BS 10102-1:2020

BSI Standards Publication

Big data

Part 1: Guidance on data-driven organizations

bsi.

Publishing and copyright information

The BSI copyright notice displayed in this document indicates when the document was last issued.

Published by BSI Standards Limited 2020

ISBN 978 0 539 04395 2

ICS 35.240.01

The following BSI references relate to the work on this document:
Committee reference MBD/1
Draft for comment 19/30393699 DC

Amendments/corrigenda issued since publication

Date	Text affected

Contents

Summary of pages
This document comprises a front cover, and inside front cover, pages i to iv, pages 1 to 26, an inside back cover and a back cover.

Foreword

Publishing information

This part of BS 10102 is published by BSI Standards Limited, under licence from The British Standards Institution, and came into effect on 29 February 2020. It was prepared by Technical Committee MBD/1, *Big Data*. A list of organizations represented on this committee can be obtained on request to its secretary.

Relationship with other publications

BS 10102, *Big data*, comprises the following parts:

- *Part 1: Guidance on data-driven organizations*; and
- *Part 2: Guidance on data-intensive projects*.

These documents explain how an organization can put in place the necessary governance and methodologies to maximize the benefits of big data and run successful data-intensive projects.

Information about this document

This publication can be withdrawn, revised, partially superseded or superseded. Information regarding the status of this publication can be found in the Standards Catalogue on the BSI website at bsigroup.com/standards, or by contacting the Customer Services team.

Where websites and webpages have been cited, they are provided for ease of reference and are correct at the time of publication. The location of a webpage or website, or its contents, cannot be guaranteed.

Use of this document

As a guide, this British Standard takes the form of guidance and recommendations. It should not be quoted as if it were a specification or a code of practice and claims of compliance cannot be made to it.

This document is intended to be read in conjunction with BS 10102-2.

Presentational conventions

The guidance in this standard is presented in roman (i.e. upright) type. Any recommendations are expressed in sentences in which the principal auxiliary verb is “should”.

Commentary, explanation and general informative material is presented in smaller italic type, and does not constitute a normative element.

Where words have alternative spellings, the preferred spelling of the Shorter Oxford English Dictionary is used (e.g. “organization” rather than “organisation”).

Contractual and legal considerations

This publication does not purport to include all the necessary provisions of a contract. Users are responsible for its correct application.

Compliance with a British Standard cannot confer immunity from legal obligations.

In particular, attention is drawn to the following specific regulations:

- The General Data Protection Regulation (GDPR) (EU) 2016/679 [1];
- The Data Protection Act 2018 [2];

- Directive 2002/58/EC on privacy and electronic communications [3];
- Directive 96/9/EC of the European Parliament and the Council of 11 March 1996 on the legal protection of databases [4];
- The Charter of Fundamental Rights of the European Union (2000) [5];
- Directive (EU) 2016/1148 concerning measures for a high common level of security of network and information systems across the Union [6];
- Directive 93/13/EEC of 5 April 1993 on unfair terms in consumer contracts [7]; and
- The Consumer Rights Act 2015 [8].

Introduction

This British Standard promotes good practice for all organizations processing and exploiting data. Depending on the context or setting of the organization, this includes data which can be viewed as "big data". Thus, while the text throughout refers to "data" generically, this includes "big data".

The term "big data" and the understanding of this continue to evolve. For the purposes of this document big data refers to single, disparate or integrated data sets potentially requiring scalable systems and technologies for efficient storage, manipulation, management and analysis for transformation into value. This standard recognizes that what is "big" for one organization is not necessarily so for another, and that technology and context are likely to change frequently, so any precise definition is likely to require frequent review.

Data can exhibit different characteristics and different levels of structure, each of which contributes to improving the insights from data. When planning, discussing, assessing or exploiting big data, seven descriptive characteristics can be used which begin with the letter 'V'. In general order of relevance, they are:

- **Volume**: awareness and understanding of the extensive amount of data available for analysis to extract valuable information;
- **Velocity**: the rate of flow at which the data are created, stored, analysed or visualized;
- **Variety**: data from various domains and data types that can be structured, semi-structured or unstructured – a wider range of data formats, logical models, timescales, and semantics complicates the integration of data;
- **Variability**: changes in data rate, format/structure, semantics and/or quality that impact the supported application, analytic or problem – impacts can include the need to refactor architectures, interfaces, processing/algorithms, integration/fusion, storage, applicability, or use of the data;
- **Volatility**: the time span in which data values remain relevant for a particular analysis, expressed as a rate of change over time – in real-time data analytics situations, it might be critical to operate on data immediately for decision-making;
- **Veracity**: the completeness and accuracy of the data which will tend to have a direct impact on the effectiveness of data exploitation; and
- **Validity**: how well data meet expected or specified data formats and patterns for a specific intended use.

Data are essential for organizational decision-making and provide ongoing benefit to an organization. Data include raw data, spreadsheets, photographs, documents and drawings, whether stored electronically or on physical media, such as paper records.

Until the middle of the 20th Century, much data collection was human intensive, and the associated storage was in physical records, such as punched cards or paper files. Those wishing to access data generally had to go to the records, rather than having the records come to them. With the widespread deployment of information and communications technology plus associated developments of real-time sensing technologies, the volume of data has burgeoned, though data quality has not always kept pace. In addition, data sets can now be copied or shared at almost zero cost and rapidly transferred worldwide.

The volume and importance of data to society and organizations will continue to grow exponentially. Organizations need to recognize this and adopt current good practice, maintaining awareness of

emerging technologies, applications [e.g. artificial intelligence (AI) and machine learning] and methodologies, and instigating changes as required.

For a typical organization, many separate systems currently create, process and store data of a range of classes, from personnel records to future product or service designs. Often these have grown to meet immediate needs rather than as part of an overarching strategy. Organizations commonly have little understanding of the totality of the data they hold and whether these are genuinely important to operation or would best be securely discarded. Data storage ranges from uncontrolled individual spreadsheets or paper files through to major centralized databases.

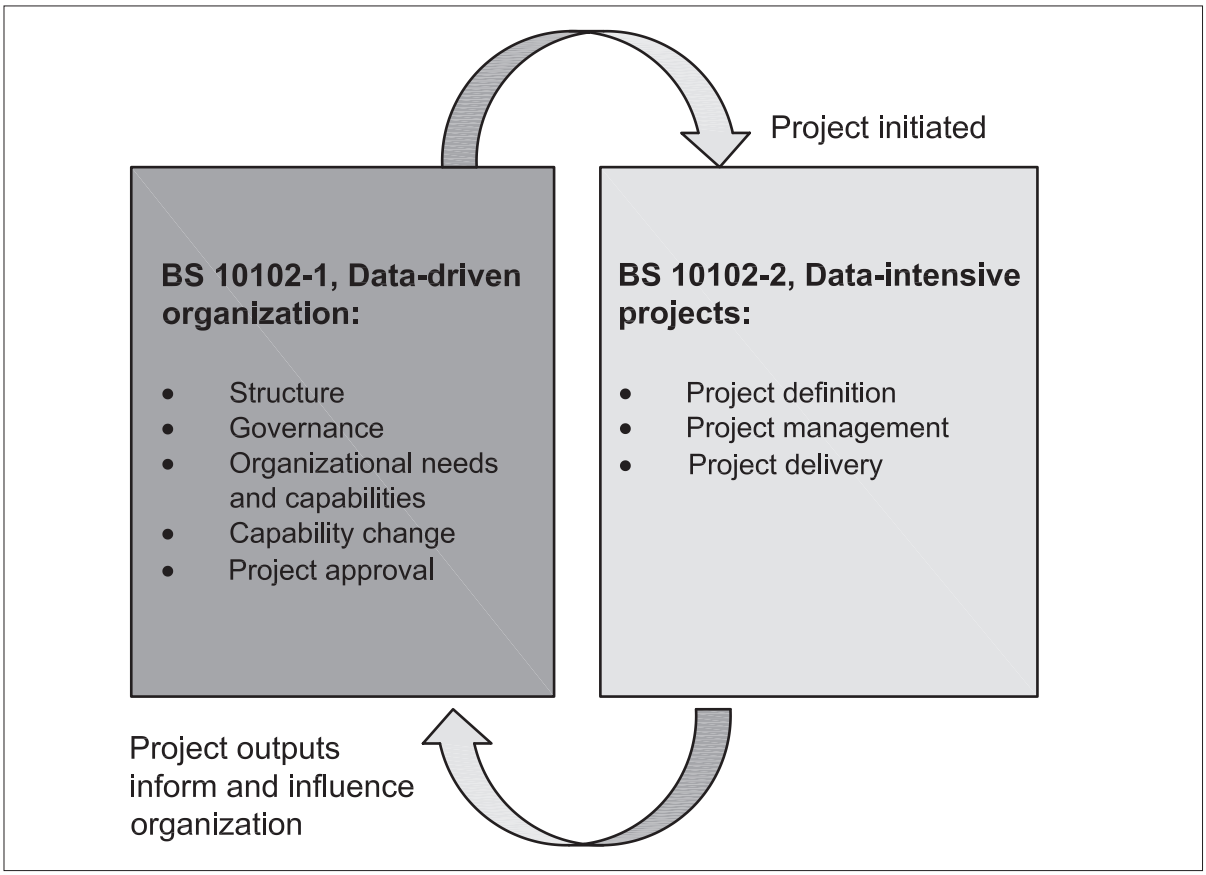
As the importance of data has come to be more fully recognized (and monetized) and the ability to link across different data sets has provided a value multiplier, the need in larger organizations for properly defined and enforced data architecture has become clear. For example, access to real-time, through-life performance data is turning former products into more profitable services. Conversely, there is increasing recognition that ineffective data exploitation introduces inefficiencies and poor decision-making to organizations.

Key issues in the management of data now include:

- data standards;
- metadata standards;
- understanding of the structure of data;
- provenance;
- timeliness;
- repeatability (potential availability of regular time series);
- quality (including accuracy, validity, consistency, completeness and uniqueness);
- potential linkage to other data sets;
- data migration and master data management;
- access controls and security of storage;
- audit trails;
- governance (including legislative and contractual compliance);
- security (including cybersecurity, minimizing hostile reconnaissance, protecting intellectual property and commercially sensitive data);
- purpose specification limits of personal information processing; and
- curation (including custodianship and long-term access).

BS 10102 has been developed in two parts to provide guidance for a data-driven organization and how data-intensive projects are initiated by the organization, with project outputs informing and influencing the organization. These two documents are intended to be used together. Figure 1 illustrates the relationship between these two parts.

Figure 1 — *Relationship between parts of BS 10102*



1 Scope

This part of BS 10102 gives guidance on realizing value from data, including big data, such as gaining insights, informing strategies, enhancing reputation, and improving compliance, efficiency and performance.

To achieve this, guidance is given on managing data efficiently, responsibly and sustainably to ensure appropriate access to data sets of the requisite quality.

This part of BS 10102 is intended for organizations forming or developing their data strategy, as well as those seeking to improve their current data exploitation and data arrangements.

It is applicable to public, private, non-profit and not-for-profit organizations of all sizes.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

For the purposes of this British Standard, the following terms and definitions apply.

3.1 anonymization

process that removes the association between the identifying data set and the data subject

[SOURCE: DD ISO/TS 25237:2008[1)], **3.2**]

[1)] Withdrawn.

