

The Road to Big Data Standardisation



Ray Walshe

Abstract This chapter covers the critical topic of standards within the area of big data. Starting with an overview of standardisation as a means for achieving interoperability, the chapter moves on to identify the European Standards Development Organizations that contribute to the European Commission's plan for the Digital Single Market. The author goes on to describe, through use cases, exemplar big data challenges, demonstrates the need for standardisation and finally identifies the critical big data use cases where standards can add value. The chapter provides an overview of the key standardisation activities within the EU and the current status of international standardisation efforts. Finally, the chapter closes with future trends for big data standardisation.

Keywords Standardisation · Strategy · Policy · European Commission · Reference architecture · Use cases · Big data · Future directions

1 Introduction

This chapter starts with an introduction to standardisation and the importance of adopting standardised services and products to effectively drive common services around the world. It identifies big data use cases for the purpose of building reference architecture. These use cases help to gather input and priority requirements more effectively to foster interoperability between legacy and new systems. Next, the chapter describes big data standardisation activities and their adoption at different levels. It discusses the trends in big data standardisation and details future plans that would leverage digital solutions to open up new opportunities and boost development. It explains that big data standards are likely to evolve with further research and

R. Walshe (✉)

ADAPT SFI Centre for Digital Content, Dublin City University, Dublin, Ireland

e-mail: Ray.Walshe@DCU.ie

© The Author(s) 2021

E. Curry et al. (eds.), *The Elements of Big Data Value*,

https://doi.org/10.1007/978-3-030-68176-0_14

333

the development of new technologies, tools and services. Finally, the chapter summarises the path to standardisation.

2 About Standardisation

In everyday life, at work, at play, at rest, we routinely use products, tools, techniques, processes and systems that are designed, tested, deployed, maintained and evolved using agreed global best practice. This agreed global best practice is the core of standardisation. It is what citizens look for when trying to determine product quality, safety, durability and interoperability. If one views standardisation as a critical input to products, services and tools, then quality and confidence are the tangible outputs.

Standards are everywhere and make it possible to carry out everyday activities as they impact our services such as communications, technology, media, healthcare, food, transport, construction and energy. Some standards have stood the test of time, being around for hundreds if not thousands of years (*Through History with Standards* 2020). The Sumerians in the Tigris/Euphrates valley devised a calendar, not very dissimilar to our modern calendar, 5000 years ago. They divided the year into 30-day months and the days into 12 h and each hour into 30 min.

Adopting standards helps ensure regularity, safety, reliability and environmental care. Standardised products and services are perceived as more dependable, raising user confidence, sales and new technology adoption. Standards are used by regulators and legislators for protecting consumer interests and to support government policies. They play a central role in the European Union's policy for a single market. Standards-compliant products and services enable devices to work together, and standardisation provides a solid foundation upon which to develop new technologies and to enhance existing practices. Standards open up market access, provide economies of scale, encourage innovation and increase awareness of technical developments and initiatives.

Standards provide the foundation for a greater variety of new products with new features and options. In a world without standards, products may be dangerous, of inferior quality, incompatible with others, lock in customers to one supplier and lead to manufacturers devising their own standards for every application or product.

The need for international standardisation in the provision of goods and services to consumers should be evident from the above and is also supported by many factual examples of success based on standards development.

The GSM™ mobile communication technology and its successors (3G, 4G) which were led by the European Telecommunications Standards Institute (ETSI) are good examples of standardisation. GSM was originally envisaged as a telecom solution for Europe, but the technologies were quickly adopted and have been deployed worldwide. Thanks to standardisation, international travellers can communicate and use common services anywhere in the world.

2.1 ICT Standardisation and the European Union

The EU supports an effective and coherent standardisation framework, which ensures that standards are developed in a way that supports EU policies and competitiveness in the global market.

Regulations on European standardisation set the legal framework in which the different actors in the standardisation system can operate. These actors are the European Commission, the European Standardization Organizations, industry, small and medium-sized industries (SMEs) and societal stakeholders.

The Commission is empowered to identify information and communications technology (ICT) technical specifications (European Commission 2020a) to be eligible for referencing in public procurement. Public authorities can therefore make use of the full range of specifications when buying IT hardware, software and services, allowing for greater competition and reducing the risk of lock-in to proprietary systems.

The Commission financially supports the work of the three European Standardization Organizations: ETSI, CEN and CENELEC.

2.1.1 ETSI: The European Telecommunications Standards Institute

ETSI, the European Telecommunications Standards Institute, produces globally applicable standards (Dahmen-Lhuissier 2020) for information and communications technologies (ICT), including fixed, mobile, radio, converged, broadcast and Internet technologies. These standards enable the technologies on which business and society rely. The ETSI standards for GSM™, DECT™, smart cards and electronic signatures have helped to revolutionise modern life all over the world.

ETSI is one of the three European Standardization Organizations officially recognised by the European Union and is a not-for-profit organisation with more than 800 member organisations worldwide, drawn from 66 countries and 5 continents. Members include the world's leading companies and innovative R&D organisations.

ETSI is at the forefront of emerging technologies, addressing the technical issues which will drive the economy of the future and improve life for the next generation.

2.1.2 CEN: The European Committee for Standardization

CEN, the European Committee for Standardization (CEN 2020), is an association that brings together the national standardisation bodies of 33 European countries. CEN is also one of three European Standardization Organizations (together with CENELEC and ETSI) that have been officially recognised by the European Union and by the European Free Trade Association (EFTA) as being responsible for developing and defining voluntary standards at European level.

CEN provides a platform for the development of European standards and other technical documents in relation to various kinds of products, materials, services and processes. It supports standardisation activities in relation to a wide range of fields and sectors including air and space, chemicals, construction, consumer products, defence and security, energy, the environment, food and feed, health and safety, healthcare, ICT, machinery, materials, pressure equipment, services, smart living, transport and packaging.

2.1.3 CENELEC: The European Committee for Electrotechnical Standardization

CENELEC is the European Committee for Electrotechnical Standardization (CENELEC 2020) and is responsible for standardisation in the electrotechnical engineering field. It prepares voluntary standards which help facilitate trade between countries, create new markets, cut compliance costs and support the development of a single European market. It creates market access at European level but also at international level, adopting international standards wherever possible, through its close collaboration with the International Electrotechnical Commission (IEC) (CENELEC n.d.), under the Dresden Agreement.

In the global economy, CENELEC fosters innovation and competitiveness, making technology available industry-wide through the production of voluntary standards. Its members, its experts, the industry federations and consumers help create European standards to encourage technological development, to ensure interoperability and to guarantee the safety and health of consumers and provide environmental protection. Designated as a European Standardization Organization by the European Commission, CENELEC is a non-profit technical organisation set up under Belgian law. It was created in 1973 as a result of the merger of two previous European organisations: CENELCOM and CENEL.

EU-funded research and innovation projects also make their results available to the standardisation work of several standards-setting organisations.

2.1.4 The European Multi Stakeholder Platform on ICT Standardisation

The European Multi Stakeholder Platform (MSP) (European Commission 2013a) on ICT standardisation was established in 2011. It advises the Commission on ICT standardisation policy implementation issues, including priority-setting in support of legislation and policies, and the identification of specifications developed by global ICT standards development organisations. The Multi Stakeholder Platform addresses:

- Potential future ICT standardisation needs
- Technical specifications for public procurements

- Cooperation between ICT standards-setting organisations
- A multi-annual overview of the needs for preliminary or complementary ICT standardisation activities in support of the EU policy activities (the Rolling Plan (European Commission 2013b))

The MSP is composed of representatives of national authorities from EU member states and EFTA countries, of the European and international ICT standardisation bodies, and of stakeholder organisations that represent industry, small and medium-sized enterprises and consumers. It meets four times per year and is co-chaired by the European Commission Directorate-General for Internal Market (European Commission 2016), Industry, Entrepreneurship and SMEs and CONNECT (*Communications Networks, Content and Technology*, 2015).

The Platform also Advises on the Elaboration and Implementation of the Rolling Plan on ICT Standardisation (European Commission 2020a)

The Rolling Plan (RP) provides a multi-annual overview of the needs for preliminary or complementary ICT standardisation activities in support of the EU policy activities. It is aimed at the broader ICT community stakeholders and outlines how practically support will be provided. It contains a distinct view of the landscape of standardisation activities in a given policy area.

The Rolling Plan puts standardisation in the policy context, identifies EU policy priorities where standardisation activities are needed, and covers ICT infrastructures and ICT standardisation horizontals. It references legal documents, available standards and technical specifications, as well as ongoing activities in ICT standardisation. The addenda to the Rolling Plan may be published alongside the Rolling Plan in order to keep current with new developments in the rapidly changing ICT sector.

Mission of the Multi Stakeholder Platform on ICT Standardisation (European Commission 2020d)

The Platform is an Advisory Expert Group on all matters related to European ICT standardisation and its effective implementation:

- Advise the Commission on its ICT standardisation work programme.
- Identify potential future ICT standardisation needs.
- Advise the Commission on possible standardisation mandates.
- Advise the Commission on technical specifications in the field of ICT with regard to its referencing in public procurement and policies.
- Advise the Commission on cooperation between standards developing organisations.

The 2016 Rolling Plan on ICT standardisation (European Commission 2020b) [13] covers all activities that can support standardisation and prioritises actions for ICT adoption and interoperability.

The Plan Offers Details on the International Contexts for each Policy

- Societal challenges: e-health, accessibility of ICT products and services, web accessibility, e-skills and e-learning, emergency communications and e-call

- Innovation for the Digital Single Market: e-procurement, e-invoicing, card/Internet and mobile payments, eXtensible Business Reporting Language (XBRL) and Online Dispute Resolution (ODR)
- Sustainable growth: smart grids and smart metering, smart cities, ICT environmental impact, European Electronic Toll Service (EETS) and Intelligent Transport System (ITS)
- Key enablers and security: cloud computing, (open) data, e-government, electronic identification and trust services including e-signatures, radio-frequency identification (RFID), Internet of things (IoT), network and information security (cybersecurity) and e-privacy

This latest Rolling Plan describes all the standardisation activities undertaken by Standard Setting Organizations (SSOs). This ensures an improved coherence between standardisation activities in the EU. This is the first time that the European Standardization Organizations and other stakeholders were involved in drafting the RP, and this improved process is a stronger guarantee that activities of standardisation-supporting EU policies in the ICT domain will be aligned.

3 Identifying Big Data Use Cases

In June 2013, the National Institute of Standards and Technology (NIST) Big Data Public Working Group (NBD-PWG) began forming a community of interested parties from all sectors, including industry, academia and government, to develop a consensus on big data definitions, taxonomies, secure reference architectures, security and privacy requirements, and ultimately a standards roadmap. Part of the work carried out by the working group identified big data use cases in NIST “Big Data Interoperability Framework: Volume 3, Use Cases and General Requirements”, which would serve as exemplars to help develop a Big Data Reference Architecture (BDRA).

The NBD-PWG defined a use case as “a typical application stated at a high level for the purposes of extracting requirements or comparing usages across fields”. They began by collecting use cases from publicly available information for various big data architecture examples. This process returned 51 use cases across nine broad areas (i.e. application domains). This list was not intended to be exhaustive, and other application domains will be considered. Each example of big data architecture constituted one use case. The nine application domains were Government Operation; Commercial; Defence; Healthcare and Life Sciences; Deep Learning and Social Media; Ecosystem for Research; Astronomy and Physics; Earth, Environmental and Polar Science; and lastly Energy.

3.1 Use Case Summaries

The initial focus of the NBD-PWG Use Case and Requirements Subgroup was to form a community of interest from industry, academia and government, with the goal of developing a consensus list of big data requirements across all stakeholders. This included gathering and understanding various use cases from diversified application domains.

The tasks assigned to the subgroup include the following:

- Gather input from all stakeholders regarding big data requirements, a goal that turned into the gathering of use cases.
- Analyse/prioritise a list of challenging general requirements derived from use cases that may delay or prevent the adoption of big data deployment.
- Develop a comprehensive list of big data requirements.

The report was produced by an open collaborative process involving weekly telephone conversations and information exchange using the NIST document system. The 51 use cases came from participants in the calls (subgroup members) and from others informed of the opportunity to contribute. The use cases are organised into nine broad sectors/areas (application domains) listed below with the number of use cases in parentheses and sample examples:

- Government Operation (4): National Archives and Records Administration, Census Bureau
- Commercial (8): Finance in cloud, cloud backup, Mendeley (citations), Netflix, web search, digital materials, cargo shipping (as in UPS)
- Defence (3): Sensors, image surveillance, situation assessment
- Healthcare and Life Sciences (10): Medical records, graph and probabilistic analysis, pathology, bioimaging, genomics, epidemiology, people activity models, biodiversity
- Deep Learning and Social Media (6): Self-driving cars, geolocate images, Twitter, crowd sourcing, network science, NIST benchmark datasets
- Ecosystem for Research (4): Metadata, collaboration, language translation, light source experiments
- Astronomy and Physics (5): Sky surveys (and comparisons to simulation), LHC at CERN, Belle Accelerator II in Japan
- Earth, Environmental and Polar Science (10): Radar scattering in atmosphere, earthquake, ocean, Earth observation, ice sheet radar scattering, Earth radar mapping, climate simulation datasets, atmospheric turbulence identification, sub-surface biogeochemistry (microbes to watersheds), AmeriFlux and FLUXNET gas
- Energy (2): Smart Grid, Home energy management

4 Big Data Standards: The Beginning

Achieving big data goals set out by business and consumers will require the interworking of multiple systems and technologies, legacy and new. Technology integration calls for standards to facilitate interoperability among the components of the big data value chain (Adolph 2013). For instance, UIMA, OWL, PMML, RIF and XBRL are key software standards that support the interoperability of data analytics with a model for unstructured information, ontologies for information models, predictive models, business rules and a format for financial reporting. The standards community has launched several initiatives and working groups on big data. In 2012, the Cloud Security Alliance established a big data working group with the aim of identifying scalable techniques for data-centric security and privacy problems. The group's investigation is expected to clarify best practices for security and privacy in big data and also to guide industry and government in the adoption of those best practices. The US National Institute of Standards and Technology (NIST) kicked off its big data activities with a workshop in June 2012 and a year later launched a public working group. The NIST (NIST 2020) working group intends to support and secure an effective adoption of big data by developing consensus on definitions, taxonomies, secure reference architectures and a technology roadmap for big data analytic techniques and technology infrastructures.

4.1 *NIST Big Data Public Working Group*

The NIST developed a Big Data Interoperability Framework (Grady et al. 2014) which consists of seven volumes, each of which addresses a specific key topic, resulting from the work of the NBD-PWG. The seven volumes are as follows.

4.1.1 **Volume 1, Definitions**

The Definitions volume addresses fundamental concepts needed to understand the new paradigm for data applications, collectively known as big data, and the analytic processes collectively known as data science. Big data has had many definitions and occurs when the scale of the data leads to the need for a cluster of computing and storage resources to provide cost-effective data management. Data science combines various technologies, techniques and theories from various fields, mostly related to computer science and statistics, to obtain actionable knowledge from data.

4.1.2 Volume 2, Taxonomies

Taxonomies were prepared by the NIST Big Data Public Working Group (NBD-PWG) Definitions and Taxonomy Subgroup to facilitate communication and improve understanding across big data stakeholders by describing the functional components of the NIST Big Data Reference Architecture (NBDRA). The top-level roles of the taxonomy are System Orchestrator, Data Provider, Big Data Application Provider, Big Data Framework Provider, Data Consumer, Security and Privacy, and Management. The actors and activities for each of the top-level roles are outlined as well. The NBDRA taxonomy aims to describe new issues in big data systems but is not an exhaustive list. In some cases, the exploration of new big data topics includes current practices and technologies to provide needed context.

4.1.3 Volume 3, Use Cases and General Requirements

The Use Cases and General Requirements document was prepared by the NIST Big Data Public Working Group (NBD-PWG) Use Cases and Requirements Subgroup to gather use cases and extract requirements.

The use cases are, of course, only representative, and do not represent the entire spectrum of big data usage. All of the use cases were openly submitted, and no significant editing was performed. While there are differences in scope and interpretation, the benefits of free and open submission outweighed those of greater uniformity.

4.1.4 Volume 4, Security and Privacy

The Security and Privacy document was prepared by the NIST Big Data Public Working Group (NBD-PWG) Security and Privacy Subgroup to identify security and privacy issues that are specific to big data. Big data application domains include healthcare, drug discovery, insurance, finance, retail and many others from both the private and public sectors. Among the scenarios within these application domains are health exchanges, clinical trials, mergers and acquisitions, device telemetry, targeted marketing and international anti-piracy. Security technology domains include identity, authorisation, audit, network and device security, and federation across trust boundaries.

4.1.5 Volume 5, Architectures White Paper Survey

The Architectures White Paper Survey was prepared by the NIST Big Data Public Working Group (NBD-PWG) Reference Architecture Subgroup to facilitate understanding of the operational intricacies in big data, and to serve as a tool for

developing system-specific architectures using a common reference framework. The Subgroup surveyed published big data platforms by leading companies or individuals supporting the big data framework and analysed the material. This effort revealed a remarkable consistency of big data architecture. The most common themes occurring across the architectures surveyed are outlined below.

- **Big Data Management:** Structured, semi-structured and unstructured data; velocity, variety, volume and variability; SQL and NoSQL; distributed file system
- **Big Data Analytics:** Descriptive, predictive and spatial; real time; interactive; batch analytics; reporting; dashboard
- **Big Data Infrastructure:** In-memory data grids; operational database; analytic database; relational database; flat files; content management system; horizontal scalable architecture

4.1.6 Volume 6, Reference Architecture

The NIST Big Data Public Working Group (NBD-PWG) Reference Architecture Subgroup prepared this NIST Big Data Interoperability Framework: Reference Architecture, to provide a vendor-neutral, technology- and infrastructure-agnostic conceptual model and examine related issues. The conceptual model, referred to as the NIST Big Data Reference Architecture (NBDRA), was crafted by examining publicly available big data architectures representing various approaches and products. Inputs from the other NBD-PWG subgroups were also incorporated into the creation of the NBDRA. It is applicable to a variety of business environments, including tightly integrated enterprise systems, as well as loosely coupled vertical industries that rely on cooperation among independent stakeholders. The NBDRA captures the two known big data economic value chains: information, where value is created by data collection, integration, analysis and applying the results to data-driven services; and the information technology (IT), where value is created by providing networking, infrastructure, platforms and tools in support of vertical data-based applications.

4.1.7 Volume 7, Standards Roadmap

The Standards Roadmap summarises the deliverables of the other NBD-PWG subgroups (presented in detail in the other volumes of this series) and presents the work of the NBD-PWG Technology Roadmap Subgroup. In the first phase of development, the NBD-PWG Technology Roadmap Subgroup investigated existing standards that relate to big data and recognised general categories of gaps in those standards.

4.2 ISO/IEC JTC1's Data Management and Interchange Standards Committee (SC32)

ISO/IEC JTC1's data management and interchange standards committee (SC32) has a study on next-generation analytics and big data (ANSI [UNITED STATES] 2020). The W3C has created several community groups on different aspects of big data.

At the June 2012 SC32 Plenary in Berlin, the SC32 Chair, Jim Melton, appointed an ad hoc committee from all four SC32 working groups: WG1 E-business, WG2 Metadata, WG3 Database Languages and WG4 Multimedia.

The original request from JTC1 referenced a report by the US industry analyst Gartner Group where both "next-generation analytics" and "big data" are identified as strategic technologies.

4.2.1 Next-Generation Analytics

Analytics is growing along three key dimensions:

- From traditional offline analytics to in-line embedded analytics. This has been the focus for many efforts in the past and will continue to be an important focus for analytics.
- From historical data to explain what happened to analysing historical and real-time data from multiple systems to simulate and predict the future.
- Over the next 3 years, analytics will mature along a third dimension, from structured and simple data analysed by individuals to the analysis of complex information of many types (text, video, etc.) from many systems supporting a collaborative decision process that brings multiple people together to analyse, brainstorm and make decisions.

Analytics is also beginning to shift to the cloud and exploit cloud resources for high performance and grid computing.

In 2011 and 2012, analytics increasingly focused on decisions and collaboration. The next step was to provide simulation, prediction, optimisation and other analytics, not simply information, to empower even more decision flexibility at the time and place of every business process action.

4.2.2 Big Data

The size, complexity of formats and speed of delivery exceed the capabilities of traditional data management technologies; the use of new or exotic technologies is required simply to manage the volume alone. Many new technologies are emerging, with the potential to be disruptive (e.g. in-memory Data Base Management System [DBMS]). Analytics has become a major driving application for data warehousing, with the use of MapReduce outside and inside the DBMS, and the use of self-service

data marts. One major implication of big data is that in the future users will not be able to put all useful information into a single data warehouse. Logical data warehouses bringing together information from multiple sources as needed will replace the single data warehouse model.

5 Big Data Standards Work

5.1 IEEE Big Data

Governance and metadata management poses unique challenges with regard to big data paradigm shift. The governance lifecycle needs to be sustainable from creation, maintenance, depreciation, archiving and deletion due to the volume, velocity and variety of big data changes, and can be accumulated whether the data is at rest, in motion or in transactions.

To facilitate and support the Internet of things, smart cities and other emerging technical and market trends, it is critical to have a standard reference architecture for Big Data Governance and Metadata Management (BDGMM) that is scalable and can enable the findability, accessibility, interoperability and reusability between heterogeneous datasets from various sources.

The goal of BDGMM is to enable data integration/mashup among heterogeneous datasets from diversified domain repositories and make data discoverable, accessible and usable through a machine-readable and actionable standard data infrastructure. The IEEE BDGMM was created jointly by the IEEE Big Data Initiative and the IEEE Standards Association.

5.2 ITU-T Big Data

Big data-driven networking (bDDN) and deep packet inspection (DPI): Deep packet inspection is essential for network operators to know the distribution of service/application traffic in the network.

- What enhancements to existing recommendations are needed to enable services/application identification/awareness/visibility and to enable traffic and resource optimisation based on deep packet inspection in future networks (including software-defined networking, network functions virtualisation, Internet of things, information-centric networking/content-centric networking and other candidate future network architecture and technology (e.g. IMT-2020))?

5.3 *ISO/IEC JTC1 WG 9 Big Data Working Group*

Standard ecosystems are required to perform analytics processing regardless of the dataset's needs in relation to the Vs (volume, velocity, variety, etc.) characteristics, underlying computing platforms and how big data analytics tools and techniques are deployed. Unified data platform architecture will support big data strategy across information management, analysis and search technology.

A standard ecosystem provides vendor, technology and infrastructure-agnostic platforms that will enable data scientists and researchers to share and reuse interoperable analytics tools and techniques. WG 9 works with academics, industry, government and various other stakeholders to understand the needs and foster such a standard big data ecosystem.

WG 9 has a three-pronged technical approach to achieve this standard ecosystem:

- Identify standard Big Data Reference Architecture (RA): this approach has already been captured in ISO/IEC 20547 to identify overall RA components and their interface descriptions.
- Identify standard Big Data Reference Architecture Interfaces: this would be a new project to investigate how data flows between RA components and define standard interfaces for such interactions. The goal is to use these validated standard interfaces to build big data applications.
- Identify standard Big Data Management Tools: this would be another new project to investigate how a collection of analytics tools and computing resources can be efficiently and effectively managed to enable standard big data enterprise computing. The goal is to provide system management tools to manage, monitor and fine-tune big data applications.

WG 9 produced the ISO/IEC 20546 (IS) Big Data Overview and Vocabulary committee draft (CD) in March 2016 with balloting results from 9 countries approved as presented, 5 countries approved with comments, 2 countries disapproved with comments and 15 countries choosing abstention. WG 9 spent two teleconferences (15 August and 30 August) reviewing, discussing and resolving all comments, and generated the Disposition of Comments and revised text for further contribution.

WG 9 produced the ISO/IEC 20547-2 Big Data Use Cases and Derived Requirements Provisional Draft Technical Report (51 use cases, 300+ pages) in July 2016 with a 2-month balloting period. All comments are expected to be reviewed, discussed and resolved at the 6th WG 9 November–December 2016 meeting.

For the 4th WG 9 meeting (7 March 2016, Ireland), WG 9 hosted a full-day programme with 16 speakers, 1 panel discussion and over 50 participants. For the 5th WG 9 meeting (11 July 2016, China), a half-day programme with 8 speakers and over 80 participants was conducted. Through outreach effort, and in addition to recruiting more big data experts, new opportunities and expansion of the big data standard foundation technologies such as Big Data Reference Architecture Standard Interface and Big Data Reference Architecture Standard Management were explored.

5.4 *JTC1 SC42: Artificial Intelligence*

5.4.1 Membership

31 Participating Members Australia SA; Austria ASI; Belgium NBN; Canada SCC; China SAC; Congo, the Democratic Republic of the OCC; Denmark DS; Finland SFS; France AFNOR; Germany DIN; India BIS; Ireland NSAI; Israel SII; Italy UNI; Japan JISC; Kenya KEBS; Korea, Republic of KATS; Luxembourg ILNAS; Malta MCCA; the Netherlands NEN; Norway SN; Russian Federation GOST R; Saudi Arabia SASO; Singapore SC; Spain UNE; Sweden SIS; Switzerland SNV; Uganda UNBS; United Arab Emirates ESMA; United Kingdom BSI; United States ANSI.

14 Observing Members Argentina IRAM, Benin ANM, Cyprus CYS, Hong Kong ITCHKSAR, Hungary MSZT, Lithuania LST, Mexico DGN, New Zealand NZSO, Philippines BPS, Poland PKN, Portugal IPQ, Romania ASRO, South Africa SABS, Ukraine DSTU.

5.4.2 Working Groups and Study Groups JTC1 SC42

The ISO/IEC standardisation committee JTC1/SC42 is structured as follows.

Working Group 1	On foundational standards that cope with AI concepts and AI terminology necessary for the full AI lifecycle.
Working Group 2	On big data that aims at vocabulary, framework and reference architecture for big data.
Working Group 3	Deals with requirements for trustworthy and bias-free AI systems that include assessment of the robustness of neural networks.
Working Group 4	Is oriented towards applications and use cases to demonstrate feasibility on AI standards.
Study Group 1	Investigates computational approaches comprising machine learning (ML) algorithms, reasoning approaches, NLP, etc.
Study Group 2	Investigates into aspects of trustworthiness and pitfalls, where the former aspects deal with system properties such as transparency, verifiability, explainability and controllability and the latter aspects deal with robustness, safety, security and privacy system properties.
New work item proposal	A new standardisation project NWIP 24300 is planned and is related to the AI process management for big data analysis (BDA) (ANSI [UNITED STATES] <i>n.d.</i>).

5.4.3 List of Published Standards in JTC1 SC42

Title	Lead editor	Co-editors
ISO/IEC 20546:2019 Information technology — Big data — Overview and vocabulary (Published)	Nancy Grady (USA)	David Boyd (USA)
ISO/IEC TR 20547-1, Information technology – Big Data Reference Architecture -- Part 1: Framework and Application Process	David Boyd (USA)	Su Wook Ha (KR), Ray Walshe (IE)
ISO/IEC TR 20547-2:2018, Information technology – Big Data Reference Architecture -- Part 2: Use Cases and Derived Requirements (Published)	Ray Walshe (IE)	Su Wook Ha (KR)
ISO/IEC 20547-3:2020, Information technology -- Big Data Reference Architecture -- Part3: Reference Architecture (Published)	Ray Walshe (IE)	David Boyd (USA), Liang Guang (CN), Toshihiro Suzuki (JP)
ISO/IEC TR 20547-5:2018(en), Information technology – Big Data Reference Architecture -- Part 5: Standards Roadmap (Published)	David Boyd (USA)	Toshihiro Suzuki (JP), Ray Walshe (IE)
ISO/IEC TR 24028:2020 Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence	Jutta Williams (USA)	

5.4.4 List of Standards in Progress JTC1 SC42

ISO/IEC WD 5059	Software engineering — Systems and software Quality Requirements and Evaluation (SQuARE)	Quality model for AI-based systems
ISO/IEC CD TR 20547-1	Information technology — Big data reference architecture — Part 1:	Framework and application process (submitted for publication)
ISO/IEC CD 22989	Artificial intelligence	Concepts and terminology
ISO/IEC CD 23053	Information Technology	Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)
ISO/IEC CD 23894	Information Technology — Artificial Intelligence	Risk Management
ISO/IEC AWI TR 24027	Information technology — Artificial intelligence (AI)	Bias in AI systems and AI-aided decision making
ISO/IEC CD TR 24029-1	Artificial Intelligence (AI)	Assessment of the robustness of neural networks — Part 1: Overview
ISO/IEC CD TR 24030	Information technology — Artificial Intelligence (AI)	Use cases

(continued)

ISO/IEC AWI TR 24368	Information technology — Artificial intelligence	Overview of ethical and societal concerns
ISO/IEC AWI TR 24372	Information technology — Artificial intelligence (AI)	Overview of computational approaches for AI systems
ISO/IEC AWI 24668	Information technology — Artificial intelligence	Process management framework for big data analytics
ISO/IEC AWI 38507	Information technology — Governance of IT	Governance implications of the use of artificial intelligence by organisations

6 Trends and Future Directions of Big Data Standards

6.1 Public Sector Information, Open Data and Big Data

A key issue for leveraging data value and data value chains in this era of continuously increasing volumes of big data and open data (European Commission 2015) is the need for interoperability. Standardisation at different levels such as metadata, data formats and licensing is essential to enable broad data integration, data exchange and interoperability with the overall goal to foster data-driven innovation. This refers to both structured and unstructured data, as well as data from different domains as diverse as geospatial data, statistical data, weather data, Public Sector Information (PSI) and research data.

On 25 April 2018, the European Commission adopted the “data package” measures to improve the availability and reusability of data (European Commission 2020c), including government data and publicly funded research results, and to foster data sharing in business-to-business (B2B) and business-to-government (B2G) settings. Data availability is crucial to enable companies to leverage the potential of data-driven innovation or develop solutions using artificial intelligence.

The key elements of the Directive on open data and the reuse of public sector information (recast of Directive 2003/98/EC (EUR-Lex 2020a) amended by Directive 2013/37/EU (EUR-Lex 2020b)) are:

- Enhancing access to and reuse of real-time data
- Lowering charges for the reuse of public sector information
- Allowing for the reuse of new types of data, including data resulting from publicly funded research
- Minimising the risk of excessive first-mover advantage in regard to certain data
- “High-value datasets” belonging to six thematic categories (geospatial, Earth observation and environment, meteorological, statistics, companies and company ownership, mobility) to be made available mandatorily free of charge

6.2 European Commission-Funded Standards Projects

Ongoing European projects ELITE-S and StandICT.eu support the training and creation of the next generation of standardisation experts needed for the Digital Single Market.

ELITE-S is a Horizon 2020 Marie Skłodowska-Curie COFUND Action based at the ADAPT Centre at Dublin City University and its Irish academic partners. It is a postdoctoral fellowship programme for intersectoral training, career development and mobility offering 16 prestigious 2-year fellowships in technology and standards development to address five EU priority areas: 5G, Internet of things, cloud computing, cybersecurity and data technologies. Experienced researchers from any country enhance their qualifications and diversify their competencies by conducting a research project at a host institution in Ireland in any of the current research and technology application areas of the programme.

StandICT.eu, “Supporting European Experts Presence in International Standardisation Activities in ICT”, addresses the need for ICT standardisation and defines a pragmatic approach and streamlined process to reinforce EU expert presence in the international ICT standardisation scene. Through a Standards Watch, it analyses and monitors the international ICT standards landscape and liaise with Standards Development Organizations (SDOs) and Standard Setting Organizations (SSOs), key organisations such as the EU Multi Stakeholder Platform for ICT standardisation, as well as industry-led groups, to pinpoint gaps and priorities matching EU Digital Single Market objectives. It provides support for European specialists:

- To contribute to ongoing standards development activities and attend SDO and SSO meetings
- To support the prioritisation of standardisation activities and build a community of standardisation experts
- To support knowledge exchange and collaboration and reinforce European presence in international ICT standardisation

6.3 The Big Data Value Association (BDVA)

The Big Data Value Association (BDVA) is a private, industry-led non-profit association with the mission of boosting European big data value research, development and innovation and fostering a positive perception of big data value. The aim is to maximise the economic and societal benefit to Europe, its businesses and its citizens, enabling Europe to take the lead in the global data-driven digital economy (Zillner et al. 2017).

BDVA membership is composed of large industries, SMEs and research organisations to support the development and deployment of the EU Big Data Value Public-Private Partnership with the European Commission representing the private

side. The BDVA organises its work in Task Forces, where its members engage and influence, and it aims to be the European big data reference point.

The BDVA is open to new members to further enrich the data value ecosystem and play an active role. These include data users, data providers, data technology providers and researchers. Membership of the Association gives the following benefits:

- Part of the European big data industry initiative which will have a high impact on the deployment of big data technologies and thus business competitiveness and economic growth
- Influencing big data challenges and needs in the following years by contributing to the Strategic Research and Innovation Agenda (SRIA)
- Direct access to discussions with EU Commission and member states, thus gaining access to and influencing strategic direction
- Networking and partnering with industrial and research partners in the European data value chain, to set up collaborative research and innovation activities

6.4 European Commission Standardisation Ongoing Activities

The success of Europe's digital transformation (European Commission 2020f) will depend on tools, techniques, services and platforms to ensure trustworthy technologies and to give businesses the confidence and means to digitise. The Data Strategy (European Commission 2020e) and the White Paper on Artificial Intelligence (European Commission 2020g) published by the European Commission endeavour to put people first in developing technology, while continuing to defend and promote European values and rights in the design, development and deployment of technology in the real economy.

The European strategy for data aims to ensure Europe's global competitiveness and data sovereignty by creating a Digital Single Market for data. Common European data spaces will ensure that more data becomes available for use in the economy and society, while keeping companies and individuals who generate the data in control.

Data is an essential resource for economic growth, competitiveness, innovation, job creation and societal progress in general. Standardisation and its impact on the economy has already been well documented (Jakobs 2017) (Blind et al. 2012). Citizens will benefit from these data-driven applications through improved health care, safer and cleaner transport systems, new products and services, reduced costs of public services, and improved sustainability and energy efficiency.

Data availability will drive innovation and necessitate practical, fair and clear rules on data access and use, which comply with European values and rules such as personal data protection.

To ensure the EU's leadership in the global data economy, this European strategy for data intends to:

- Adopt legislative measures on data governance, access and reuse
- Open up high-value publicly held datasets across the EU for free
- Invest €2 billion in a European high-impact project to develop data processing infrastructures, data sharing tools, architectures and governance mechanisms for thriving data sharing and to federate energy-efficient and trustworthy cloud infrastructures and related services
- Enable access to secure, fair and competitive cloud services
- Empower users to stay in control of their data and invest in capacity building for small and medium-sized enterprises and digital skills
- Foster the roll-out of common European data spaces in crucial sectors such as industrial manufacturing, green deal, mobility and health

As part of data strategy, the European Commission has published a report on business-to-government (B2G) data sharing. The report, which comes from a high-level Expert Group (European Commission 2018), contains a set of policy, legal and funding recommendations that will contribute to making B2G data sharing in the public interest a scalable, responsible and sustainable practice in the EU.

6.5 Open Consultation AI White Paper and Data Strategy

The European Commission has adopted a new digital strategy for a European society powered by digital solutions that puts people first, opens up new opportunities for businesses and boosts the development of trustworthy technology. The Commission also presented a White Paper on Artificial Intelligence setting out its proposals to promote the development of AI in Europe whilst ensuring respect of fundamental rights.

Commission President Ursula von der Leyen stated: “Today we are presenting our ambition to shape Europe’s digital future. It covers everything from cybersecurity to critical infrastructures, digital education to skills, democracy to media. I want that digital Europe reflects the best of Europe – open, fair, diverse, democratic and confident”.

The Commission published on 15th December 2020 the proposal for a Regulation on a Single Market For Digital Services (Digital Services Act) and on 3rd December 2020 its European Democracy Action Plan to empower citizens and build more resilient democracies across the EU. The Regulation on electronic identification and trust services for electronic transactions in the internal market (eIDAS Regulation) allows use of national electronic identification schemes (eIDs) to access public services available online in other EU countries. The EU aims to enhance cyber defence cooperation and cyber defence capabilities, building on the work of the European Defence Agency. Europe will also continue to build alliances with global partners, leveraging its regulatory power, capacity building, diplomacy and finance to promote the European digitalisation model.

The White Paper on Artificial Intelligence was open for public consultation until 19 May 2020. The Commission is also gathering feedback on its data strategy. Using the feedback received, the Commission will take further action to support the development of trustworthy AI and the data economy.

7 Future (Big) Data Standardisation Actions

Standards are living documents. They coevolve with technology and, as such, go through similar phases. ICTs, tools and services go through innovation cycles with ideation, research and development, standardisation and disruption. Standards documents go through ideation, consensus building, publication and obsolescence where in many cases obsolescence is a step change where a new technology will replace existing standards. (Big) Data-related technological changes are on the horizon for the short to medium term as we come to terms with the expected 463 GB/day of digital data by 2025. Future standards work in JTC1 includes the following.

7.1 ISO/IEC JTC1: Data Usage Advisory Group—AG9

- **Frameworks for Data Sharing Agreements:** To address the intersection of the value chain and data sharing.
- **Decision to Share Issue:** Where transformation of digital services requires data to be shared, exchanged or exploited to deliver benefits and value, and needs to determine on what basis the decision to use data should be authorised.
- **Data Quality:** Data quality is an important element of data usage. Further work is needed to determine if JTC1 data usage needs are met.
- **Appropriate Use of Analytics Outputs:** Whilst restrictions to data use are often cited as concerns related to privacy, many of the concerns relate to unintended consequences of the use of data.
- **Terminology and Use Cases:** Data use is relevant to many JTC1 standards. Standardised terminology and harmonised use cases are needed for wider data usage and to unlock the value of data sharing, exchange and exploitation.
- **Metadata:** AG9 recognised the importance of metadata definition and use, especially to facilitate the utility to underpin data usage, kinds of metadata, models of metadata and metamodels of repositories.

7.2 ISO/IEC JTC1 SC42 AI WG2 Data

SC42 WG2 Data is investigating the following data topics related to data, data analytics and machine learning:

- Data Quality (DQ)
- Data Quality: Overview, Terminology and Examples
- Data Quality: Measurement
- Data Quality: Management Requirements and Guidelines
- Data Quality: Process Framework
- Data Quality: Assurance – potential new part
- Data Quality: Governance – potential new part
- Big Data: Data Analytics – leverage 20547-3 Big Data Reference Architecture
- Big Data: Data Governance, Usage, Curation, Contextualisation
- Data Mining: Management

8 Summary

This chapter has outlined the case for standardisation, the path to big data standardisation and exemplar activities ongoing in big data standards ecosystems. Projects completed and under way nationally, within European and global initiatives, have been mentioned and sample big data use case scenarios are listed, and some of the initiatives in the evolution of big data standards are described.

The digital ecosystems are global and do not stop at state or regional boundaries. Standardisation is the glue that holds the digital ecosystems together, the gravity of the digital universe. Standardisation in data is central to cloud, big data, IoT, AI and smart city technologies. ISO/IEC JTC1 committees are developing such standards on AI and data, data usage and data interoperability. Standardisation is the foundation stone of certification, regulation and legislation, and in this global digital age, in order to achieve digital sovereignty, we need to synergise the relationships between digital standardisation, digital innovation and digital research.

Acknowledgements This chapter is supported in part by the ADAPT SFI Centre for Digital Content Technology, which is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

References

- Adolph, M. (2013). *Big data: Big today, normal tomorrow*. ITU-T Technology Watch Report 2013.
- ANSI [UNITED STATES]. (2020). *ISO/IEC JTC 1/SC 32 – Data management and interchange*.
- ANSI [UNITED STATES]. (n.d.). *ISO/IEC JTC1/SC42 WG2 N1504 NWIP 24300 IT AI process management framework for big data analysis*.

- Blind, K., Jungmittag, A., & Mangelsdorf, A. (2012). *The economic benefits of standardisation. An update of the study carried out by DIN in 2000.*
- CEN. (2020). *CEN: Who we are.*
- CENELEC. (2020). *Welcome to CENELEC – European committee for electrotechnical standardization.*
- CENELEC. (n.d.). *CENELEC – about CENELEC – Who we are – Global partners.*
- Communications Networks, Content and Technology. (2015).
- Dahmen-Lhuissier, S. (2020). ETSI.
- EUR-Lex. (2020a). EUR-Lex - 32003L0098 - EN - EUR-Lex.
- EUR-Lex. (2020b). EUR-Lex - 32013L0037 - EN - EUR-Lex.
- European Commission. (2013a). *European multi stakeholder platform on ICT standardisation – shaping Europe’s digital future.*
- European Commission. (2013b). *Rolling plan for ICT standardisation – shaping Europe’s digital future.*
- European Commission. (2015). *The 2016 rolling plan for ICT standardisation is released – internal market, industry, entrepreneurship and SMEs.*
- European Commission. (2016). *About us – internal market, industry, entrepreneurship and SMEs.*
- European Commission. (2018). *High-level expert group on artificial intelligence – shaping Europe’s digital future.*
- European Commission. (2020a). *Advanced technologies – Internal market, industry, entrepreneurship and SMEs.*
- European Commission. (2020b). *DocsRoom.*
- European Commission. (2020c). *Public sector information, open data and big data.*
- European Commission. (2020d). *Register of Commission expert groups and other similar entities.*
- European Commission. (2020e). *The European data strategy.*
- European Commission. (2020f). *The European digital strategy – Shaping Europe’s digital future.*
- European Commission. (2020g). *White Paper on Artificial Intelligence: a European approach to excellence and trust.*
- Grady, N. W., Underwood, M., Roy, A., & Chang, W. L. (2014). Big data: Challenges, practices and technologies: NIST Big Data Public Working Group workshop at IEEE Big Data 2014. 2014 IEEE International Conference on Big Data (Big Data), 11–15. <https://doi.org/10.1109/BigData.2014.7004470>
- Jakobs, K. (2017). Two dimensions of success in ICT standardization – A review. *ICT Express*, 3, 85–89. <https://doi.org/10.1016/j.ict.2017.05.008>
- NIST. (2020). *Big data information.*
- Through History with Standards. (2020).
- Zillner, S., Curry, E., Metzger, A., Auer, S., & Seidl, R. (Eds.). (2017). *European big data value strategic research & innovation agenda.* Retrieved from Big Data Value Association website: www.bdva.eu.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

