

I n t e r n a t i o n a l T e l e c o m m u n i c a t i o n U n i o n

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

Series Y
Supplement 40
(07/2016)

SERIES Y: GLOBAL INFORMATION
INFRASTRUCTURE, INTERNET PROTOCOL ASPECTS
AND NEXT-GENERATION NETWORKS, INTERNET OF
THINGS AND SMART CITIES

Big data standardization roadmap

ITU-T Y-series Recommendations – Supplement 40

ITU-T

ITU-T Y-SERIES RECOMMENDATIONS

GLOBAL INFORMATION INFRASTRUCTURE, INTERNET PROTOCOL ASPECTS AND NEXT-GENERATION NETWORKS, INTERNET OF THINGS AND SMART CITIES

GLOBAL INFORMATION INFRASTRUCTURE

General	Y.100–Y.199
Services, applications and middleware	Y.200–Y.299
Network aspects	Y.300–Y.399
Interfaces and protocols	Y.400–Y.499
Numbering, addressing and naming	Y.500–Y.599
Operation, administration and maintenance	Y.600–Y.699
Security	Y.700–Y.799
Performances	Y.800–Y.899

INTERNET PROTOCOL ASPECTS

General	Y.1000–Y.1099
Services and applications	Y.1100–Y.1199
Architecture, access, network capabilities and resource management	Y.1200–Y.1299
Transport	Y.1300–Y.1399
Interworking	Y.1400–Y.1499
Quality of service and network performance	Y.1500–Y.1599
Signalling	Y.1600–Y.1699
Operation, administration and maintenance	Y.1700–Y.1799
Charging	Y.1800–Y.1899
IPTV over NGN	Y.1900–Y.1999

NEXT GENERATION NETWORKS

Frameworks and functional architecture models	Y.2000–Y.2099
Quality of Service and performance	Y.2100–Y.2199
Service aspects: Service capabilities and service architecture	Y.2200–Y.2249
Service aspects: Interoperability of services and networks in NGN	Y.2250–Y.2299
Enhancements to NGN	Y.2300–Y.2399
Network management	Y.2400–Y.2499
Network control architectures and protocols	Y.2500–Y.2599
Packet-based Networks	Y.2600–Y.2699
Security	Y.2700–Y.2799
Generalized mobility	Y.2800–Y.2899
Carrier grade open environment	Y.2900–Y.2999

FUTURE NETWORKS

Y.3000–Y.3499

CLOUD COMPUTING

Y.3500–Y.3999

INTERNET OF THINGS AND SMART CITIES AND COMMUNITIES

General	Y.4000–Y.4049
Definitions and terminologies	Y.4050–Y.4099
Requirements and use cases	Y.4100–Y.4249
Infrastructure, connectivity and networks	Y.4250–Y.4399
Frameworks, architectures and protocols	Y.4400–Y.4549
Services, applications, computation and data processing	Y.4550–Y.4699
Management, control and performance	Y.4700–Y.4799
Identification and security	Y.4800–Y.4899
Evaluation and assessment	Y.4900–Y.4999

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

Supplement 40 to ITU-T Y-series Recommendations

Big data standardization roadmap

Summary

Supplement 40 to ITU-T Y-series Recommendations provides the standardization roadmap for big data in the telecommunication sector. It describes the landscape and conceptual ecosystem of big data from an ITU-T perspective, related technical areas, activities in standards development organizations (SDOs) and gap analysis.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T Y Suppl. 40	2016-07-08	13	11.1002/1000/13022

Keywords

Big data, big data ecosystem, data analytics, roadmap.

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

Compliance with this publication is voluntary. However, the publication may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the publication 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 publication is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this publication 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 publication development process.

As of the date of approval of this publication, ITU had not received notice of intellectual property, protected by patents, which may be required to implement this publication. 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 2016

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	Definitions	1
	3.1 Terms defined elsewhere	1
	3.2 Terms defined in this Supplement	1
4	Abbreviations and acronyms	1
5	Conventions	2
6	Landscape of big data from an ITU-T perspective	2
	6.1 Characteristics and general concepts of big data.....	2
	6.2 Benefits of big data.....	4
7	Related technical areas of big data	4
	7.1 Cloud computing	4
	7.2 Internet of things.....	5
	7.3 Security and privacy	5
	7.4 Software-defined networking	5
	7.5 Deep packet inspection.....	6
	7.6 Big data-driven networking	6
	7.7 Open data.....	6
	7.8 Standardization areas of big data.....	6
8	Conceptual model of big data ecosystem	7
9	Big data SDO activities	8
	9.1 ITU-T.....	8
	9.2 ISO/IEC JTC 1	9
	9.3 W3C.....	10
	9.4 OASIS.....	11
	9.5 Data Mining Group.....	12
	9.6 TM Forum	12
10	Gap analysis in big data standardization	12
	Appendix I – Summaries of referenced standardization work items	15
	I.1 ITU-T references and associated summaries.....	15
	I.2 ISO/IEC JTC 1 References and associated summaries	18
	I.3 W3C references and associated summaries.....	19
	I.4 OASIS references and associated summaries.....	20
	I.5 Data Mining Group references and associated summaries.....	21
	I.6 TM Forum references and associated summaries	21
	Bibliography.....	22

Supplement 40 to ITU-T Y-series Recommendations

Big data standardization roadmap

1 Scope

This Supplement provides the standardization roadmap for big data area in the telecommunication sector. It addresses the following subjects:

- landscape of big data from an ITU-T perspective;
- related technical areas of big data;
- conceptual model of big data ecosystems;
- big data activities in standards development organizations (SDOs);
- standardization gap analysis.

2 References

- [ITU-T Y.2060] Recommendation ITU-T Y.2060 (2013), *Overview of the Internet of things*.
- [ITU-T Y.3300] Recommendation ITU-T Y.3300 (2014), *Framework of software-defined networking*.
- [ITU-T Y.3500] Recommendation ITU-T Y.3500 (2014), *Information technology – Cloud computing – Overview and vocabulary*.
- [ITU-T Y.3600] Recommendation ITU-T Y.3600 (2015), *Big data – Cloud computing based requirements and capabilities*.

3 Definitions

3.1 Terms defined elsewhere

This Supplement uses the following term defined elsewhere:

3.1.1 big data [ITU-T Y.3600]: A paradigm for enabling the collection, storage, management, analysis and visualization, potentially under real-time constraints, of extensive datasets with heterogeneous characteristics.

NOTE – Examples of datasets characteristics include high-volume, high-velocity, high-variety, etc.

3.2 Terms defined in this Supplement

None.

4 Abbreviations and acronyms

This Supplement uses the following abbreviations and acronyms:

AMQP	Advanced Message Queuing Protocol
API	Application Program Interface
BDaaS	Big Data as a Service
BDC	Big Data service Customer
bDDN	big Data-Driven Networking
BDSP	Big Data Service Provider

CSV	Comma-Separated Values
DCAT	Data Catalogue Vocabulary
DMG	Data Ming Group
DPI	Deep Packet Inspection
HTTP	Hypertext Transfer Protocol
ICT	Information and Communications Technology
IEC	International Engineering Consortium
IoT	Internet of Things
ISO	International Organization for Standardization
JSON	Java Script Object Notation
JTC 1	Joint Technical Committee 1
KVDB	Key-Value Database Application Interface
LDP	Linked Data Platform
M2M	Machine to Machine
MQTT	Message Queuing Telemetry Transport
PMML	Predictive Model Markup Language
RDF	Resource Description Framework
SC	Subcommittee
SDN	Software-defined Networking
SDO	Standards Development Organization
SG	Study Group
TC	Technical Committee
URL	Uniform Resource Locator
W3C	World Wide Web Consortium
WG	Working Group
XMILE	XML Interchange Language
XML	Extensible Markup Language

5 Conventions

None.

6 Landscape of big data from an ITU-T perspective

6.1 Characteristics and general concepts of big data

[ITU-T Y.3600] describes the characteristics and general concepts of the big data ecosystem.

With the rapid development of information and communications technology (ICT), Internet technologies and services, huge amount of data are generated, transmitted and stored with explosive growth. Data are generated by many sources and not only sensors, cameras, network devices, web pages, email systems, social networks and many other sources. Datasets are becoming so large and

complex or are arriving so fast that traditional data processing methods and tools are inadequate. Efficient analytics of data within tolerable elapsed times becomes very challenging. The paradigm being developed to resolve the above issues are called big data [ITU-T Y.3600].

Within big data ecosystem, data types include structured, semi-structured and unstructured data. Structured data are often stored in databases which may be organized in different models, such as relational model, document model, key-value model, graph model etc. Semi-structured data does not conform to the formal structure of data models, but contain tags or markers to identify data. Unstructured data do not have a pre-defined data model and are not organized in any defined manner. Within all data types data can exist in formats, such as text, spreadsheet, video, audio, image, map, etc. [ITU-T Y.3600].

Big data is used in many fields, where data processing is characterized by scale (volume), diversity (variety), speed (velocity) and possibly others like credibility (veracity) or business value, if traditional methods and tools are not efficient. These characteristics, usually called v's, can be explained as following [ITU-T Y.3600]:

- **Volume:** refers to the amount of data collected, stored, analyzed and visualized, which big data technologies need to resolve;
- **Variety:** refers to different data types and data formats that are processed by big data technologies;
- **Velocity:** refers to both how fast the data is collected and how fast the data is processed by big data technologies to deliver expected results.

NOTE – Additionally, veracity refers to the uncertainty of data, and value refers to the business results from gaining new information using big data technologies. Other v's can be considered as well.

Taking into account the described above v's characteristics, big data technologies and services can resolve many new challenges, and can also create more new opportunities than ever before [ITU-T Y.3600]:

- **Heterogeneity and incompleteness:** data processed using big data can miss some attributes or introduce noise into data transmission. Even after data cleaning and error correction, some incompleteness and some errors in data are likely to remain. These challenges can be managed during data analysis [b-CRA].
- **Scale:** processing of large and rapidly increasing volumes of data is a challenging task. Using data processing technologies, the data scale challenge is mitigated by evolution of processing and storage resources. However, nowadays data volumes are scaling faster than resources are evolving. Technologies such as parallel databases, in-memory databases, non-SQL databases and analytical algorithms resolve this challenge.
- **Timeliness:** the acquisition rate and timeliness, to effectively find elements in a limited-time period that meet a specified criterion in a large dataset, are new challenges faced by data processing. Other new challenges are related to the types of criteria specified, and need to devise new index structures and responses to the queries having tight response-time limits.
- **Privacy:** data about human individuals, such as: demographic information, Internet activities, commutation patterns, social interactions, energy or water consumption, are being collected and analyzed for different purposes. Big data technologies and services are challenged to protect personal identities and sensitive attributes of data throughout the entire data processing process, while respecting applicable data retention policies.

Positive resolution of the above challenges opens new opportunities to discover new data relationships, hidden patterns or unknown dependencies [ITU-T Y.3600].