

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

Version 1 January 2022



BSI Flex 236 v1.0:2022-01



Department for
Business, Energy
& Industrial Strategy

Office for Product Safety & Standards

bsi.

Publishing and copyright information

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

© The British Standards Institution 2022.

Published by BSI Standards Limited 2022.

ICS: 01.120; 01.140.40; 03.100.40; 33.040.40

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

BSI's 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/our-services/standards-services/flex/>

Release history

First published January 2022.

Contents

Foreword ii

0 Introduction..... iv

1 Scope..... 1

2 Normative references 1

3 Terms and definitions 2

4 Principles for data use in standards development 4

5 Good practice for data use in standards development..... 11

Bibliography..... 13

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

Foreword

This BSI Flex was sponsored by the UK Department for Business, Energy and Industrial Strategy. 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 January 2022.

Acknowledgement is given to Caitlin Kraft-Buchman of Women at the Table as the technical author, and the following organizations that were involved in the development of this BSI Flex as members of the Advisory Group:

- Ada Lovelace Institute
- Association of Convenience Stores
- Centre for Data Ethics & Innovation
- Consumer and Public Interest Network
- Data 2X, United Nations Foundation
- Engineering Design Centre, University of Cambridge
- Kingston University
- Market Research Society
- Open Data Institute
- Internet Institute, Oxford University
- Prospect Union
- Women's Engineering Society

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 British Standard. It will be withdrawn in the event it is superseded by a PAS or a British Standard.

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

Information about this document

This release is Version 1 of BSI Flex 236, which has been released to enable industry to engage with the initial content and feedback comments. All comments received will be reviewed to inform the future direction of this standard. At present there are no confirmed plans to develop further versions. Interested parties are encouraged to check the BSI website for any updates regarding future versions.

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 BSI Flex takes the form of guidance and recommendations. It should not be quoted as if it were a specification or a code of practice.

Presentational conventions

The guidance in this BSI Flex 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 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

All standards are based on some form of data. The analogue and digital data we collect and/or use, whether in memory, numerical, text, images, audio, associated processing, or lived experience, the models we build from that data and the standards we make have a profound impact on the lives of individual people and groups. Our understanding of linkages and impacts between humans, data, code and systems and the role of data and data analysis, helps us to create more inclusive standards that better reflect our shared values.

The first steps in this understanding begin with questions. Who defines the problem that the data is 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 is an integral part of almost every step of modern processes.

Data is embedded everywhere in the new economy, and is the analogue foundation for 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 is not neutral, nor does it stand alone as “reality”. Much or most data is historical data which has the context of having been gathered on often small homogeneous subsets of the global population. This historical data is incomplete, and socially and technically biased due to its incompleteness.

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 for minority group members.

Older, unrepresentative data forms 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]) to newer, larger, yet also unrepresentative benchmark data 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 harm, from transport and trade, to finance, health, medicine and criminal justice. Data works at a velocity and scale in this century that touches on all corners of modern life. 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 if the original data has excluded all but a small generally homogenous (e.g. young, white, male, educated, heterosexual) dataset, 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. Of great concern for high-risk applications, it should also be noted that low risk applications also can have profound social impact on the quality of life and well-being of individuals, beginning with forms of access to information from targeted advertisements to search algorithms to financial access.

Data, even if “debiased”, behaves differently in different machine learning models, and identical machine learning models behave differently from one another in real world situations “in the wild”. Debaised data is also interpreted differently in different human decision-making processes. This is not exclusive to machines. Who or what does the debiasing influences the outcome of any such exercise.

A growing body of research has shown how exclusion of 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 we develop, from the selection, assessment and usage of incomplete data to create provisions and requirements, to the unquestioned usage 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 datasets, there is a risk of bias being replicated at scale and embedded in the systems of our daily lives and governance.

In addition, our notions of what is “representative” is itself evolving, such that data is always lagging behind our evolutions in understanding and application. The processes of standards bodies determines whether standards are developed and so the representative nature of these groups could determine whether a standard is 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.

0.2 Aims and objectives

With the use of inclusive data in standards development, there is an opportunity to not only mitigate, but correct for historic inequities buried in previous data. The opportunity, as new systems proliferate and are being designed, as new inclusive standards are created, is to revisit old assumptions and conceive of standards, datasets and data models with inclusion, efficacy and equity at the core, and to expand mono-notions of good practice to ways of making the data life cycle, including the presentation of data, and interpretation of data at the point of decision and its utility, more effective, inclusive and transformational, including the way we represent and share standards.

BSI Flex 236 is intended to build the capability of standards developers to work with 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 leverage a series of 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 we produce become representative and include communities traditionally excluded from benchmark datasets or disregarded during the standards development process. Representative datasets 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.

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

This page is deliberately left blank.