DRAFT INTERNATIONAL STANDARD

ISO/IEC DIS 20547-4

ISO/IEC JTC **1**/SC **27**

Secretariat: **DIN**

Voting begins on:
2019-12-30

Voting terminates on:
2020-03-23

Information technology — Big data reference architecture —

Part 4:
Security and privacy

Technologies de l'information — Architecture de référence des mégadonnées —

Partie 4: Sécurité et Confidentialité

ICS: 35.020

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
https://standards.iteh.ai/catalog/standards/sist/22cc47cd-9d99-4e5a-ad90-062792041b4c/iso-iec-dis-20547-4

THIS DOCUMENT IS A DRAFT CIRCULATED FOR COMMENT AND APPROVAL. IT IS THEREFORE SUBJECT TO CHANGE AND MAY NOT BE REFERRED TO AS AN INTERNATIONAL STANDARD UNTIL PUBLISHED AS SUCH.

IN ADDITION TO THEIR EVALUATION AS BEING ACCEPTABLE FOR INDUSTRIAL, TECHNOLOGICAL, COMMERCIAL AND USER PURPOSES, DRAFT INTERNATIONAL STANDARDS MAY ON OCCASION HAVE TO BE CONSIDERED IN THE LIGHT OF THEIR POTENTIAL TO BECOME STANDARDS TO WHICH REFERENCE MAY BE MADE IN NATIONAL REGULATIONS.

RECIPIENTS OF THIS DRAFT ARE INVITED TO SUBMIT, WITH THEIR COMMENTS, NOTIFICATION OF ANY RELEVANT PATENT RIGHTS OF WHICH THEY ARE AWARE AND TO PROVIDE SUPPORTING DOCUMENTATION.

This document is circulated as received from the committee secretariat.

Reference number
ISO/IEC DIS 20547-4:2019(E)

© ISO/IEC 2019

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
https://standards.iteh.ai/catalog/standards/sist/22cc47cd-9d99-4e5a-ad90-062792041b4c/iso-iec-dis-20547-4

COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2019

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

© ISO/IEC 2019 – All rights reserved

Contents

Page

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
https://standards.iteh.ai/catalog/standards/sist/22cc47cd-c197-4e5a-a49e-0067210abbc/iso-iec-dis-20547-4

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
https://standards.iteh.ai/catalog/standards/sist/22cc47cd-9d99-4e5a-ad90-062792041b4c/iso-iec-dis-20547-4

© ISO/IEC 2019 – All rights reserved

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see the following URL: www.iso.org/iso/foreword.html.

The committee responsible for this document is Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 27, *Information security, cybersecurity and privacy protection*.

A list of all parts in the ISO/IEC 20547 series can be found on the ISO website.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

© ISO/IEC 2019 – All rights reserved

Introduction

Big data refers to the massive amount of digital information collected in various forms from different sources of digital and physical environments. This data is not only generated by traditional means of information exchange, but also from sensors embedded in physical environments, such as city surroundings, transportation vehicles, critical infrastructures, etc. The collection and processing of big data provides additional challenges not inherent in the traditional digital information exchange setting.

This document was developed in response to worldwide demand for a common baseline security and privacy aspects for big data architectures to facilitate interoperability in big data systems without compromising privacy, confidentiality, or integrity.

The big data paradigm blurs the security boundaries between data collection, storage, and access – areas traditionally addressed independently – that must now be confronted holistically with a comprehensive security and privacy foundation, tightly coupled to all architecture components.

Effective standardization of security and privacy is paramount to the development of mutual trust and cooperation amongst big data stakeholders.

Identification of patent holders, if any.

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
https://standards.iteh.ai/catalog/standards/sist/22cc47cd-9d99-4e5a-ad90-062792041b4c/iso-iec-dis-20547-4

© ISO/IEC 2019 – All rights reserved

Information technology — Big data reference architecture —

Part 4: Security and privacy

1 Scope

This document specifies the security and privacy aspects applicable to the Big Data Reference Architecture (BDRA) including the big data roles, activities, and functional components, and also provides guidance on security and privacy operations for big data.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 20547-3, *Information technology — Big data reference architecture — Part 3: Reference architecture*

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
https://standards.iteh.ai/catalog/standards/sist/12cc47ed-a499-45a-a190-00792041b4c/iso-iec-dis-20547-4

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 20546 apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at http://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

4 Symbols and abbreviated terms

The following abbreviations apply to this document.

APT	Advanced Persistent Threat
BDRA	Big Data Reference Architecture
BD-S&P	Big Data Security and Privacy
DDoS	Distributed Denial of Service
PII	Personally Identifiable Information

© ISO/IEC 2019 – All rights reserved

5 Overview

5.1 Big data security and privacy concerns

Subclause 5.1 addresses three types of concerns:

— Risks caused by big data characteristics

— Security and privacy challenges from big data

— Ecosystem coordination capabilities needed in the context of big data

Big data has the key data characteristics of volume, velocity, variety and variability, and also the key data processing characteristics of volatility, veracity and value. These characteristics introduce additional risks and thus challenges on the security and privacy aspect of big data.

— Volume of data is at risk associated with massive amounts of data in various layers. For example, multi-tiered, distributed storages and transmission on various networks with different protocols.

— Velocity of data has risks associated with faster flow at which the data is created, stored, analysed or visualized. Security controls might be a burden on velocity and easily omitted.

— Variety of data brings more complexity from a diversity of sources under control of various actors. Complexity inevitably leads to vulnerability. An emergent phenomenon introduced by big data variety is its ability to infer identity from anonymized datasets by correlating with apparently innocuous public databases.

— Variability of data has risks associated with faster changes in data rate, format/structure, semantics, and/or quality. It might become more difficult to apply security controls on data security and privacy.

— Volatility of data could affect to keep audit trails and make security management difficult.

— Veracity of data brings higher requirements to integrity, consistency and accuracy. Associated risks could be aggregated and magnified.

— Value of data brings more attacks for a variety of purposes and interests.

Big data application boom brings more serious security and privacy issues on data, such as frequent incidents of data loss and personal data leakage, illegal data transactions underground, which cause data abuse and Internet fraud, and endanger social stability and national security.

From technology platform perspective, due to the continuous emergence of a variety of big data technology, new technical architecture and support platform, and big data software, the following security and privacy capabilities are needed in big data context:

— Traditional security controls in big data context are needed.

  Massive, multi-sourced, heterogeneous, dynamic and other big data characteristics lead to the difference of data application security from a closed environment. Big data applications generally use the open distributed computing and storage architecture with complex underlying support to provide massive data distributed storage and high-performance computing services. These new technologies and architectures make the network boundaries of big data applications become blurred, so that the boundary-based traditional security protection measures are no longer valid. Meanwhile, under this new situation, the Advanced Persistent threat (APT), Distributed Denial of Service (DDoS), machine learning-based data mining and privacy discovery, and other attacks make the traditional defence, detection and other security controls expose serious deficiencies. For instance, providing secure data management and threat intelligence, providing secure data storage for big data as well as secure log data generation, transmission, storage, analysis and disposal becomes very difficult. Additionally novel technical approaches for privacy-preserving, machine learning, cryptographic mechanisms for data-centric security and access control are necessary.

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
https://standards.iteh.ai/catalog/standards/sist/12fc0e7cd-9d99-e55a-ad9e-0627b2041b4c/iso-iec-dis-20547-4

 © ISO/IEC 2019 – All rights reserved

For more information on the requirements of big data security and privacy, refer the security and privacy technical considerations in different use cases provided by ISO/IEC TR 20547-2:2018.

— Security and privacy need to be provided for the distributed computation and data store infrastructure of big data.

This requires privacy-preserving and secure distributed computation and information dissemination. Big data requires scalable and distributed solutions for secure data storage as well as for audits and investigations for data provenance. Data integrity for streaming influx of data from various sensors and other end-points has to be provided. Real-time analytics for threat intelligence require processing of large amounts of security related information such as traffic streams and log information.

— Platform security mechanisms need to be improved in the context of big data security and privacy.

In general, existing big data applications use the big data management platforms and technology, such as Hadoop-based HBase/Hive, Cassandra/Spark, MongoDB. At the beginning of design, these platforms and technology are mostly considered to be used in the trusted internal network, with little consideration of authentication, authorization, key services and security audit. Although some software are improved, such as adding Kerberos authentication mechanism, the overall security capability is still relatively weak. Meanwhile, the third party open source components are often used in big data applications. Due to the lack of rigorous test management and security certification of these components, the ability to prevent software vulnerabilities and malicious backdoors in big data applications is insufficient.

— Application access control capabilities in big data context are needed.

Because of the variety of data types and the wide range of applications of big data, it is often used to provide multiple services to users with different identities and purposes from different organizations or departments. In general, access control is an effective means to achieve controlled access to data. However, due to a large number of unknown data users and data to be accessed, it is very difficult to pre-set roles and permissions of data access. Although user's rights to access data can be classified in advance, because of the numerous roles, it is difficult to define the control of each role's actual permissions in a fine-grained way. So, it is difficulty to accurately specify the range of data for each user to access without deploying newer access control model such as ABAC (Attribute-Base Access Control). This also causes the issue with the data minimization principle in ISO/IEC 29100.

— Scalable security and privacy mechanisms are needed.

When designing and applying security and privacy mechanisms such key management, identity and access management, de-identification, etc., in big data environment, not only the security and privacy functionalities need to be considered, but also scalability of these mechanisms need to be taken into account in order to support processing of high volume and high velocity of the data.

From data application perspective, due to the big Vs (Volume, Variety, Velocity and Variability) characteristics of big data, and huge value in big data, the following security and privacy capabilities are needed in big data context:

— Data protection capabilities in big data context are needed.

In the open network society, the huge volume of big data with immeasurable potential value makes it more favoured and easier to become a significant target of network attacks. In recent years, information security incidents frequently occur, for example, leakage of email accounts, social security information and bank card numbers. The distributed system deployment, the open network environment, the complex data application and the large amount of user accesses, all cause the big data to face the bigger challenges in the confidentiality, integrity, availability and so on.

iTeh STANDARD PREVIEW
(standards.iteh.ai)
https://standards.iteh.ai/catalog/standards/sist/22cd47ed-9d99-4a53-a000-0627924f4c4f/iso-iec-dis-20547-4

© ISO/IEC 2019 – All rights reserved

— Personal data protection capabilities in big data context are needed.

  Due to the large amount of personal data in big data systems, when the security incidents such as data abuse, internal theft and network attacks occur, the consequences of personal data leakage will be more serious than ordinary information systems. On the other hand, the advantage of generating value from the analysis and usage of large amounts of data could be compromised by the more risk of personal data leakage during the comprehensive analysis of multi-source data where analysts are easier to explore more personal data through correlation analysis.

— Data authenticity capabilities in big data context are needed.

  Data in big data systems have a wide range of sources that could be a variety of sensors, active uploads and public websites. In addition to reliable data sources, there are a large number of untrusted data sources. Some attackers even deliberately falsify data in an attempt to induce data analysis results. Therefore, it is very important to verify the authenticity of data and their sources. However, there are many difficulties in verifying all the data authenticity because of the performance limitation of data acquisition terminals, the lack of technology, the limited amount of information, and the variety and complexity of sources.

— Data owner's right protection capabilities in big data context are needed.

  During the application of big data, the data could be accessed by a variety of users, flow from one controller to another, and even be mined to produce new data. Therefore, in the process of data exchange and sharing, there are the circumstance where the data ownership of a data owner and the data use right of a data manager are separated, which implies that data could be out of data owner's control, and brings the risks such as data abuse, vague ownership of data, unclear responsibilities of data security supervision, so that the rights and interests of data owners could be seriously damaged.

iTeh STANDARD PREVIEW
(standards.iteh.ai)
https://standards.iteh.ai/catalog/standards/sist/2cce7c1-9d9a-4eca-ad00-0279204f4c/iso-iec-dis-20547-4

Big data involves ecosystems, or networks of organizations which collaborates to collect, analyse and share data. The following collaborations are needed in a big data context:

— collaboration between stakeholders to ensure that overall ecosystem security and privacy requirements and individual organization's security and privacy requirements are consistent;

— collaboration between stakeholders to ensure that the overall ecosystem risk management and the individual organization's risk management are consistent; and

— collaboration between stakeholders to ensure that the individual organizations ensure a consistent treatment of the assets to protect.

5.2 Security and privacy objectives

Big data applications have security and privacy objectives. Table 1 describes examples of security objectives. Table 2 describes examples of privacy objectives. Note that Table 2 describes alternative ways to look at objectives.

 © ISO/IEC 2019 – All rights reserved

Table 1 — Security objectives

Objective		Description	Examples
Security protection goals [From ISO/IEC 27000]	Confidentiality	Ensures that information is not made available or disclosed to unauthorized individuals, entities, or processes.	Protected transmission of collected data, protected access with suitable authentication schemes, protected processing of data, and protected storage.
	Integrity	Ensures the accuracy and completeness of data over its entire life cycle.	Protection of integrity during transmission, processing of data, as well as at storage level using schemes such as digital signatures
	Availability	Ensures accessibility and usability upon demand by an authorized entity	Preventing service disruptions due to power outages, hardware failures, or denial of service attacks using schemes such as redundant systems.

Table 2 — Privacy objectives

Objective		Description
Privacy principles [From ISO/IEC 27550]	Consent and choice	Providing PII principals with clear, prominent, easily understandable, accessible and affordable mechanisms to exercise choice and to give consent in relation to the processing of their PII at the time of collection
	Purpose legitimacy and specification	Ensuring that the purpose(s) complies with applicable law and relies on a permissible legal basis
	Collection limitation	Limiting the collection of PII to that which is within the bounds of applicable law and strictly necessary for the specified purpose
	Data minimization	Strictly minimizes the processing of PII
	Use retention and disclosure limitation	Limiting the use, retention and disclosure (including transfer) of PII to that which is necessary in order to fulfil specific, explicit and legitimate purposes
	Accuracy and quality	Ensuring that the PII processed is accurate, complete, up-to-date (unless there is a legitimate basis for keeping outdated data), adequate and relevant for the purpose of use
	Openness, transparency and notice	Providing PII principals with clear and easily accessible information about the PII controller's policies, procedures and practices with respect to the processing of PII
	Individual participation and access	Giving PII principals the ability to access and review their PII, provided their identity is first authenticated with an appropriate level of assurance and such access is not prohibited by applicable law
	Accountability	The processing of PII entails a duty of care and the adoption of concrete and practical measures for its protection
	Information security	Protecting PII under its authority with appropriate controls at the operational, functional and strategic level to ensure the integrity, confidentiality and availability of the PII, and protect it against risks such as unauthorized access, destruction, use, modification, disclosure or loss throughout the whole of its life cycle;
	Privacy compliance	Verifying and demonstrating that the processing meets data protection and privacy safeguarding requirements by periodically conducting audits using internal auditors or trusted third-party auditors

iTeh STANDARD PREVIEW (standards.iteh.ai)

© ISO/IEC 2019 – All rights reserved

Table 2 *(continued)*

Objective		Description
Privacy protection goals [From ISO/ IEC 27550]	Unlinkability	Ensures that a PII principal may make multiple uses of resources or services without others being able to link these uses together EXAMPLE A customer uses two different accounts for a service which involves big data analysis.
	Transparency	Ensures that an adequate level of clarity of the processes in privacy relevant data processing is reached so that the collection, processing and use of the information can be understood and reconstructed at any time EXAMPLE Understandable documentation covering technology, organization and responsibilities accessible to the PII principal.
	Intervenability	Ensures that PII principals, PII controller, PII processors and supervisory authorities can intervene in all privacy-relevant data processing EXAMPLE Processes for influencing or stopping the data processing fully or partially, manually overturning an automated decision, data portability precautions.
Privacy Engineering objectives [From ISO/ IEC 27550]	Predictability	Providing a reliable understanding about what is occurring with PII processing within a system EXAMPLE Developing a PII processing capability that is repeatable. Related to a the transparency protection goal.
	Manageability	Administration of PII with sufficient granularity so that the right level of control can be applied EXAMPLE Development of a privacy preference management that allows PII principal the right level of control. Related to the intervenabililty protection goal.
	Disassociability	Actively protect or “blind” an individual’s identity or associated activities from unnecessary exposure during transactions EXAMPLE Development of a health data processing capability that is anonymous. Related to the unlinkability protection goal.

All the following clauses of this document are under the context of big data systems which have the above security and privacy (S&P) concerns and objectives.

6 Security and privacy aspects of BDRA user view

6.1 Overview

Clause 6 specifies the security and privacy activities from BDRA user view.

6.2 Governance activities

6.2.1 Purpose

The data governance defined in ISO/IEC 20547-3 is the governance of an organization that focuses on the data aspect of the organization. The purpose of BD-S&P Governance activities is to establish and maintain BD-S&P's coherence and alignment with objectives and constraints with respect to the current and future security & privacy needs of an organization and its stakeholders or interested parties.

NOTE Governance of BD S&P can apply to an organization internally, or to an ecosystem. For instance, an organization in charge of a big data market place can provide governance policies to its participants.

 © ISO/IEC 2019 – All rights reserved

6.2.2 Prepare for and plan BD-S&P governance effort

a) Establish BD-S&P strategy derived from organizational objectives, strategy and vision.

b) Establish role, responsibilities, accountabilities, authorities, and organizational structures to support BD-S&P governance effort and reporting requirements.

c) Establish BD-S&P governance organizational structure that is consistent with defined roles, authorities, responsibilities and accountabilities.

d) Establish guiding principles and work instructions for performing BD-S&P governance.

NOTE These work instructions will delineate the steps to be performed by those who execute the governance activities. This is to ensure the decision processes are transparent and consistently followed. Sometimes a secretariat is used to administer these work instructions and to ensure they are properly followed.

e) Establish decision forums to carry out BD-S&P governance work instructions.

f) Define procedures for identifying, managing, auditing and disseminating information related to BD-S&P governance decisions.

1) Link these procedures to BD-S&P strategy.

2) Map these procedures to resources and constraints to support strategy, planning and decision making.

g) Plan BD-S&P governance effort

1) Establish the scope of BD-S&P governance effort.

2) Establish metrics for BD-S&P governance effort.

3) Identify the data and information needed for BD-S&P governance effort.

4) Identify and define BD-S&P governance work elements and associated resources.

5) Develop BD-S&P governance schedule and define associated milestones.

h) Obtain necessary approvals and funding for the plan.

i) Collect the data and information needed for BD-S&P governance effort.

6.2.3 Monitor, assess and control BD-S&P governance activities

a) Monitor and assess metrics for BD-S&P governance effort.

b) Identify and assess risks and opportunities associated with BD-S&P governance effort.

c) Ensure that other processes (e.g. enterprise life cycle processes, system life cycle processes) are properly using BD-S&P governance directives and guidance.

d) Report BD-S&P governance activity plans and status in accordance with reporting requirements.

e) Assess and control BD-S&P governance effort.

f) Manage risks associated with BD-S&P governance.

6.2.4 Establish BD-S&P governance objectives

a) Examine current and future big-data-related business needs when known.

1) Examine current and future big-data-related enterprise objectives that need to be achieved, such as maintaining competitive advantage.

iTeh STANDARD PREVIEW
(standards.iteh.ai)
Full standard:
https://standards.iteh.ai/catalog/standards/sist/22ec47cd-3d89-4e5a-ad0-062920204b4c/iso-iec-dis-20547-4