

Enabling the development of inclusive standards – Understanding the role of data and data analysis – Guide

August 2024 Version 3



BSI Flex 236 v3.0:2024-08



Office for Product
Safety & Standards



Publishing and copyright information

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

© The British Standards Institution 2024
Published by BSI Standards Limited 2024

ISBN 978 0 539 29571 9

ICS 01.120; 03.100; 03.120.30; 35.040.50

No copying without BSI permission except as permitted by copyright law.

BSI Flex standards provide a new, flexible way to develop consensus-based good practice that dynamically adapts to keep pace with fast-changing markets. For more information, go to: <https://www.bsigroup.com/en-GB/capabilities/standards-development>

Release history

First version January 2022
Second version June 2023
Third version August 2024

Contents

Foreword ii

0 Introduction..... iv

1 Scope..... 1

2 Normative references 1

3 Terms and definitions 2

4 Data use in standards development..... 6

4.1 Purpose and scope of a standards development project..... 6

4.2 Stakeholder participation..... 7

4.3 Requirements of data for developing standards 8

4.4 Identifying implications of data use in standards development..... 12

4.5 Assessment methods for identifying data-related implications..... 14

4.6 Mitigating unwanted implications of data use in standards development 16

5 Good practice for data use in standards development..... 18

5.1 Data governance..... 18

5.2 Transparency..... 19

5.3 Accountability 19

Bibliography 21

List of figures

Figure 1 – Standards development process overview vi

Foreword

This BSI Flex was sponsored by the Office for Product Safety and Standards, part of the UK Department for Business and Trade. Its development was facilitated by BSI Standards Limited and it was released under licence from The British Standards Institution. It came into effect on 31 August 2024.

Acknowledgement is given to Sania Haq of Bye 2 Stereotypes as the technical author, and the following organizations that were involved in the development of this BSI Flex as members of the Advisory Group:

- Engineering Design Centre, University of Cambridge
- ISO Working Group ISO/TC 184/SC4/WG13, Industrial data quality
- Sports Technology Institute, Loughborough University
- Market Research Society
- Nottingham University Business School
- Towards Vision
- University of Nottingham
- Women's Engineering Society

Acknowledgement is also given to the identified broader stakeholders who were consulted in the development of this version of the BSI Flex.

The British Standards Institution retains ownership and copyright of this BSI Flex. BSI Standards Limited, as the publisher of the BSI Flex, reserves the right to withdraw or amend this BSI Flex on receipt of authoritative advice that it is appropriate to do so.

This BSI Flex is not to be regarded as a PAS or British Standard.

The BSI Flex process enables a standard to be rapidly developed, on an iterative basis, in order to fulfil an immediate stakeholder need. A BSI Flex can be considered for further development as a PAS or British Standard, or constitute part of the UK input into the development of a European or international standard.

The content in this version is part of an iterative process. It is likely to change from time to time with subsequent iterations.

Supersession

This version of BSI Flex 236 supersedes BSI Flex 236 v2.0:2023-06, which is withdrawn.

Information about this document

This is Version 3 of BSI Flex 236, which is the last planned version of the document. All comments received will be reviewed to inform the future direction of this document.

This is a full update of the document and introduces the following principal changes:

- expansion of 4.3 on types of data and data analysis that can be used in standards development;
- expansion of 4.4.2 on the understanding of bias; and
- expansion of 4.4.5 on marginalized, excluded or vulnerable groups.

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 knowledge.bsigroup.com, 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.

Copyright is claimed in terms 3.1, 3.2, 3.5, 3.6, 3.11, 3.13, 3.17, 3.18, 3.19, 3.20, 3.21, 3.23, 3.24 and 3.26.

The content has been reproduced with the permission of the International Organization for Standardization (ISO). Copyright remains with ISO. All rights reserved.

Use of this document

As a guide, this BSI Flex takes the form of guidance and advisory recommendations. It is not to be quoted as if it were a specification or a code of practice.

Presentational conventions

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

Additional commentary, explanation and general informative material is presented in smaller italic type.

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 has been prepared in good faith, however no representation, warranty, assurance or undertaking (express or implied) is or will be made, and no responsibility or liability is or will be accepted by BSI in relation to the adequacy, accuracy, completeness or reasonableness of this publication. All and any such responsibility and liability is expressly disclaimed to the full extent permitted by the law.

This publication is provided as is, and is to be used at the recipient's own risk.

The recipient is advised to consider seeking professional guidance with respect to its use of this publication.

This publication is not intended to constitute a contract. Users are responsible for its correct application.

Compliance with a BSI Flex cannot confer immunity from legal obligations.

0 Introduction

0.1 Background

Standards and the data that inform them have a profound impact on the lives of individual people and groups. Data can be analogue or digital and can be numerical, text, images and/or audio. The data are the result of processing or lived experience.

To create more inclusive standards that better reflect the shared values of humankind, it is necessary to understand:

- the linkages and impacts between human interpretation, data, code and systems;
- the role of data and data analysis;
- the source, scope and purpose of data used; and
- any biases that the data contain.

The first steps in this understanding begin with questions around why, when and how data are used. Who defines the problem that the data are intended to solve? Who decides what data to collect and/or use? Who collects the data and how? Who analyses and questions the data? Who uses the data and for what purposes? Indeed, data are an integral part of almost every step of digitalized processes.

Data are embedded everywhere in the digitally enabled economy, and are the foundation for analogue protocols and guidelines, as well as newer machine learning models ranging from automated decision making to neural networks. However, the data upon which this new economy is based are not neutral, nor do data stand alone as “reality”. Much or most data are historical data, which have the context of having been gathered on small homogeneous subsets of the global population or by non-representative methods. By reflecting only subsets of the human population, historical data are socially and technically biased. They also do not reflect changes in social and cultural norms that occur over time.

The nature of data, inherent limitations within data and the way in which data are collected and interpreted can introduce bias that leads to unfair or non-inclusive outcomes. Bias can occur either intentionally or unintentionally. Scarcity of useful representative data is a major issue, having been found to compromise the quality of health information available to women, for example, as well as the healthcare they receive. Critical healthcare provision and safety are made precarious by the consistent lack of representative data, especially for those from marginalized groups.

Older, unrepresentative data form the basis of many guidelines that continue to drive data decisions, from metabolic rates (see *Nature* article [1]), airplane cockpit design (see *Science, Technology, & Human Values* article [2]), car safety (see *Discover Magazine* article [3]) and radiology protocols (see *Report of the Task Group on Reference Man* [4]), while newer and larger, yet also unrepresentative benchmark data are used in the accelerating world of algorithmic decision making and artificial intelligence (AI). These decisions range from low- to high-risk applications with the ability to cause harm, from transport and trade, to finance, health, medicine and criminal justice.

Data are fundamental to processes and technologies such as AI, robotics and machine learning, the outputs of which impact the day-to-day lives of individuals all over the world. Part of AI's utility is in the machine's ability to perceive patterns in data, and then derive a set of rules from those patterns. However, often the original data exclude all but a small, generally homogenous (e.g. lacking sufficient data across characteristics such as age, ethnicity, sex, education, and sexual orientation) data set; therefore, the AI perceives that data to be the only reality that exists (because it is the only reality the machine has been exposed to and trained on). This exclusion at scale has dangerous consequences for the entire population. While this is of great concern for high-risk applications, low-risk applications can also have a profound social impact on the quality of life and well-being of individuals, including forms of access to information from targeted advertisements, search algorithms and financial access.

Even when attempts have been made to debias, various problems can occur. For example, identical machine learning models behave differently from one another in real-world situations. Debaised data are also interpreted differently in different human decision-making processes. This is not exclusive to machines. Who or what does the debiasing, as well as why, influences the outcome and how inclusive this is.

A growing body of research has shown how the failure to use representative or inclusive data has had negative consequences for the development of hiring algorithms, university allocation, credit scoring, medical decision making, bail decisions and facial recognition. Similarly, data can affect the standards being developed, from the selection, assessment and use of incomplete data to create provisions and requirements, to the unquestioned use of received wisdom, unchallenged assumptions and unconscious bias. AI technologies that build on such data used for standards are already widely used in real-world applications. Given the velocity of AI uptake and the risk of using non-inclusive, incomplete, discriminatory and unrepresentative data sets, there is a risk of bias being replicated at scale and embedded in the systems of our daily lives and governance.

In addition, our notion of what is representative is itself evolving, such that data always lag behind our evolutions in understanding and application. Furthermore, the diversity and understanding of organizations that develop standards could also influence whether a standard is inclusive or even considered for development.

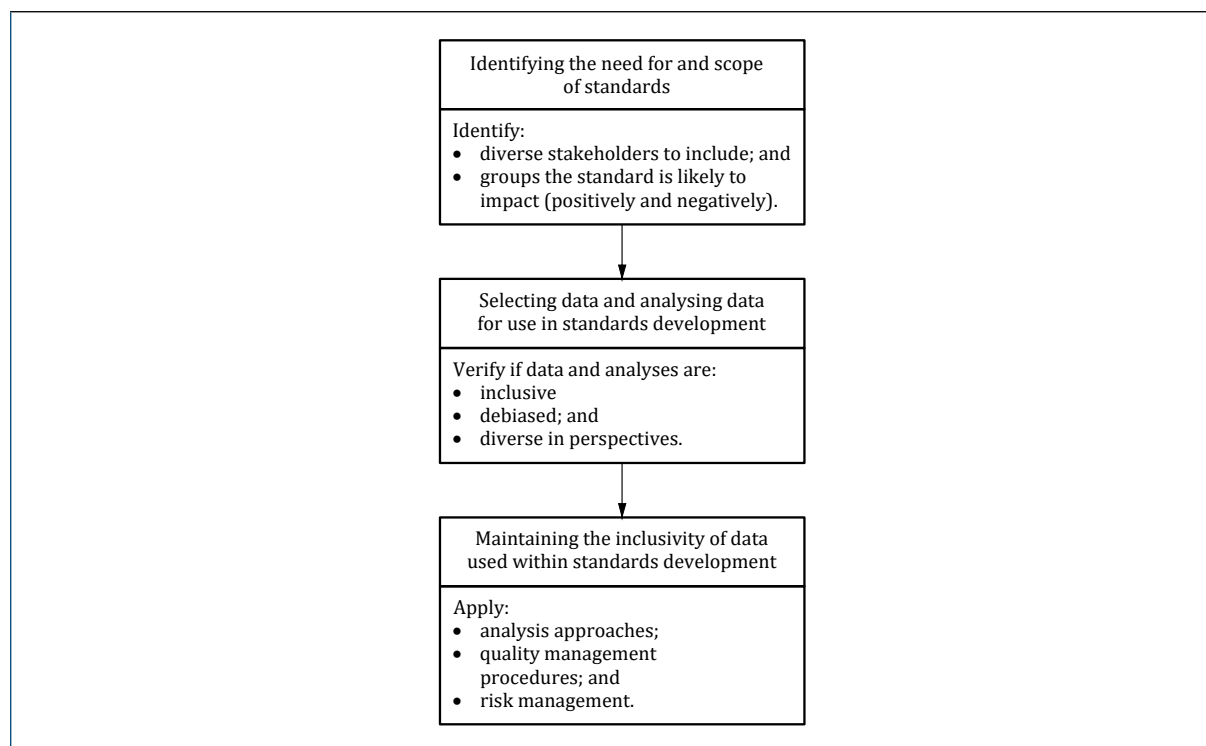
Good practices relating to data in the development and implementation of standards alone are not enough – they need to be embedded in broader good practice, including shared human judgement on how to achieve societal objectives, such as inclusion and equality.

0.2 Aims and objectives

By using inclusive data in standards development, there is an opportunity to mitigate and correct for historic inequities buried in previous data. The opportunity is to revisit old assumptions when new systems proliferate and are being designed and new, inclusive standards are created. This will allow standards makers to conceive of standards and, where applicable, the data sets and analysis approaches that inform these standards, with inclusion, efficacy and equity at the core. Furthermore, it will allow standards makers to expand notions of good practice to ways of making the data life cycle, including the presentation of data, and interpretation of data at the points of decision and use, more effective, inclusive and transformational. This includes the way we represent and share standards.

BSI Flex 236 is intended to expand the capability of standards makers to use data with inclusion in mind, to understand the limitations of the data that they rely on, and meaningfully apply that understanding to standards. This BSI Flex suggests a process to help standards makers apply critical questions and analyses aimed at understanding and interacting with data and the data life cycle. Understanding the role of data and the data life cycle is critical so that data and the standards produced become representative and include communities traditionally excluded from benchmark data sets or disregarded during the standards development process. Representative data sets and consideration of a wider range of communities and individuals can help to minimize harm, result in more robust products and services, and serve a more diverse population more inclusively.

Figure 1 summarizes the guidance given in this BSI Flex on data use relating to inclusivity in the standards development process.

Figure 1 – Standards development process overview

1 Scope

This BSI Flex gives guidance on how data are to be used in the development of all types of standards, including data used to create standards, data used to decide which standards to develop, what to include and exclude from standards, and to inform development decision making.

This BSI Flex covers:

- the ways data are collected, assessed, selected and used in standards development, including different data types and processing;
- how to deal with gaps in, and/or limitations of, existing data when developing inclusive standards;
- analysis of types of human bias;
- how biases can change the way data are interpreted and incorporated into decision-making processes;
- how data can introduce bias or exclusion into the standards development process; and
- the impacts of data bias or exclusion and the communities affected – this can include impacts related to protected characteristics, as well as non-protected characteristics.

It does not cover:

- specific aspects of product or service design; or
- data privacy or security.

This BSI Flex is for standards makers and organizations that develop standards, which might include small and medium-sized enterprises, and organizations that do not have National Standards Body status.

This BSI Flex could also be used by government and other organizations for the development of policies, procedures and regulations.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

For the purposes of this BSI Flex, the following terms and definitions apply.

3.1 algorithm

process or set of rules to be followed in calculations or other problem-solving operations

[SOURCE: BS ISO 19731:2017, 3.2]

3.2 artificial intelligence (AI)

capability of a functional unit to perform functions that are generally associated with human intelligence such as reasoning and learning

[SOURCE: ISO/IEC 2382:2015, **2123770**, modified – notes removed]

3.3 civil society organization

non-state, not-for-profit, voluntary organization formed by people in the social sphere, where the organization is separate from the State and the market

NOTE For further information, see <https://www.un.org/en/civil-society/page/about-us>.

3.4 cognitive bias

deviations from optimal, intentional or rational behaviour, as a result of psychological mechanisms

3.5 data

reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing

[SOURCE: ISO/IEC 2382:2015, **2121272**, modified – notes removed]

3.6 data governance

development and enforcement of policies related to the management and use of data

[SOURCE: ISO 8000-2:2022, **3.16.1**, modified – notes removed; “use of” added to definition]

3.7 data set

identifiable collection of data

3.8 debias

process of identifying and mitigating biases in data

3.9 discrimination

treating an individual or a group of individuals in a way that puts them at a disadvantage or at an advantage compared to others

[SOURCE: PAS 1948:2023, 3.6]

3.10 diversity

ideal state of not being dominated by any social, cultural or physical group but including representatives from a number of different social, cultural and physical groups

[SOURCE: PAS 1948:2023, 3.7, modified – notes removed; “physical” added to definition]

3.11 equality

state of being equal, especially in status, rights, opportunities or outcomes

[SOURCE: BS ISO 30415:2021, 3.9]

3.12 equity

system of justice and fairness designed to give everyone the same opportunities by acknowledging and making allowances for differences that would otherwise place them at a disadvantage compared to others

[SOURCE: PAS 1948:2023, 3.11, modified – notes removed]

3.13 fairness

principle that everyone is subject to procedures that are equitable and, as far as possible, free from systematic bias

[SOURCE: BS ISO 30415:2021, 3.11]

3.14 gender

socially constructed identities, attributes and roles of persons in relation to their sex and the social and cultural meanings attached to biological differences based on sex

[SOURCE: United Nations Human Rights, Office of the High Commissioner [5]]

3.15 harm

negative impacts on people, communities or environments

3.16 inclusion

allowing or inviting participation for all, irrespective of differences

NOTE *Stakeholders and individuals from different groups to be accepted and welcomed (e.g. offered opportunities on the basis of abilities, talents and skills).*

[SOURCE: PAS 1948:2023, 3.14, modified – notes amended]

3.17 inclusive culture

values, beliefs and practices that influence the conduct and behaviour of people and organizations by including and valuing the perspectives and contributions of the broadest range of stakeholders

[SOURCE: BS ISO 30415:2021, 3.13]