TECHNICAL REPORT

ISO/IEC TR 20547-5

First edition
2018-02

Information technology — Big data reference architecture —

Part 5: Standards roadmap

Technologies de l'information — Architecture de référence des big data —

Partie 5: Feuille de route pour les normes

Reference number
ISO/IEC TR 20547-5:2018(E)

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

Published in Switzerland

Contents

Page

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 document 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 and IEC 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 voluntary nature of standards, 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.

This document was prepared by Technical Committee ISO/IEC JTC 1, *Information technology*.

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

Introduction

There is broad agreement among commercial, academic, and government leaders about the remarkable potential of big data to spark innovation, fuel commerce, and drive progress. big data is the common term used to describe the deluge of data in today's networked, digitized, sensor-laden, and information-driven world. The availability of vast data resources carries the potential to answer questions previously out of reach, including the following:

— How can a potential pandemic reliably be detected early enough to intervene?

— Can new materials with advanced properties be predicted before these materials have ever been synthesized?

— How can the current advantage of the attacker over the defender in guarding against cyber-security threats be reversed?

There is also broad agreement on the ability of big data to overwhelm traditional approaches. The growth rates for data volumes, speeds, and complexity are outpacing scientific and technological advances in data analytics, management, transport, and data user spheres.

Despite widespread agreement on the inherent opportunities and current limitations of big data, a lack of consensus on some important, fundamental questions continues to confuse potential users and stymie progress. These questions include the following:

— What attributes define big data solutions?

— How is big data different from traditional data environments and related applications?

— What are the essential characteristics of big data environments?

— How do these environments integrate with currently deployed architectures?

— What standards are in place to support big data and how does big data affect existing standards?

— What are the central scientific, technological, and standardization challenges that need to be addressed to accelerate the deployment of robust big data solutions?

This document is focused on providing at least some portion of the answers to the last two questions.

Information technology — Big data reference architecture —

Part 5: Standards roadmap

1 Scope

This document describes big data relevant standards, both in existence and under development, along with priorities for future big data standards development based on gap analysis.

2 Normative references

The following documents, in whole or in part, are normatively referenced in this document and are indispensable for its application. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 20546:—[1], *Information technology — Big data — Definition and vocabulary*

3 Terms, definitions and abbreviations

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

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

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

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

3.1 Terms defined elsewhere

3.1.1
big data
extensive datasets — primarily in the characteristics of volume, variety, velocity, and/or variability — that require a scalable architecture for efficient storage, manipulation, and analysis

[SOURCE: ISO/IEC 20546:—[1]]

3.2 Terms defined in this document

3.2.1
standard implementer
component that enables the provision of services based on the standards

Note 1 to entry: For example, a developer who need to comply with SQL commands would be an implementer of that standard.

1) Under preparation. Stage at the time of publication: ISO/IEC DIS 20546:2018.

3.2.2
standard user
person or component that interacts with a service via the standard or that accepts/consumes/decodes data represented by the standard

3.3 Abbreviations

ANSI	American National Standards Institute
AP	Application Provider layer
BDRA	Big Data Reference Architecture
BSI	British Standards Institute
DC	Data Consumer layer
DIN	Deutsches Institut für Normung e.V. (German Institute for Standardization)
DMTF	Distributed Management Task Force, Inc.
DP	Data Provider layer
ISO	International Organization for Standardization
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
INF	Infrastructure Layer
INT	Integration Layer
ITU-T	International Telecommunication Union – Telecommunication standardization sector
JISC	Japanese Industrial Standards Committee
MGT	Management Layer
OASIS	Organization for the Advancement of Structured Information Standards
OGC	Open Geospatial Consortium
OGF	Open Grid Forum
OSS-Association	Open Security Standards Association
PL	Platform Layer
PR	Processing layer
SAC	Standardization Administration of China
S&P	Security and Privacy Layer
SDO	Standards Development Organization
W3C	World Wide Web Consortium

4 Rationale

Identifying and locating relevant standards developed by ISO/IEC and other organizations and determining their applicability to big data and the BDRA (Big Data Reference Architecture) is a continual process. This roadmap provides standard implementers and users the pointers and links to other standards which would apply to or inform their implementation of the BDRA.

5 Relationship to BDRA

The Big Data Reference Architecture (BDRA) specified in ISO/IEC 20547-3 describes multiple viewpoints of the big data paradigm and how those viewpoints fit together. Because the big data paradigm integrates a wide variety of existing technologies, it is useful to identify the standards behind those technologies.

6 Standards development organizations

Big data has generated interest in a wide variety of multi-stakeholder, collaborative organizations, including those involved in the de jure standards process, industry consortia, and open source organizations. These organizations may operate differently and focus on different aspects, but they all have a stake in big data. Integrating additional big data initiatives with ongoing collaborative efforts is a key to success. Identifying which collaborative initiative efforts address architectural requirements and which requirements are not currently being addressed is a starting point for building future multi-stakeholder collaborative efforts. Collaborative initiatives include, but are not limited to the following:

— International Standard development organizations e.g.: ISO, IEC, ITU-T;

— National Standard development organizations e.g: ANSI, BSI, DIN, JISC, SAC;

— Industry consortium e.g.: W3C, OASIS, DMTF;

— others e.g. OSS-Association specification.

Some of the leading SDOs and industry consortia working on big data related standards include:

— International Organization for Standardization (ISO)—de jure standards process;

— Institute of Electrical and Electronics Engineers (IEEE)—de jure standards process;

— International Electrotechnical Commission (IEC);

— Internet Engineering Task Force (IETF);

— World Wide Web Consortium (W3C)—Industry consortium;

— Open Geospatial Consortium (OGC®)—Industry consortium;

— Organization for the Advancement of Structured Information Standards (OASIS)—Industry consortium;

— Open Grid Forum (OGF)—Industry consortium.

NOTE The organizations and initiatives referenced in this document do not form an exhaustive list. It is anticipated that as this document is more widely distributed, more standards efforts addressing additional segments of the big data mosaic will be identified.

There are many government organizations that publish standards relative to their specific problem areas. Many of these are based on other standards (e.g., ISO, IEEE, ANSI) and could be applicable to the big data problem space. However, a fair, comprehensive review of these standards would exceed the available document preparation time and may not be of interest to much of the audience for this document. Readers interested in domains covered by the government organizations and standards, are encouraged to review the standards for applicability to their specific needs.

Open source implementations are providing useful new technology that is being used either directly or as the basis for commercially supported products. These open source implementations are not just individual products. One needs to integrate an eco-system of products to accomplish ones goals. Because of the ecosystem complexity, and because of the difficulty of fairly and exhaustively reviewing open source implementations, such implementations are not included in this section. However, it should be noted that those implementations often evolve to become the de facto reference implementations for many technologies.

7 Existing standards

This section presents a list of existing standards from the above listed organizations that are relevant to big data and the BDRA. Determining the relevance of standards to the big data domain is challenging since almost all standards in some way deal with data. Whether a standard is relevant to big data is generally determined by impact of big data characteristics (e.g. volume, velocity, variety) on the standard or, more generally, by the scalability of the standard to accommodate those characteristics. A standard can also be applicable to big data depending on the extent to which that standard helps to address one or more of the big data characteristics. Finally, a number of standards are also very domain or problem specific and, while they deal with or address big data, they support a very specific functional domain and developing even a marginally comprehensive list of such standards would require a massive undertaking involving subject matter experts in each potential problem domain, which is beyond the scope of this document.

Documents included in Table 1 focus on standards that would do the following:

— facilitate interfaces between BDRA components;

— facilitate the handling of data with one or more big data characteristics;

— represent a fundamental function needing to be implemented by one or more BDRA functional components or activities;

— be commonly available standards which facilitate big data, regardless of the application domain.

Table 1 represents a portion of potentially applicable standards from a portion of contributing organizations working in big data domain.

As most standards represent some form of interface between components, Table 1 is annotated with whether the BDRA component would be an Implementer or User of the standard. The definitions of Standard Implementer and Standard User are provided in Clause 3.

NOTE While the above definitions provide a reasonable baseline, for some standards the difference between implementation and use can be negligible or non-existent for others.

The BDRA functional layers and multilayer functions are abbreviated in the table columns as follows:

— DP = Data Provider layer;

— DC = Data Consumer layer;

— AP = Application Provider layer;

— PR = Processing layer;

— PL = Platform Layer;

— INF = Infrastructure Layer;

— INT = Integration Layer;

— S&P = Security and Privacy Layer;

— MGT = Management Layer.

Please refer to ISO/IEC 20547-3[4] for the complete descriptions of the layers and the names of the types of functional components within the layers.

Within the table, each standard is annotated as to whether that layer would be an Implementer or User of the standard. The definitions of a Standard Implementer and Standard User are provided in Part 3. Standards are ordered by the SDO and industry consortia, and then alphabetically/numerically by standard name/number.

Table 1 — Existing big data standards

Standard Name/ Number	Description	BDRA Functional Layers								
		DP	DC	AP	PR	PL	INF	INT	S&P	MGT
ISO 6709:2008	Standard representation of geographic point location by coordinates	I	U	IU	IU	I				
ISO/IEC 9075-*	ISO/IEC 9075 defines SQL. The scope of SQL is the definition of data structure and the operations on data stored in that structure. ISO/IEC 9075-1, ISO/IEC 9075-2 and ISO/IEC 9075-11 encompass the minimum requirements of the language. Other parts define extensions.	I	IU	U	U	I			U	U
ISO/IEC TR 9789 (Technical Report)	Guidelines for the organization and representation of data elements for data interchange	IU	IU	IU	IU	IU		IU		
ISO/IEC 9798-*	Information technology — Security techniques — Entity authentication	IU	U	U	U	U	IU	U	IU	U
ISO/IEC 10728-*	Information Resource Dictionary System (IRDS) Services Interface			U	I	I		I		I
ISO/IEC 11770-*	Information technology — Security techniques — Key management	IU	U	U	U	U	I	U	IU	U
ISO/IEC 11179-*	The 11179 standard is a multipart standard for the definition and implementation of metadata registries. The series includes the following parts: Part 1: Framework Part 2: Classification Part 3: Registry metamodel and basic attributes Part 4: Formulation of data definitions Part 5: Naming principles Part 6: Registration Part 7: Metamodel for data set registration	I	IU	IU	U	IU			U	
ISO/IEC 13249-*	Database languages – SQL multimedia and application packages Part 1: Framework Part 2: Full-Text Part 3: Spatial Part 5: Still image Part 6: Data mining	I	IU	U	U	I		U		

DP = Data Provider layer; DC = Data Consumer layer; AP = Application Provider layer; PR = Processing layer
PL = Platform Layer; INF = Infrastructure Layer; INT = Integration Layer; S&P = Security and Privacy Layer
MGT = Management Layer
I = Standard implementer; U = Standard user