

# BSI Big Data standards development

## 1. Introduction

In early 2016 BSI published a [Big Data and standards market research report](#). The main purpose of the research was to determine the requirements for standards development associated with Big Data, focussed on the needs of the UK market. These findings were then developed into a series of potential 'scopes' for new standards. The scopes were then refined in three workshops that took place in March and May 2016 with 30 stakeholders from government, consumer groups, large corporates and SMEs. The scopes below are the draft scopes developed from the feedback in that workshop.

## 2. Draft standard scopes

### 2.1 Draft standard scope - Terms and Conditions

Title: *Big Data. Guidance on terms and conditions (T&Cs) for obtaining and using data.*

Context: Big datasets typically comprise vast quantities of transactional data (sometimes provided indirectly as part of using a 'free' service). The consent and trust of the data provider (often the public providing personal data) are vital to underpin both big data projects Big Data projects and the value derived from them. As such data providers (individuals) need to understand clearly what data they are contributing and what they get in return - including any potential adverse impacts. This standard will define best practice for T&Cs that are simple to understand and to optimise informed consent prior to data being used in Big Data projects.

This standard will ensure that consumers providing data:

- Understand how their data is being used (including sharing) and retained
- Understand who owns their data. The standard will provide policies for provision on access to data to consumers.
- Understand their rights concerning products derived from their data
- Understand the benefits that the service provides
- Understand where there is a 'social contract' (and provide a definition for social contract), including guidance on 'implicit consent' for their data use

This standard also:

- Describes governance frameworks for T&Cs
- Provides template T&Cs, including new approaches to communicating the data consent clauses
- Describes standard T&Cs for different service types
- Provides guidance on organisational data aggregation (allowing T&C's to be combined)
- Provides "dashboard" examples for managing 'my data consents'
- Covers cross over links to General Data Protection Regulation obligations

...making excellence a habit.™

- Management of Data Protection Subject Access Request

Standard type: Guidance - Guidance to develop initial progress, to be followed up with a Specification when there is greater acceptance of the approach

Users: Data Scientists, Legal teams, Consumer groups, Privacy managers, Chief Data Officers

PAS/BS/EN/ISO: PAS initially given the complexity of the topic. With the aim of the document then becoming the basis of an ISO.

Urgency in the market: Very high

## 2.2 Draft standard scope - Big Data Organizations

Title: *Big Data. Specification for a data driven organisation.*

Context: Data-centric organisations that exploit the benefits of big data do not follow traditional organisation structures. This standard provides a framework to allow an organisation to be data centric and hence maximise the benefits they obtain from Big Data and at the same time ensure compliance with the regulatory framework around the use of Big Data.

The standard would cover:

- Big Data Strategy
  - Recognition of data value in the organisation
  - Value generation at the intra and inter organisational level (different use cases)
  - Developing a data driven culture including guidance on transition / change management
- Roles and responsibilities
  - Cataloguing and managing data
  - Corporate Social Responsibility around data including Developing and maintaining ethical frameworks
- Process and Procedures
  - Governance and Security
  - Cost-benefit analysis of Big Data projects
  - Data management frameworks
- Enabling Technologies
  - Determining technological requirements based on the Data needs of the organisation
  - The standard will not specify particular technologies, but will consider generic issues around particular classes of technology (e.g. storage, transfer, archive)
- Audit and Review
  - Methods to ensure lessons learned are taken from projects
  - Internal communications

Standard type: Specification

Users: Data Scientists, Legal teams, Project Managers, Data Scientists, Project Managers, Chief Data Officers, CIO, Information Security, Developers, B.A.s, Marketing, (Customer) intelligence professions, Chief Social Scientist, Ethics committees, Technical architects

PAS/BS/EN/ISO: PAS/BS initially given the complexity of the topic. With the aim of the document then becoming the basis of an ISO.

Urgency in the market: High

## 2.3 Draft standard scope - Communications

*Big Data. Guidance on Big Data project communications.*

Context: There are examples where Big Data projects have failed to deliver, not because of the quality of the project, but due to a lack of understanding of the benefits from the perspective of those whose data has been used. This standard should aid such projects to define and articulate benefits to ensure public trust and societal/economic benefits. This standard will provide guidance (primarily for projects using personal data – whether or not the data is ‘anonymised’) on how to communicate externally before, during and after such projects to data subjects and other relevant stakeholders.

There is a relationship between this standard and the standard on Terms and Conditions, but this standard is concerned with communication of project aims and objectives.

The standard would cover:

- Transparency of data activities
- Engagement with data users, including templates for communication
- Communications frameworks around data breach and remedies (linking to GDPR requirements)
- Success stories (benefits) of using data
- Guidance on providing contextual use of data (“your data will be used for XX”)

Standard type: Code of practice

Users: Data Scientists, Legal teams, Project Managers, privacy, marketing/PR, CDO/Business Continuity, FOI teams

PAS/BS/EN/ISO: PAS/BS initially given the complexity of the topic. With the aim of the document then becoming the basis of a European standard or an ISO.

Urgency in the market: Medium

## 2.4 Draft standard scope - Metadata

*Title: Big Data. Guidance on Big Data metadata management.*

*Context:* Metadata has always been important to data management and several standards for metadata exist across different domains. Big data is an emerging domain that would benefit from its own metadata vocabulary especially for resource discovery and analytics and to address directly the practitioners in the Big Data domain.

This standard will articulate the best practices for using metadata in the big data context covering topics such as:

- Metadata models
  - Context [use case] of metadata (discovery [really simple discovery], use, evaluation...)
  - Cross-community understanding of big data / metadata (terminology)
  - Validation
- Metadata content
  - Context of origination/originator and provenance management
  - Approaches to metadata classification, including:
    - aggregation of datasets
    - spectrum of 'data v metadata'
    - mapping data to ontology
    - Metadata v metadata services (catalogue services, asset registers, etc.)
- Managing Metadata
  - Access rights to metadata (including sharing of metadata internally/externally)
  - Managing metadata quality
  - Collecting/storing/retention/updating metadata
  - Enhancing metadata (when new knowledge becomes apparent)

Standard type: Guidance or Specification. It is anticipated that this standard would initially be developed as guidance on using existing metadata approaches. A specification may subsequently be developed from this, however at this stage it is not clear if the specification will look at a new metadata standard, or specify how to use existing standards.

Users: Data Scientists, Project Managers, Chief Data Officers, Information Security, Developers, Business Analysts, Marketing, (Customer) intelligence professions, lawyers/digital forensics

PAS/BS/EN/ISO: PAS initially given the complexity of the topic. With the aim of the document then becoming the basis of an ISO.

Urgency in the market: Medium

## 2.5 Draft standard scope - Data Analytics

*Title: Big Data. Project guidance on undertaking Big Data analytics.*

*Context:* Big Data projects are new to most organisations and hence these do not know where to begin and how to avoid issues. Some projects may be research led that generates new insights for us in existing business process, whilst others may lead to a new business process. Projects often need to be agile, yet they cannot be a weak link in an organisation's information security policy.

This standard will provide guidance on how to manage Big Data projects/initiatives including:

- Governance and legal framework
- Project management methodology (*referencing existing frameworks/standards*)
- Risk management
- Involvement of the correct stakeholders
- Big Data project roles and responsibility
- Managing transparency and trust
- Capturing purpose of research (e.g. What-if analysis – data driven research)
- Definitions of data types - Structured, unstructured data models, encodings, etc
- Moving project outcomes from research to production (or Business As Usual)
  - Working with delivery partners
  - 'Standard' for products (validation)
- Managing T&C's of data
- Maintaining data validity
- Defining the analytics methods (including Workflow documentation)
- Repeatability of project
- Ensuring data quality
- Data retention, including Proactive disposal (lifecycle management)

Standard type: Code of practice

Users: Data Scientists, Legal teams, Project Managers, Data Scientists, Project Managers, Chief Data Officers, CIO, Information Security, Developers, B.A.s, Marketing, (Customer) intelligence professions, Chief Social Scientist, Ethics committees, Technical architects

PAS/BS/EN/ISO: PAS/BS initially given the complexity of the topic. With the aim of the document then becoming the basis of an ISO.

Urgency in the market: Medium

*CRISP-DM Should be referred to as a potential source/reference*

### 3. Other standards ideas

These topics were raised under the AOB of the workshops. These could be allocated to any of the five potential standards if appropriate; however for accurate reflection of the workshops they are presented separately:

- Specification for a data context description. This is a standard to intelligently determine how data can (and should) be used. It will set a framework for a context description for individual data sets.
- Determining how the data can be used in different contexts (protecting privacy, commercial rights, etc.)
- Guidance on the ethics of Big Data
- Management of 'data rights' when data is transferred between organisations
- Standard for 'data as currency' to put a value on the data that customers provide for 'free apps', etc.
- Guidance on MoUs and Data Sharing agreements
- Traceability and audit trails to make explicit data provenance
- DRM and value management in the context of big data
- Using API's to regulate and control access/rules. Rather than grant access to data, access is granted to an API to access the data. This can promote greater traceability in supply chains.

### 4. Next steps

The expectation is that potentially at least one further workshop will be required to develop these scopes further and determine the appropriate standardisation routes.

### 5. Contact

If you wish to contact BSI about this work please contact: Tim McGarr at [tim.mcgarr@bsigroup.com](mailto:tim.mcgarr@bsigroup.com) or Keiran Millard at [keiran.millard@bsigroup.com](mailto:keiran.millard@bsigroup.com)