to hide the identity of that data principal.

3.1.15 quasi-identifier [b-ISO/IEC 20889]: Attribute in a dataset that, when considered in conjunction with other attributes in the dataset, singles out a data principal.

3.1.16 record [b-ISO/IEC 20889]: Set of attributes concerning a single data principal.

3.1.17 re-identification [b-ISO/IEC 20889]: Process of associating data in a de-identified dataset with the original data principal.

NOTE – A process that establishes the presence of a particular data principal in a dataset is included in this definition.

3.1.18 single out [b-ISO/IEC 20889]: Isolate records belonging to a data principal in the dataset by observing a set of characteristics known to uniquely identify this data principal.

3.1.19 third party [b-ISO/IEC 29100]: Privacy stakeholder other than the personally identifiable information (PII) principal, the PII controller and the PII processor, and the natural persons who are authorized to process the data under the direct authority of the PII controller or the PII processor.

3.1.20 Trusted Third Party [b-ISO/IEC 18014-1]: Security authority, or its agent, trusted by other entities with respect to security related activities.

3.1.21 k-anonymity [b-ISO/IEC 20889]: Formal privacy measurement model that ensures that for each identifier in a dataset there is a corresponding equivalence class containing at least K records.

3.1.22 l-diversity [b-ISO/IEC 20889]: Formal privacy measurement model that ensures that for a selected attribute each equivalence class has at least L well-represented values.

NOTE – L -diversity is a property of a dataset that gives a guaranteed lower bound, L , on the diversity of values shared by an equivalence class for a selected attribute.

3.1.23 t-closeness [b-ISO/IEC 20889]: Formal privacy measurement model that ensures that the distance between the distribution of a selected attribute in an equivalence class and the distribution of this attribute in the entire table is no more than a threshold T

NOTE – A table is said to have T -closeness with respect to a selected attribute if all equivalence classes containing this attribute have T -closeness.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 data controller: Stakeholder (or privacy stakeholder) that determines the purposes and means for processing data other than natural persons who use data for personal purposes.

3.2.2 data processor: Stakeholder that processes data on behalf of and in accordance with the instructions of a data controller.

3.2.3 data protection officer: Person appointed by the personally identifiable information (PII) controller to ensure, in an independent manner, compliance with the privacy law or regulation requirements.

NOTE – "PII controller" is synonym of "data controller".

3.2.4 data subject: Entity to which data relates.

NOTE – "data subject" is synonym of "PII principal" and "data principal".

3.2.5 process: In relation to information or data, this means obtaining, recording, or holding the information or data or carrying out any operation or set of operations on the information or data, including:

- organization, adaptation or alteration of the information or data,
- retrieval, consultation or use of the information or data,
- disclosure of the information or data by transmission, dissemination or otherwise making available, or
- alignment, combination, blocking, erasure or destruction of the information or data.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

DP	Differential Privacy
DPO	Data Protection Officer
PII	Personally Identifiable Information
TTP	Trusted Third Party

5 Conventions

None.

6 Overview of de-identification process

The purpose of de-identification process is to protect the confidentiality of a subjects' data. As these data may include personally identifiable information (PII), before and after data analytics with the purpose of extracting meaningful information, a data analyst must include security considerations.

This clause defines data analysis environments, data lifecycle model, roles of entities in the de-identification process and other de-identification considerations.

6.1 Data lifecycle model and de-identification phase

Typically, an organization sets de-identification objectives for privacy and security aims. This clause defines a data life cycle and describes when to consider a de-identification process based on this data life cycle model.